Published in final edited form as:

Exp Eye Res. 2012 March ; 96(1): 163–170. doi:10.1016/j.exer.2011.11.014.

Composition and proteolytic processing of corneal deposits associated with mutations in the *TGFBI* gene

Henrik Karring^a, Kasper Runager^{b,*}, Ida B. Thøgersen^b, Gordon K. Klintworth^c, Peter Højrup^d, and Jan J. Enghild^b

^aInstitute of Chemical Engineering, Biotechnology and Environmental Technology, Faculty of Engineering, University of Southern Denmark, Niels Bohrs Allé 1, DK-5230 Odense M, Denmark

^bCenter for Insoluble Protein Structures (inSPIN) and Interdisciplinary Nanoscience Center (iNANO), Department of Molecular Biology, Science Park, University of Aarhus, Gustav Wieds Vej 10c, DK-8000 Aarhus C, Denmark

^cDepartments of Pathology and Ophthalmology, Duke University Medical Center, Durham, North Carolina

^dDepartment of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, Odense M, DK-5230, Denmark

Abstract

Different types of granular corneal dystrophy (GCD)1 and lattice corneal dystrophy (LCD) are associated with mutations in the transforming growth factor beta induced gene (*TGFBI*). These dystrophies are characterized by the formation of non-amyloid granular deposits (GCDs) and amyloid (LCD type 1 and its variants) in the cornea. Typical corneal non-amyloid deposits from GCD type 2 (R124H), amyloid from a variant of LCD type 1 (V624M) and disease-free tissue controls were procured by laser capture microdissection and analyzed by tandem mass spectrometry. Label-free quantitative comparisons of deposits and controls suggested that the non-amyloid sample (R124H) specifically accumulated transforming growth factor beta induced protein (TGFBIP/keratoepithelin/βig-h3), serum amyloid P-component, clusterin, type III collagen, keratin 3, and histone H3-like protein. The amyloid (V624M) similarly accumulated serum amyloid P-component and clusterin but also a C-terminal fragment of TGFBIP containing residues Y571-R588 derived from the fourth fasciclin-1 domain (FAS1-4), apolipoprotein E and apolipoprotein A-IV. Significantly, analyses of the amyloid sample also revealed the presence of the serine protease Htr (High-temperature requirement) A1 and a number of proteolytic cleavage

¹The abbreviations used are: GCD, granular corneal dystrophy; LCD, lattice corneal dystrophy; TBCD, Thiel-Behnke corneal dystrophy; *TGFBI*, transforming growth factor beta induced gene; TGFBIP, transforming growth factor beta induced protein; FAS1-4, fourth fasciclin-1 domain; LCM, Laser Capture Microdissection; CAPDan, Comparison of Abundances using exogenous Proteins for Dataset-normalization; emPAI, exponentially modified Protein Abundance Index; emPAIn, emPAI, normalized; enp, endogenous protein; cexp, common exogenous protein; tryptic peptide; peptide generated exclusively by proteolysis at the C-terminal side of Lys or Arg residues; semi-tryptic peptide; peptide generated by proteolysis on the C-terminal side of a Lys or Arg residue in one terminus and another type of cleavage event at the other terminus. Peptides with only non-tryptic cleavages were included in the group of semi-tryptic peptides.

© 2011 Elsevier Ltd. All rights reserved.

Corresponding authors: Henrik Karring, Niels Bohrs Allé 1, DK-5230 Odense M, Denmark, Tel.: +45-2135 6350; Fax: +45-6550 7354; hka@kbm.sdu.dk, Jan J. Enghild, Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark, Tel.: +45-8942 5062; Fax: +45-8942 5063; jje@mb.au.dk.

*Present address: Department of Dermatology, University of North Carolina, Chapel Hill, North Carolina.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

sites in the FAS1-4 domain of TGFBIp. These cleavage sites were consistent with the ligand binding and proteolytic activity of HtrA1 suggesting that it plays a role in the proteolytic processing of the amyloidogenic FAS1-4 domain. Taken together, the data suggest that the amyloidogenic-prone region of the fourth FAS1 domain of TGFBIp encompasses the Y571-R588 peptide and that HtrA1 is involved in the proteolytic processing of TGFBIp-derived amyloid *in vivo*.

Keywords

corneal dystrophy; *TGFBI*; mutation; laser capture microdissection; proteomics; emPAI; amyloidosis; serine protease HtrA1

1. Introduction

The transforming growth factor beta induced protein (TGFBIp/keratoepithelin/ β ig-h3) is ubiquitously expressed in connective tissues but is especially abundant in the cornea (Klintworth et al., 1994). Mutations in the transforming growth factor beta induced gene (*TGFBI*), which encodes this protein are responsible for several phenotypically distinct corneal dystrophies identified by the extracellular deposition of protein-rich material within the cornea (Munier et al., 1997; Weiss et al., 2008). The phenotypes include lattice corneal dystrophy (LCD) type 1 and its variants, different types of granular corneal dystrophy (GCD), and Thiel-Behnke corneal dystrophy (TBCD) (Weiss et al., 2008). The particular clinical and histopathologic phenotypes relate to specific mutations in *TGFBI* and the nature of the extracellular accumulations (Kannabiran and Klintworth, 2006). Thus, LCD type 1 and its variants are defined by the corneal deposition of amyloid while non-amyloid crystalloid aggregates that stain red with the Masson trichrome stain accumulate in the corneas of the GCDs. TBCD is characterized by the appearance of “curly fibers” in the superficial cornea (Munier et al., 1997; Ridgway et al., 2000). TGFBIp is anticipated to be a major component of the corneal lesions, but the detailed compositions of the aggregates remain unknown.

