Identification of the Amino Terminus of Human Filaggrin Using Differential LC/MS Techniques: Implications for Profilaggrin Processing†

Craig D. Thulin* and Kenneth A. Walsh

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

Received February 1, 1995; Revised Manuscript Received May 1, 1995⊗

ABSTRACT: Filaggrin, the intermediate filament aggregating protein of epidermis, is the product of proteolytic processing of the precursor profilaggrin, which consists of 10−20 tandem filaggrin domains. The proteolytic processing sites in mouse and rat profilaggrin have been previously reported. Mouse filaggrin is N-terminally blocked. Rat filaggrin is N-terminally ragged, making it heterogeneous. Human filaggrin, in addition to being N-terminally blocked and potentially ragged at the amino terminus, is heterogeneous due to sequence variation between one filaggrin domain and another along the profilaggrin gene. This complexity has made more difficult the analysis of processing sites in human profilaggrin. We have identified the amino terminus of human filaggrin by applying a general method we have developed for the recognition of amino-terminal peptides in digests of N-terminally blocked proteins. This method compares the peptides in an acetylated and an unacetylated tryptic digest of the protein during their separation by liquid chromatography on-line with electrospray mass spectrometry. In this comparison only the original blocked amino-terminal peptides appear unchanged between the two profiles. Human filaggrin was found to have a heterogeneous N-terminus, as a result both of sequence heterogeneity and of ragged processing; it is blocked by a pyrrolidonecarboxyl group derived from glutamine. By comparison to the termini of rat and mouse filaggrins, implications for the processing of human profilaggrin are discussed.

The amino termini of most proteins are not correctly predicted by the corresponding DNA sequences. Most lose at least the initiating methionine (Kendall et al., 1990), and the majority of eukaryotic proteins are covalently blocked at the N-terminus (Brown, 1979). Other examples are found in proteins cleaved from precursor forms such as zymogens and polyproteins. An extreme case is found in the epidermal protein filaggrin, which is initially synthesized as a large precursor, profilaggrin, containing 10−12 tandem homologous filaggrin domains flanked by non-filaggrin amino- and carboxy-terminal domains (Presland et al., 1992). At a regulated time during the terminal differentiation of epidermal cells, profilaggrin is converted to filaggrin by limited proteolysis (Dale et al., 1990). Filaggrin then serves in the maturing epidermis as the intermediate filament associated protein that aggregates keratin filaments into macrofibrils (Dale et al., 1990).

The events that convert profilaggrin to filaggrin include dephosphorylation and limited proteolysis, each of which is tightly regulated and highly specific (Resing et al., 1993a). A chymotrypsin-like protease, "profilaggrin endoproteinase 1", has been isolated from mouse epidermis (Resing et al., 1993b) and cleaves both mouse and rat profilaggrin into filaggrin domains. However, knowledge of its specificity and control is limited. The termini of mouse (Resing et al., 1989) and rat profilaggrins (Resing et al., 1993b) have been previously identified, but human filaggrin has presented a special problem because of its greater heterogeneity. Among rodent profilaggrins, identity from one filaggrin domain to another is about 97% (in mouse; Rothnagel & Steinert, 1990) to 99.5% (in rat; Resing et al., 1993b). In contrast, the cloned human filaggrin sequences vary at 40% of the amino acid positions (Gan et al., 1990). Human filaggrin, like mouse filaggrin, is N-terminally blocked. We have undertaken a study of the structure of human filaggrin, using the resolving power of liquid chromatography/mass spectrometry (LC/MS).[1] Using a general method developed to identify amino-terminal peptides of simpler N-terminally blocked proteins (Thulin & Walsh, 1995), we report here three alternative blocked amino termini of human filaggrin and compare them to those of the rodent proteins.

EXPERIMENTAL PROCEDURES

Human filaggrin was purified from foreskin epidermis which had been separated from tissue specimens by heating foreskins at 55 °C for 5 min in a 15 mM EDTA, 85% phosphate-buffered saline solution. The epidermis from one or two foreskins was then homogenized in 0.5 mL of 9 M urea and 50 mM Tris (pH 8.0) to which several crystals of PMSF had been added. A typical filaggrin prep was made from 10 such homogenates (about 20 foreskins). The pooled homogenate (approximately 5 mL) was pumped into a DE52 column (5 × 1 cm), and the flow-through was collected. The protein was precipitated by adding 3 vol of cold acetone at 0 °C for 10 min and centrifuging for 20 min at 12 000*g*. This precipitate was resuspended in 5 mL of 50 mM Tris (pH 8.0), to which 250 μL of 100 mM EDTA (pH 7) and 50 μL of 100 mM PMSF in methanol had been added. The

† This work was supported by Public Health Service National Research Service Award T32 GM07270 awarded to C.D.T.

