Characterization of Protease Processing Sites during Conversion of Rat Profilaggrin to Filaggrin†

K. A. Resing,* R. S. Johnson, and K. A. Walsh

Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received April 27, 1993; Revised Manuscript Received July 9, 1993•

ABSTRACT: Profilaggrin is an intermediate filament-associated protein of cornified epithelia. It consists of multiple copies of similar filaggrin domains joined by peptide linker regions; during terminal differentiation of the epidermis, the linker regions are processed away in a regulated manner. In order to characterize the sites of proteolysis in rat profilaggrin, tryptic peptides of filaggrin and profilaggrin were fractionated by reverse-phase HPLC, and the HPLC fractions were analyzed by nebulization-assisted electrospray ionization mass spectrometry. Peptide sequences were confirmed or corrected by tandem mass spectrometry; in several cases, this was achieved by collisional activation of multiply charged precursor ions of peptides exceeding 3 kDa in mass. The tryptic peptides accounted for all of the sequence predicted by a partial cDNA sequence, with the exception of six arginines or dipeptides. Although the cDNA sequence predicted eight sites of heterogeneity among the filaggrin domains, only one of these was observed. An additional unpredicted site of heterogeneity was also seen. Comparison of the peptides from filaggrin with those of profilaggrin revealed several peptides unique to filaggrin, specifically at the new amino- and carboxyl-termini, that result from proteolytic processing of the linker region of profilaggrin. Both the amino- and carboxyl-termini were "ragged", suggesting that processing may involve exopeptidase action after an initial endopeptidase cleavage. The average mass of this mixture of filaggrins was determined by electrospray mass spectrometry to be 42 452 Da, in reasonable agreement with that predicted from the mass spectrometric analysis of the terminal sequences. The linker peptide of rat profilaggrin was found in two forms, which differed only in the phosphorylation state of serine 22.

One of the last events occurring in terminal differentiation of mammalian epidermis is the synthesis of profilaggrin and processing of profilaggrin to filaggrin [for a review, see Resing and Dale (1991)]. Synthesis of profilaggrin involves extensive phosphorylation by several kinases and deposition in the cytosol in a paracrystalline granule (the keratohyalin granule). Profilaggrin has an unusual repeating structure, with each domain of filaggrin joined by a linker peptide that is removed during processing (Mckinley-Grant et al., 1989; Resing et al., 1989). The monomer repeat units of profilaggrin display a high degree of similarity to one another within an animal species (95% identity in mouse, 85% in human), but vary greatly between species (30–42% between mouse and human) (Gan et al., 1990; Haydock & Dale, 1990; Rothnagel & Steinert, 1990). As a cell terminally differentiates, the granules disperse as the profilaggrin is both dephosphorylated and proteolyzed in a multistep process to yield filaggrin (Resing et al., 1989).

Because filaggrin is involved in aggregating intermediate filaments as the cell terminally differentiates, the release of filaggrin must be tightly regulated to avoid premature aggregation. Processing of profilaggrin occurs in the cytoplasm, where regulated proteolysis is not well understood. Proteolytic processing of mouse profilaggrin is the most extensively characterized (Resing et al., 1989), where the first stage of processing at a unique linker yields intermediates with several filaggrin domains. A second stage of processing involves proteolysis at the remaining linkers. The two mouse linker regions differ in that only those processed in the first stage have an FYPVY insert. A chymotrypsin-like protease, "profilaggrin protease 1" (P1 protease), has been identified in mouse epidermis and implicated as the protease in the first stage of the pathway.

Little is known about the specificity determinants of P1 proteases in other organisms. Although there is a high degree of sequence identity between mouse, rat, and human profilaggrin in the regions of the protease processing sites (Haydock & Dale, 1990), the primary sequences at the immediate sites of proteolysis are different. To understand these events, it would be useful to compare the proteolytic target sites of several species; however, the unusual composition and multidomain nature of these proteins have complicated these anaylses.

Recent advances in mass spectrometry have permitted the analysis of nonvolatile molecules of high mass such as peptides and proteins (Barber, 1981; Covey et al., 1988; Fenn et al., 1989; Karas, 1988; Smith et al., 1991; Sundqvist, 1985; Tanaka, 1988). In the present study, nebulization-assisted electrospray ionization (ESI-MS)[1] (Covey et al., 1988) is used to ionize proteins and peptides for mass analysis in a triple-quadrupole mass spectrometer. Sequence information can be obtained for peptides, even in mixtures, by tandem mass spectrometry (MS/MS) (Biemann, 1990a,b; Hunt et al., 1986). This involves selection of a precursor ion of a particular mass to charge ratio (m/z), expressed in daltons per unit

† Supported by NIH Grants AR39730 (K.A.R.), RR05543, and HL40990 (K.A.W.).

* Author to whom correspondence should be addressed. Current address: Department of Chemistry and Biochemistry, Campus Box 215, University of Colorado, Boulder, CO 80309. Telephone: 303-492-4604. FAX: 303-492-3586.

• Abstract published in *Advance ACS Abstracts,* September 1, 1993.

[1] Abbreviations: Da/e, daltons per unit charge; ESI-MS, electrospray ionization mass spectrometry; HPLC, high-pressure liquid chromatography; LC/MS, liquid chromatography coupled to mass spectrometry; MS/MS, tandem mass spectrometry; m/z, mass to charge ratio.

0006-2960/93/0432-10036$04.00/0

charge (Da/e), in the first of three quadrupole mass analyzers (Q1), collisional activation of the selected ion in the second quadrupole (Q2), and mass analysis of the resulting fragments (product ions) in the third quadrupole (Q3). Collisional activation induces cleavages along the peptide backbone to yield a series of product ions from which the sequence can be deduced. The application of these methods to a study of the events of profilaggrin processing is described herein.

METHODS

Profilaggrin was extracted from newborn rat epidermis with 1 M NaSCN, 10 μg/mL PMSF, 0.1 mg/mL aprotinin, 10 mM EDTA, 10 mM glycerol phosphate, and 50 mM Hepes (pH 6.8). Profilaggrin and filaggrin precipitated upon dilution with 9 vol of H_2O at 0 °C. They were redissolved in 9 M urea and 50 mM Tris (pH 7.8) and separated on DE52 in the same buffer, where filaggrin does not bind and gradient elution to 0.4 M NaCl releases profilaggrin. The unbound protein was applied to Sephacryl S300 (1.5 M × 1.6 cm) in 6 M urea and 50 mM phosphoric acid, and the fractions containing filaggrin were dialyzed against 5% formic acid and lyophilized. Profilaggrin from the DE52 fractions was precipitated with 50% MeOH. Both the profilaggrin precipitate and the lyophilized filaggrin (approximately 150 μg each) were suspended in 200 mM Tris and 1 mM $CaCl_2$ (pH 8) and digested with 1% TPCK treated trypsin (Worthington, 2× crystallized) on a weight basis at 38 °C for 2 h. Digests were acidifed with phosphoric acid and analyzed by reverse-phase high-performance liquid chromatography (HPLC) using a Synchropak C18 column with a Beckman Model 332 pump developed with 0.1% TFA and with acetonitrile and with 0.08% TFA. Peptides that did not bind to the Synchropak column were separated on an Ultrasphere C18 column in the same buffers. For N-terminal analysis of filaggrin by gas-phase sequencing, the lyophilized S200 purified filaggrin was resuspended in H_2O, and 50 pmol was analyzed on an Applied Biosystems Model 470 sequencer with on-line HPLC analysis. For mass spectrometric analysis of the whole protein, approximately 200 μg of filaggrin was desalted on a Whatman C18 HPLC column, lyophilized, and dissolved in 0.1% formic acid and 50% methanol.