In this study we investigated the composition of corneal stromal amyloid and typical non-amyloid granular deposits from individuals with a variant of LCD type 1 and GCD type 2, respectively. To determine the protein composition we isolated deposits and internal control tissue samples from the corneas by laser capture microdissection and performed MS-based proteomic profiling. We developed an “easy to use” method termed Comparison of Abundances using exogenous Proteins for Dataset-normalization (CAPDan) to identify accumulated and attenuated proteins within dissimilar samples using label-free indexes such as exponentially modified Protein Abundance Index (emPAI) values. The emPAI-based CAPDan analyses of the most abundant proteins in the deposits revealed that some proteins co-deposit in both the amyloid and non-amyloid aggregates while others were deposit-specific. The composition of proteins and peptide specifically enriched in the corneal deposits gave insight about the assembly of the phenotypically distinct aggregates. In addition, the accumulation of different TGFBIp fragments and the diverse proteolytic cleavage sites suggested variations in the molecular mechanisms of amyloid fibril - or amorphous aggregate formation. Significantly, the identification of the serine protease HtrA1 in the amyloid deposits and the specificity of the *in vivo* generated cleavage sites within TGFBIp suggested that HtrA1 is involved in the disease mechanism.

2. Materials and Methods

See the supplemental Materials and Methods in Supplemental Data for detailed descriptions.

2.1. Capture, preparation and LC-MS/MS analyses of corneal tissue samples

Corneal deposits and control samples were captured from formalin-fixed paraffin-embedded tissue sections by laser capture microdissection. Amyloid deposits and corresponding control samples were obtained from an individual with a variant of LCD type 1 associated with a homozygous V624M mutation in *TGFBI* (Afshari et al., 2008). Typical non-amyloid GCD deposits and corresponding control samples were captured from the cornea of an individual with GCD type 2 and heterozygous R124H mutation in *TGFBI*. Thus, due to the rareness of these genetic diseases the analyzed samples were from single individuals. Proteins were extracted from the microdissected tissue samples using the Liquid Tissue MS Protein Prep Kit (Expression Pathology Inc., Gaithersburg, USA) and digested with trypsin (Sequencing Grade Modified Porcine Trypsin, Promega, Madison, WI). Recovered peptides were dried, dissolved in 0.1% formic acid, and analyzed by nanoLC-MS/MS using an LTQ-Orbitrap XL mass spectrometer. Studies on the human material were approved by the Internal Review Board of Duke University Medical Center.

2.2. Identification and rough relative quantitation of proteins

Peak lists from all MS/MS spectra were used to query the human Swiss-Prot database using an in-house Mascot search engine (Matrix Science, London, UK; version 2.2.06) (Perkins et al., 1999). Peptides with scores lower than 20 were rejected and only proteins with total scores higher than 40 were accepted (supplemental Tables S1, S2, S3 and S4). The identified proteins were classified as non-contaminating endogenous protein (enp) and contaminating common exogenous protein (cexp) observed in mass spectrometry-based proteomic analyses before listed according to their emPAI values (supplemental Table S5) (Ishihama et al., 2005).

2.3. Identification of accumulated and attenuated proteins in deposits by emPAI-based CAPDan analyses

Since the protein deposits and corresponding control samples were captured from the same tissue sections and processed in parallel, the contaminating common exogenous proteins found in both samples can be used as internal standards to normalize the emPAI values (emPAIn: emPAI, normalized) according to Equation 1, and evaluate relative abundances of the non-contaminating endogenous proteins between the very dissimilar deposit and control tissue samples. Proteins that were only identified in the deposit or control tissue were given an emPAIn value in the other sample corresponding to the lowest emPAIn value calculated for that sample.

$$\text{emPAIn} = \frac{\text{emPAI}}{\sum(\text{emPAI}^{\text{cexp}})} \times 100 \quad (\text{Eq. 1})$$

To determine which proteins specifically accumulate and attenuate in the amyloid and non-amyloid GCD deposits, the emPAIn^{enp} values for the deposits were compared with those in the control tissues and outliers relative to the emPAIn^{cexp} correlation were concentrated or attenuated in the deposit, provided that the outliers have the same size and modifications in the deposit and control samples. Thus, this emPAI-based CAPDan is a further development of the emPAI approach suggested by Ishihama *et al.* (Ishihama et al., 2005) and is an “easy to use” procedure for detection of highly concentrated and attenuated proteins especially between dissimilar samples with no or few common endogenous proteins. To ensure valid identification of specifically accumulated endogenous proteins only significant outliers with emPAIn values in the deposits higher than 1.0 were accepted.

2.4. Spectral counting for tryptic and semi-tryptic TGFB1p peptides

For each dataset (supplemental Tables S1, S2, S3 and S4) the ion spectra for identified TGFB1p peptides obtained by the MS/MS searches were divided into groups of tryptic peptides and semi-tryptic peptides. To detect large differences in the TGFB1p peptide abundances and proteolytic processing between the deposits and control tissues the number of spectra for each of the identified TGFB1p peptides were counted. The spectral count ratios for each of the MS-observable tryptic TGFB1p peptides (mass range 600 – 5400 Da) with no missed cleavages were calculated according to Equation 2, where $\sum(\text{spectral counts})$ is the total number of spectra for tryptic TGFB1p peptides or for semi-tryptic TGFB1p peptides. The annotation i indicates a specific peptide.

$$\text{Spectral count ratio } (i) = \frac{\text{spectral count } (i)}{\sum(\text{spectral counts})} \times 100 \quad (\text{Eq. 2})$$

