A model of FAS1 domain 4 of the corneal protein βig-h3 gives a clearer view on corneal dystrophies

Naomi J. Clout, Erhard Hohenester

Department of Biological Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

**Purpose:** A progressive alteration of the cornea resulting in loss of transparency occurs in a set of hereditary diseases known as corneal dystrophies. A number of these dystrophies have been linked to mutations in the 5q31-linked gene product βig-h3 (TGFβIP, kerato-epithelin, MP78/70, RGD-CAP) although the mechanism by which the mutations cause disease remains unknown. Here we investigate the structural basis for the different corneal dystrophies caused by mutations of the βig-h3 protein. The integrin binding properties of βig-h3, described in several recent studies, have been analysed with respect to the βig-h3 structure.

**Methods:** The recently determined structure of a FAS1 domain pair from fasciclin I, an insect cell adhesion molecule and βig-h3 homologue, was used to generate a homology model of the βig-h3 FAS1 domain 4.

**Results:** The structural analysis of FAS1 domain 4 of βig-h3 predicts that the common mutations at positions 124 and 555 do not substantially alter the βig-h3 structure. In contrast, the rare missense mutations appear incompatible with the FAS1 fold. A number of residues implicated in integrin binding by previous mutagenesis are mostly buried and appear to have important structural roles.

**Conclusions:** The common mutations of βig-h3 at positions 124 and 555 are likely to affect protein-protein interactions directly, whereas the rare mutations are likely to cause misfolding of the protein within the cell. Previously identified integrin binding residues are unlikely to be directly involved in receptor binding.

βig-h3 (TGFβIP, kerato-epithelin, MP78/70, RGD-CAP) was first identified as a TGF-β inducible gene in a lung adenocarcinoma cell line [1]. The cDNA sequence encodes a secreted protein of 68 kDa molecular mass, consisting of four regions of internal homology of approximately 140 amino acids and a carboxy terminal Arg-Gly-Asp (RGD) motif, which in other proteins mediates binding to receptors of the integrin family [2]. The tandemly repeated regions are known as FAS1 domains due to the original discovery of four such domains in the insect cell adhesion molecule fasciclin I [3] (Figure 1). Interest in βig-h3 has increased sharply since genetic studies discovered that a group of hereditary corneal dystrophies are caused by specific missense mutations in the BIGH3 (TGFβI) gene [4]. Despite the genetic characterization of these, BIGH3 related corneal dystrophies, the biochemical mechanisms responsible for the alteration of protein behavior remain unclear. The likely effects of amino acid substitutions can be predicted only if a three-dimensional structure of the affected protein is available. The recently determined structure of a fasciclin I fragment, consisting of two FAS1 domains, provides a template for the entire FAS1 domain superfamily [5]. The βig-h3 protein closely resembles fasciclin I in that it consists of an uninterrupted tandem of 4 FAS1 domains. Here this new structural information was used to build a homology model of the βig-h3-FAS1 domain 4 (the disease-causing mutations are clustered in this domain), and a detailed examination was carried out. Before describing the results of this analysis an introduction is given to the βig-h3 protein and the corneal dystrophies resulting from BIGH3 mutations.

**Expression and biochemical properties of βig-h3:** The BIGH3 gene is expressed by a large number of human tissue types, including the corneal epithelium of the eye [1,6]. Despite the initial characterization of βig-h3 as a TGF-β-induced gene product [1], the expression of βig-h3 is not induced by TGF-β in corneal epithelial cells, where there is an undetectable level of TGF-β mRNA [6]. A study of the embryonic expression of the mouse BIGH3 gene has shown that it is expressed in the mesenchyme of most developing organs and the developing tissues of the heart, pancreas and eye [7]. The expression pattern of βig-h3 in developing bovine tissues resembles that of type VI collagen and, although less widely distributed than type VI collagen, the βig-h3 protein is prominently present in the extracellular matrix of a wide range of developing and mature tissues [8]. For example, βig-h3 is found in skin, especially the papillary dermis [9,10]. In normal corneas (Figure 2), the βig-h3 protein is detected throughout the stroma, in Bowman’s layer and in endothelial cells [6,8,11-13]. In healing and diseased corneas, βig-h3 has been identified in the retrocorneal fibrous membrane (which develops posterior to the pre-existing Descemet’s membrane) in the stromal keratocytes [13,14], and as deposits in the corneal stroma in the case of corneal dystrophies [12].