Mass and sequence determinations were carried out on a triple-quadrupole mass spectrometer (API-III, Sciex, Thornhill, Ontario, Canada) equipped with an ion spray source. Lyophilized HPLC fractions (containing approximately 3 nmol of peptide) were resuspended in 100 μL of 1:1 or 1:3 methanol/water with 0.1% formic acid and continuously infused at a flow of 1.7 μL/min through the ion spray needle, which was held at a potential of 5 kV and surrounded by a concentric flow of nebulizing air at 0.6 L/min. Charged droplets were directed through a curtain of nitrogen (1.2 L/min), which served to desolvate ions entering the mass spectrometer. A mixture of polypropylene glycols was used to calibrate both Q1 and Q3; ions were detected with a continuous dynode electron multiplier operating in a pulse counting mode. Ion spray mass spectra were acquired at unit resolution, whereas for MS/MS the resolution of Q1 allowed a 2-3 Da/e window to be transmitted, and the resolution of Q3 was similarly reduced to increase sensitivity. Precursor ions were accelerated to a kinetic energy of 25 eV and collisionally activated with argon in Q2 at a thickness of 3.5×10^{14} atoms/cm^3. Ion spray mass spectra were typically scanned from 200 to 1500 Da/e at a rate of 26 s/scan; three to five spectra were summed. For on-line mass spectrometric analysis of HPLC eluates (LC/MS), an Applied Biosystems Model 140A HPLC and a Hewlett-Packard 2-mm C_{18} reverse-phase column were used at a flow rate of 200 μL/min. Tryptic digests were acidified with TFA just before application to the column, and 10% of the HPLC effluent was directed to the mass spectrometer; the remainder was directed to an ABI Model 785A UV detector and fractions were collected by hand.

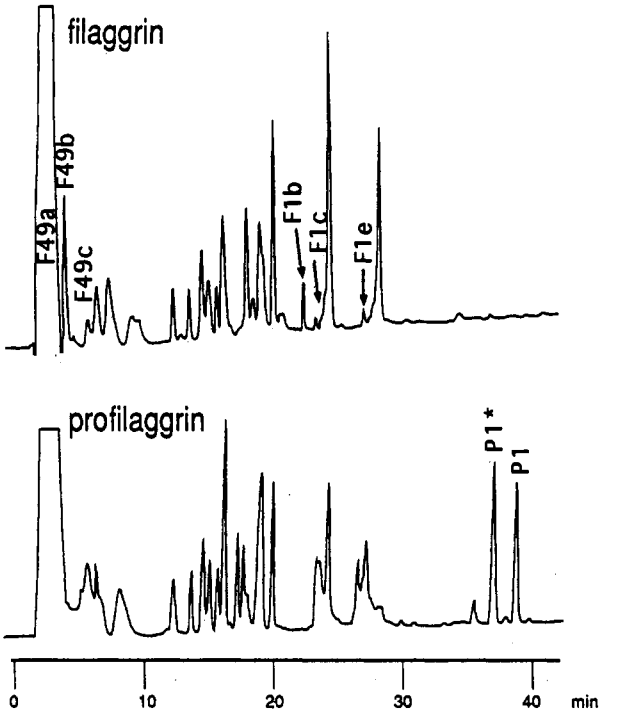

FIGURE 1: Comparison of the HPLC elution patterns (Synchropak C18) of tryptic peptides from rat profilaggrin and filaggrin. Although the peaks seem well resolved, those peaks eluting before 20 min actually contain several peptides. The elution positions of the peptides that are relevant to protease processing are identified by prefixes F and P for filaggrin and profilaggrin, respectively. P1 and P1* denote nonphosphorylated and phosphorylated forms of the linker peptide; F1b,c,e and F49a–c denote variants of the amino- and carboxyl-terminal peptides of filaggrin; F49a was demonstrated only after rechromatography of the breakthrough fraction on Ultrasphere C18. F1a and F1d were not found in this experiment but were observed in a subsequent LC/MS study. The absorbance full scale is 2.0 at 210-nm wavelength.

Data collection and reduction were carried out on a Macintosh IIfx computer using computer programs supplied by Sciex for that purpose. The masses of tryptic peptides and their expected product ions following collisional activation and dissociation were calculated from the published sequence using the program MacProMass (Lee & Vemuri, 1991). Interpretation of the MS/MS spectra was facilitated by two computer programs written in-house. The program PEPID searches a known protein sequence (in this case a translated cDNA sequence) for all hypothetical peptides of a specified mass and compares the expected product ion masses of each of these sequences with the MS/MS data. The other program, LUTEFISK, generates a list of possible peptide sequences from the product ion masses generated by collisional activation (MS/MS), the mass of the peptide, and the precursor ion charge (Johnson et al., 1991). The possible sequences are then ranked according to the percentage of product ion current (ion abundance) that each sequence can account for, on the basis of the known low-energy collision-induced dissociations of multiply charged peptide precursor ions (Burlet et al., 1992; Tang & Boyd, 1992). Product ions of signal to noise ratio greater than 5 were used to generate the input file for LUTEFISK.

Table I: Analysis of Tryptic Peptides

peptide no.[a]	predicted mass[f] (Da)	observed mass (Da)	sequence score[g] (%)
P1[b]	4159.2	4157.7	96
F1a[a]	2407.4	2407.4	
F1b	2536.2	2536.4	85
F1c	2635.4	2635.3	94
F1d[c]	2692.7	2693.3	
F1e	2855.9	2857.0	93
F2	558.3	558.5	87
F3, F10	302.2	302.1	90
F4	962.4	962.6	90
F5	1499.5	1499.6	89
F6	718.3	718.5	88
F7	412.2	412.1	75
F8[d]	316.2	316.2	100
F9	1087.5	1087.5	91
F11	1036.5	1037.0	90
F12	845.4	845.5	76
F13, F18, F30, F34[d]	231.1	231.1	95
F14	1199.6	1200.0	97
F15	445.2	445.1	93
F16	1015.4	1015.6	93
F17	691.4	691.4	85
F19	700.3	700.4	91
F20	626.3	626.3	86
F21	855.4	855.4	94
F22	620.3	620.4	86
F23	1729.7	1729.3	89
F24[d]	575.3	575.4	93
F25	1537.7	1538.0	84
F26	506.3	506.5	90
F27	773.4	773.7	80
F28	1027.5	1027.6	90
F29	419.2	419.2	88
F31	1227.6	1227.8	99
F32	245.2	245.1	97
F33	640.3	640.3	88
F35	275.2	275.2	98
F36[b]	2938.9	2938.8	88
F37	835.4	835.0	88
F38	1177.5	1177.6	95
F39	1189.6	1189.9	93
F40	389.2	389.3	92
F41	261.2	261.1	100
F42	743.4	743.5	96
F43	1023.5	1023.6	96
F44	1794.9	1794.7	96
F45	427.3	427.2	77
F46	3389.4	3389.6	90
F47[e]	568.4	568.2	87
F48[e]	358.2	358.2	93
F49a	532.2	532.3	87
F49b	669.3	669.3	93
F49c	766.3	766.3	87

[a] Tryptic peptides from amino (F1) to carboxyl (F49) termini of filaggrin (cf. Figure 3). a–e refer to variants in order of elution from HPLC. P1 of profilaggrin encompasses F1 + F49 + linker segments (cf. Figure 6). [b] Sequence confirmed by MS/MS analysis of subdigested peptides. [c] Seen in LC/MS, no MS/MS data. [d] A peptide of this sequence with Arg at the amino-terminus was also found. [e] Minor cleavage between F47 and F48 occurs at an RP bond; the overlap peptide, F47/48, was also seen as a major ion (obs m = 908.4, pred m = 908.6). [f] MW was calculated using monoisotopic masses for peptides less than 1200 Da and average masses for larger peptides. [g] Sequence scores are the percent of the total product ion current that can be accounted for according to the ion types shown in Figure 2.

