Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 849 (2007) 181–189
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

The proteomics of keratin proteins☆

Jeffrey E. Plowman*

Canesis Network Ltd., Private Bag 4749, Christchurch, New Zealand

Received 31 May 2006; accepted 26 July 2006
Available online 22 August 2006

Abstract

Keratin proteins are widespread in nature, being found in the nuclei and cytoplasm of almost all differentiated eukaryote cells. However, they are best known as the principal structural proteins in hair, wool and skin. Because of difficulties associated with their extraction from biological samples, high sequence homology and the presence of numerous post-translational modifications, they have been less well studied than other protein families. Thanks to the advent of modern proteomic techniques we now have available a good suite of tools to study this neglected family of proteins.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Cytokeratins; Hair; Wool; Fish; Cancer; Cystic fibrosis; Ocular keratins

Contents

1. Introduction

The term keratin protein covers a class of proteins which includes intermediate filament proteins (IFPs) and intermediate filament associated proteins (IFAPs) that are known to occur in nature in a variety of different cell types [1]. The IFPs are so named because they have been found to be associated in intermediate filaments (IFs), a class of intracellular filamentous structures that are intermediate in size between microtubules and microfilaments. The IFPs have been found to occur in both the nuclei and cytoplasm of almost all differentiated eukaryotic cells, including epithelial, neuronal and glial cells, in addition to the cells of hair fibres, horns and nails. At least five types of IFPs are known. The acidic Type I and neutral-basic Type II chains are found in the hard α-keratins, such as in wool and hair fibres, and keratin in filaments composed of keratin-like proteins (cytokeratins) in epithelial cells. Type III chains occur in cells of mesenchymal origin, in smooth, cardiac and skeletal myogenic cells, glial cells and astrocytes and neuronal cells; Type IV chains are found solely in neuronal cells; and Type V chains are found within the nuclear envelope of mammalian cells [1].

The IFPs are peculiar in having in common a secondary structural motif consisting of four α-helical segments, separated by non-helical linker regions, with amorphous head and tail regions (Fig. 1), the total amount of α-helix comprising some 50% of the molecule [2]. Furthermore, this α-helix is noted for a heptad repeat of amino acid residues with a particular arrangement of hydrophobic residues that results in a hydrophobic stripe running around the right-handed α-helix in a left-handed manner. This

☆ This paper is part of a special volume entitled "Analytical Tools for Proteomics", guest edited by Erich Heftmann.

* Tel.: +64 3 325 2421; fax: +64 3 325 2717.
E-mail address: plowman@canesis.com.

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.07.055

Ovine K1.1

SFNFCLPNLSFRSSCSSRPCVPSSCCGTTLPGACNIPANVGSCNWFCEGS
FDGNEKETMQFLNDRLASYLEKVRQLERENAELESRILERSQQQEPLVCP
NYQSYFRTIEELQQKILCAKSENARLVVQIDNAKLAADDFRTKYETELGL
RQLVESDINGLRRILDELTLCKSDLEAQVESLKEELICLKSNHEEEVNTL
RSQLGDRLNVEVDAAPTVDLNRVLNETRAQYEALVETNRRDVEEWYIRQT
EELNKQVVSSSEQLQSCQTEIIELRRTVNALEVELQAQHNLRDSLENTLT
ETEARYSCQLNQVQSLISNVESQLAEIRGDLERQNQEYQVLLDVRARLEC
EINTYRGLLDSEDCKLPCNPCATTNACGKTITPCISSPCAPAAPCTPCVP
RSRCGPCNSYVR

Ovine K2.10

CGFSTVGSGFGSRAFSCVSACGPRPGRCCITAAPYRGISCYRGLTGGFGS
RSVCGGFRAGSCGRSFGYRSGGVCGPSPPCITTVSVNESLLTPLNLEIDP
NAQCVKQEEKEQIKCLNNRFAAFIDKVRFLEQQNKLLETKLQFFQNRQCC
ESNLEPLFEGYIETLRREAECVEADSGRLSSELNHVQEVLEGYKKKYEQE
VALRATAENEFVALKKDVDCAYVRKSDLEANSEALIQEIDFLRRLYQEEI
RVLQANISDTSVIVKMDNSRDLNMDCIVAEIKAQYDDIASRSRAEAESWY
RSKCEEIKATVIRHGETLRRTKEEINELNRVIQRLTAEVENAKCQNSKLE
AAVTQAEQQGEVALNDARCKLAGLEEALQKAKQDMACLLKEYQEVMNSKL
GLDIEIATYRRLLEGEEQRLCEGVGAVNVCVSSSRGGVVCGDLCVSGSRP
VTGSVCSAPCSGNLAVSTGLCAPCGQLNTTCGGGSCSLGRC

Human K18

SFTTRSTFSTNYRSLGSVQAPSYGARPVSSAASVYAGAGGSGSRISVSRS
TSFRGGMGSGGLATGIAGGLAGMGGIQNEKETMQSLNDRLASYLDRVRSL
ETENRRLESKIREHLEKKGPQVRDWSHYFKIIEDLRAQIFANTVDNARIV
LQIDNARLAADDFRVKYETELAMRQSVENDIHGLRKVIDDTNITRLQLET
EIEALKEELLFMKKNHEEEVKGLQAQIASSGLTVEVDAPKSQDLAKIMAD
IRAQYDELARKNREELDKYWSQQIEESTTVVTTQSAEVGAAETTLTELRR
TVQSLEIDLDSMRNLKASLENSLREVEARYALQMEQLNGILLHLESELAQ
TRAEGQRQAQEYEALLNIKVKLEAEIATYRRLLEDGEDFNLGDALDSSNS
MQTIQKTTTRRIVDGKVVSETNDTKVLRH