* To whom correspondence should be addressed. Current address: Vollum Institute, Oregon Health Sciences University, 3181 SW Sam Jackson Park Rd., Portland, OR 97201-3098.

⊗ Abstract published in *Advance ACS Abstracts,* June 15, 1995.

[1] Abreviations: LC/MS, liquid chromatography/mass spectrometry; CID, collision-induced dissociation; PCA, pyrrolidonecarboxylic acid; PMSF, phenylmethanesulfonyl fluoride; TFA, trifluoroacetic acid; MS/MS, tandem mass spectrometry; *m/z*, mass to charge ratio.

0006-2960/95/0434-8687$09.00/0

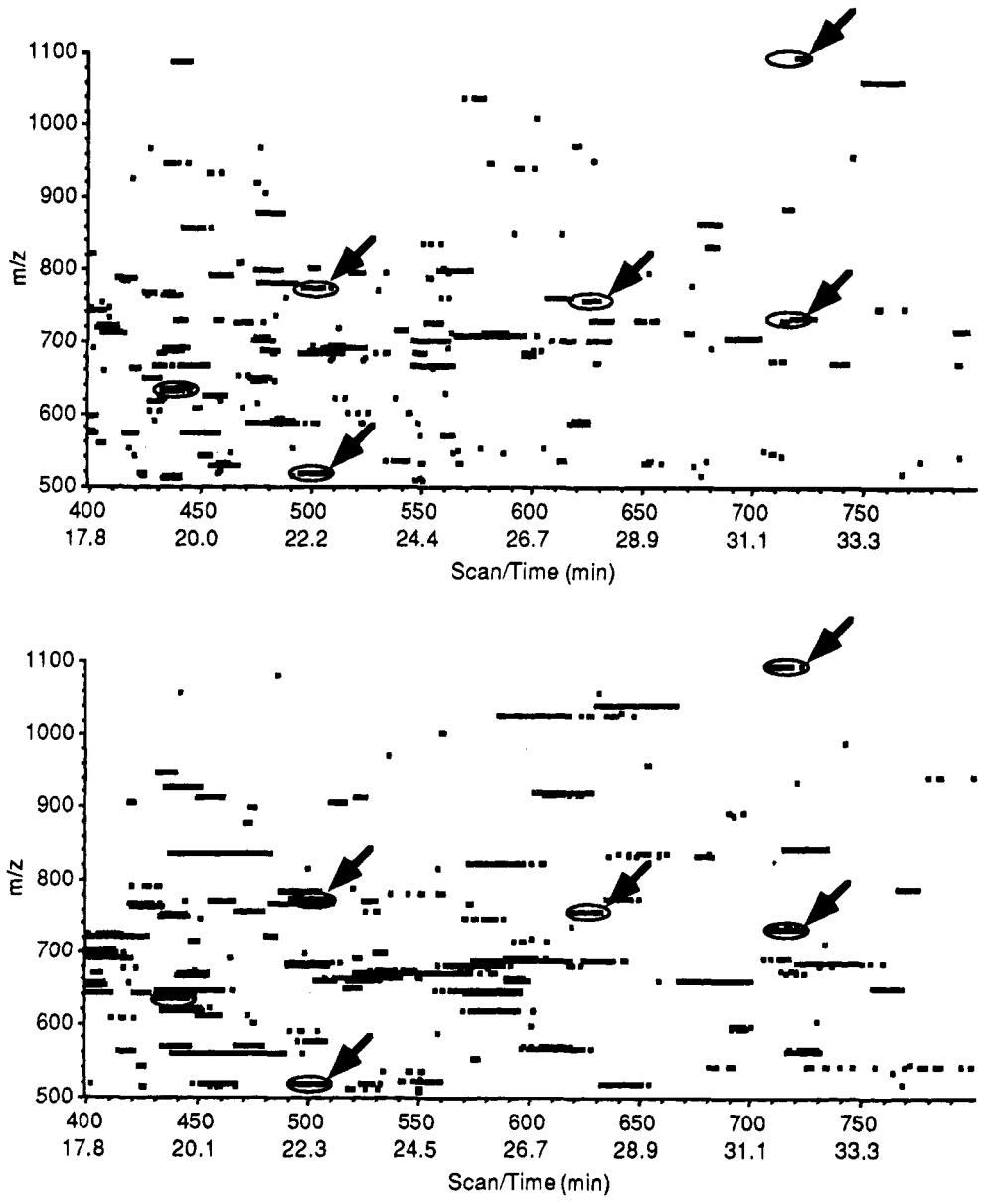

FIGURE 1: Region of the LC/MS contour plot of acetylated (top) and unacetylated (bottom) tryptic digests of human filaggrin. On the horizontal axis, the lower set of numbers indicates the time of elution, and the upper set refers to the scan number (see Experimental Procedures). The display was simplified by viewing in black-and-white mode. Six of the peaks common to both profiles are enclosed in ovals (the leucine enkephalin internal standard is not in this region of the profile; see text for exact *m/z* values). Ions identified as N-terminal fragments are indicated with arrows. Other apparent overlaps differ in *m/z* by more than 1.