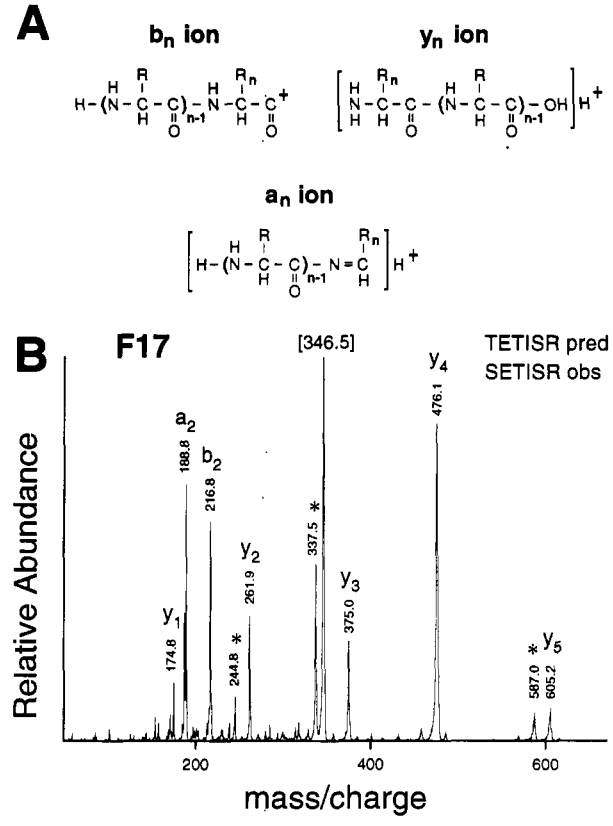

FIGURE 2: (A) Nomenclature of the principal fragment ions produced by low-energy collisional activation of peptides (Biemann, 1990a,b). The most common fragmentation occurs at the peptide bond, resulting in b and y ions; a-type ions are derived from b ions by the loss of CO. The mass of each singly charged fragment ion is the sum of its residue masses plus 19 for y ions, plus 1 for b ions, or minus 27 for a ions. In addition, each of these ions can lose water or ammonia. Internal ions derived from secondary fragmentations are often observed. A subscripted number refers to the number of residues from the amino-terminus (a or b ion) or from the carboxyl-terminus (y ion). Multiply charged fragment ions result from the protonation of amino acid side chains. (B) MS/MS spectrum of the filaggrin peptide F17 (the predicted and observed sequences are shown). The y axis represents the mass to charge ratio with units Da/e. The doubly charged precursor ion (346.5 Da/e) was selected for collisional activation. The observed masses of the a, b, and y fragment ions and their ion designations are shown. Ions indicated by an * represent secondary loss of water (18 from y_5 or 9 from the precursor) or ammonia (17 from y_2).

RESULTS

Identification and Characterization of Peptides from Rat Filaggrin and Profilaggrin. Tryptic peptides of filaggrin and profilaggrin were partially resolved by reverse-phase HPLC (Figure 1). There were several peptides in each peak; furthermore, several *pro*filaggrin peptides are phosphorylated, whereas their filaggrin counterparts are not (Resing et al., 1985). To clarify the comparison, the masses of the components in each HPLC fraction were determined by ESI-MS. Table I summarizes the filaggrin peptides identified in a single digest of 4 nmol, with peptides numbered by their position from the amino-terminus of filaggrin. Each peptide is also given a letter designation indicating its source, P for profilaggrin and F for filaggrin, so that, for example, F2 corresponds to P2. A study of the phosphorylation sites of profilaggrin will be presented in a separate communication (K. A. Resing, R. S. Johnson, and K. A. Walsh, manuscript in preparation). With the exception of P1, F1, and F49, the peptides in Table I were identified in both profilaggrin and filaggrin, although the data are shown only for filaggrin. Comparison with the cDNA-derived sequence of rat profilaggrin revealed that F1 and F49 are contained within the linker regions of profilaggrin (P1), suggesting that F1 and F49 are the N- and C-termini of filaggrin.

In most cases, the observed masses of the tryptic peptides of filaggrin matched those predicted from the cDNA sequence (Haydock & Dale, 1990), and these identifications were confirmed by tandem mass spectrometry (cf., Figure 2). Fragmentation of multiply charged tryptic peptides in tandem

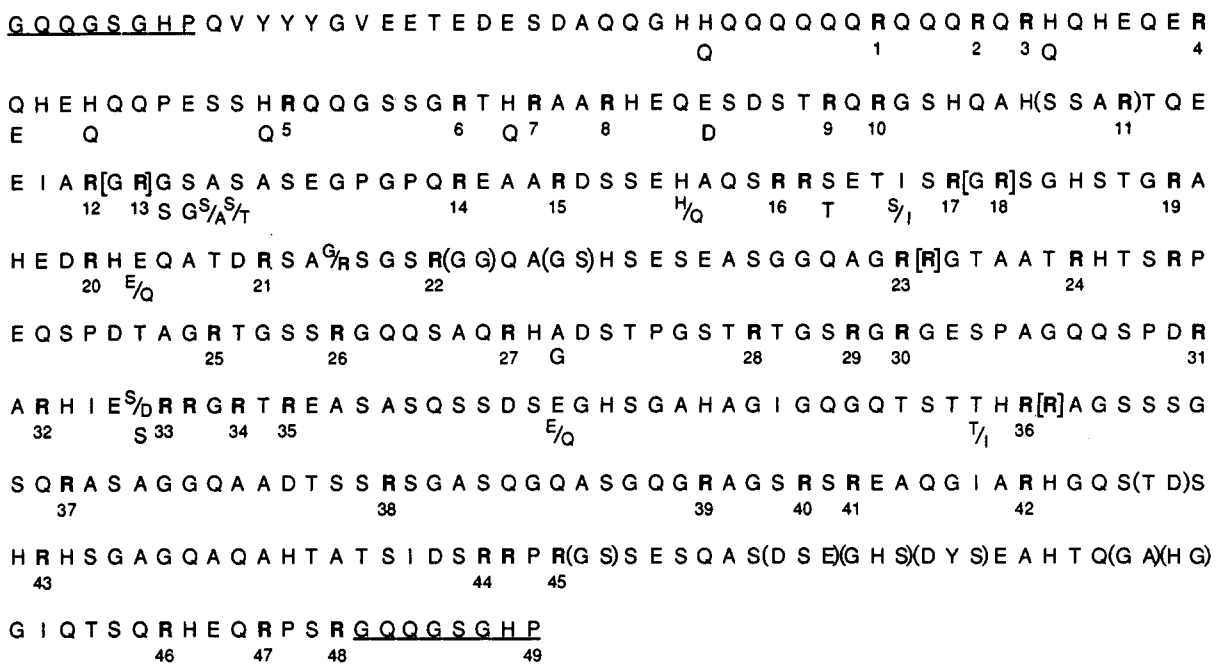

FIGURE 3: Comparison of the sequence predicted by the cDNA of rat profilaggrin (Haydock & Dale, 1990) with those observed in tryptic peptides by tandem mass spectrometry. One repeat unit of profilaggrin is represented, beginning with the linker peptide. The carboxyl-terminal region includes eight residues of overlap into the next linker peptide (underlined). Because the reported cDNA data include two filaggrin domains with occasional differences in sequence, both residues are shown. Hypothetical tryptic peptides are numbered by the arginines (there are no lysines), where the first peptide in the filaggrin domain is numbered 1, the second is 2, and so on; however, only the first arginine in an Arg-Arg sequence is numbered. All tryptic peptides predicted from the cDNA were found by mass spectrometry. The sequences in brackets (GR and R) are predicted to be at more than one location within the sequence and could not be identified unambiguously. Sequence differences from the cDNA were noted at 19 loci (in these cases, the cDNA assignments are listed *beneath* the continuous sequence). The sequences in parentheses denote regions where sequence-specific fragments were not seen in the MS/MS spectra.

MS/MS experiments is known to produce primarily y ions (Tang & Boyd, 1992), with occasional a, b, and internal product ions from multiple fragmentation events (the nomenclature for MS/MS spectra is described in Figure 2). Generally, either a complete set of y ions or an overlapping set of b and y ions was identified. Two peptides (P1 and F36) required subdigestion to establish their sequences by MS/MS. The computer program LUTEFISK was used to assess the reliability of the predicted sequences by comparing the MS/MS data with the predicted product ions. Generally the predicted sequences account for 75–100% of the ion current (Table I). The unaccounted ion current may represent anomalous, sequence-specific cleavages or heterogeneity in the peptide that does not alter the mass or the HPLC retention time of the parent ion (for example, the presence of both GS and SG in the population of repeating domains).