Human Filensin

MYRRSYVFQTRKEQYEHADEASRAAEPERPADEGWAGATSLAALQGLGER
VAAHVQRARALEQRHAGLRRQLDAFQRLGELAGPEDALARQVESNRQRVR
DLEAERARLERQGTEAQRALDEFRSKYENECECQLLLKEMLERLNKEADE
ALLHNLRLQLEAQFLQDDISAAKDRHKKNLLEVQTYISILQQIIHTTPPA
SIVTSGMREEKLLTEREVAALRSQLEEGREVLSHLQAQRVELQAQTTTLE
QAIKSAHECYDDEIQLYNEQIETLRKEIEETERVLEKSSYDCRQLAVAQQ
TLKNELDRYHRIIEIEGNRLTSAFIETPIPLFTQSHGVSLSTGSGGKDLT
RALQDITAAKPRQKALPKNVPRRKEIITKDKTNGALEDAPLKGLEDTKLV
QVVLKEESESKFESESKEVSPLTQEGAPEDVPDGGQISKGFGKLYRKVKE
KVRSPKEPETPTELYTKERHVLVTGDANYVDPRFYVSSITAKGGVAVSVA
EDSVLYDGQVEPSPESPKPPLENGQVGLQEKEDGQPIDQQPIDKEIEPDG
AELEGPEEKREGEERDEESRRPCAMVTPGAEEPSIPEPPKPAADQDGAEV
LGTRSRSLPEKGPPKALAYKTVEVVESIEKISTESIQTYEETAVIVETMI
GKTKSDKKKSGEKSS

Fig. 1. The amino acid sequences of four IFPs: K1.1, an ovine Type I IFP; K2.10, an ovine Type II IFP; K18, a human cytokeratin and human filensin. The proteins are characterised by a common secondary structural motif consisting of four α-helical segments (dark grey background) separated by three linker regions (white background) and amorphous N- and C-terminal segments of variable length (light grey background). (Note that in the human K18 and filensin sequences the third linker region has not yet been identified.).

allows two such helices to associate to produce a coiled coil—the basic building block of the IF. Studies have shown that this association is driven by the dephosphorylation of specific serine and threonine residues located in the N- or C-terminal domains of the IFPs, by enzymes such as alkaline phosphatase [1]. These coiled-coils can then associate further, in conjunction with the IFAPs (by crosslinking through disulfide bonds), to form macrofibrils in wool and hair. At least three types of IFAPs are known, those high in cysteine, the high sulfur proteins (HSPs) and the ultra-high sulfur proteins (UHSPs), and the high glycine–tyrosine

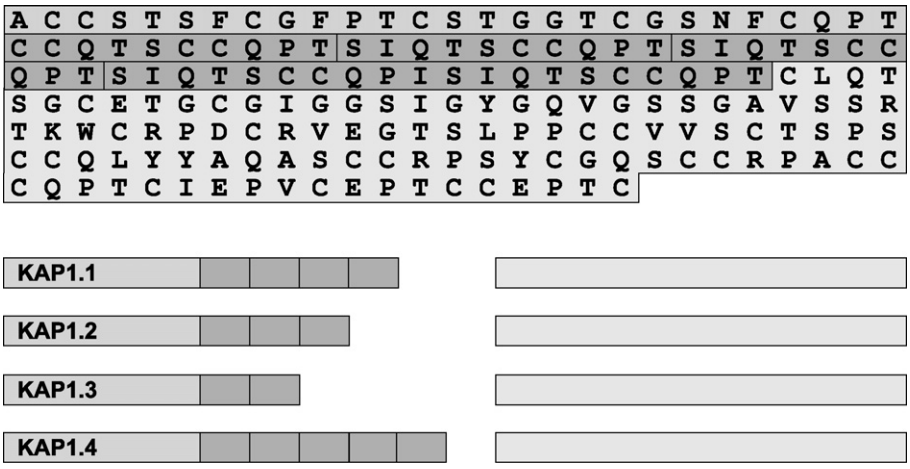

Fig. 2. Sequence homologies in the ovine KAP1 HSP family. The proteins are characterised by a short N-terminus (mid-grey background), a long C-terminus (light grey background) and series of decapeptide repeat sequences (dark grey background), KAP1.4 having five of these repeats, KAP1.1 four, KAP1.2 three and KAP1.3 two.

proteins (HGTPs). Some families of IFAPs such as the ovine KAP1 HSPs are noted for having a highly conserved sequence including a 10-residue repeat sequence, the members of this family differing in the number of these repeat sequences present in the molecule (Fig. 2) [3]. Alternatively the sheep KAP5 UHSP family is characterised by a series of alternating long glycine-rich and short cysteine-rich sequences (Fig. 3), which are present in differing amounts in the members of this protein family [3]. In contrast other IFAPs, such as the sheep KAP3 HSP family or the KAP7 and KAP8 HGTPs show little in the way of repetitive sequences [3].

While most of the initial emphasis of the proteomics of keratin proteins has been on wool and hair, there has been an increasing interest in studying the keratins from other tissues by this approach, because of the recognition of their role in many diseases and the potential for the development of diagnostic tools for these conditions.

2. Wool keratins

Among the earliest approaches to proteomic studies of wool keratins was electrophoretic separation of proteins which had been extracted with high concentrations of urea (8 M), under reducing conditions at pHs of around 9.3 and then converted to their S-carboxymethyl form with iodoacetic acid. Separation was by way of two-dimensional electrophoresis (2DE), using isoelectric focusing with carrier ampholytes in the first dimension, followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension [4], which for the purposes of this review will be referred to as IEF-2DE. Although, at least, one of the proteins in the map was identified by spiking the protein extract with a purified sample of the IFP Type II keratin protein as K2.12/Component 5 [5,6] (Table 1), identification of the other proteins appears to have been based on knowledge of their electrophoretic mobilities and molecu-