It appears that βig-h3 has a role in corneal development and healing, as well as additional functions in other tissues [15-17], but the precise physiological role of βig-h3 is still largely unknown. Biochemical studies suggest that βig-h3 may...
be a component of certain extracellular matrices. βig-h3 obtained from pig cartilage binds to type I, II and IV collagens [18]. Also, βig-h3 copurifies, and is disulphide linked, with type VI collagen from rabbit cornea [19]. Recombinant βig-h3 binds to type I collagen and fibronecitin [20].

A number of studies have shown that βig-h3 supports cell adhesion and spreading and is an integrin ligand [9,21-23]. Chondrocyte and fibroblast spreading is enhanced by βig-h3 and this enhancement requires divalent cations (Mg$^{2+}$ or Mn$^{2+}$) and is reduced by EDTA. This activity does not require the RGD motif near the βig-h3 carboxy terminus and antibody inhibition showed that cell spreading is dependent on α$\text{II}β$1 integrin [21]. Another study, using human corneal epithelial cells and site-directed mutagenesis, showed that an Asp-Ile motif, conserved in both the second and fourth FAS1 domain of βig-h3 (Figure 1), is essential for cell adhesion. In this case the functional receptor for βig-h3 was identified as α$\text{III}β$1 integrin [22]. Yet another integrin, α$\text{II}β$5, has been implicated in the adhesion of fibroblasts to βig-h3. All four FAS1 domains were independently capable of mediating cell adhesion and mutation of two amino acid residues, an adjacent tyrosine and histidine conserved in all FAS1 domains (Figure 1), abolished integrin binding [23]. While there can be no doubt that βig-h3 is cell-adhesive, the mutagenesis data have to be treated with caution as no FAS1 domain structure was available to guide the mutagenesis (see below).

**BIGH3-Linked corneal dystrophies:** Genetic studies have shown that a group of hereditary corneal dystrophies are caused by a number of missense mutations in the *BIGH3* gene, and the different mutations are linked to particular forms of the disease [4,24,25]. These BIGH3-linked corneal dystrophies are all autosomal dominant and may present amyloid deposits, granular deposits, or a mixture of both. Homozygous BIGH3 mutations lead to severe corneal dystrophies with early onset [26-30].

Lattice corneal dystrophy (LCD) is the most common corneal dystrophy and is usually bilateral amyloidosis characterized by refractile lattice lines. Ten types have been characterized phenotypically but there is genetic redundancy between these (LCD type I, IA, II and IIIa, IIIb, IV, V, VI, and VII) six of these are due to mutations in *BIGH3* (LCD type I, IA, IIIa, IV, VI, and VII, Table 1) [25,31,32]. Lattice corneal dystrophy type I (LCDI) is a bilaterally symmetrical corneal disorder that is characterized by numerous translucent fine lattice lines that are associated with white dots and faint haze in the superficial and middle layers of the central stroma. The symptoms appear during the first or second decades of life. In LCDIIIa, visual disturbances occur in the fourth decade of life. In LCDIIIa, visual disturbances occur in the fourth decade of life.
to sixth decades and the lattice lines, which extend from limbus to limbus, are located in the mid to deep stroma of the cornea. LCDIIIA has frequent recurrent corneal erosion.

There are other autosomal dominant corneal dystrophies caused by mutations in \textit{BIGH3} which are granular in nature [25,32]. Granular corneal dystrophy (GCD) is characterized by the deposition of sharply demarcated opacities in the anterior central stroma. As the disease advances the opacities increase in size and number and may coalesce to form larger lesions. The limbus is typically spared. Dystrophies in this group are granular corneal dystrophy Groenouw type I (GCDGI) and granular corneal dystrophy type II (GCDII), which is a combined LCD-GCD caused by mutations in \textit{BIGH3} and was originally known as Avellino corneal dystrophy.