In some cases, the MS/MS data did not allow unambiguous identification of the sequence (residues contained within parentheses in Figure 3). The two amino acids at the N-termini of peptides F14, F23, F27, F36, and F46 were ambiguous, because b_1 ions generally are not seen and these peptides did not yield a complete set of y ions. F14 was purified and its N-terminus determined by Edman degradation. The N-terminal GQ of F27 was identified by the observation of the c_1 ion, which is often observed when glutamine is the second residue (K. A. Resing and R. S. Johnson, unpublished results). Subdigestion of F36 produced smaller peptides which gave unambiguous MS/MS data for the N-terminus. F23 and F46 were recovered in low yields and could not be purified sufficiently for gas-phase sequence analysis. F25, F31, and F43 each had an *internal* two-residue sequence that was not unambiguously identified by MS/MS. The sequence of F31 was confirmed by gas-phase sequence analysis. The RP sequence in F25 was supported by the fact that the arginine was not cleaved by trypsin. Cleavage at the TD in F43 would produce a y ion that would be obscured by the precursor ion; however, no fragment ions indicating an alternative sequence at this site were observed. In all cases, the masses of the unsequenced pairs of amino acids of these three peptides correspond to those predicted by the cDNA. However, it should be noted that peptide F14 was predicted from the cDNA sequence to have an N-terminal sequence of SG, but was shown by gas-phase sequencing to be GS.

Although most of the cDNA-derived sequence of a profilaggrin repeat was confirmed or corrected *within* the tryptic peptides (Table I), the alignment of these peptides relied on the cDNA sequence (Figure 3), with the exception of a few cases where incomplete proteolytic products were observed. The presence of five RR, four RGR, and two RQR sequences in the predicted filaggrin sequence allows for some ambiguity, even though peptides with sequences R, GR, and QR were identified. Nevertheless, partial proteolytic products were identified that confirmed all except F13, F18, and arginines N-terminal to F37 and F24.

The partial cDNA clone of multidomain profilaggrin encodes the sequence of 1.5 monomer repeat units of profilaggrin (beginning with F14 and ending with F37). Of the 48 tryptic peptides predicted from this cDNA clone (excluding the RR sequences), 36 displayed both the predicted masses and MS/MS spectra that matched the predicted sequence. The remaining peptides fell into three groups: (1) peptides where heterogeneity was predicted by the cDNA sequence, but only one of the two possible sequences was observed; (2) peptides which differed in one or two residues from the sequence predicted by the clone (data for these is available on microfiche); and (3) peptides found as sets of truncated variants, derived from the amino- and carboxyl-termini of filaggrin (discussed in the next section).

Heterogeneity was predicted at eight loci by the cDNA sequence (Figure 3). Of these, F16 and F21 were seen only

as DSSEHAQSR and HEQATDR, respectively; no peptides of the mass of the other form were detected, although the resolution of the instrument was not sufficient to identify minor signals of the Gln variant of F21 (Gln and Glu differ by 1 Da and by 0.5 Da/e for doubly charged ions). In contrast, F22 was seen both as SAGSGSR and as two peptides resulting from replacement of the underlined glycine by arginine. There appeared to be approximately equal yields of the two forms.

Three other peptides (F14, F17, and F36) displayed sequences that were different from both variants predicted from the cDNA. Only one sequence was found for F14, where the MS/MS spectrum exhibited a continuous series of y_n ions in support of the partial sequence SASEGPGPQR. In addition, a pair of ions taken to be b_2 and b_3 suggested that F14 had the sequence (G,S)ASASEGPGPQR, which was a hybrid of the two sequences expected from the cDNA. The amino-terminus of F14 was determined by Edman degradation to be GSASA. Thus, the first four residues of F14 differed from both cDNA sequences.

In the case of F17, there were two predicted sequences, TETISR and TETSSR, but we observed only the sequence SETISR (Figure 2B) and no ions corresponded in mass to the two predicted. A peptide, RSETISR, was also detected, as expected, because trypsin, after cleavage between two arginines, usually does not then remove the N-terminal arginine. A peptide was identified with a mass corresponding to that of F36 with glutamate (not Gln) at residue 11 and threonine (not Ile) at residue 28 (a hybrid of two predicted sequences); however, the MS/MS spectrum of this 30-residue peptide was ambiguous in the regions of the proposed changes. The peptide was subdigested with thermolysin and the products were analyzed by MS/MS, where the ambiguous regions were confirmed, although a small amount of Q at residue 11 could have been overlooked as mentioned in the case of F21 (data not shown).

The possibility was considered that further heterogeneity not predicted by the cDNA sequence might be present. Only one such clear instance was detected—a minor variant of F33 in which Asp replaced a Ser at the 5% level, as judged by the ion current. This is consistent with one of the 22 repeats within profilaggrin having Asp in place of Ser. A rigorous search for other heterogeneity was not made; however, every peptide which gave a signal greater than 5% of the most abundant ion was identified. Because of the small difference in m/z between amides and acids when multiply charged, we cannot preclude minor heterogeneity from such interchanges. As isoleucine and leucine have the same mass, interchange of those residues is not detected; amino acid analysis of filaggrin showed no detectable leucine, but replacement of 5% of the isoleucine with leucine cannot be ruled out.

The replacement by arginine of other amino acids or vice versa presents a special problem in tryptic digests, because such changes create two new smaller peptides or a new larger peptide rather than a small mass shift of a single peptide in other replacements. The G/R heterogeneity in F22 was clearly documented when both the smaller peptides (SAR and SGSR) and the larger combined peptide (SAGSGSR) were identified and sequenced. F36 may have such replacement sites, because minor peptides were identified with the sequences AGIGQGQTSTTHR, SGAHAGIGQGQTSTTHR, and SSDSEGHSGAHAGIGQGQTSTTHR, suggesting that, in some repeat units, F36 might have an extra Arg in place of Gln or His residues 6, 13, or 17. F44 also exhibited such heterogeneity because peptides of sequences SIDSR and TATSIDSR were identified. As in the case of F36, they seemed minor because

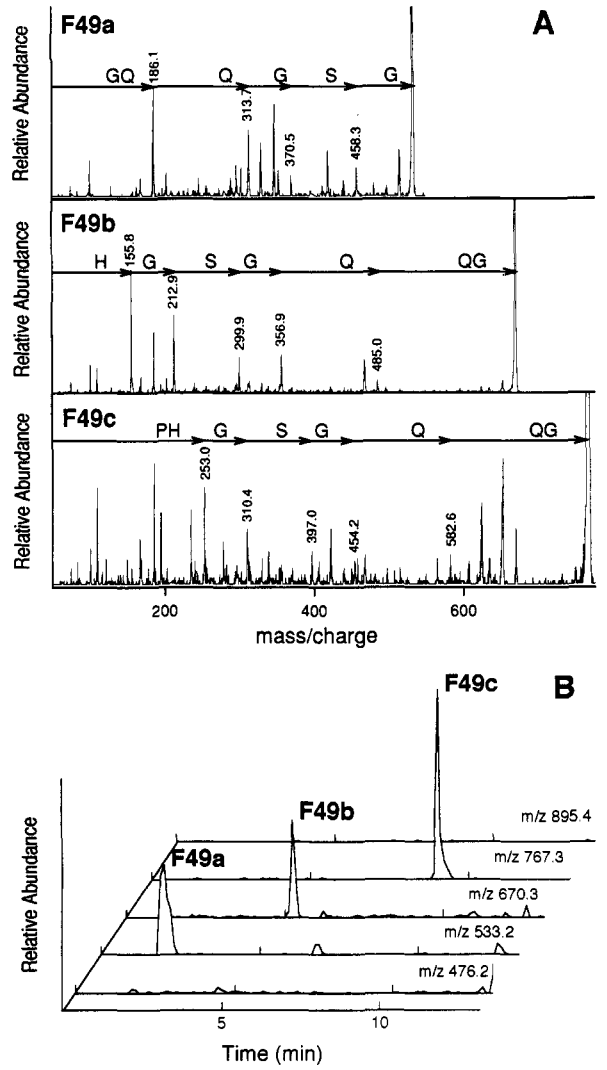

FIGURE 4: Identification of carboxyl-terminal peptides of filaggrin. (A) MS/MS spectra of peptides F49a, F49b, and F49c. The mass increments between the y ions (indicated by arrows) correspond to amino acid residue masses, which are denoted by a single-letter code above each arrow. The sequence is deduced largely from the singly charged y ions for F49b and F49c (arrows indicate the sequence from the C- to N-termini) and from the singly charged b ions for F49a (arrows in the opposite direction). (B) Ion chromatograms for MH^+ ions corresponding to GQGQS (m/z 476.2), GQGQSG (F49a; m/z 533.2), GQGQSGH (F49b; m/z 670.3), GQGQSGHP (F49c; m/z 767.3), and GQGQSGHPQ (m/z 895.4).

these peptides resulted in ions of low abundance. The possibility of nonspecific proteolytic cleavages cannot be ruled out, as the corresponding peptides at the N-terminal side of the cleavage were not identified.