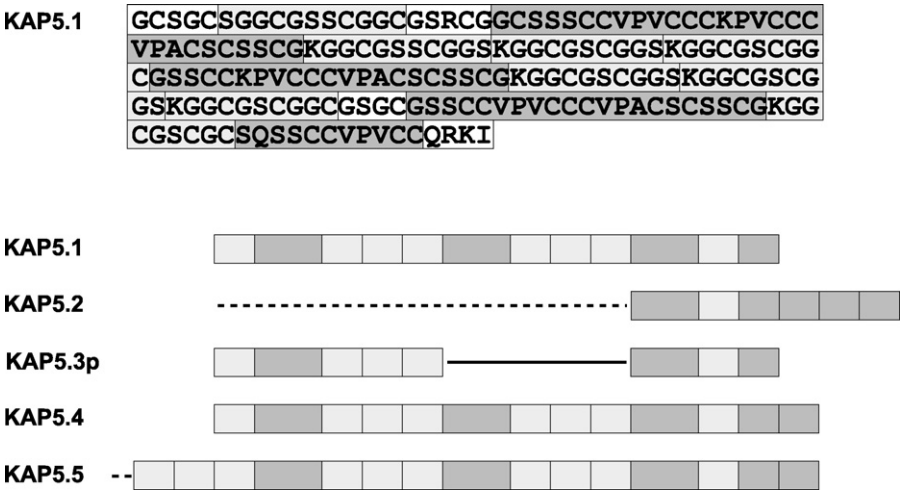

Fig. 3. Sequence homologies in the ovine KAP5 UHSP family. The family is characterised by glycine-rich (dark shading) and cysteine-rich (light shading) repeat sequences.

Table 1
Nomenclature of wool IFPs

	Swiss-Prot	Crewther et al. [5,6]	Powell and Rogers [2]	Plowman et al. [23]
Type I	K1M1	Component 8a Component 8b Component 8c1 Component 8c2	K1.1	oHa1
	K1M2		K1.2	
				oHa5
				oHa8
Type II	K2M3	Component 5 Component 7a Component 7b	K2.12	oHb5
	K2M2	Component 7c	K2.10	

Note: In the nomenclature of Plowman and co-workers [23], o stands for ovine, H for hair, a for the acidic proteins and b for the neutral-basic proteins.

lar weights. This latter approach appears to have been applied to identifying proteins in the more commonly employed non-equilibrium 2D-electrophoretic (NE-2DE) approach, whereby the S-carboxymethylated proteins were separated in the first dimension by exploiting differences in protein charge, using alkaline PAGE (pH 8.9), and separating by apparent molecular weight using SDS-PAGE in the second dimension [7] (Fig. 4). However, at least one of these protein classes appears to have been identified by running enriched UHSP fractions on the NE-2DE system [8], while the location of the BIIIB HSP was determined by the use of a translation product of a cloned gene [9]. Most of these studies have been applied to whole fibre extracts, however proteins extracted from isolated cuticle were also separated by NE-2DE and found to be dominated by UHSPs [10].

More recently IEF-2DE, performed in immobilised pH gradients in the first dimension, has been applied to wool keratin proteins in which the cysteine residues were converted to their S-carboxyamidomethyl form with iodoacetamide. In the resulting gels, the Type II IFPs, which focused at low p*I* in the carboxymethylated form [4], were observed to spread out into long trains of spots between p*I* 5.0 and 6.5, while the Type I IFPs were clustered in four major trains of spots between p*I* 4.8 and 5.2 (Fig. 5) [11]. Similar results were obtained when the cysteine residues in the proteins were converted by oxidative sulfitolysis to their S-sulfonate form [12]. An initial attempt to identify seven spots in the Type II IFP train in these 2DE gels, using amino acid compositional analysis found two major proteins, K2.10/Component 7c at low p*I* and K2.12/Component 5 at high p*I* (Table 1) [13]. This was subsequently confirmed using mass spectral peptide fingerprinting from matrix-assisted laser desorption ionisation time of flight (MALDI TOF) spectra [14]. In addition, the four major trains in the Type I region were identified as being either of the two Type I IFPs for which sequences are available from web-based databases [15]. A similar train of spots at low molecular weight was identified as being a member of the KAP 6 HGTP family (Fig. 2), though no specific protein in the family could be identified because the one prominent peak in the mass spectrum was for a sequence common to the two known sequences and no other peaks were observed. In contrast, at least three proteins from the HSP region were identified as being from the B2/KAP1 family, based on the masses of their C-terminal sequences. While the minor proteins from the HSP region of the gel map produce short trains of spots [13], these prominent HSPs from the B2/KAP1 family usually only exhibit a few spots per train, two in Merino sheep and some minor spots at lower p*I* in Romney and Corriedale sheep [16].

Application of IEF-2DE, using immobilised pH gradients, to the separation of proteins from cultured wool follicles, where the proteins, labelled with ^{35}S-methionine, were detected by autoradiography, produced a very different pattern of proteins from the keratinised wool fibre [17]. IFPs between 40 and 65 kDa fell into two groups, an acidic group between p*I* 3.0 and 5.0 and a basic group around p*I* 6.0. Some of the proteins in the both groups had lower molecular weights (between 30 and 40 kDa) than previously observed for IFPs. One interesting observation was the larger number of spots in the region from 10 to 26 kDa, corresponding to HSPs, and, although the majority were acidic, a discrete number were observed in the basic region of the p*I* gradient.

Until recently, it was believed that there were only four Type I and four Type II IFPs in the wool fibre [3]. One thing notable from the earlier studies, particularly in relation to the Type I IFPs, is that of all of the spots on the NE-2DE and 2DE maps were assigned to the four known Type I and four known II IFPs. In contrast, more recent attempts using modern proteomic techniques, have only been able to assign the four major trains of spots in the Type I IFP region to two proteins for which sequences are available on publicly accessible databases [18]. This highlights one of the problems associated with the application of mass spectroscopy to the identification of keratin proteins from gels. These proteins are highly homologous (sequence identities of around 80–90%), hence identification is entirely dependent on the availability of known sequences to obtain a match. Thus, more known sequences are needed to allow the proteins in these trains of spots to be correctly assigned.