Reis-Bücklers corneal dystrophy (RBCD) is another autosomal dominant dystrophy caused by mutations in \textit{BIGH3}. In the first two decades of life Bowman’s layer is replaced with fibrocellular scar tissue resulting in corneal opacification. Rod-like deposits are seen. Thiel-Behnke corneal dystrophy (TBCD) is phenotypically closely related to RBCD, but curly fibers are deposited in Bowman’s layer.

Other stromal corneal dystrophies have been characterized, which have similar clinical features, but are not genetically related to LCDI, LCDIIIA or GCDGI. Also characterized are a number of corneal dystrophies that are not associated with \textit{BIGH3} mutations and affect the epithelial, endothelial or Descemet’s membrane layers of the cornea [32].

The amino acid substitutions in the β-h3 protein, which result in autosomal dominant corneal dystrophies, are listed in Table 1. A single missense mutation, or in some cases a number of single mutations, are responsible for each distinct corneal dystrophy. The corneal dystrophies, categorized by clinical characteristics, do not necessarily occur by the same biochemical mechanisms. Deposits have different rates of onset and occur at varying locations within the cornea, depending on the particular mutation. A number of different amino acid mutations can individually cause the same type of dystrophy, for example R555W and R124S both cause GCDGI. Conversely, different mutations of the same amino acid residue can cause different types of corneal dystrophy. For example, R124S causes GCDGI but a change of Arg124 to Cys causes LCDI. Two major hotspots for mutation, amino acid residues

\begin{table}[h]
\centering
\caption{Missense mutations in \textit{BIGH3} that cause corneal dystrophies}
\begin{tabular}{ llllll }
\hline
Corneal phenotype & Mutation in \textit{BIGH3} & Location & Onset/patient age (years) & Lattice/amyloid & Predicted secretion \\
\hline
LCDI/IA & R124C & sub-epithelial and stromal & <20 & L & + + \\
LCDIIIA & N622K & Bowman’s layer and stromal & <40 & L & + - \\
LCDIIIA/IIIB & N622H & stromal & >40 & L & + ? \\
LCDIIIA/V & A546T & sub-Bowman’s layer & >40 & L & + ? \\
LCDI/IIIA & G623D & sub-epithelial and stromal & >30 & L & + - \\
 & H626P & stromal & >30 & L & + ? \\
 & T538R & sub-epithelial & <20 & L & + - \\
 & L518R & sub-epithelial and stromal & >40 & L & + - \\
 & L518P & sub-epithelial and stromal & <20 & L & + ? \\
 & N544S & & & L & + ? \\
LCDI/IIIA/IIIB/VII & H626R & stromal & >40 & L & + - \\
LCD-deep/IV & V631D & stromal-Descemet’s membrane & >40 & L & + - \\
 & L527R & stromal-Descemet’s membrane & >60 & L & + ? \\
GCDGI & R555W & sub-epithelial and stromal & <20 & G & - + \\
 & R124S & stromal & >50 & G & - + \\
GCDII & R124H & stromal & >50 & L + G & + + \\
RBCD & R124L & Bowman’s layer and stromal & <20 & G & - + \\
TBCD & R555Q & Bowman’s layer & <20 & G & - + \\
\hline
\end{tabular}
\end{table}

The abbreviations used are: LCD is lattice corneal dystrophy, GCDGI is granular corneal dystrophy Groenouw type I, GCDII is granular corneal dystrophy type II, RBCD is Reis-Bücklers corneal dystrophy, and TBCD is Thiel-Behnke corneal dystrophy.
Arg124 and Arg555, have been recognized in the βig-h3 protein. Between them, mutations at these two positions cause five clinically distinct hereditary corneal dystrophies and are responsible for approximately 50% of all βig-h3 related corneal dystrophies that are diagnosed [25]. A mutation in βig-h3 of Arg555 to Gln is responsible for TBCD (a phenotype also caused by non-βig-h3 genetic mutations), and Arg555 to Trp for GCDGI. Arg124 seems to be particularly critical in corneal dystrophies as four different phenotypes are associated with four different mutations of this residue. These mutations are: Arg124 to Cys in LCDI, Arg124 to Ser in GCDGI, Arg124 to His in ACD, and Arg124 to Leu in RBCD. The reason(s) for the alteration of protein behavior following changes of residues Arg124 and Arg555 remain unknown.