suspension was again centrifuged, and the pellet was discarded. The supernatant was applied to a POROS S/M 4.6 × 100 mm strong cation-exchange column (PerSeptive Biosystems, Inc.). A chromatogram was developed on a BioCAD Sprint system (PerSeptive Biosystems, Inc.) at a flow rate of 10 mL/min as follows: After an initial wash of 2 column vol in 100 mM phosphate, pH 3.5, and a further 2 column vol wash in the same buffer containing 500 mM NaCl, a gradient to 1 M NaCl was developed over 10 column vol, followed by a step to 1.5 M NaCl in the phosphate buffer. One milliliter fractions were collected, and aliquots were analyzed by SDS–PAGE and Western blots to identify the filaggrin-containing fractions and to determine their purity. Fractions containing filaggrin were pooled and desalted on a POROS R1/M 2.1 × 30 mm reversed-phase column (PerSeptive Biosystems, Inc.) using a step gradient from 0 to 80% acetonitrile in 0.05% trifluoroacetic acid (TFA). Fractions were collected and lyophilized. Since filaggrin has no known enzymatic activity, a purification table is not appropriate. Purity was established on the basis of the presence of a single band on silver-stained SDS–PAGE, as well as a single (though somewhat broad) peak on matrix-assisted laser desorption ionization time-of-flight mass spectrometry (data not shown). The protein (approximately 30 μg) was dissolved in 200 μL of 100 mM Tris (pH 8.0) and 1 mM $CaCl_2$. Four microliters of 0.1 mg/mL trypsin in millimolar HCl was then added, and the sample was incubated at 37 °C for 6 h. One-half of the tryptic digest was removed, and the newly created amino termini were acetylated using a modification of the method of Fraenkel-Conrat (Fraenkel-Conrat, 1957). First, 100 μL of 0.4 M sodium acetate was added, and then three 0.5-μL aliquots of acetic anhydride were added over the course of 1 h while the sample sat on ice. Following acetylation, the samples were frozen for subsequent analysis by liquid chromatography/mass spectrometry (LC/MS).

Peptides in the unacetylated and in the acetylated aliquots from the digest were compared by LC/MS using an Applied