Subsequently, the ratios were compared between the deposits and corresponding control samples for the identification of enriched/attenuated peptides and proteolytically processed regions. The spectra for tryptic peptides were used to compare peptide abundances while spectra for semitryptic peptides were used to compare levels of non-tryptic *in vivo* cleavages. For both tryptic and semi-tryptic peptides, ion spectra of peptides with methionine oxidations were included in the counting. In the group of tryptic peptides the spectral counts were registered for the corresponding MS-observable tryptic peptide. Spectra of peptides with missed trypsin cleavages that otherwise would have generated two MS-observable tryptic peptides were registered with one spectral count for each of the MS-observable tryptic peptides. In the group of semi-tryptic peptides the spectral counts were registered for the MS-observable tryptic peptide containing the identified non-tryptic cleavage site. Spectral counts for TGFB1p peptides with the mutations V624M and R124H were obtained from the error-tolerance searches of the MS/MS datasets against human TGFB1p (Swiss-Prot entry BGH3_HUMAN, Q15582) and the number of spectrum counts were included in the counting. Differential spectral count ratios for the semitryptic TGFB1p peptides were calculated according to Equation 3, where annotation i indicates a specific peptide.

$$\text{Differential spectral count ratio } (i) = \left(\frac{\text{spectral count } (i, \text{deposit})}{\sum(\text{spectral counts}) (\text{deposit})} - \frac{\text{spectral count } (i, \text{control})}{\sum(\text{spectral counts}) (\text{control})} \right) \times 100 \quad (\text{Eq. 3})$$

Frozen shortly after collection the tissue sections prepared for microdissection were obtained from formalin-fixed paraffin embedded material and heated before digestion with trypsin, thus eliminating the risk of proteolytic activity during sample preparation. In addition, according to the product description (Promega, Madison, WI), the trypsin used for the digestions was modified by reductive methylation to reduce autolysis and chymotrypsin-like specificity and treated with Tosyl Phenylalanyl Chloromethyl Ketone, which will inactivate any chymotrypsin.

3. Results

3.1. Laser capture microdissection of deposits from corneal tissue sections

Amyloid and the characteristic non-amyloid stromal GCD deposits were excised from corneal tissue sections using laser capture microdissection (Fig. 1). This allowed us to limit the contribution from the surrounding normal tissue and thus facilitate the specific identification of the component in the corneal lesions. Sections of amyloid from the LCD tissue were easily procured without contamination from the surrounding normal tissue.

Since corneal deposits that typify GCD accumulate within the extracellular matrix of the corneal stroma, the captured nonamyloid deposits not surprisingly contained some collagenous material (Fig. 1F). Peptides generated by trypsin digestion were recovered from the dissected samples and analyzed by LC-MS/MS using an LTQ-Orbitrap XL mass spectrometer. The R124H mutation in TGFBIp associated with GCD type 2 (Weiss et al., 2008) was identified in the non-amyloid GCD deposits by mass spectrometry (Fig. S1, Supplemental Data).

3.2. Proteomic profiling and identification of accumulated proteins in deposits using emPAI-based CAPDan analyses

The proteomic profiling resulted in the identification of 36 proteins derived from the amyloid residues in the LCD type 1 variant (supplemental Table S1), 46 proteins from the LCD control tissue (supplemental Table S2), 43 proteins from the non-amyloid GCD deposits (supplemental Table S3), and 62 proteins from the GCD control tissue (supplemental Table S4). For each of the samples the identified proteins were listed according to their emPAI values (supplemental Table S5). The identified endogenous proteins and their relative abundances in the normal tissues procured adjacent to the LCD and the GCD deposits were very similar suggesting that *TGFBI* mutations do not affect the composition of the deposit-free regions of the cornea (supplemental Tables S2, S4 and S6, Supplemental Data). However, the protein compositions and relative protein abundances of the amyloid deposits (supplemental Table S1) and non-amyloid GCD aggregates (supplemental Table S3) differed significantly from the control tissues (supplemental Table S6, Supplemental Data). The analyses showed that several proteins were relatively more abundant in the deposits than in the adjacent unaffected control tissues indicating that they accumulated in the deposits (supplemental Table S6, Supplemental Data). However, identification of specifically accumulated proteins in a background of common corneal proteins required further analyses. Because the protein profile of the deposits and the control tissues were very different, the $\sum(\text{emPAI})$ values and molar fractions (protein content (mol %)) (Ishihama et al., 2005) could not be used for normalization and dataset comparisons. Consequently, the emPAI analyses did not allow us to conclude that the most abundant proteins (supplemental Table S6, Supplemental Data) were indeed specifically accumulated in the deposits and not just appeared to be enriched due to a reduction in the concentration of other proteins. Alternatively, the CAPDan procedure was employed. Endogenous proteins with emPAIn coordinates ($\text{emPAIn}^{\text{enp}}$ (control); $\text{emPAIn}^{\text{enp}}$ (deposit)) positioned above the correlation line for the $\text{emPAIn}^{\text{cexp}}$ values were concentrated in the deposits compared to the control tissue, while proteins with emPAIn coordinates positioned below the line were attenuated in the deposits compared to the control tissue (Fig. 2). This comparison was based on the assumption that the size of each protein was the same in the deposit and control samples and, therefore, that the theoretical number of observable peptides (N_{obsbl}) for a protein was the same in both samples (Ishihama et al., 2005; Rappsilber et al., 2002). Although TGFBIp was identified as the most abundant protein in the amyloid deposit (supplemental Table S6, Supplemental Data), the emPAIn coordinate derived was positioned below the emPAIn correlation line for the common exogenous proteins failing to show an accumulation of the full-length TGFBIp in the amyloid deposits (Fig. 2A). However, since emPAI is based on the number of observed peptides per protein (N_{obsd}) divided by N_{obsbl} the accumulation of a particular peptide region could be missed. To address this concern spectral counting was employed as an alternative relative quantification method (Liu et al., 2004; Old et al., 2005). In addition, since several of the identified keratin 3 peptides are shared by other keratins (supplemental Table S3) the emPAIn value (2.02) for keratin 3 in the non-amyloid GCD deposits (supplemental Table S5) may be lower than estimated ($\text{emPAIn}^{\text{enp}}$ coordinate number 4n in Fig. 2D.).