Many of these corneal dystrophies are difficult to diagnose in a clinical examination. Historically they have been categorized in anatomical terms according to the layer(s) of the cornea involved. This has led to disputes among clinicians. For example, there has been confusion in the literature over TBCD, which has been described as RBCD, but is a separate type of corneal dystrophy affecting Bowman’s layer [33]. Diagnosis would be more accurate if it were based on the underlying genetic cause rather than the clinical description of the disease [25,34]. Classification schemes based on the genetic or molecular defect responsible for the dystrophy have been initiated [31,32,35] and will be important for the long-term therapeutic approach to dystrophies [36]. To make full use of the genetic information, and, in particular, to relate the genetic defect to the biochemical disease mechanism, it is important to know the three-dimensional structure of the affected protein. Until very recently, however, no structural information on the FAS1 domain was available.

The FAS1 domain structure: The structure of a FAS1 domain pair from fasciclin I has recently been reported [5]. The FAS1 domain fold (Figure 3) includes two orthogonal β-sheets arranged as a wedge and sharing the curved β-strand, β6. One β-sheet is central to the domain (β2-β1-β7-β6) and one β-sheet is solvent exposed (β4-β5-β6-β7). There are two N-terminal α-helices which cover the central β-sheet and two solvent exposed helices which connect β-strands in the central β-sheet. A fifth helix connects the two β-sheets and closes the open end of the β-wedge.

βig-h3 and fasciclin I are members of a large protein superfamily containing FAS1 domains. Proteins containing FAS1 domains have been identified and their sequences aligned in the SMART and Pfam databases. It is unlikely that function is conserved throughout the FAS1 domain superfamily. However, homologous sequences adopt the same protein fold. This is illustrated by the close structural similarity of the two FAS1 domains in the crystal structure of fasciclin I. Domain 3 and 4 of fasciclin I superimpose with a root-mean-square deviation of 1.7 Å for 105 Cα atoms despite a sequence identity of only 16% [5]. The βig-h3 protein closely resembles fasciclin I in that it consists of an uninterrupted tandem of four FAS1 domains (Figure 1). In fact, the four internal repeats in the sequences of βig-h3 and fasciclin I have higher homology with each other than the domains of each protein do compared internally. For example, domain 3 and 4 of fasciclin I share 16% sequence identity while domain 4 of βig-h3 and domain 4 of fasciclin I share 19% sequence identity. The βig-h3 protein has approximately 80 additional residues at the N-terminus compared to fasciclin I.

There is a substantial domain interface between domain 3 and 4 of fasciclin I which bury a total of 1,700 Å² of solvent accessible surface area with two N-acetylglucosamine moieties at Asn441 accounting for 140 Å² of this area. This residue is not conserved in the βig-h3 protein but some other features of the linker/interface region are conserved, for example ωL. It is not possible to predict whether the arrangement of domain 3 and 4 seen in the fasciclin I crystal structure is conserved in βig-h3.

All mutations of βig-h3 that cause corneal dystrophy are located in FAS1 domain 4, with the single exception of Arg124 in FAS1 domain 1. From sequence alignment, Arg124 is located in the turn between helices α1 and α2 of FAS1 domain 1 (Figure 1) [5]. In the initial report on the FAS1 domain structure, a brief analysis of βig-h3 mutations was presented [5]. Here, we have generated a homology model of βig-h3 domain 4, using fasciclin I domain 4 as a template and we examine, in detail, the causal amino acids in BIGH3-linked corneal dystrophies. Also, the amino acid residues implicated in integrin binding have been analysed and their availability for binding assessed.