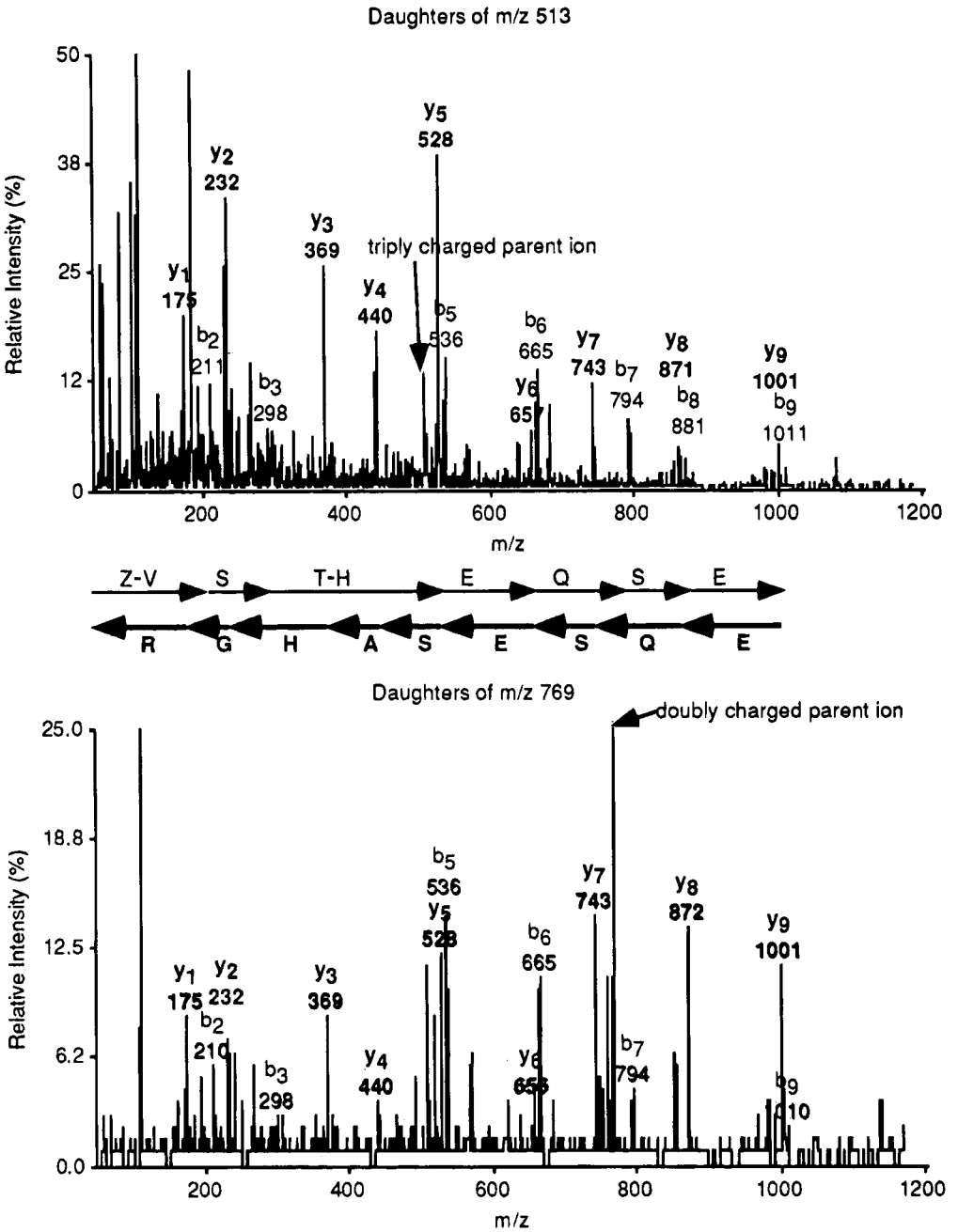

FIGURE 2: CID spectra from related parent ions of *m/z* 513 and 769. Fragment ions of the b series (amino-terminal fragments) and the y series (carboxy-terminal fragments) are indicated and interpreted by the sequence between the two spectra, with y ions indicated in bold type. The positions of the arrowheads locate the corresponding fragment ion peaks. The peptide sequence is deduced to be ZVSTHEQSESAHGR, where Z denotes a pyrollidonecarboxyl residue.

Biosystems Model 140A HPLC with an UpChurch 2 mm i.d. C_{18} reversed-phase column and a PE Sciex API III triple-quadrupole ion spray mass spectrometer. At a flow rate of 200 μL/min the chromatogram was developed from 0.05% TFA by a linear gradient of 0 to 30% acetonitrile and 0.04% TFA over 30 min, followed by a gradient to 60% acetonitrile and 0.04% TFA over the next 15 min. Ten percent of the HPLC effluent was directed to the mass spectrometer, and the remainder was directed through an ABI Model 785A UV detector, where fractions were collected by hand. Ions were scanned from *m/z* 300 to 1500 with a 0.5-ms dwell time at intervals of 0.25 *m/z*. Comparison of the LC/MS data before and after acetylation sought peptides that were identical in both mass and chromatographic mobility. Alignment of the two sets of data was facilitated by matching an internal standard of leucine enkephalin. Data analysis was performed using the MacSpec software from PE Sciex, as well as an in-house program, Sherpa, written by J. Alex Taylor, which identifies and relates peptide *m/z* values in an LC/MS experiment to masses predicted from a given protein sequence. MacBioSpec (PE Sciex) was also used to predict masses for some modified peptides. Fractions containing ions of interest were infused into the mass spectrometer at 1.7 μL/min, and ions selected in the first quadrupole were analyzed by interpreting collision-induced dissociation (CID) spectra. Sherpa was used to identify the most likely sequence among known filaggrin sequences which accounted for the ions of a given CID spectrum. Both Sherpa and MacBioSpec were used to generate lists of expected mass spectral fragments from known filaggrin sequences for comparison to the observed data. Final analysis of each CID spectrum was completed by visual inspection. All masses, both

calculated and observed, use average isotope distribution (as opposed to monoisotopic masses).