A similar situation exists for the Type II IFPs, where only three full sequences are available [2,18–22]. However at least two new sequences, oHa5 and oHa8 (Table 1) have been uncovered [23], which have strong homology to human hair proteins. This suggests that wool may have as many IFPs as human hair,

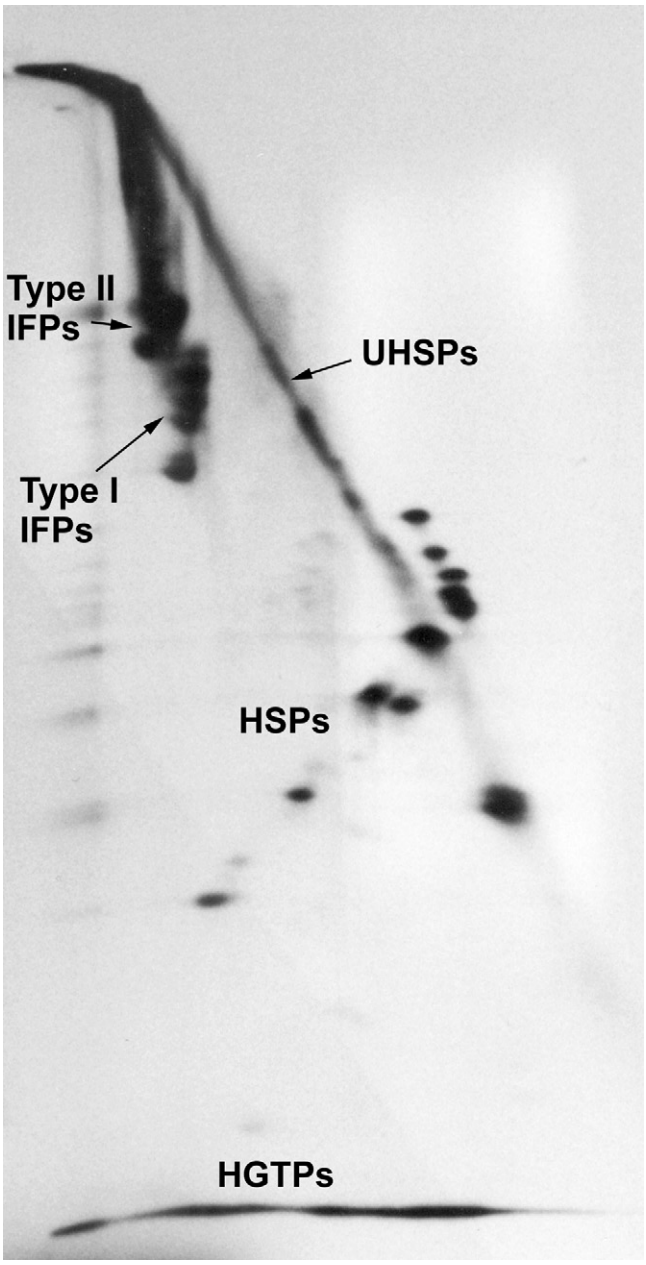

Fig. 4. A Coomassie blue stained NE-2DE map of carboxymethylated keratin proteins from Merino wool separated at p*I* 8.9 in the first dimension, showing the relative positions of the main protein classes: intermediate filament proteins (IFPs), high sulfur proteins (HSPs), ultra-high sulphur proteins (UHSPs) and high glycine–tyrosine proteins (HGTPs).

for which nine Type Is and six Type IIs have been identified [24,25]. This study also indicated that there may be a degree of polymorphism among the Types I and II IFPs.

One significant difference between the protein maps obtained by NE-2DE and IEF-2DE is the increased number of protein spots apparent in the latter, yet proteomic analysis indicates that the trains of spots are composed of the same protein. Originally, this was thought to be due to post-translational modification of the proteins in the trains. One study claimed to show differences in the proteins on the IEF-2DE map after treatment of the proteins with alkaline phosphatase [12]. Subsequently, there were claims that both phosphoserine and phosphotyrosine had been detected by HPLC in Type II IFP spots after excision from a 2DE gel and partial gas-phase hydrolysis [13]. More recent work, using the specific stains for phosphoproteins and glycoproteins, found no evidence for either phosphorylation or glycosylation in any wool keratin proteins [26]. Further support for this comes from mass spectral studies, where no glycosylations were detected in Type II IFP spots excised from an IEF-2DE gel [26]. However, when the proteins were eluted from spots excised from the gels and re-run on another IEF-2DE gel, the train was observed to re-establish itself, suggesting that it is the result of an equilibrium between different conformations of the protein. Further support for this has come from studies in which the proteins were progressively alkylated for up to 48 h, resulting in the gradual coalescence of the trains of IFP spots into a single spot [27].

One particular interest in studying keratin proteins in wool has been to locate possible markers for particular wool traits, such as fibre curvature and wool strength. An early study used wools from chimaeric sheep, developed from fusing fertilised Merino and Lincoln embryos at the blastomere stage, because the expression of genotypes in one animal neutralised the effects of physiology, age, nutrition and weather in the animals. Although there was evidence of a correlation between the B2/KAP1 HSP expression and fibre curvature, the analysis was limited by the small size of the population available [14]. Wools from more sheep were examined in a recent study, which indicated that the expression of these proteins is complex and shows considerable within-breed, as well as between-breed, variation [16]. There also appeared to be no correlation with curvature, except possibly for one breed.

3. Human hair keratins

Similar patterns of protein spots to those of wool were observed when NE-2DE was applied to keratin proteins extracted from human hair and nails [28,29]. No proteins were identified on the gels at this stage but in one study, spanning three generations of one family [29], areas of variability in the electrophoretic pattern appeared to be due to genetic variations between individuals, as well as cosmetic manipulation and environmental damage to the fibre. The application of IEF-2DE, whereby S-carboxyamidomethylated hair proteins were separated in the first dimension by isoelectric focusing in carrier ampholyte solutions, produced protein maps similar to those of wool, with a cluster of IFPs at low p*I* and a long train of proteins at higher molecular weight between p*I* 5 and 7 [30].