3.3. Accumulation of C-terminal TGFBIp fragment containing peptide Y571-R588 in amyloid deposits

To determine if a specific region of TGFBIp accumulated in the amyloid aggregates or amorphous deposits the ratios of total parent ion spectra observed for each of the MS-observable tryptic TGFBIp peptides were compared between the deposits and corresponding control samples (Fig. 3A and Fig. S2A, Supplemental Data). Basically, divergent spectral count ratios between the samples can in this context be explained by differences in (i), abundance or (ii), the three dimensional structure as more compact regions are less susceptible to proteolysis and thus produce fewer peptides. Common for all four samples were the high number of spectral counts observed for tryptic peptides originating from the inter-domain regions (Fig. 3 and Fig. S2, Supplemental Data). This suggested that the fasciclin-1 (FAS1) domains were resistant to proteolysis and may retain some structure. In total, 65 and 56 tryptic peptide ion spectra were observed for TGFBIp from the amyloid deposits and control tissue, respectively (supplemental Tables S1 and S2). The analyses revealed that all the relative spectral counts were about the same for the amyloid and the control tissue except for the Y571-R588 peptide. The spectral count ratio of this peptide was approximately 10% higher for the amyloid deposit than that of the control tissue (Fig. 3A). For the amyloids and corresponding control sample most of the identified peptides were derived from the FAS1-4 domain. For both samples the highest relative spectral count was obtained for the T378-R396 peptide, located roughly between the FAS1-2 and FAS1-3 domains indicating high proteolytic susceptibility. In comparison, 169 and 71 spectra for tryptic TGFBIp peptides were observed in the non-amyloid deposits and control tissue, respectively (supplemental Tables S3 and S4). Relatively fewer spectral counts were obtained for peptides derived from the amorphous GCD FAS1-4 domain (Fig. S2A, Supplemental Data) compared to the amyloid deposits and corresponding control samples (Fig. 3A). In the GCD samples, the highest ratios of total peptide ion spectra were observed for peptides T378-R396 and F515-R533 of the inter-domain regions between FAS1-2 and FAS1-3, and FAS1-3 and FAS1-4, respectively. Except for some variations in the relative spectral counts for peptides positioned between the FAS-1 domains no major differences were detected between the spectral count ratios for the non-amyloid GCD deposits and control tissue samples (Fig. S2A, Supplemental Data). All together these results suggest that the wild-type FAS1 domains are compact while the regions between FAS1-2, FAS1-3 and FAS1-4 are rather susceptible to proteolysis. However, spectral count ratios of semi-tryptic peptides were also compared to identify variations in spectral count ratios of tryptic TGFBIp peptides caused by (i) differences in abundance or (ii) differences in proteolytic susceptibility due to the mutations and subsequent aggregation.

3.4. Differential proteolytical processing of TGFBIp from deposits and control tissues

By using the semi-tryptic peptide spectra for the calculation of spectral count ratios we were able to suggest the TGFBIp regions predominantly cleaved *in vivo* (Fig. 3B and Fig. S2B, Supplemental Data). In total, 23 and 29 semi-tryptic TGFBIp peptide spectra were observed from the amyloid deposits and control tissue, respectively (supplemental Tables S1 and S2). For the GCD cornea, 99 and 42 semi-tryptic TGFBIp peptide spectra were observed from the nonamyloid deposits and control tissue, respectively (supplemental Tables S3 and S4). Differential spectral count ratios for the semi-tryptic peptides revealed that the C-terminal region of TGFBIp (covering residues V497-R642) from the LCD amyloid deposit contained a cluster of non-tryptic cleavage sites which were significantly less frequent or absent in TGFBIp from the control tissue (Fig. 3B). Thus, region Y571-R588 from the amyloid deposits is flanked by non-tryptic *in vivo* cleavage sites as expected for an accumulated fragment. In comparison, TGFBIp from the nonamyloid GCD deposit has an N-terminal sequence (residues L128-R172) located just after the mutated residue R124 containing significantly more *in vivo* cleavage sites than TGFBIp from the control sample (Fig. S2B,

Supplemental Data). Interestingly, this region is also frequently more proteolyzed but to a lesser extent in TGFBIP identified from the amyloid deposit (Fig. 3B) suggesting that this region is rather susceptible to proteolysis. Thus, for both types of deposits the highest differential spectral count ratios (“proteolytic hotspots”) are in proximity to the relevant mutation in TGFBIP (R124H and V624M).

Taken together, our results suggest that intact TGFBIP is present but does not accumulate in the amyloid deposits while a C-terminal fragment of TGFBIP containing residues Y571-R588 specifically accumulates relative to the control tissue. Therefore, the emPAIn value (17.24) for TGFBIP in the amyloid material is estimated too low explaining its position below the emPAIn^{cexp} correlation line (Fig. 2A). Thus, the spectral counting comparisons together with the emPAIn-based CAPDan analyses suggest that TGFBIP-fragment containing residues Y571-R588, serum amyloid P-component, apolipoprotein E, apolipoprotein A-IV and clusterin accumulate in the amyloid deposits from the LCD cornea, while intact TGFBIP, serum amyloid P-component, type III collagen, keratin 3, clusterin and histone H3-like protein have accumulated in the non-amyloid deposits that typify GCD (Table 1).