RESULTS

Mouse filaggrin termini have been characterized using standard biochemical techniques which relied on purification of the tryptic fragments to homogeneity (Resing et al., 1989). However, the complexity of the tryptic peptide mixture resulting from a digest of human filaggrin makes this impractical (a theoretical digest based on 13 reported human filaggrin DNA sequences predicts 179 unique tryptic peptides, many of which have very similar sequences). Rat filaggrin termini have been identified by exhaustive analysis of the molecular ions seen in an LC/MS separation of the tryptic digest (Resing et al., 1993b). Again, complexity makes this approach formidable in the case of the human protein. To overcome these difficulties, an approach was developed which allows the specific identification of blocked N-terminal peptides in an LC/MS peptide separation without necessitating identification of all of the peaks [see Thulin and Walsh (1995) for a generalization of the method which involves the preliminary blocking of lysine side-chain amino groups with deuterated acetic anhydride]. In the present investigation, we took advantage of the lack of lysine in filaggrin and compared LC/MS separations of tryptic peptide mixtures before and after treatment with acetic anhydride. In this comparison, each peptide except one already blocked should become N-terminally acetylated, thus increasing each molecular weight (by 42) and increasing each peptide's hydrophobic behavior during HPLC. We sought ions that appeared to have identical characteristics in the two halves of the digest.

There were seven ions in the LC/MS contour plots of the acetylated tryptic digest of human filaggrin that corresponded within 1 *m/z* unit (five of them matched exactly) and corresponded in elution time within 1 min (see Figure 1). One of these, having *m/z* 556.3 and eluting at about 37 min, is an internal standard, leucine enkephalin, used to align the chromatograms. The other six were candidates for the N-terminal peptide(s). The two signals appearing at about 22 min, with *m/z* 512.8 and 768.7, correspond in mass to the $(M + 3H)^{3+}$ and $(M + 2H)^{2+}$ ions, respectively, of a molecular species of 1535.4 Da. Likewise, the two peaks eluting from the acetylated digest at about 32 min, with *m/z* 728.2 and 1091.6, correspond to the triply and doubly charged ions of a molecular species of 2181.5 Da. All four ions were matched in the unacetylated digest. The ion appearing at about 28 min with an *m/z* of 750.7 corresponds to a 1499.4-Da peptide in the doubly charged form, as determined by the isotope distribution within the signal. Ion signals appearing between 19 and 20 min elution time have an *m/z* of 628.6 in the unacetylated digest and 628.1 in the acetylated digest. Both of these signals display isotope distributions consistent with doubly charged ions, corresponding to molecular species of 1254.2 and 1255.2 Da, respectively. Each of the seven ions shared between the two runs was subjected to tandem mass spectrometry (MS/MS) to identify it by its fragmentation pattern.

Figure 2 shows the collision-induced dissociation (CID) spectra of the two ions that elute at 22 min, which confirm that these ions represent two charge states of the same molecular species. The CID spectra from the acetylated and

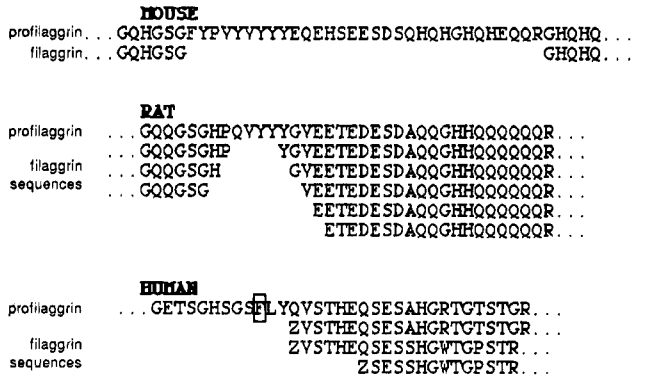

FIGURE 3: Termini of human filaggrin compared to filaggrins from mouse (Resing et al., 1989) and rat (Resing et al., 1993b). A pyrrolidonecarboxyl residue is represented by Z. Human profilaggrin sequence is from Gan et al. (1990), who reported that the phenylalanine (enclosed in box) is removed during the processing of profilaggrin to filaggrin. This figure is modified after Resing et al. (1993b).

the unacetylated digest were virtually identical, showing successive b fragment ions [corresponding to the amino-terminal fragments; see Biemann (1990)] for each residue from the b_2 ion ZV, where Z denotes pyrrolidonecarboxylic acid, to the b_9 ion (ZVSTHEQSE) with the single exception of the b_4 ion (ZVST), indicating that there was no fragmentation between the threonine and the histidine. Successive y fragment ions (carboxy-terminal fragments) are seen for each residue from the y_1 arginine to the y_9 ion (EQSESAHGR). Thus the series of b ions and y ions can be interpreted as sequences that overlap for four residues (EQSE). These data identify the peptide as ZVSTHEQSESAHGR, which has a predicted molecular mass of 1535.5 Da. If glutamine is substituted for the pyrrolidonecarboxyl group this sequence is readily recognized in a cloned human filaggrin gene (see Figure 3).