A second group of peptides showed one or two residues different from those predicted by the cDNA. These included F4, F5, F7, F9, F14 (discussed above), F17, and F28. The various forms of F1 also differed at one residue from the predicted sequence, and they are discussed in the next section. In most cases, the MS/MS spectra of these peptides provided unambiguous sequence information covering the residues in question, some of which are shown in Figure 4. In the case of F5, the y product ions were consistent with histidine instead of two of the glutamines predicted by the cDNA sequence, but the predicted mass was still 1 Da too small. The b product ions were also 1 Da too small, suggesting that the amino-terminus is most likely glutamine, not glutamate; as already discussed for F21 and F36, it is not possible to unambiguously

assert that there was not some glutamate at that position as well.

In summary, of the 406 residues in each repeat, only six residues (1.4%) lacked information from which the predicted sequence (RR and RGR) could be checked; although free R and GR were identified in the Ultrasphere fractions, these peptides were present several times in the complete sequence, preventing unambiguous assignment. Other short sequences encompassing a total of 25 residues (6.2%) did not fragment well and were confirmed only in total mass. Thus, 92% of the predicted sequence was unambiguously verified or corrected in this study. Of the residue differences observed, the following interchanges were observed: five Q/H, one E/D, two S/G, one T/S, one G/R, and one S/D. If the fragments of F36 and F44 are due to (unobserved) variants containing arginine, then there are, in addition, three H/R, one T/R, and one Q/R interchanges, as well as the R/G and S/D heterogeneity in F22 and F17. Most of the differences could be accounted for by single base pair changes in the nucleic acid sequence.

Amino- and Carboxyl-Termini of Filaggrin. The native cleavage sites involved in the processing of profilaggrin to yield filaggrin were explored by comparing the tryptic peptides of profilaggrin with those of filaggrin. Seven peptides unique to filaggrin were identified in the original Synchropak-analyzed tryptic digest. Three of these, with masses of 532.2, 669.3, and 766.3, were found from their MS/MS spectra to have the sequences GQQGSG (F49a), GQQGSGH (F49b), and GQQGSGHP (F49c) (Figure 4A). These sequences could be aligned on the P1 sequence (Figure 6B) and appear to represent three alternative C-terminal peptides of filaggrin. To determine the relative amounts of these three peptides, a sample of the entire filaggrin tryptic digest was injected onto a reverse-phase HPLC column, and a portion of the effluent was directed on-line to the mass spectrometer (LC/MS). Ion-specific chromatograms were extracted from the LC/MS data set by plotting the signal intensity corresponding to a selected mass (Figure 4B). F49a, F49b, and F49c were observed at 25%, 17%, and 58% of the total ion signal for these three peptides, respectively. Although the mass spectrometric response for two completely different peptides cannot be used to quantify their amounts relative to each other (some peptides may ionize more readily than others), the data may serve as a semiquantitative guide for the relative yields of peptides that are nearly identical, in this case differing by one amino acid. Shorter or longer variants of the C-terminus, such as GQQGS or GQQGSGHPQ, were not detected, although examination of Figure 4B suggests that the putative GQQGS peptide would probably elute with the salt from this column and might be difficult to detect. However, no peptides of that sequence or any shorter variants were observed in the Ultrasphere fractions, which should have included shorter variants of F49.

Three additional peptides unique to filaggrin had masses of 2857.0, 2635.3, and 2536.4 Da. Comparison of the MS/MS spectra of their triply charged precursor ions (Figure 5A) identified their N-termini as YGVEETE..., VEETE..., and EETE.... However, the masses of these peptides did not agree with the masses of truncated forms of P1 as predicted by the cDNA; MS/MS analyses revealed that a cDNA-predicted glutamine in P1 is replaced by histidine, as discussed below. Therefore, these three sequences correspond to the region following Tyr-Tyr-Tyr in profilaggrin (residues 11–13 in Figure 3), which is part of the region linking the filaggrin repeats and excised during processing (Haydock & Dale, 1990; Resing et al., 1989). Thus, these three peptides are likely

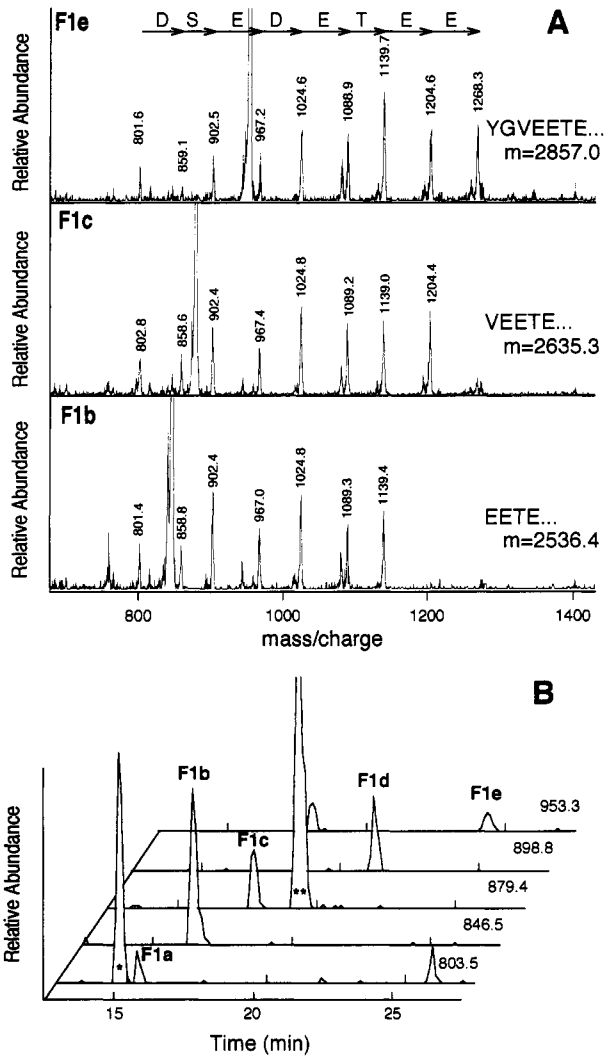

FIGURE 5: Identification and quantitation of amino-terminal peptides of filaggrin. (A) MS/MS spectra of the triply charged ions (large unlabeled peaks) of peptides found in filaggrin (Figure 2, peptides F1b,c,e), but not in profilaggrin. Doubly charged y ions are the major (labeled) product ions. The sequence of F1c was confirmed by gas-phase sequencing; each peptide is an N-terminally truncated form of the sequence YYGVEETEDESDA—R (peptide 1 in Figure 3). (B) Ion chromatograms of the various MH_3^{3+} ions of the amino-terminal peptides as they elute from an HPLC chromatograph (the m/z of the specific ion analyzed is indicated). The LC/MS data were searched for all possible charged species and truncated forms of the linker peptide, and only the indicated peptides were found (the selected ion m/z is given at the right of each chromatogram). The peaks indicated with * and ** are the MH^+ of F12 and the MH_2^{2+} of F44, respectively (cf. Table I), which coincide in mass/charge ratio with certain triply charged amino-terminal peptides.

candidates for a "ragged" N-terminus of filaggrin (F1e, F1c, and F1b in Figure 6B). Verification of this hypothesis was obtained by Edman degradation of intact filaggrin, which yielded 37, 51, and 18 pmol of Y, V, and E, respectively, in the first cycle as well as 32 pmol of G, suggesting that a GVEETE... form had been overlooked in the peptides analyzed by MS/MS. A detailed search of the LC/MS data revealed a peptide (F1d) with the mass expected for GVEETE..., as well as a minor peptide (F1a) of mass corresponding to that of ETE... (Figure 5B). Because the Edman degradation yields are not a reliable measure of the relative amounts of the various N-termini, the specific ion chromatograms were also examined (Figure 5B). Based on their ion abundances, the peptides ETE..., EETE..., VEETE..., GVEETE..., and YGVEETE... were recovered in yields of 4%, 41%, 23%, 28%, and 3%, respectively. Carbamylated GVEETE... was also recovered