METHODS

The homology model of the βig-h3 FAS1 domain 4 was generated by the MODELLER program [37]. MODELLER implements a fully automated comparative protein structure modelling by satisfaction of spatial restraints [37,38]. The core modelling procedure begins with an alignment of the sequence to be modelled (‘target’) with a related known three-dimensional atomic structure (‘template’). This alignment is the input to the program along with the atomic coordinates of the template structure. The output is a three-dimensional homology model for the target sequence containing all main chain and side chain non-hydrogen atoms. In the present case, the template was the crystal structure of domain 4 of Drosophila fasciclin I and the target was domain 4 of human βig-h3. The sequence alignment used as input to MODELLER was based on the structural superposition of fasciclin I domains 3 and 4 (Figure 1) [5]. The local environment of each residue, involved in corneal dystrophy formation, has been illustrated using the programs MOLMOL [39] and POVRay.

RESULTS & DISCUSSION

Quality of the homology model: The homology model of FAS1 domain 4 of βig-h3, generated by MODELLER, is necessarily very similar to the experimentally determined structure of FAS1 domain 4 of fasciclin I. The crystal structure (template) and homology model superimpose with a root mean squared deviation of 1.2 Å for the backbone Cα atoms (19% sequence identity). By comparison, the two FAS1 domains in the crystal structure of fasciclin I superimpose with an root mean squared deviation of 1.8 Å for the backbone Cα atoms (16%
sequence identity. The degree of change between the fold of the target homology model compared to the template crystal structure is quite small, the amino acid residues of the target structure are sensibly accommodated in the FAS1 domain fold. Analysis by the program PROCHECK [40] showed that 85.2\% of residues appear in most favored regions of the Ramachandran plot, while 12.2\% appear in allowed regions, and only two residues appear in disallowed regions. Visual examination of the structure showed that amino acid residues are distributed as expected, with hydrophobic residues packed in the core and hydrophilic residues located on the surface. Helix α5 is rather hydrophobic in the βig-h3 homology model of domain 4. Since the sequence is less well conserved in FAS1 family members in this region (Figure 1) the model may be incorrect at α5.

**Figure 3. Location of missense mutations in βig-h3 causing corneal dystrophies.** A: Shown is a cartoon representation of the model of domain 4 of βig-h3. Secondary structure elements of the fold are labelled. The positions of amino acid residues mutated in corneal dystrophies are marked by yellow spheres and labelled. **B:** Same as A, but with residues shown in atomic detail. Carbon atoms are yellow, oxygen atoms red, and nitrogen atoms blue. Hydrogen bonds are indicated by black broken lines. The web version of this article contains an animation providing a 360° view of the FAS1 domain fold, represented here by the model of domain 4 of βig-h3. A representative frame from the movie is included here as a still image.

** Likely effect of βig-h3 mutations in corneal dystrophies:** In principle, disease mutations in βig-h3 could affect corneal opacity by interfering with protein folding or by altering specific binding interactions of βig-h3. A mutation in the core of the protein could have a subtle effect on protein stability manifesting over time, or a dramatic effect where the protein would not fold properly and not be secreted [41-44]. In the former case, the βig-h3 protein may become insoluble and aggregate, perhaps with other extracellular matrix proteins. An Arg to His mutation in the lens protein γD-crystallin causes a loss of solubility and a peculiar crystallization of the protein leading to cataract [43]. If mutant βig-h3 is not secreted at all, there may be adverse effects on binding partners due to the decreased amount of βig-h3 available for binding (in a heterozygous patient there would be half the amount of βig-h3). In the following, we will discuss each individual mutation shown to cause a corneal dystrophy (Figure 3 and Figure 4).

More than half of the βig-h3 mutations affect one of two arginine residues, Arg124 in FAS1 domain 1, and Arg555 in FAS1 domain 4 [25]. From sequence alignment Arg124 is located in the turn between α1 and α2 of domain 1 [5]. Arg555 is located in the α3-α4 linker of domain 4 (Figure 3), which is the most variable region of the FAS1 fold family and has an additional α-helix in domain 4 of fasciclin I compared to domain 3. Both Arg124 and Arg555 are predicted to be solvent exposed and it is unlikely that mutation of either arginine would result in misfolding. Arg124His and Arg124Cys cause amyloid deposits, whereas the other Arg124 and Arg555 mutations, mostly to hydrophobic residues, cause granular deposits [28]. Assuming mutant βig-h3 is secreted, a decrease in protein solubility would be expected to cause the extracellular aggregation of mutant βig-h3. It is also possible that specific binding properties mediated by Arg124 and Arg555 are abolished. Interestingly, in corneal dystrophy patients, mutant forms of βig-h3 do not form amyloid deposits in the skin, indicating that βig-h3 may be involved in corneal specific protein-protein interactions [10].