Similar analysis of the CID spectra of the two ions eluting at 32 min (Figure 4) confirm that these ions also represent a single molecular species closely related to that above. Successive b ions are evident for each residue from the b_2 ion (ZV) to the b_{11} ion (ZVSTHEQSESS), again lacking the fragment that would be created by a cleavage between the threonine and the histidine. Successive y fragment ions are seen for each residue from the y_1 arginine to the y_{13} ion (SESSHGWTGPSTR), which overlaps four residues of the b ion sequence (SESS). The peptide is thus identified as ZVSTHEQSESSHGWTGPSTR, which has a predicted mass of 2181.2 Da, agreeing with the mass observed. This sequence differs from that above at the 11th residue (S for A) and by substitution of tryptophan for arginine at residue 14. This sequence has also been reported in cloned filaggrin sequences (see Figure 3).

A CID spectrum of the ion eluting at 28 min, of *m/z* 750.7, is consistent with an $(M + 2H)^{2+}$ ion (data not shown). A series of successive b ions was seen from the b_2 ion ZS to the b_{10} ion ZSESSHGWTG, except fragments representing cleavages both before and after the third serine in the sequence. A full series of y ions was seen for all of the successive residues from the y_1 arginine to the y_{10} ion (SHGWTGPSTR), producing an overlap of six residues (SHGWTG) with the b ion series. The peptide is thus seen to be ZSESSHGWTGPSTR, in accord with the observed mass of 1499.4 Da. It corresponds to a segment within the