3.5. TGFBIP from amyloid deposits was predominantly cleaved at the C-termini of aromatic and hydrophobic residues

The semi-tryptic TGFBIP peptides (supplemental Tables S1–S4) were generated by the action of a Lys or Arg specific protease at one terminus, most likely trypsin during sample preparation. Since trypsin is highly specific for Arg or Lys (Olsen et al., 2004) the other cleavage event producing the semi-tryptic peptide is likely to be the result of an endogenous protease. An examination of the semi-tryptic datasets derived from the deposits or control tissues suggested TGFBIP proteolytic cleavage sites likely to have been generated *in vivo* (Fig. S3, Supplemental Data). For both corneas several *in vivo* TGFBIP cleavage sites were detected only in the deposits or in the control tissue, while others were identified in both samples. Three TGFBIP cleavage sites were specific for the amyloid deposits (Fig. S3A, Supplemental Data) and located within the FAS1-4 region in a cluster of relatively more *in vivo* cleavage sites (Fig. 3B) and flank the accumulated Y571-R588 peptide (Fig. 3A). In contrast, cleavage sites identified in TGFBIP derived from the non-amyloid GCD deposits were distributed over the entire TGFBIP sequence (Fig. S3B, Supplemental Data). To further investigate the processing of TGFBIP derived from the amyloid deposits, the cleavage sites generated *in vivo* were compared (Fig. 4). The alignment reveals that all cleavage sites in the amyloidogenic FAS1-4 domain are at the C-termini of aromatic residues (F, Y) or hydrophobic side chains (A, L, M) thus reflecting a chymotrypsin-like activity (Fig. 4). For all four corneal samples (deposits and controls) the predominant *in vivo* cleavage sites in TGFBIP were at the C-termini of residues F, Y, A, L and N (the P₁ position) (Fig. S3, Supplemental Data).

Significantly, the chymotrypsin-like serine protease HtrA1 (High-temperature requirement A1) was detected in the amyloid deposits (supplemental Table S1). Actually, HtrA1 identified in the amyloid deposits was the only endogenous protease identified in all four samples (supplemental Tables S1–S4) suggesting that HtrA1 is responsible for the proteolytic processing of the amyloidogenic TGFBIP fragment observed.

4. Discussion

Until now the detailed composition of amyloid and non-amyloid corneal deposits associated with mutations in the *TGFBIP* gene have remained unknown. In the present work we have determined the protein composition and identified the major specifically accumulated proteins in human corneal amyloid and granular aggregates associated with mutations V624M and R124H in the *TGFBIP* gene, respectively. As part of our work we have

developed an “easy to use” label-free method abbreviated CAPDan for comparison of relative protein abundances between dissimilar samples using emPAI values obtained from mass spectrometry data. In addition, we have mapped and investigated the proteolytic processing of TGFBIp from the phenotypically distinct corneal deposits. Although our results are obtained from only two individuals with rare mutations in the *TGFBI* gene we believe this study shed light on the molecular pathobiology of corneal protein misfolding diseases linked to the *TGFBI* gene and provide a basis for future proteomic analyses of corneal aggregates associated with other *TGFBI* mutations.

4.1. The pathogenesis of corneal TGFBIp amyloidosis involves proteolytical processing and protein co-deposition events

Our results suggest that full-length TGFBIp is present in the amyloid deposits but that the region containing residues Y571-R588 is significantly enriched. This suggests that the *in vivo* amyloidogenic-prone sequence of TGFBIp contains the peptide Y571-R588 in the FAS1-4 domain. This conclusion is supported by the observation that the Y571-R588 peptide is flanked by non-tryptic *in vivo* cleavage sites suggesting that the accumulated region is produced by trimming at both the N- and C-termini. The accumulated sequence 571-YHIGDEILVSGGIGALVR-588 is dominated by hydrophobic and non-polar residues indicating that it is buried in the native FAS1-4 domain structure. It is worthwhile to note that the accumulated peptide does not include the mutated residue 624 and that no mutations, except for residue H572, associated with corneal dystrophies are located within the Y571-R588 sequence (Kannabiran and Klintworth, 2006; Weiss et al., 2008). Based on these observation and our recent results concerning the thermodynamic stabilities of mutant FAS1-4 domains (Runager et al., 2011) we propose that the amyloidogenic mutations in the FAS1-4 domain are not part of the amyloid fibrillation prone sequence but destabilize the native domain structure leading to partly unfolded FAS1-4 domains. Exposure of the amyloidogenic prone FAS1-4 sequence and subsequent amyloid fibril formation is concomitant with the pathobiology of other amyloid diseases such as the transthyretin and lysozyme amyloidoses (Hou et al., 2007; Kelly, 1998; Pepys et al., 1993; Quintas et al., 2001). This view is further supported by the fact that the relative spectral count ratios of tryptic peptides from the FAS1-4 domain were generally higher for the LCD samples than for the GCD samples indicating that the amyloidogenic V624M mutation destabilises the FAS1-4 domain structure making it more susceptible to tryptic digestion (Fig. 3A and Fig. S2A, Supplemental Data). Interestingly, the identified Y571-R588 peptide of TGFBIp accumulated in the amyloid deposits is included in the multiple C-terminal fragments of TGFBIp observed in the normal human cornea (Karring et al., 2010) suggesting that it is part of a structural motif resistant to proteolysis *in vivo*. Our results are in accordance with the observations by Stix *et al.* (Stix et al., 2005) showing that corneal amyloid caused by mutation P540S in the FAS1-4 domain of TGFBIp reacts strongly with antiserum raised against C-terminal residues 426-682 but not antiserum against the N-terminus of TGFBIp (residues 69-364) and that they contain proteolytic fragments of TGFBIp. The HtrA1 protease identified in the amyloid is a member of the widely conserved HtrA family of serine proteases and is implicated in various disorders including Alzheimer's disease, arthritis, cancer and age-related macular degeneration (AMD) (Dewan et al., 2006; Yang et al., 2006; Zurawa-Janicka et al., 2010). Recent studies have shown that HtrA1 degrades various fragments of amyloid precursor protein (APP) and co-localizes with β -amyloid deposits in Alzheimer disease (Grau et al., 2005) and contribute to articular cartilage deterioration by degrading extracellular proteins such as aggrecan, fibromodulin, decorin, type II collagen (Tsuchiya et al., 2005), fibronectin (Grau et al., 2006) and TGF- β 1 (Launay et al., 2008). In general, the HtrA proteases function in protein quality control to prevent protein unfolding and accumulation of misfolded proteins (Clausen et al., 2011). HtrA1 consists of a chymotrypsin-like proteolytic domain followed by a C-terminal PDZ domain,