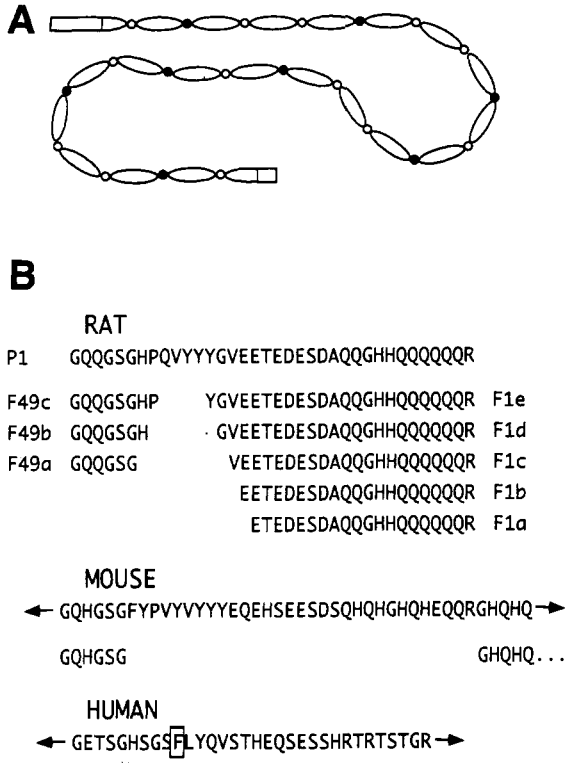

FIGURE 6: (A) Diagram of profilaggrin structure [reviewed in Resing and Dale (1991)] showing the individual filaggrin domains (ovals), connected by two types of linker regions (filled or open circles). Cleavage of the filled circles will yield intermediates with two or three filaggrin domains that are still connected by linkers. The N- and C-termini of profilaggrin consist of regions that are different from the filaggrin domains and are represented by boxes. (B) Summary of the linker regions of profilaggrin and the heterogeneous processing of rat profilaggrin. The region from the bold line box in panel A is represented; the sequence shown is that corrected in this study. The amino- and carboxyl-termini of filaggrin are shown below the linker region sequence of rat profilaggrin. One of two reported mouse linker regions is shown; below it are the putative amino- and carboxyl-termini of mouse filaggrin determined previously (Resing et al., 1989). Gan et al. (1990) indicated that the (boxed) phenylalanine is removed during the processing of human profilaggrin, but the amino- and carboxyl-terminal sequences of human filaggrin have not been determined.

at 2% yield; carbamylated forms of the other peptides were not observed.

Characterization of the Profilaggrin Peptide Encompassing the Processing Sites. The unique sequences identified could be aligned with the P1 peptide in a manner consistent with the removal of a short linker segment (Figure 6B). Because of the importance of this peptide in proteolytic processing, the peptide was purified and characterized more extensively. As mentioned above, only profilaggrin released the tryptic peptides P1 and P1* (Figure 1), both of which contained the Tyr-Tyr-Tyr linker region of profilaggrin. The earlier eluting peak (P1*) had a mass of 4237.4 Da, and the later eluting peak had a mass of 4157.7 Da (Figure 7A); the mass difference of 80 suggested that these peptides have the same sequence, but differ by a single phosphate group. This was confirmed by MS/MS, and the location of the phosphorylation site was deduced to be the serine at residue 22 of this peptide (Figure 7B). The upper panel of Figure 7B shows a region of the MS/MS spectrum of the MH_4^{4+} form of the unphosphorylated peptide, where doubly charged y ions delineate the predicted sequence, EETEDESD (14–23 of the P1 peptide). The lower panel shows the corresponding region of the MS/MS spectrum of the phosphorylated peptide. Whereas both peptides exhibit ions at 801 and 859 Da/e, corresponding to doubly charged

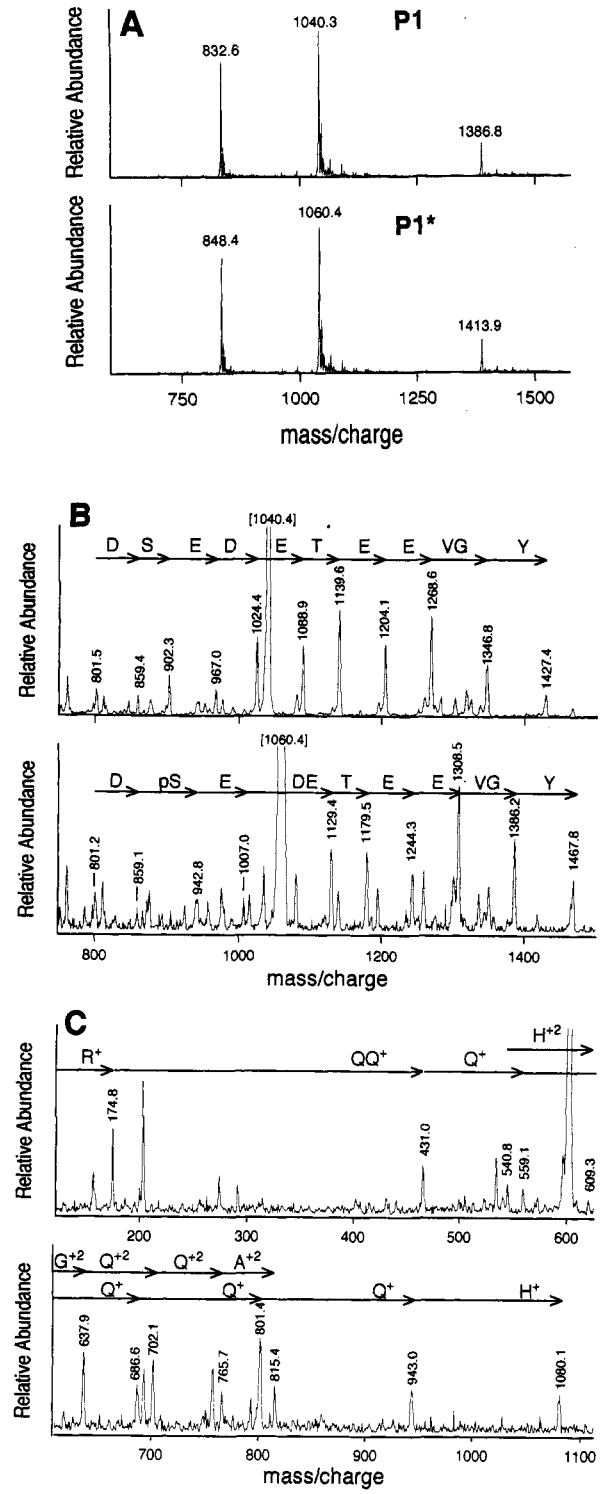

FIGURE 7: Identification and characterization of the tryptic peptide "P1" from profilaggrin, which encompasses the linker region. (A) The two peptides labeled P1 and P1* in Figure 2 were each observed as three major ions, MH_5^{5+}, MH_4^{4+}, and MH_3^{3+}, as well as in association with Na^+. The observed molecular weights of the two peptides, 4157.2 and 4237.6 Da, differed by 80 Da, suggesting that they could represent identical sequences if P1* was phosphorylated. (B) Segments of the MS/MS spectrum of the MH_4^{4+} ions of P1 and P1*, showing the major doubly charged y ions that identify the location of a phosphoserine residue at Ser_{22} in the sequence ESD (Figure 3). (C) MS/MS spectrum of a subpeptide (generated with *S. aureus* V8 protease) from unphosphorylated P1, indicating the sequence SDAQQGHHQQQQQQR and identifying a histidine (underlined) instead of a glutamate predicted by the cDNA at residue 29 in Figure 3. The two sets of arrows indicate the sequence assignments for the singly charged y ions (lower set of arrows) and the doubly charged y ions (upper set).

y_{13} and y_{14} ions, the doubly charged y_{15}–y_{21} ions are each heavier by approximately 40 Da/e in the phosphorylated peptide. Furthermore, none of the mass to charge ratios for y_{15}–y_{21} in the MS/MS spectrum of the unphosphorylated peptide are found in the MS/MS spectrum of the phosphorylated peptide. Thus, the site of phosphorylation was identified as Ser_{22} in the sequence ESD.

