This is a repository copy of Structure and function of a bacterial Fasciclin I Domain Protein elucidates function of related cell adhesion proteins such as TGFBIp and periostin.

White Rose Research Online URL for this paper: http://eprints.whiterose.ac.uk/75432/

**Article:**
Moody, Robert G. and Williamson, Mike P. (2013) Structure and function of a bacterial Fasciclin I Domain Protein elucidates function of related cell adhesion proteins such as TGFBIp and periostin. FEBS Open Bio, 3. pp. 71-77. ISSN ISSN: 2211-5463

https://doi.org/10.1016/j.fob.2013.01.001

**Reuse**
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Structure and function of a bacterial Fasciclin I Domain Protein elucidates function of related cell adhesion proteins such as TGFBIp and periostin

Robert G. Moody, Mike P. Williamson

Dept. of Molecular Biology and Biotechnology, Firth Court, Western Bank, University of Sheffield, Sheffield S10 2TN, UK

Abstract

Fasciclin I (FAS1) domains have important roles in cell adhesion, which are not understood despite many structural and functional studies. Examples of FAS1 domain proteins include TGFBIp (βig-h3) and periostin, which function in angiogenesis and development of cornea and bone, and are also highly expressed in cancer tissues. Here we report the structure of a single-domain bacterial fasciclin I protein, Fdp, in the free-living photosynthetic bacterium Rhodobacter sphaeroides, and show that it confers cell adhesion properties in vivo. A binding site is identified which includes the most conserved region and is adjacent to the N-terminus. By mapping this onto eukaryotic homologues, which all contain tandem FAS1 domains, it is concluded that the interaction site is normally buried in the dimer interface. This explains why corneal dystrophy mutations are concentrated in the C-terminal domain of TGFBIp and suggests new therapeutic approaches.

© 2013 The Authors. Published by Elsevier B.V. on behalf of Federation of European Biochemical Societies. All rights reserved.

1. Introduction

Members of the fasciclin I family of proteins (FAS1) occur in a wide range of vertebrates, invertebrates and microorganisms. A bioinformatics study concluded that the domain fold is ancient, traceable back to the Last Universal Common Ancestor [1], implying a likely common function across all phyla. They are generally cell-surface and membrane-anchored proteins involved in homophilic cell adhesion or symbiotic processes. One of the earliest and best studied link that is developmentally regulated, resulting in variable levels and membrane-anchored proteins involved in homophilic cell adhesion or symbiotic processes. One of the earliest and best studied examples is Drosophila FAS1, which is expressed during embryonic development, and guides axons from axon-generating neural cells to other target neurons or muscle cells [2–4]. FAS1 domains do not span the membrane, but are attached to the membrane via a lipid link that is developmentally regulated, resulting in variable levels of soluble and membrane-anchored proteins during embryogenesis [5,6]. Examples in mammals include transforming growth factor-β-induced gene product (TGFBIp, formerly known as βig-h3) [7], periostin [8–10], also known as osteoblast-specific factor 2 (OSF-2) [11], and stabilins 1 and 2, also known as scavenger receptor FEEL-1 and -2 proteins [12]. Mutations in TGFBIp are linked to corneal dystrophies, while periostin is required for development of tooth, bone and heart [13]. Many of these mammalian proteins are found expressed at high levels by tumour cells, presumably because of their roles in cell adhesion and angiogenesis, and they have been proposed both as tumour markers and therapeutic targets [13–15]. Several have been shown to bind to integrin cell surface receptors [8,10,16,17] including periostin which is suggested to be a ligand for αvβ3 integrin [16]. Knock-out mutations seldom exhibit discernible phenotypes. However, when combined with mutations in other linked signal transduction loci, distinct phenotypes can be observed, as shown by accompanying mutations in the abl tyrosine kinase in Drosophila, which results in defective axon tracts [2]. Amongst plants, fasciclin I-like domains occur widely as a major subgroup of the cell surface arabinoxylan proteins required for plant growth and development [18,19], and as the Arabidopsis thaliana SOSS protein required for normal cell expansion [20,21]. Microbial fasciclin I proteins include the antigenic MPB70 protein secreted by Mycobacterium bovis, identical to M. tuberculosis MPT70 [22], and proteins important for symbiotic relationships of cyanobacteria [23] and in cnidarian–algal associations [24]. MPB70 is homologous to OSF-2, and adhesion of MPB70 to bone in neonates has been implicated in osteitis following BCG vaccination [25]. In symbiotic rhizobia such as Sinorhizobium meliloti, the fasciclin I protein Nex18 is required for normal nodule formation with leguminous plant partners [26].