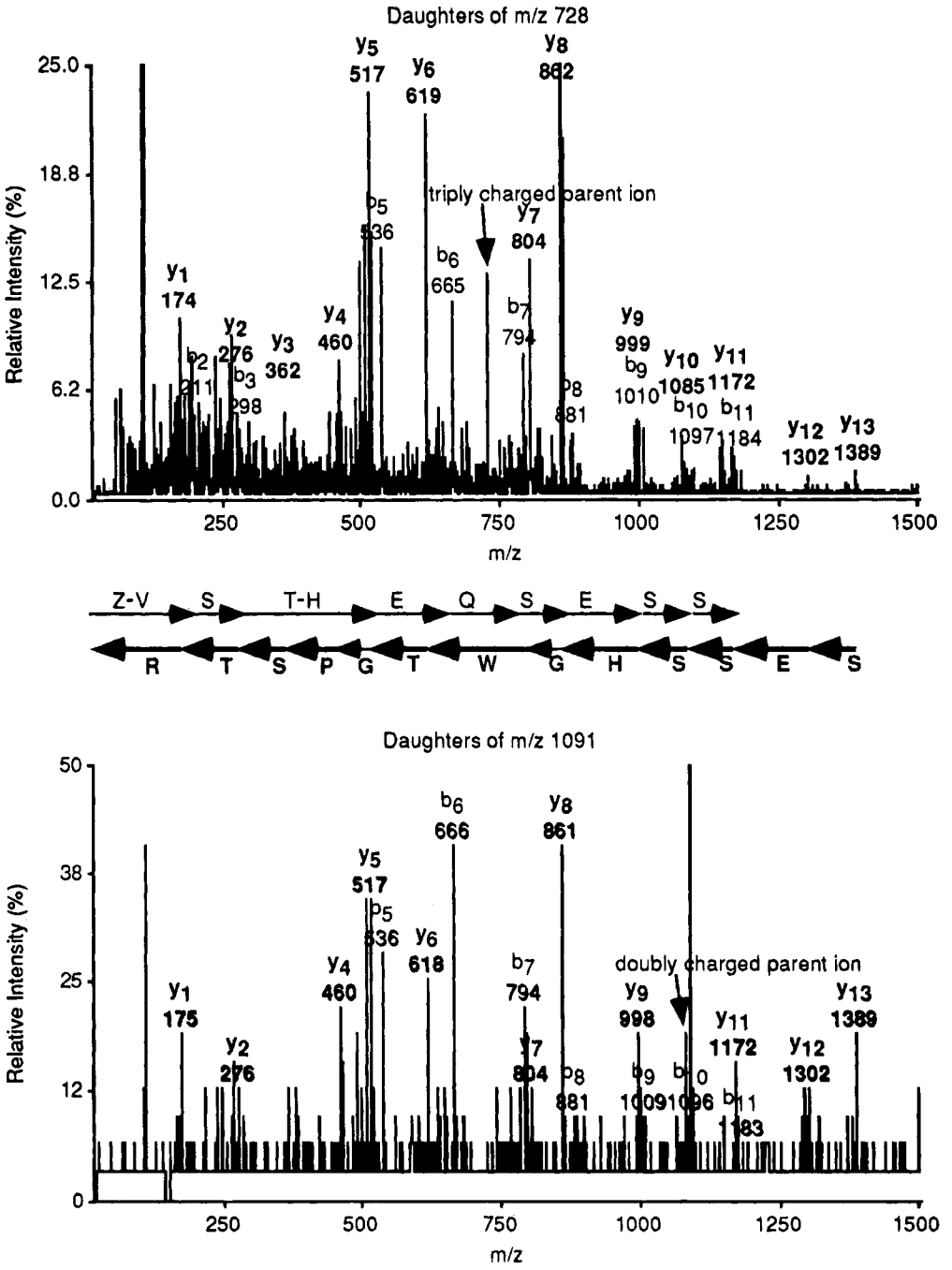

FIGURE 4: CID spectra from related parent ions of *m/z* 728 and 1091. Fragment ions of the b series and the y series are indicated in the spectra and interpreted by the sequence between the two spectra, with y ions in boldface type. The peptide sequence is deduced to be ZVSTHEQSESSHGWTGPSTR.

sequence of the previous peptide (Figure 3). Although electrospray mass spectrometry is not a rigorously quantitative method, we find it significant that the molecular ion at 750.7 is much less abundant (only about 15%) than the sum of the signals of the two ions representing the full-length peptide of mass 2181.2 Da, indicating that this smaller peptide is a minor variant of the amino terminus.

The two ions at 19.5 min (*m/z* 628.1 in the acetylated and 628.6 in the unacetylated run) differ more in *m/z* than any of the other matches noted above. CID spectra were generated and were found to differ markedly from one another (data not shown), indicating that the corresponding peptide sequences are not identical. The ion at *m/z* 628.1, from the acetylated digest, was identified as the $(M + 3H)^{3+}$ ion of the N-terminally acetylated peptide HTQTSSGGQAASSHEQAR (data not shown), which is predicted as an internal tryptic fragment by one of the filaggrin cDNAs (McKinley-Grant et al., 1989). The ion at *m/z* 628.6, from the unacetylated digest, did not correspond to this peptide. Additionally, the amount of the ion representing this acetylated peptide (as measured by the area under the primary ion signal) was 12-fold that of the other ion, in marked contrast with the pairs of N-terminal peptides (above), which were observed to be of the same order of magnitude. We conclude that the near coincidence of the elution time and the *m/z* does not in this case result from identical peptides in the acetylated and unacetylated preparations, but rather from an acetylated peptide in one aliquot behaving like a different unacetylated peptide in the other. With the large number of peaks seen in this complex LC/MS contour plot, such a random and meaningless coincidence would be a rare but possible expectation.

DISCUSSION

Because of the degree of its heterogeneity, human filaggrin is more difficult to examine structurally than are the filaggrins from rat and mouse. The tryptic fragments of mouse filaggrin were identified by standard techniques of protein chemistry (Resing et al., 1989). Rat filaggrin is nearly twice as large as mouse filaggrin, and it was necessary to employ a two-dimensional analysis of LC/MS (chromatographic and mass dimensions) and the power of tandem MS for structural information (Resing et al., 1993b). In the rodent proteins the identity among the filaggrin domains is about 97%; human filaggrins are much more heterogeneous, even from one individual, complicating their analysis. For example, no single tryptic peptide is conserved in all of the cloned human filaggrin sequences. Each human profilaggrin allele codes for 10−12 tandem versions of the filaggrin domain (Presland et al., 1992); hence 24 different filaggrin protein sequences are potentially represented in the epidermis of a single individual. We are currently analyzing the extent of this heterogeneity as observed in protein purified from single individuals. Given this complexity, even the two-dimensional analysis of LC/MS does not adequately simplify the problem of analyzing structural features. Since human filaggrin, like the murine protein, was known to be N-terminally blocked, we developed a method to identify amino-terminally blocked peptides in LC/MS profiles of proteolytic digests (Thulin & Walsh, 1995) and have now refined that method and applied it to the complex case of human filaggrin. This led to three different but related blocked amino-terminal sequences.

It is useful to compare the filaggrins from rat, mouse, and humans, which have been shown to be homologous (Haydock & Dale, 1990). When the sequences are aligned, the proteolytic cleavage sites are seen to be in very similar loci (see Figure 3). In the three profilaggrins, virtually the only aromatic residues in the protein are found here. The three cleavage sites do differ in detail, however. Cleavage in the mouse is reported to remove a rather large "linker" region of 31 amino acids, leaving a single blocked amino terminus and a single carboxy terminus. The rat cleavage removes linkers of four to ten residues, leaving ragged amino and carboxy termini. In the present study of human filaggrin two alternative points of cleavage create two amino termini. The carboxy terminus has not yet been identified. These comparisons indicate that the specific site of cleavage is not stringent; rather the significance of the proteolytic processing event may be simply to separate filaggrin domains, whether by a single, double, or nibbled cleavage. Since the C-terminus of human filaggrin has not been identified, it is not clear whether the primary cleavage sites in human profilaggrin are X-Gln bonds or whether exopeptidases are involved as suggested for rat filaggrin.