The mass of P1 was 9 Da higher than that predicted from the cDNA sequence, suggesting that a histidine might replace one of the 12 glutamine residues. The MS/MS data in Figure 7B (nonphosphorylated) demonstrate that this extra mass is in the C-terminal 13 amino acids of the peptide; however, the quality of the MS/MS data for this large peptide was not sufficient to determine this C-terminal sequence unambiguously. Hence, the unphosphorylated peptide was subdigested with *S. aureus* V8 protease to produce three peptides. Two, of mass 1897.8 (GQQGSGHPQVYYYGVEE) and 492.2 (TEDE), were as predicted by the clone and confirmed by MS/MS. A third peptide of mass 1803.8 was formed that corresponded to the remaining mass of P1, which contained the putative Gln to His replacement. The MS/MS spectrum of the triply charged ion provided sufficient data (Figure 7C) to identify position 29 of P1 as histidine.

Heterogeneity of Filaggrin Reflects the Ragged N- and C-Termini. The electrospray mass spectrum of the filaggrin population (Figure 8A) was difficult to interpret, because the observed peaks were unusually broad. The data were deconvoluted so as to display a single peak with an apex at 42 454 Da (Figure 8B) and a width of approximately 850–900 Da. The peak broadening expected from the natural isotopic distribution should be about 20 Da. Additional peak broadening due to metal ion adduction is often observed in ESI-MS as peak tailing toward higher mass to charge ratios. The peaks shown in Figure 8A, however, are uniformly broad. It is therefore reasonable to conclude that much of the peak width reflects the variation in the sizes of the filaggrins, rather than metal addition. The observed variations in the termini of filaggrin (F49 and F1 peptides) are aligned in Figure 6B below P1, the linker region from which they are excised; the excised sequences have masses ranging from 390 for QVY to 1235 Da for HPQVYYYGVE, so that a range of heterogeneity of 845 Da would be consistent with the ragged amino- and carboxyl-termini. In addition, the observed heterogeneity among the tryptic peptides would add slightly to the peak width. Thus, the experimental observation of 850–900 Da of peak broadening is in reasonable agreement with the "ragged ends" of filaggrin summarized in Figure 6B. The theoretical population distribution was calculated from the yields of the various N- and C-termini found in the LC/MS. The resulting *weighted* average mass was 42 430 Da, which compares well with the observed value, 42 454 Da.

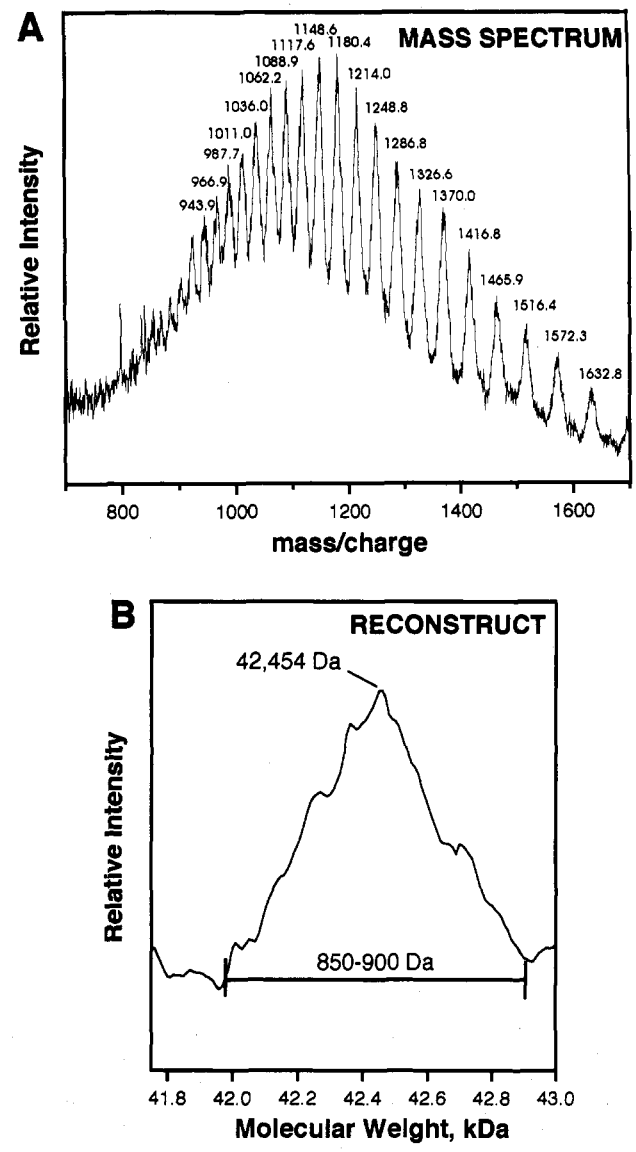

FIGURE 8: Determination of the average molecular weight of rat filaggrin, (A) ESI mass spectrum of filaggrin. The apexes of the multiply charged peaks corresponded to a molecular weight of 42 454 Da. However, each peak is broader than would be expected from the normal isotopic distribution; hence, heterogeneity is indicated. (B) Computer-reconstructed profile of the filaggrin population from the data of A. The width of apparent mass distribution is 850–900 Da, whereas isotopic variation predicts a spread of ca. 28 Da. The difference, 820–870 Da, can be accounted for by the sequence heterogeneity deduced in Figures 4 and 5 and summarized in Figure 6.

DISCUSSION

Conventional techniques of protein chemistry were inadequate for analyzing the similar structural domains of filaggrin within profilaggrin and assessing the complexity of filaggrin processing; the wider range of detection capability of the mass spectrometer and the ability to analyze impure peptides provide an invaluable tool. In this study, mass spectrometric analysis of proteolytically generated peptides of rat filaggrin and profilaggrin extends the structural hypothesis deduced earlier from analysis of tryptic peptides of mouse profilaggrin, both with regard to the boundaries of processed filaggrin and with regard to the conservation of sequence among the tandem domains of profilaggrin. These data provide sequence information for 98.6% of the amino acid residues in the sequence predicted by the cDNA clone (Figure 3). In a few cases (6.2% of the amino acid residues), there were dipeptide or tripeptide sequences that did not fragment following collisional activation. The remaining 1.4% of the residues represented short (R or GR) peptides which were repeated in the sequence, so that unambiguous data were not possible. Thus, 92% of the predicted sequence was unambiguously verified or corrected in this study, demonstrating the power of this technique.

Because of the unusual polymeric nature of profilaggrin, it would not be surprising to see heterogeneity and differences from the sequence predicted by the cDNA, as only 1.5 of the monomer repeat units (out of 22 present in each profilaggrin molecule in this strain of rats) were sequenced in the earlier study of the cDNA clone (Haydock & Dale, 1990). Surprisingly, only one of the eight variants predicted by the cDNA

(in peptide F22) and one additional minor variant (F33) were documented. Although the data illustrate the major components clearly, most minor sequences should also be seen, even at the 5% (single domain) level. Because we could not isolate individual repeat units, it was not possible to determine whether there were errors in the particular carboxyl-terminal monomer repeat unit that was characterized in the cDNA study.