Unlike the common mutations of Arg124 or Arg555, the much rarer βig-h3 missense mutations [25] alter conserved core residues in FAS1 domain 4 of βig-h3. Since regions of high sequence conservation can be modelled reliably, it is possible to assess the likely effect of a given mutation with confidence. Leu518, which is altered in one form of LCDL is situated in helix α1 and is surrounded by hydrophobic residues which are part of a stable hydrophobic core between the N-terminal helices αL, α1, and α2, helices α3 and α4, and the central β-sheet (Figure 3 and Figure 4A). Residues surrounding Leu518 are Phe515 in αL, Ile522 in α1, Val539 in α2, Ala541 in β1, Phe547 in α3, and Leu569 in α4. Due to the tight packing of hydrophobic residues around Leu518, the large and charged side chain of an arginine could not be accommodated at this position without drastic structural changes. It seems unlikely, therefore, that L518R βig-h3 could fold properly and be secreted. Substitution of Leu518 by proline, while not necessarily leading to steric side chain clashes, would disrupt helix α1.
Leu527 is part of a hydrophobic groove on the surface of FAS1 domain 4 of βig-h3 (Figure 3 and Figure 4B). The residues that line the groove are all hydrophobic in nature; Ile522 in α1, Ala525 in the turn between α1 and α2, Leu531 in α2, Ala566 in α4, Leu569 in α4, Ile573 in β2, and Leu592 in β3. Due to the tight packing of hydrophobic residues around Leu527, an arginine at this position could not be accommodated without drastic structural changes. If the protein rearranged to accommodate the large and charged side chain of arginine, it would suffer decreased stability. It is possible that the hydrophobic groove has some functional significance, in which case the mutation would disrupt a protein-protein interaction.

Thr538, situated in β1, makes a stabilizing hydrogen bond across secondary structure elements of the FAS1 domain fold (Figure 3 and Figure 4C). Thr538 hydrogen bonds with the backbone nitrogen of the extended segment preceding α5. This residue is completely conserved throughout the FAS1 superfamily, implying that it is crucial for the domain fold. A positively charged arginine at position 538 would clash with a number of residues of β2, the extended segment preceding α5, β7, or the side chain of Leu589 in β3. T538R βig-h3 is not expected to fold and be secreted. In agreement with the predicted dramatic effect of the T538R mutation, amyloid deposits appear in the first decades of life.

Asn544 is located in α3 and makes two hydrogen bonds with backbone atoms in αL and α1 (Figure 3 and Figure 4D). These hydrogen bonds reinforce the local structure at the turning point between αL and α1 and the start of α3. Mutation of this residue to a serine abolishes these hydrogen bonds. Again, the conservation of this residue in the FAS1 superfamily suggests its importance for the integrity of the FAS1 fold. It is possible, however, that N544S βig-h3 is secreted, but is significantly less stable than wild-type protein.

Ala546 is situated at the start of α3 (Figure 3 and Figure 4E). The corresponding residue is an alanine or glycine in all FAS1 domain family members. It is obvious from the model that there is restricted space for a side chain at position 546. The surrounding residues, Pro542 and Thr543 at the end of β1, Ala546 and Phe547 in α3, as well as Asn630 and Val631 in the C-terminal extended section, would clash with any residue larger than alanine. The A546T mutation either abolishes secretion altogether, or results in severely destabilized protein.

The situation is similar for Gly623, located in β7, which is hemmed in by residues Val505 in αL, Val536 in β1, Glu576 in the extended section preceding α5, and Ala620 at the end of β6 (Figure 3 and Figure 4F). A mutation of this glycine to the much larger and charged aspartic acid is incompatible with the FAS1 fold. G623D βig-h3 is unlikely to be secreted.