With more careful analysis, separating the proteins by IEF-2DE using immobilised pH gradients in the first dimension, it was found that the proteins of the human Type I IFP family could be partly resolved, despite being clustered close together [24]. It was also possible to identify their position on IEF-2DE gels by Western blot analysis using antibodies specific to each member of the family. Of the nine members of the family the hair keratins hHa1, hHa3-I/II, hHa4 and hHa5 (where ‘h’ stands for human, ‘H’ for hair and ‘a’ for the acidic proteins) were found to be strongly expressed in the fibre and Coomassie-stainable,

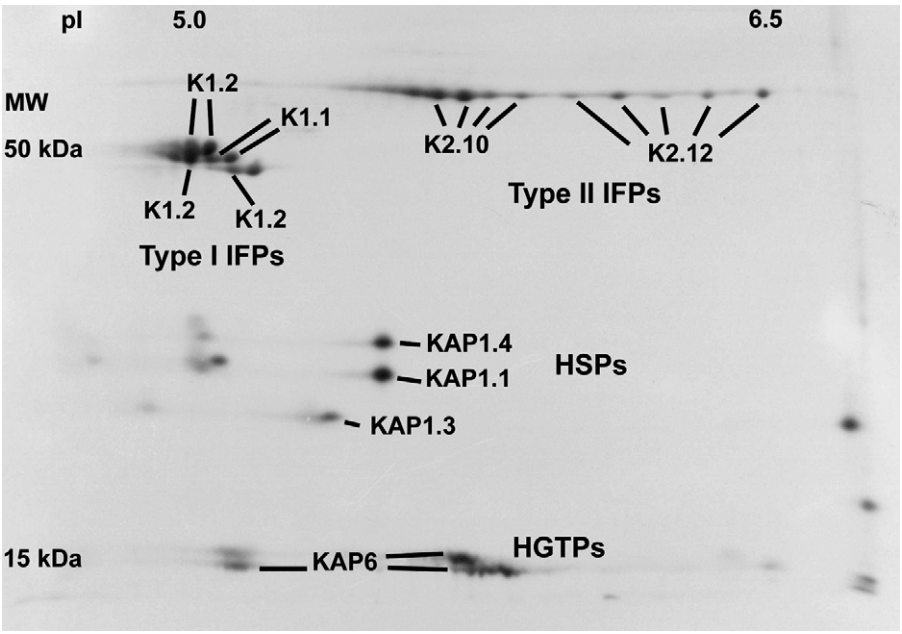

Fig. 5. A Coomassie blue stained IEF-2DE map of proteins from Merino wool, from p*I* 4 to 7 and molecular weight 15 to 50 kDa, showing the relative positions of the main protein classes: intermediate filament proteins (IFPs), high sulfur proteins (HSPs) and high glycine–tyrosine proteins (HGTPs) and the proteins identified to date using peptide mass fingerprinting [15].

while hHa2, hHa6, hHa7 and hHa8 were weakly expressed and only immunodetectable. Not all of these proteins were found in terminal hair (the main hair type in humans). *In situ* hybridisation studies on scalp follicles demonstrated that hHa7 is found only in vellus hair (fine, non-pigmented hairs), while hHa2 is found only in the cuticle. Unfortunately, there was considerable overlap between some of the members on the gel, which has the potential to create difficulties identifying them with the mass spectral fingerprinting approach or peptide sequencing using MS–MS.

The location of the human Type II IFPs on the IEF-2DE gel map was also determined using the same approach [25]. One protein, hHb5 (where b stands for neutral-basic), was found to focus in a long string between p*I* 6.5 and 7.5, while four other proteins, hHb1, hHb2, hHb3 and hHb6, were found to focus very close together at lower pIs and molecular weights. The remaining protein in the family, hHb4 was not detected in the hair follicle at all but in cytoskeletal extracts of human tongue. Once again, the strong overlap between the four Type II IFPs at lower p*I* would create difficulties in identification by mass spectrometry.

If the number of IFP proteins in sheep wool is similar to that in human hair, the process of identifying these proteins on 2DE gel maps by approaches such as mass spectral peptide mapping is potentially complicated, particularly for the low p*I* train of Type II IFPs, and especially if more than four unique sequences are found.

As an alternative the conventional 2DE approach using polyacrylamide gels Lee et al. [31] employed the multi-dimensional protein identification technology (MudPIT) approach to examine the human hair proteome. The soluble protein component was extracted by repeatedly pulverising hair in a solution of 2% SDS, 20 mM dithioerythritol and 50 mM sodium phosphate at pH 7.8, which was found to remove approximately 80% of the protein from the fibre. Following this both it and the insoluble component were digested with TPCK-treated trypsin. The resulting peptides were separated by two-dimensional chromatography, using strong cation exchange chromatography in the first dimension and capillary reverse phase chromatography, before being injected directly onto a ion-trap mass spectrometer where survey scans were collected for each fraction and MS/MS spectra obtained for the three most intense ions. Sequest software was then used to find sequences in a human proteome database. While the soluble fraction was found to contain a mixture of intracellular proteins (IFPs and KAPs) the insoluble fraction was composed of a variety of proteins. Prominent among these were the cuticular IFPs, such as hHa2 and hHb2, and cuticular KAPs but also found were desmosomal proteins that link the keratins to proteins from the transmembrane region, transglutaminases that are involved in the formation of ε-(γ-glutamyl)lysine crosslinks, as well as proteins involved in regulation and metabolism, such as the 14-3-3 proteins that are reportedly involved in keratin filament organisation.

The presence of post-translational modifications in human hair proteins was also examined by Nakamura et al. [32] after separation by one-dimensional electrophoresis and detection by Western blot using antibodies to phosphoserine, phosphotyrosine and phosphothreonine. In contrast to the studies for wool when no phosphorylations were observed [26] they found evidence for phosphorylations in both the acidic Type I IFPs and the Type II IFPs. More recently, evidence has been found for ubiquitination and methylation of proteins from the hair fibre with some keratin proteins showing signs of dimethylation and trimethylation, however no signs of phosphorylation were found [31].

4. Cytokeratins

Studies of filaments have shown that cytokeratins are characteristic of epithelial cells; vimentin filaments occur in mesechymally derived cells, astrocytes, Sertoli cells and vascular smooth muscle cells; desmin filaments are typical of most types of myogenic cells; neural filaments are typical of neuronal cells and glial filaments are typical of astrocytes [33]. During cell transformation and tumour development, the cell type specificity of IFs is conserved and, as a result, classification of tumours by the specific type of IFs has become a very valuable tool in clinical histodiagnosis. Thus, there has been an increasing interest in the application of proteomics to the separation and identification of cytokeratins.