In addition to confirming (and making minor corrections in) the sequence predicted by the cDNA clone, the putative amino- and carboxyl-terminal peptides of filaggrin were identified, and filaggrin was shown to be "ragged" on both the N- and C-termini. The ragged N-terminus was confirmed by gas-phase sequencing of the intact protein. Consistent with this, an ESI-MS spectrum showed that filaggrin consisted of a heterogeneous population ranging in size from 42 100 to 42 800 Da (Figure 8B), with the most common mass in the filaggrin population being 42 452 Da. As the predicted mass of one monomer repeat unit of profilaggrin is 43 225 Da, this population distribution appears to reflect the removal of a linker segment of a size varying from 575 to 1125 Da. Examination of the sequences of the amino- and carboxyl-termini identified in filaggrin indicates that these sequences would be produced by the removal of residues whose mass would range from 409 to 1255 Da, similar to the extent of heterogeneity observed in the filaggrin population. If we assume that the recovery of the ions from the LC/MS experiment represents the distribution of the amino- and carboxyl-termini in the filaggrin population, the predicted weighted average mass of the filaggrin population is 42 430 Da. The slightly larger observed mass may be due to sodium binding, to incomplete dephosphorylation (small amounts of phosphorylated peptides were observed), and/or to minor heterogeneity that was undetected.

The proteolytic processing sites found in this study of rat filaggrin differ significantly from those seen in earlier studies of mouse filaggrin (Figure 7). First of all, the N-terminus was not blocked. Previously, the amino-terminal sequence was not observed in rat filaggrin (Lonsdale-Eccles et al., 1982), possibly due to carbamylation resulting from extensive exposure to urea. An amino-terminal peptide apparently blocked by a carbamoyl group was identified in this study, probably derived from urea during purification. The N-terminal peptide of mouse filaggrin could not be sequenced, suggesting that it was blocked even though not exposed to urea (Resing et al., 1985). Studies are in progress to clarify the nature of the N-terminus of mouse filaggrin. A second major difference was in the extent of the linker region, in that 3361 or 3811 Da is removed from mouse compared to 1255 Da from rat. The simplest explanation for this lies in the hypothesis that there is an initial endoproteinase cleavage followed by aminopeptidase processing to produce the N-terminus of filaggrin. The extent of activity of the aminopeptidase may vary in different species or under different conditions. For instance, in cultured rat keratinocytes, the size of filaggrin is up to 2000 Da smaller than that seen in epidermis (Resing et al., 1993), suggesting that aminopeptidase processing may be more extensive in culture. A third difference was that ragged carboxyl-termini were not observed in mouse filaggrin, where a single peptide corresponding to the sequence of F49a was observed (Resing et al., 1985). Ragged C-termini are difficult to detect by the methods used in the mouse study, although the C-terminal peptide of mouse filaggrin was recovered in high yield, suggesting that other C-termini would have been minor.

Previous studies on in vivo processing (Resing et al., 1989) demonstrated that intermediates are produced in a two-stage process in mouse epidermis and rat keratinocytes. A structural basis for the two stages was proposed for the mouse system, and two linker regions with different primary sequences were identified, where only one was cleaved in the first stage of processing. In view of the two-stage processing of rat filaggrin seen in cultured keratinocytes (Resing et al., 1993), it was surprising that all of the linker peptides recovered from the rat protein were derived from a single primary sequence near the cleavage site. A similar situation is seen in human profilaggrin, although there is significant heterogeneity in the region just amino-terminal to the predicted cleavage site (Gan et al., 1990). In order to generate the intermediates, an alternative linker sequence should be present at 30–50% frequency (assuming that only one is cleaved in the first stage); it is unlikely that such an alternative linker peptide could have been missed in this study. The more likely explanation is that unphosphorylated and phosphorylated linker peptides differ in their susceptibility to the first stage of proteolysis. It should be noted that the phosphorylated linker peptide eluted earlier from the reverse-phase column than did the unphosphorylated linker (Eigure 2); the earlier elution was greater than would be predicted for simple phosphorylation, suggesting that these peptides have significant differences in their secondary structure. We also noted that the phosphoserine in this peptide was resistant to the β-elimination normally seen in collisional activation of phosphopeptides, and it is possible that the peptide is stabilized in the collisional cell be secondary structure (although more studies with model peptides are necessary to test that hypothesis).

Whatever the explanation, preliminary studies with the stage one endoproteinase (profilaggrin proteinase I) indicate a strong preference for the phosphorylated linker. Thus, the digesting protease may recognize this structural feature on rat profilaggrin rather than simply the primary sequence itself as in mouse profilaggrin. The different recognition structures in mouse and rat may represent two extremes—primary structure in mouse and phosphorylation in rat. More studies in other species and with model peptides are necessary to clarify the range of determinants of proteolytic specificity.

The comparison of the processing of profilaggrin in rat and mouse systems has provided insight into the parameters involved, as well as suggesting hypotheses about the mechanism. Earlier studies on mouse profilaggrin proposed that two endoproteinases were involved, followed by an amino peptidase (Resing et al., 1989). The presence of ragged C- and N-termini in rat filaggrin now suggests that a carboxypeptidase may also be involved. Furthermore, the differences in the details of cleavage of mouse and rat linker regions indicate that the importance of the various proteases may not be the same in both species. These differences may relate to the activation state of the processing enzymes or to transcriptional control events when the cell is in its transition stage. Analyses of the human system and of in vitro processing with the purified proteases should provide additional insight into this complex process.

SUPPLEMENTARY MATERIAL AVAILABLE

MS/MS spectra of these filaggrin peptides that differed in sequence from those predicted by the cDNA (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Barber, M., Bordoli, R. S., Elliot, G. J., Sedgwick, R. D., & Tyler, A. N. (1981) *J. Chem. Soc., Chem. Commun.* 325–327.

Biemann, K. (1990a) *Methods Enzymol. 193*, 455–79.
Biemann, K. (1990b) *Methods Enzymol. 193*, 886–7.
Burlet, O., Orkiszewski, R. S., Ballard, K. D., & Gaskell, S. J. (1992) *Rapid Commun. Mass Spectrom. 6*, 658–662.
Covey, T. R., Bonner, R. F., Shushan, B. I., & Henion, J. (1988) *Rapid Commun. Mass Spectrom. 2*, 249–256.
Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989) *Science 246*, 64–71.
Gan, S.-Q., McBride, O. W., Idler, W. W., Markova, N., & Steinert, P. M. (1990) *Biochemistry 29*, 9432–9440.
Haydock, P. V., & Dale, B. A. (1990) *DNA Cell Biol. 9*, 251–261.
Hunt, D. F., Yates, J. R. I., Shabanowitz, J., Winston, S., & Hauer, C. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6233–7.
Johnson, R. S., Ericsson, L., & Walsh, K. A. The 39th ASMS Conference Mass Spectrometry and Allied Topics, Nashville, TN, 1991, 1233–1234.
Karas, M., & Hillenkamp, F. (1988) *Anal. Chem. 60*, 2299–2301.
Lee, T. D., & Vemuri, S. The 39th ASMS Conference on Mass Spectrometry and Allied Topics, Nashville, TN, 1991, 1229–1230.
Lonsdale-Eccles, J. D., Teller, D. C., & Dale, B. A. (1982) *Biochemistry 21*, 5940–5948.
Mckinley-Grant, L. J., Idler, W. W., Bernstein, I. A., Parry, D. A. D., Cannizzaro, L., Croce, C. M., Huebner, K., Lessin, S. R., & Steinert, P. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4848–4852.
Resing, K. A., & Dale, B. A. (1991) in *Biochemistry and Physiology of the Skin* (Goldsmith, L. A., Ed.) pp 148–167, Oxford University Press, New York.
Resing, K. A., Dale, B. A., & Walsh, K. A. (1985) *Biochemistry 24*, 4167–4175.
Resing, K. A., Walsh, K. A., Haugen-Scofield, J., & Dale, B. A. (1989) *J. Biol. Chem. 264*, 1837–1846.
Resing, K. A., Al-Alawi, N., Blomquist, C., Fleckman, P., & Dale, B. A. (1993) *J. Biol. Chem.* (in press).
Rothnagel, J. A., & Steinert, P. M. (1990) *J. Biol. Chem. 265*, 1862–1865.
Smith, R. D., Loo, J. A., Ogorzalek, R. R., Busman, M., & Udseth, H. R. (1991) *Mass Spectrom. Rev. 10*, 359–451.
Sundqvist, B. U. R., & MacFarlane, R. D. (1985) *Mass Spectrom. Rev. 42*, 421–460.
Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., & Yoshida, T. (1988) *Rapid Commun. Mass Spectrom. 2*, 151–153.
Tang, X., & Boyd, R. K. (1992) *Rapid Commun. Mass Spectrom. 6*, 651–657.