which can mediate the binding of target proteins. The specific role of the HtrA1 PDZ domain is not clear but it is known to recognize C-terminal and internal sequences of general hydrophobic character. Thus, in a two-hybrid screening performed by Murwantoko and colleagues (Murwantoko et al., 2004) the consensus sequence for the protein C-terminal ends recognized by the mouse HtrA1 PDZ domain was determined to be Phi-X-Phi-[V/L/F/A]-COOH sequence (where Phi is a hydrophobic/non-polar amino acid and X is any amino acid). In addition, in a recent study by Truebestein *et al.* (Truebestein et al., 2011), the HtrA1 proteolytic domain showed preference for proteolysis at the C-termini of leucine, valine and alanine using unfolded citrate synthase and malate dehydrogenase as model substrates. The fact that the amyloidogenic FAS1-4 domain is cleaved in accordance with the ligand binding and proteolytic activity of the serine protease HtrA1 (Murwantoko et al., 2004; Runyon et al., 2007; Truebestein et al., 2011) found in the amyloid deposits indicate that this protease is responsible for the proteolytic processing of the corneal TGFBIp amyloid.

The CAPDan analyses revealed that the nonfibrillar and ubiquitously amyloid-associated proteins serum amyloid P-component, apolipoprotein E, and clusterin specifically accumulate in TGFBIp amyloid deposits. The fact that the TGFBIp amyloid deposits accumulate common amyloid-associated proteins suggests that the *TGFBI*-linked corneal amyloidoses share fundamental properties and formation pathways with other amyloidoses. In addition we show that the amyloidogenic apolipoprotein A-IV accumulates in the corneal TGFBIp amyloid deposits.

4.2. Co-deposition of type III collagen in the formation of corneal granular TGFBIp deposits

We have demonstrated by LC-MS/MS analyses that the patient with characteristic stromal deposits of GCD type 2 had a heterozygote R124H mutation in TGFBIp (Fig. S1, Supplemental Data). Noteworthy, peptides with mutation R124H were only detected in the typical GCD deposits and not in the control tissue supporting the hypothesis that the mutant TGFBIp species rather than wild-type TGFBIp accumulate in the corneal deposits. Our emPAI-based CAPDan analyses in combination with comparison of spectral count ratios support previous observations suggesting that intact or nearly full-length TGFBIp isoforms accumulate in the typical non-amyloid GCD deposits (Klintworth et al., 1998; Korvatska et al., 2000). However, the differences in spectral count ratios for tryptic peptides from the inter-domain regions in TGFBIp could indicate that the aggregated TGFBIp has an altered domain arrangement compared to that of TGFBIp from the control tissue (Fig. S2A, Supplemental Data). In addition, the fact that the highest differential spectral count ratio for semi-tryptic TGFBIp peptides for the granular deposits was for sequence L128-R172 right next to the mutated R124 residue could indicate that local structural instabilities have induced an increase in cleavages of this region *in vivo*. Our recent studies of the thermodynamic stability of intact mutant TGFBIp variants show that R124H and other FAS1-1 mutations do not alter the overall stability of the molecules, but does not exclude local changes in structural stabilities (Runager et al., 2011). The CAPDan analysis suggested that, in addition to serum amyloid P-component and clusterin, which also accumulated in the amyloid and are often associated with aggregated proteins, type III collagen, keratin 3, and histone H3-like protein were enriched in the GCD deposits. Accumulation of type III collagen in the GCD deposits suggests that the corneal TGFBIp aggregates damage the tissue and induces a tissue repair response.

In conclusion, our results show that significant insight into disease mechanisms can be obtained by detailed analyses of label-free MS/MS data generated from isolated protein deposits and control samples in the cornea. We propose that the CAPDan method can be used to identify protein enrichments and attenuations between other types of dissimilar protein samples isolated from tissues (e.g. protein misfolding diseases, cancers and

neurological diseases) or cell cultures. Our results show that the TGFBIp-linked amyloid disease and GCD analyzed in this study involve several proteins associated with other protein misfolding disorders suggesting that they share fundamental molecular processes.

Highlights

TGFBIp, serum amyloid P-component, type III collagen, keratin 3, clusterin, and histone H3-like protein accumulate in granular corneal deposits. > C-terminal fragment of TGFBIp (containing Y571-R588), serum amyloid P-component, apolipoprotein E, apolipoprotein A-IV, and clusterin accumulate in corneal amyloid. > The amyloidogenic FAS1-4 domain of TGFBIp is cleaved at the C-termini of aromatic residues and hydrophobic side chains. > Serine protease HtrA1 is associated with amyloid in the cornea.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by grants from the National Eye Institute (R01 EY012712), the Danish National Research Foundation, and the Carlsberg Foundation.