Despite the existence of many cytokeratins, they are well resolved by molecular weight in the second dimension, as well as manifesting themselves as trains of spots [34,35]. Unlike the hard α-keratins of wool and hair [26], this appears to be due to phosphorylation. Liao et al. [34] used IEF-2DE gel analytical techniques to study the post-translational phosphorylation of epithelia/glandular keratins, particularly keratins 8 and 18 (K8 and K18). Both were purified from whole cell extracts after metabolic labelling with $^{32}PO_4$, followed by electroelution from preparative SDS-PAGE gels and then precipitated by acetone. The post-translationally modified forms of the proteins were then separated by IEF-2DE, using a mix of carrier ampholytes of pH 5–7 and pH 3–10 in the first dimension and SDS-PAGE in the second dimension. The stoichiometry of the phosphorylation of a given phospho-isoform at a particular amino acid residue was determined by immunoblotting K8/18 immunoprecipitates that had been isolated from *in vivo* $^{32}PO_4$-labelled cells, followed by separation by IEF-2DE with the antibody that specifically recognises the particular K8/18 epitope [35]. Using this approach, the antibody reactivity with a given isoform was correlated with the intensity of the *in vivo* phosphorylation detected by autoradiography and the intensity of the Coomassie staining of the corresponding protein isoform, thereby indicating the extent of protein phosphorylation at a particular site for a subset of molecules. Thus, the application of 2DE to study glandular keratin phosphorylation helped to characterise different phosphorylation states which depended on the biological state of the cell and also the phosphorylation gradient between cytosolic, membrane-associated and filamentous keratin pools. Using this approach, they were also able to establish that the cytokeratins K8/K18 are expressed in insect ovarian cells and that the glycosylation and phosphorylation are very similar to that in humans [36].

Moll et al. [33] were able to distinguish between 19 cytokeratins from various human epithelium carcinomas and cultured epithelial cells using a 2DE approach. Highly purified fractions of cytokeratins were obtained from epithelial tissues using the detergent Triton X-100 in combination with high salt buffers. These cytokeratins were separated either by NE-2DE or IEF-2DE, the latter using carrier ampholytes in the first dimension. The proteins were detected by either Coomassie or silver staining or by reactions of the proteins with antibodies blotted on to nitrocellulose paper and visualised with ^{125}I-labelled protein or ^{125}I-labelled goat anti-mouse immunoglobulin followed by autoradiography [37]. Most of the cytokeratins were found to focus between p*I* 5 and 7 and were well separated in the second dimension between 40 and 70 kDa and, as with K8, K18, K19 and K20 [36], all produced multiple spots as a result of post-translational modification by phosphorylation. It is evident from the study that these diverse cytokeratins are expressed in specific sets of polypeptides that are characteristic of particular cell types. Thus, it was possible to distinguish and classify epithelial cell types on the basis of their cytokeratin pattern.

Analyses of cytokeratins of human carcinomas also showed that tumours derived from different types of epithelia display different cytokeratin polypeptide patterns that are characteristic of certain types of tumours [33]. Furthermore, the patterns of cytokeratins of a given type of tumour appear to be identical in the primary tumours and the metastases. In general, tumours derived from the epithelium appear to continue to express many of the same polypeptides typical of the non-transformed cells and in some cases, such as the digestive tract, the patterns are identical.

Proteomic techniques have been applied to the analysis of cytokeratins associated with lung adenocarcinoma [38]. Proteins extracted from tissue samples of lung adenocarcinomas were solubilised with a standard lysis buffer and separated by IEF-2DE using carrier ampholytes in the first dimension. The proteins were visualised by either silver staining, to show the general protein composition of the extracts, or Western blot analysis using antibodies to the specific cytokeratins of interest. For mass spectral identification of proteins, silver staining was still used for spot visualisation but the loading on the gel was increased by 30% and peptide mass fingerprinting of tryptic peptides was employed using either MALDI-TOF or electrospray tandem mass spectroscopy (ESI MS/MS). At least 300 proteins were differentially expressed between lung adenocarcinoma tumour and normal cell lines. Among these, four cytokeratins (K7, K8, K18 and K19) were identified, each of which exhibited multiple isoforms, though no attempt was made to determine their nature (ie. whether they were the result of post-translational modifications). In addition, one or more of the isoforms for each cytokeratin showed significant increases in lung adenocarcinomas compared to normal lung. In the case of K19, two isoforms were over-expressed and one underexpressed in lung adenocarcinomas. Of particular interest was the fact that one K8, one K19 and all K7 isoforms were significantly associated with patient survival.

Cystic fibrosis is a genetic disease characterised by an abnormal chloride secretion in epithelia due to a dysfunction of a mutated cystic fibrosis transmembrane conductance regulator (CTFR). One of the most frequent mutations leads to an aberrantly folded protein which causes a dysfunction of a cAMP-dependent chloride channel. Proteomics was employed to identify proteins whose expression was dependent on the either the wild-type or the mutated gene [39].

Proteins were extracted from the HeLa cells stably transfected with the wild-type or mutant gene, using a standard lysis buffer, though in order to improve the reproducibility of the protein preparations from one cell culture to another, a new

protocol was introduced. This involved sequential addition of sodium dodecyl sulfate at 95 °C and spermin to precipitate DNA. The proteins were subsequently separated by IEF-2DE, using immobilised pH gradients in the first dimension. Differences in protein expression were then determined by 2D-image analysis of gels stained with a MS-compatible silver stain. Spots of interest were then excised from the gel, peptides obtained from their trypic digests were identified by peptide mass fingerprinting using MALDI-TOF MS and their sequences determined by ESI-MS/MS on a Q-TOF mass spectrometer. The gels were characterised by short trains of spots at molecular weights ranging from 40 to 150 kDa, between p*I* 4 and 7.

Using this approach, four spots were found to be differentially expressed between the normal and cystic fibrosis cell line; in all cases these proteins were over-expressed in the mutant strain. One spot in one train was found to be cytokeratin K18 and three in another train to due to K8, though no attempt was made to determine whether these were due to post-translational modifications. Unfortunately the study was not able to determine why these proteins were over-expressed in the mutant cystic fibrosis cell line.