The neighbor of Gly623, Asn622 makes hydrogen bonds with backbone atoms of αL and α2, thus holding in place the two α-helices that cover the central β-sheet (Figure 3 and Figure 4F). A lysine or histidine at this position would not be able to make equivalent hydrogen bonds, and there is no room for the larger side chain of lysine. In fasciclin I, the residue corresponding to Asn622 is buried in the domain interface to the preceding FAS1 domain 3. If this domain arrangement is conserved in βig-h3, mutation of Asn622 may have a dramatic effect on the overall protein structure.

His626 makes a stabilizing hydrogen bond linking β7 to the extended section preceding α5 (Figure 3 and Figure 4G). Replacement of this residue with arginine or proline would abolish this hydrogen bond. The conservation of this residue in the FAS1 superfamily suggests it is important for the fold. It is extremely unlikely that H626R βig-h3 would fold and be secreted since, in all accessible conformations, an arginine side chain would clash with Phe540 in β1, Val608 in β5, Val579 in the extended section preceding α5, or Pro616 in β6. A proline could be accommodated sterically at position 626, but would disrupt β7 and severely destabilize the domain.

Finally, Val631 in the C-terminal section of the domain contributes to the hydrophobic core between the central β-
sheet and the N-terminal $\alpha$-helices (Figure 3 and Figure 4H). Pro542 in $\beta_1$, Ala546 and Phe547 in $\alpha_3$, Ile568 and His572 in $\alpha_4$, and Ile628 in $\beta_1$ all tightly pack around Val631. Mutation to aspartic acid would introduce a negative charge into the hydrophobic core and is very unlikely to be tolerated. We predict that V631D $\beta$ig-h3 is not able to fold.

In summary, it is possible to distinguish at least two routes for corneal dystrophy formation resulting from point mutations in the $\beta$ig-h3 protein. The common mutations at positions 124 and 555 may directly affect protein-protein interactions (either homo- or heterophilic), whereas the rare mutations are likely to cause misfolding within the cell. In some of the latter cases, it is possible that mutant protein escapes the quality control of the secretory pathway, but it will be severely destabilized and unable to withstand the harsh extracellular milieu.

**Integrin binding:** Mutation of the Asp-Ile sequence motif in FAS1 domains 2 and 4 (residues 617 and 618 in domain 4) abolishes the ability of $\beta$ig-h3 to support cell adhesion via $\alpha_3\beta_1$ integrin [22]. The aspartic acid residue is not buried in the FAS1 domain core and is, in principle, available for receptor binding (Figure 5). However, in the fasciclin I structure the Asp-Ile motif in domain 4 is buried in the interface to the preceding FAS1 domain 3. Even if the relative domain orientation is not the same in $\beta$ig-h3, the close proximity of the N-terminus (residue Met502) of FAS1 domain 4 to the Asp-Ile motif (Figure 5) makes it almost certain that domain 3 contacts domain 4 in this region. Thus, it appears unlikely that the Asp-Ile motif is directly involved in integrin binding. Mutation of the motif may have corrupted the protein structure sufficiently to abrogate integrin binding to an unidentified site distant to the Asp-Ile motif. This interpretation is further supported by the high conservation of the motif across functionally unrelated molecules, which are unlikely to all bind $\alpha_3\beta_1$ integrin. The aspartic acid of the motif forms a crucial hydrogen bond linking $\beta_6$ to $\alpha_5$ and, despite its surface location, appears to be a structurally important residue in most FAS1 domains.