References

- Afshari NA, Bahadur RP, Eifrig DE Jr, Thogersen IB, Enghild JJ, Klintworth GK. Atypical asymmetric lattice corneal dystrophy associated with a novel homozygous mutation (Val624Met) in the TGFBI gene. *Mol Vis*. 2008; 14:495–499. [PubMed: 18385782]
- Clausen T, Kaiser M, Huber R, Ehrmann M. HTRA proteases: regulated proteolysis in protein quality control. *Nat Rev Mol Cell Biol*. 2011; 12:152–162. [PubMed: 21326199]
- Dewan A, Liu M, Hartman S, Zhang SS, Liu DT, Zhao C, Tam PO, Chan WM, Lam DS, Snyder M, Barnstable C, Pang CP, Hoh J. HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science*. 2006; 314:989–992. [PubMed: 17053108]
- Grau S, Baldi A, Bussani R, Tian X, Stefanescu R, Przybylski M, Richards P, Jones SA, Shridhar V, Clausen T, Ehrmann M. Implications of the serine protease HtrA1 in amyloid precursor protein processing. *Proc Natl Acad Sci U S A*. 2005; 102:6021–6026. [PubMed: 15855271]
- Grau S, Richards PJ, Kerr B, Hughes C, Caterson B, Williams AS, Junker U, Jones SA, Clausen T, Ehrmann M. The role of human HtrA1 in arthritic disease. *J Biol Chem*. 2006; 281:6124–6129. [PubMed: 16377621]
- Hou X, Aguilar MI, Small DH. Transthyretin and familial amyloidotic polyneuropathy. Recent progress in understanding the molecular mechanism of neurodegeneration. *FEBS J*. 2007; 274:1637–1650. [PubMed: 17381508]
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics*. 2005; 4:1265–1272. [PubMed: 15958392]
- Kannabiran C, Klintworth GK. TGFBI gene mutations in corneal dystrophies. *Hum Mutat*. 2006; 27:615–625. [PubMed: 16683255]
- Karring H, Runager K, Valnickova Z, Thogersen IB, Moller-Pedersen T, Klintworth GK, Enghild JJ. Differential expression and processing of transforming growth factor beta induced protein (TGFBIp) in the normal human cornea during postnatal development and aging. *Exp Eye Res*. 2010; 90:57–62. [PubMed: 19788893]
- Kelly JW. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol*. 1998; 8:101–106. [PubMed: 9519302]

- Klintworth GK, Enghild JJ, Valnickova Z. Discovery of a novel protein (big-h3) in normal human cornea. *Invest Ophthalmol Vis Sci*. 1994; 35 (Sup.):1938.
- Klintworth GK, Valnickova Z, Enghild JJ. Accumulation of beta ig-h3 gene product in corneas with granular dystrophy. *Am J Pathol*. 1998; 152:743–748. [PubMed: 9502416]
- Korvatska E, Henry H, Mashima Y, Yamada M, Bachmann C, Munier FL, Schorderet DF. Amyloid and non-amyloid forms of 5q31-linked corneal dystrophy resulting from kerato-epithelin mutations at Arg-124 are associated with abnormal turnover of the protein. *J Biol Chem*. 2000; 275:11465–11469. [PubMed: 10753964]
- Launay S, Maubert E, Lebeurrier N, Tennstaedt A, Campioni M, Docagne F, Gabriel C, Dauphinot L, Potier MC, Ehrmann M, Baldi A, Vivien D. HtrA1-dependent proteolysis of TGF-beta controls both neuronal maturation and developmental survival. *Cell Death Differ*. 2008; 15:1408–1416. [PubMed: 18551132]
- Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004; 76:4193–4201. [PubMed: 15253663]
- Munier FL, Korvatska E, Djemai A, Le Paslier D, Zografos L, Pescia G, Schorderet DF. Kerato-epithelin mutations in four 5q31-linked corneal dystrophies. *Nat Genet*. 1997; 15:247–251. [PubMed: 9054935]
- Murwantoko, Yano M, Ueta Y, Murasaki A, Kanda H, Oka C, Kawaichi M. Binding of proteins to the PDZ domain regulates proteolytic activity of HtrA1 serine protease. *Biochem J*. 2004; 381:895–904. [PubMed: 15101818]
- Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinisky JR, Resing KA, Ahn NG. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics*. 2005; 4:1487–1502. [PubMed: 15979981]
- Olsen JV, Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*. 2004; 3:608–614. [PubMed: 15034119]
- Pepys MB, Hawkins PN, Booth DR, Vigushin DM, Tennent GA, Soutar AK, Totty N, Nguyen O, Blake CC, Terry CJ, Feest TG, Zalin AM, Hsuan JJ. Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature*. 1993; 362:553–557. [PubMed: 8464497]
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999; 20:3551–3567. [PubMed: 10612281]
- Quintas A, Vaz DC, Cardoso I, Saraiva MJ, Brito RM. Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J Biol Chem*. 2001; 276:27207–27213. [PubMed: 11306576]
- Rappilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. *Genome Res*. 2002; 12:1231–1245. [PubMed: 12176931]
- Ridgway AE, Akhtar S, Munier FL, Schorderet DF, Stewart H, Perveen R, Bonshek RE, Odenthal MT, Dixon M, Barraquer R, Escoto R, Black GC. Ultrastructural and molecular analysis of Bowman's layer corneal dystrophies: an epithelial origin? *Invest Ophthalmol Vis Sci*. 2000; 41:3286–3292. [PubMed: 11006215]
- Runager K, Basaiawmoit RV, Deva T, Andreassen M, Valnickova Z, Sorensen CS, Karring H, Thogersen IB, Christiansen G, Underhaug J, Kristensen T, Nielsen NC, Klintworth GK, Otzen DE, Enghild JJ. Human Phenotypically Distinct TGFBI Corneal Dystrophies Are Linked to the Stability of the Fourth FAS1 Domain of TGFBIp. *J Biol Chem*. 2011; 286:4951–4958. [PubMed: 21135107]
- Runyon ST, Zhang Y, Appleton BA, Sazinsky SL, Wu P, Pan B, Wiesmann C, Skelton NJ, Sidhu SS. Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. *Protein Sci*. 2007; 16:2454–2471. [PubMed: 17962403]
- Stix B, Leber M, Bingemer P, Gross C, Ruschoff J, Fandrich M, Schorderet DF, Vorwerk CK, Zacharias M, Roessner A, Rocken C. Hereditary lattice corneal dystrophy is associated with corneal amyloid deposits enclosing C-terminal fragments of keratoepithelin. *Invest Ophthalmol Vis Sci*. 2005; 46:1133–1139. [PubMed: 15790870]