In non-teleost (cartilaginous) fish, such as sharks, rays and skates, keratin filaments assembled from Type I/II heterodimers are almost exclusively restricted to epithelial cells, whereas in teleost (bony) fish they are also found in a variety of mesenchymal cell. In simple terms, although IFPs resembling the vertebrate Types I, II and III have been found in lower chordates, only true orthologs of human keratins (K8 and K18) and vimentin and desmin have been found within the teleost fish. Nevertheless, the system is much more complex than this and it was with this in mind that Schaffeld and Markl [40] have been applying proteomic techniques to separate and identify cytokeratins in teleost and non-teleost fish to map phylogenic trees in fish. Their ultimate aim was to extend this into land based vertebrates.

Keratins were extracted on ice using a low salt buffer (150 mM NaCl), followed by extraction of the resulting protein pellet with a high salt buffer (1.5 M KCl) containing polymethyl sulfonyl fluoride and other protease inhibitors, including pepstatin, leupeptin and benzamidin, followed by separation of the keratin proteins by IEF-2DE or NE-2DE. The latter was found to be preferable when acidic to neutral keratins and basic keratins were to be separated, while IEF-2DE was more useful for separating keratins when their pIs were all below 7.0. Generally, however, IEF-2DE was sufficient for examining keratins from fish, NE-2DE being necessary for keratins obtained from mammals. The proteins were visualised on the gels by staining with Coomassie Blue R250, immunoblotting or a complimentary keratin blot binding (CKBB) assay, the latter two on nitrocellulose sheets. The CKBB assay was found to be more efficient than immunoblotting for identifying keratins and had the advantage of allowing them to be classified immediately into Type I or II polypeptides. Proteins were also identified using the peptide mass fingerprinting approach using spectra generated by MALDI TOF MS and these results were used to confirm the sequences derived from cDNA sequencing. The application of peptide mass fingerprinting enabled the grouping of the polypeptides separated by 2DE into clusters each representing a genuine keratin and its modifications.

By combining data from this study with that from other vertebrates, including amphibians and mammals, they were able to demonstrate that keratins are consistently expressed in epithelial cells, and that keratins expressed in the epidermis and other stratified epithelia can be distinguished from keratins expressed in simple epithelia. The sequences thus derived from the various species were used to generate a phylogenic tree, which, amongst other things, was able to demonstrate that the keratins from the lower chordates are distinct from the Types I and II keratins found in vertebrates. It also showed that keratins produced by the lamprey originated from an independent radiation event and that separate branches are formed by IF proteins from the lancelet, the tunicate and by mammalian hair keratins.

5. Ocular keratins

There has been some interest in the expression of keratins in lens tissue in the eye because of its relationship to the formation of cataracts later in life. Studies have shown that in the embryonic eye there is a transient expression of the cytokeratins K8 and K18, which has been found to persist until the 8th fetal week, after which vimentin is the only IFP to be found in the lens epithelium [41]. In contrast, adult lens tissue is characterised by the down-regulation of vimentin and the expression of two lens specific IFPs, phakinin and filensin (Fig. 1). These two IFPs are known to form filaments with each other, known as beaded filaments because they appear under a TEM as fine filaments covered along their length with fine globular beads. They do not form filaments with vimentin, however.

IEF-2DE has been used to identify keratins in lens tissues of rats [42]. Membrane fractions from ocular lenses were extracted with 8 M urea and then separated by carrier ampholyte isoelectric focusing in the first dimension. Proteins were electroblotted from the gels on to PVDF membranes and detected by Western blotting with either a monoclonal antibody to vimentin or a polyclonal antibody to phakinin. Both proteins were detected on the gels, as well as a peptide with immunoreactivity to both vimentin and phakinin. Analysis of this peptide by MALDI-TOF MS revealed a match to a 68 kDa Type II IFP, cytokeratin 1. This was further investigated by probing proteins electroblotted on to PVDF membranes with a monoclonal pancytokeratin antibody which reacted with cytokeratins 1, 5, 6 and 8 or a monoclonal antibody specific for cytokeratin 1, both of which reacted with the unknown peptide. While this may have been the first time cytokeratin 1 had been detected in lens tissue the authors had to concede that further work would need to be undertaken to confirm this finding.

The effect of ultraviolet light on bovine lens tissue was examined by irradiating it for 75 min at 365 nm with a UV lamp in a medium of Earle's salt, 6% fetal calf serum, penicillin and streptomycin [43]. After the epithelium was removed from the lens, the cortex was extracted with water and then 5 M urea. The proteins from both the water- and urea-soluble fractions were separated by mini IEF-2DE, using carrier ampholytes in the first dimension. The proteins from the water soluble fraction

were detected by Coomassie Blue staining, while those from the urea-soluble fraction were electroblotted on to nitrocellulose paper and Western blotting analysis was performed with an antibody specific for vimentin. The results from this study indicated that the vimentin had been partly degraded by the irradiation.

6. Conclusions

Keratins are ubiquitous proteins, being found in everything from hair, nails or rhino horn to the nuclei of cells. Because of difficulties relating to extractability, particularly from biological samples, such as hair and wool, extensive sequence homology and the presence of multiple species on gels due to post-translational modifications, keratins present a challenge for analysis using proteomic approaches. Fortunately, thanks to the advent of modern technologies, such as mass spectrometry and immunochemistry there are now a wide range of tools available to aid the detection and identification of these proteins.

To some extent keratin proteins have been neglected because cytosolic proteins are more easily analysed. Nevertheless there are potential benefits from the application of proteomics to the identification and classification of this class of proteins. Studies conducted over the past few years have revealed potential applications for determination of marker proteins for wool quality traits, or as diagnostic tools for the detection of such ailments as cancer and cystic fibrosis or the degradation of lens tissue by ultraviolet light.

Acknowledgements

I am grateful to Joy L. Woods for her help in the preparation of this review and for NZ Government funding via the Foundation for Research, Science and Technology, Contract Number WROX0301.