Of the three species examined, only rat filaggrin has a free amino terminus. The amino-terminal blocking group in the mouse protein has not been identified. It may be significant that PCA is found as the amino-terminal blocking group in human filaggrin. PCA, a derivative of glutamine, is a markedly hygroscopic compound which, in its free form, is believed to contribute to the hydration of epidermis (Laden & Spitzer, 1967), where it is found in extremely high concentrations relative to other tissues (Wolfersberger et al., 1973). PCA is one of the important catabolites of filaggrin, which is degraded in the stratum corneum after having aggregated keratin filaments (Scott et al., 1988). PCA is known to form spontaneously in peptides with N-terminal glutamine, catalyzed by weak acids (Bodansky, 1988). It is not probable, however, that PCA was generated during the chromatographic purification of human filaggrin, because partially purified human filaggrin that had not been subjected to acidic conditions was also found to have a blocked amino terminus.

Comparing the processing of human filaggrin with that of the rat and mouse proteins has enabled us to identify conserved characteristics (such as cleavage following the aromatic residue cluster) amidst divergent features (such as the raggedness and presence or absence of an N-terminal blocking group). This information has been realized due to a method we have developed that has been proven capable of recognizing and identifying blocked amino-terminal peptides even in this complex mixture of heterogeneous but homologous proteins. It should be noted that the method used in the present study would have been less definitive if the blocking group had been the more common *N*-acetyl group. However, a more general method using *deuterated* acetic anhydride and LC/MS is applicable in such cases (Thulin & Walsh, 1995).

ACKNOWLEDGMENT

The authors gratefully acknowledge the laboratory of Dr. Philip Fleckman for donation of the human foreskins which were used in these investigations, Dr. Richard Johnson and Lowell Ericsson for critical discussion concering mass spectrometric techniques, and J. Alex Taylor for sharing the Sherpa data analysis program.

REFERENCES

Biemann, K. (1990) *Methods Enzymol. 193*, 455−480.
Bodansky, M. L. (1988) in *Peptide Chemistry*, p 108, Springer-Verlag, Berlin.
Brown, J. L. (1979) *J. Biol. Chem. 254*, 1447−1449.
Dale, B. A., Resing, K. A., & Haydock, P. V. (1990) in *Cellular and molecular biology of intermediate filaments*, pp 393−412, Plenum Publishing Corp., New York.
Fraenkel-Conrat, H. (1957) *Methods Enzymol. 4*, 247−269.
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.
Kendall, R. L., Yamada, R., & Bradshaw, R. A. (1990) *Methods Enzymol. 185*, 398−407.
Laden, K., & Spitzer, R. (1967) *J. Soc. Cosmet. Chem. 18*, 351−360.
McKinley-Grant, L. J., Idler, W. W., Bernstein, I. A., Parry, D. A., Cannizzaro, L., Croce, C. M., Huebner, K., Lessin, S. R., & Steinert, P. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4848−4052.
Presland, R. B., Haydock, P. V., Fleckman, P., Nirunsuksiri, W., & Dale, B. A. (1992) *J. Biol. Chem. 267*, 23772−23781.
Resing, K. A., Walsh, K. A., Haugen, S. J., & Dale, B. A. (1989) *J. Biol. Chem. 264*, 1837−1845.
Resing, K. A., al-Alawi, N., Blomquist, C., Fleckman, P., & Dale, B. A. (1993a) *J. Biol. Chem. 268*, 25139−25145.
Resing, K. A., Johnson, R. S., & Walsh, K. A. (1993b) *Biochemistry 32*, 10036−10045.
Rothnagel, J. A., & Steinert, P. M. (1990) *J. Biol. Chem. 265*, 1862−1865.
Scott, I. R., Richards, S., Harding, C., Liddell, J. E., & Curtis, C. G. (1988) *Ann. N.Y. Acad. Sci. 548*, 125−136.
Thulin, C. D., & Walsh, K. A. (1995) in *Techniques in Protein Chemistry VI* (Crabb, J. W., Ed.) pp 55−62, Academic Press, San Diego.
Wolfersberger, M. G., Tabachnick, J., Finkelstein, B. S., & Levin, M. (1973) *J. Invest. Dermatol. 60*, 278−281.

BI950226S