A different two-residue motif, Tyr-His, has been implicated in $\alpha_5\beta_5$ integrin binding to $\beta$ig-h3 [23]. The histidine (residue 572 in domain 4) makes a hydrogen bond linking $\beta_2$ to the C-terminal region of domain 4 and is completely buried by hydrophobic residues in the FAS1 domain core (Figure 5). Clearly, this residue cannot be involved in integrin binding. Further, this residue is completely conserved across FAS1 superfamilies, which are unlikely to all bind integrin $\alpha_5\beta_5$. The tyrosine (residue 571 in domain 4) is slightly more exposed near the C-terminus of the domain but there are no additional surface features, such as a cluster of hydrophobic or charged amino acid residues, implying the presence of a protein-protein interface. It is also worth noting that the domain boundaries of some constructs used in the integrin binding studies are unlikely to give rise to stable, folded domains. For example, D-I [22,23] consists of residues 129-241, this construct starts half-way through $\alpha_2$ leaving the hydrophobic surface of the central $\beta$-sheet exposed where it would naturally be covered by $\alpha_1$. Finally, the possibility that $\alpha_5\beta_5$ integrin could bind $\beta$ig-h3 via the RGD motif was not addressed thoroughly in these studies. In summary, the structural analysis of mutations previously used to map integrin binding sites in $\beta$ig-h3 shows that it is unlikely that such sites have been identified. Our critical analysis also highlights the dangers of using site-directed mutagenesis in mapping protein-protein binding sites in the absence of any structural information. In particular, mutation of residues conserved across a functionally divergent superfamily is unlikely to be a promising strategy to map specific interaction sites.

**Concluding remarks and outlook:** Mutations in the gene coding for the extracellular matrix protein $\beta$ig-h3 are responsible for a group of hereditary corneal dystrophies [4,24,25]. These dystrophies are characterized by a progressive alteration of the corneal structure, resulting in loss of its transparency and eventually blindness. The current treatments for corneal dystrophies are keratoplasty and phototherapeutic laser keratectomy. Unfortunately, the benefits of these therapies are transient since the protein deposits that characterize the $BIGH3$-linked corneal dystrophies generally are recurrent [45,46]. In order to develop more successful treatments for corneal dystrophies, an understanding of the disease mechanism is essential. Recent achievements, notably the accurate diagnosis by genetic means, have been a great advance towards this goal.

The first structural information about the FAS1 domain [5] has provided a foundation to analyse likely routes of dystrophy formation resulting from missense mutations in the $\beta$ig-

---

*Figure 5. Residues implicated in integrin binding viewed on the surface of domain 4 of $\beta$ig-h3. A and B: Space-filling representation of domain 4 of $\beta$ig-h3. The two views are related by a 180° rotation about a vertical axis. The N- and C-termini of the domain are colored pink and labelled (Met502 and Pro634 respectively). The postulated integrin binding amino acids are in yellow (carbon atoms) and red (oxygen atoms) and are labelled. C and D: Electrostatic surface representations in identical orientations to A and B, respectively. Positive and negative potential is indicated by blue and red coloring, respectively. Residues implicated in integrin binding are colored yellow and red as in A and B.*
h3 protein. It is postulated that the rare mutations affecting FAS1 domain 4 of βig-h3 cause misfolding of the protein within the cell, whereas the common mutations at positions 124 and 555 affect protein-protein interactions directly. Clearly, these structure-based predictions now will have to be tested by recombinant expression of normal and mutant βig-h3. The observation that mutant forms of βig-h3 do not form amyloid deposits in the skin of corneal dystrophy patients [10] indicates that βig-h3 may require other cornea-specific factors to form the abnormal deposits that are the hallmark of BIGH3-linked corneal dystrophies. One established biological activity of βig-h3 that may be affected in corneal dystrophies is its ability to mediate cell adhesion via integrins [9,21-23]. For reasons outlined in this paper, we believe that these studies have not identified the integrin binding sites in βig-h3. In the absence of an experimental structure of βig-h3, the homology model presented in this paper will be useful to inform future studies into the physiological roles of βig-h3 and its malfunction in corneal dystrophies.

ACKNOWLEDGEMENTS

We thank Dr. Christine Slingsby for critical reading of the manuscript. This work was supported by a Wellcome Senior Research Fellowship in Basic Biomedical Sciences to E.H.

REFERENCES


32. Klintworth GK. The molecular genetics of the corneal dystrophies—current status. Front Biosci 2003; 8:d687-713.


44. Wu WW, Molday RS. Defective discoidin domain structure, subunit assembly, and endoplasmic reticulum processing of retinoschisin are primary mechanisms responsible for X-linked retinoschisis. J Biol Chem 2003; 278:28139-46.


The print version of this article was created on 11 Sep 2003. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.