References

[1] D.A.D. Parry, P.M. Steinert, Intermediate Filament Structure, Springer-Verlag, New York, 1995.
[2] B.C. Powell, G.E. Rogers, in: P. Jollès, H. Zahn, H. Höcker (Eds.), Formation and Structure of Human Hair, Birkäuser Verlag, Basel, 1997, p. 59.
[3] B.C. Powell, Wool Technol. Sheep Breed 44 (1996) 100.
[4] R.C. Marshall, R.J. Blagrove, J. Chromatogr. 172 (1979) 351.
[5] W.G. Crewther, L.M. Dowling, K.H. Gough, A.S. Inglis, N.M. McKern, L.G. Sparrow, E.F. Woods, Proceedings of the 5th International Wool Textile Research Conference, vol. II, Aachen, 1975, p. 233.
[6] W.G. Crewther, L.M. Dowling, K.H. Gough, R.C. Marshall, L.G. Sparrow, in: D.A.D. Parry, L.K. Creamer (Eds.), Fibrous Proteins: Scientific, Industrial and Medical Aspects, 2, Academic Press, London, 1980, p. 151.
[7] R.C. Marshall, Text. Res. J. 51 (1981) 106.
[8] B.C. Powell, G.E. Rogers, in: J. Bereiter-Hahn, A.G. Matoltsy, K. Sylvia-Richards (Eds.), Biology of the Integument 2: Vertebrates, Springer-Verlag, Berlin, 1986, p. 695.
[9] P.J. MacKinnon, Molecular analysis of the ultra-high-sulphur Keratin proteins, Ph.D. thesis, University of Adelaide, 1989.
[10] R.C. Marshall, K.F. Ley, Text. Res. J. 55 (1986) 772.
[11] B.R. Herbert, J.L. Woods, Electrophoresis 15 (1994) 972.
[12] B.R. Herbert, A.L.P. Chapman, D.A. Rankin, Electrophoresis 17 (1996) 239.
[13] B.R. Herbert, M.P. Molloy, J.X. Yan, A.A. Gooley, W.G. Bryson, K.L. Williams, Electrophoresis 18 (1997) 568.
[14] J.E. Plowman, W.G. Bryson, T.W. Jordan, Electrophoresis 21 (2000) 1999.
[15] J.E. Plowman, W.G. Bryson, L.M. Flanagan, T.W. Jordan, Anal. Biochem. 300 (2002) 221.
[16] L.M. Flanagan, J.E. Plowman, W.G. Bryson, Proteomics 2 (2002) 1240.
[17] J.J. Bond, P.C. Wynn, G.P.M. Moore, Acta Derm. Venereol. 78 (1998) 337.
[18] Swiss-Prot Database. http://www.expasy.ch/.
[19] L.G. Sparrow, C.P. Robinson, J. Caine, D.T.W. McMahon, P.M. Strike, Biochem. J. 282 (1992) 291.
[20] L.G. Sparrow, C.P. Robinson, D.T.W. McMahon, M.R. Rubira, Biochem. J. 261 (1989) 1015.
[21] L.M. Dowling, W.G. Crewther, A.S. Inglis, Biochem. J. 236 (1986) 695.
[22] B.W. Wilson, K.J. Edwards, M.J. Sleigh, C.R. Bryne, K.A. Ward, Gene 73 (1988) 21.
[23] J.E. Plowman, W.G. Bryson, Z.-D. Yu, S.W. Gordon, A.J. Pearson, R.J. Kelly, Proc. N.Z. Soc. Anim. Prod. 66 (2006) 133.
[24] L. Langbein, M.A. Rogers, H. Winter, S. Praetzel, U. Beckhaus, H.R. Rackwitz, J. Schweizer, J. Biol. Chem. 274 (1999) 19874.
[25] L. Langbein, M.A. Rogers, H. Winter, S. Praetzel, J. Schweizer, J. Biol. Chem. 276 (2001) 35123.
[26] L.N. Paton, J. Gerard, W.G. Bryson, Unpublished manuscript (2006).
[27] J.E. Plowman, L.M. Flanagan, L.N. Paton, A.C. Fitzgerald, N.I. Joyce, W.G. Bryson, Proteomics 3 (2003) 942.
[28] R.C. Marshall, J. Invest. Dermatol. 80 (1983) 519.
[29] R.C. Marshall, J.M. Gillespie, J. Forensic Sci. Soc. 22 (1980) 377.
[30] K. Katsuumi, T. Ito, Y. Sato, Arch. Dermatol. Res. 281 (1989) 495.
[31] Y.J. Lee, R.H. Rice, Y.M. Lee, Mol. Cell. Proteomics 5 (2006) 789.
[32] A. Nakamura, M. Arimoto, K. Takeuchi, T. Fujii, Biol. Pharm. Bull. 25 (2002) 569.
[33] R. Moll, W.W. Franke, D.L. Schiller, B. Geiger, R. Krepler, Cell 31 (1982) 11.
[34] J. Liao, N.-O. Ku, M.B. Omary, Electrophoresis 17 (1996) 1671.
[35] C.-F. Chou, A.J. Smith, M.B. Omary, J. Biol. Chem. 267 (1992) 3901.
[36] N.-O. Ku, M.B. Omary, Exp. Cell Res. 211 (1994) 24.
[37] H.W. Heid, R. Moll, W.W. Franke, Differentiation 37 (1988) 137.
[38] T.G. Gharib, G. Chen, H. Wang, C.-C. Huang, M.S. Prescott, K. Shedden, D.E. Misek, D.G. Thomas, T.J. Giordano, J.M.G. Taylor, S. Kardia, J. Yee, M.B. Orringer, S. Hanash, D.G. Beer, Neoplasta 4 (2002) 440.
[39] N. Davezac, D. Tondelier, J. Lipecka, P. Fanen, F. Demaugre, J. Debski, M. Dadlez, A. Schrattenholz, M.A. Cahill, A. Edelman, Proteomics 4 (2004) 3833.
[40] M. Schaffeld, J. Markl, Methods Cell Biol. 78 (2004) 627.
[41] S. Pitz, R. Moll, Prog. Retin. Eye Res. 21 (2002) 241.
[42] J. Brooks, C.R. Fleschner, Ophthalmic Res. 35 (2005) 8.
[43] O. Weinreb, F.A. van Rijk, H.T. Steely, A. Dovrat, H. Bloemendal, Ophthalmic Res. 32 (2000) 195.