- Truebestein L, Tennstaedt A, Monig T, Krojer T, Canellas F, Kaiser M, Clausen T, Ehrmann M. Substrate-induced remodeling of the active site regulates human HTRA1 activity. *Nat Struct Mol Biol.* 2011; 18:386–388. [PubMed: 21297635]
- Tsuchiya A, Yano M, Tocharus J, Kojima H, Fukumoto M, Kawaichi M, Oka C. Expression of mouse HtrA1 serine protease in normal bone and cartilage and its upregulation in joint cartilage damaged by experimental arthritis. *Bone.* 2005; 37:323–336. [PubMed: 15993670]
- Weiss JS, Moller HU, Lisch W, Kinoshita S, Aldave AJ, Belin MW, Kivela T, Busin M, Munier FL, Seitz B, Sutphin J, Bredrup C, Mannis MJ, Rapuano CJ, Van Rij G, Kim EK, Klintworth GK. The IC3D classification of the corneal dystrophies. *Cornea.* 2008; 27 Suppl 2:S1–S83. [PubMed: 19337156]
- Yang Z, Camp NJ, Sun H, Tong Z, Gibbs D, Cameron DJ, Chen H, Zhao Y, Pearson E, Li X, Chien J, Dewan A, Harmon J, Bernstein PS, Shridhar V, Zabriskie NA, Hoh J, Howes K, Zhang K. A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration. *Science.* 2006; 314:992–993. [PubMed: 17053109]
- Zurawa-Janicka D, Skorko-Glonek J, Lipinska B. HtrA proteins as targets in therapy of cancer and other diseases. *Expert Opin Ther Targets.* 2010; 14:665–679. [PubMed: 20469960]

Fig. 1. Isolation of corneal deposits by laser capture microdissection

Amyloid deposits from a cornea with LCD (**A**, **B**, and **C**) and non-amyloid amorphous deposits from a cornea with GCD (**D**, **E**, and **F**) were captured from fixed tissue-sections by laser capture microdissection and analyzed by LC-MS/MS. **A**, **D**: tissues before laser capture microdissection. **B**, **E**: tissues after laser capture microdissection. **C**, **F**: captured sections of deposits.

Fig. 2. Identification of accumulated proteins in cornea deposits by emPAI-based CAPDan analyses

The emPAIn values for each of the proteins identified in the amyloid deposits (A) and non-amyloid GCD deposits (C) were plotted against the corresponding emPAIn values for the control tissues. The black lines represent the best-fit for the emPAIn correlations (emPAIn^{cexp}; black diamonds) for the common exogenous proteins (cexp) which were used for dataset normalization. Endogenous proteins (enp) with emPAIn-coordinates (emPAIn^{enp}; crosses) positioned above the emPAIn^{cexp} correlation line are considered to be accumulated in the deposit compared to the corresponding control tissue. The diagrams in panels B and D clearly show the relations for the lower emPAIn values of the amyloid deposits (A) and non-amyloid GCD deposits (C), respectively. The numbers associated with the emPAIn^{enp} coordinates correspond to the numbers in Table 1.

Fig. 3. Spectral count ratios for tryptic and semi-tryptic TGFBIp peptides from the LCD cornea
 (A) Comparison of spectral count ratios for tryptic TGFBIp peptides detected in the amyloid deposit (black bars) and control tissue (gray bars). The spectral count ratio for peptide Y571-R588 is significantly higher for the amyloid than for the control sample. (B) Differential spectral count ratios for semi-tryptic TGFBIp peptides detected in the amyloid deposit and control tissue. TGFBIp peptide regions with relatively more and less non-tryptic cleavages in the amyloid deposits compared to the control tissue have positive and negative bar values, respectively. The C-terminal region of TGFBIp (covering residues V497-R642) contains a cluster of non-tryptic cleavage sites which are more abundant in the amyloid deposits than in the control tissue. Thus, the FAS1-4 domain is significantly more proteolytically processed

in the amyloid deposit than in the control tissue. In addition, the N-terminal peptide covering residues L128-R172 is more cleaved in the deposit than in the control tissue. Spectral count ratios are plotted for each MS-observable tryptic TGFBIp peptide with no missed cleavages. The schematic drawing above the bar charts shows the domain structure of TGFBIp to illustrate the locations of the peptides. The structure includes the N-terminal EMILIN domain (EMI, residues 45–99) and the four consecutive FAS1 domains (FAS1-1: residues 103-235; FAS1-2: residues 242-371; FAS1-3: residues 376-498; and FAS1- 4: residues 505-632). The arrow indicates that the major isoform of TGFBIp in the human cornea is cleaved at the C-terminus of residue A657.

Fig. 4. Alignment of the *in vivo* cleavage sites in TGFBIp from amyloid deposits

The sequences of the 18 non-tryptic cleavage sites in TGFBIp from the amyloid deposits were aligned (from position -5 to +5) and residues with aromatic side chains (F, Y, W, H) and hydrophobic side chains (A, I, L, M, F, W, Y, V) are marked with red bold letters and gray background, respectively. The residue numbers in TGFBIp of the aligned sequences are indicated and the black arrow shows the cleavage sites. All 10 cleavage sites positioned in the FAS1-4 domain (residues 505-632) are at the C-termini of aromatic or hydrophobic residues (F, Y, A, L and M).

Table 1
Proteins accumulated in corneal deposits

The table shows the identified specifically accumulated proteins in the amyloid deposits from the LCD cornea and the nonamyloid amorphous deposits from the GCD cornea compared to the control tissues according to the CAPDan analyses. The numbers (No.) refer to those indicated for emPAIn^{enp} coordinates in Fig. 2.

No.	Amyloid deposits	No.	Non-amyloid deposits
1a	TGFB1p (fragment containing Y571-R588)	1n	TGFB1p (intact)
2a	Serum amyloid P-component	2n	Serum amyloid P-component
3a	Apolipoprotein E	3n	Type III collagen
4a	Apolipoprotein A-IV	4n	Keratin 3 ^(a)
5a	Clusterin	5n	Clusterin
		6n	Histone H3-like

^(a) Note that several of the identified keratin 3 peptides are shared by other keratins. Thus, the emPAIn value for keratin 3 may be lower than estimated.