FAS1 domains in animals almost always occur in pairs: Drosophila FAS1 has two tandem pairs, as do TGFBIp and periostin, while the stabilins have seven tandem copies [27]. The best characterized system is TGFBIp, where a large number of mutations have been identified that lead to corneal dystrophies [28,29]. Over half of these derive

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: FAS, fasciclin; Fdp, Fasciclin I Domain Protein

* Corresponding author. Tel.: +44 114 222 4224; fax: +44 114 222 4243.
E-mail address: m.williamson@sheffield.ac.uk (M.P. Williamson).

2211-5463/$36.00 © 2013 The Authors. Published by Elsevier B.V. on behalf of Federation of European Biochemical Societies. All rights reserved.
http://dx.doi.org/10.1016/j.fob.2013.01.001
from only two sites, one in FAS1 domain 1 (FAS1-1) and one in domain 4 (FAS1-4). However, almost all the other mutations are found in FAS1-4, the exception being one in the interface between FAS1-3 and FAS1-4.

Despite their low overall sequence conservation, fasciclin I domains are easily identifiable due to the presence of two conserved sequence motifs called H1 and H2. Several FAS1 structures have been reported, namely the crystal structure of a FAS1 domain pair from Drosophila [30], NMR and crystal structures of the FAS1-4 domain from TFGBp [31] (Yoneyama et al., unpublished), and the single-domain MBP70 [32]. No clear binding site or mode of action has emerged [27,30], although a conserved Asp-Ile sequence was shown to be important [8]. In view of the growing clinical importance of FAS1 domains, a greater understanding of the function of these domains is urgently required.

Here we report on the identification of a new member of the fasciclin I family, Fdp (Fasciclin I Domain Protein), a simple single-domain protein found in the photosynthetic bacterium Rhodobacter sphaeroides, which is confirmed as a member of this protein family by determination of its structure. Our study defines a possible role for Fdp in adhesion properties of whole cells, which may be of significance for the bacterium in its natural environment. We identify a probable binding site on Fdp. On comparison to animal FAS1, we conclude that the physiological binding site of FAS1 is buried in a domain interface, and discuss therapeutic implications.

2. Materials and methods

2.1. Bacterial strains and growth conditions

Escherichia coli strains were cultured aerobically in LB. Where appropriate, media were supplemented with 50 µg ml⁻¹ ampicillin and/ or 50 µg ml⁻¹ kanamycin, or 500 µg ml⁻¹ carbenicillin. Plasmid transfer into R. sphaeroides was by conjugative transfer from E. coli S17-1 [33]. R. sphaeroides NBIB 2853 was cultured at 34 °C in M22+-medium [33]; fdp mutants were cultured in M22+ containing 20 µg ml⁻¹ kanamycin. Complementation plasmid pRKfdp was constructed by inserting a 1.1 kb BamHI fragment possessing the intact fdp gene into replicative pRK415 [34], and verified by sequencing.

2.2. Expression of recombinant fdp

Regions 57 to 470 (relative to ATG, where A is position 1) of fdp were amplified by PCR using primers 5'-TCAGCCATATGGAACCGGAGACATCGTGGA-3' (Ndel site underlined) and 5'-GCTAGCATATGATCCGCCGATCCGAC-3' (BamHI site underlined), using pSUP202 as template. The 413-bp fragment was isolated and cloned into Smal-digested pBluescript-SK to give pBLfdp. The presence of inserts with correct sequence was verified by sequencing.

2.3. NMR studies

The fdp gene was cloned into a pET14b vector and expressed in E. coli BL21(DE3). Labelled protein was produced by growth and IPTG induction in M9 minimal medium containing 13C and 15N. Cells were disrupted by sonication and the protein was purified using Ni-NTA chromatography (Qiagen). NMR experiments were recorded on Bruker DRX-500, 600 and 800 spectrometers at 298 K, using 1–2 mM protein in 50 mM sodium phosphate pH 7.0, 0.03% NaN₃, in H₂O containing 10% D₂O. Processing and analysis of the spectra was carried out using Felix (Felix NMR Inc., San Diego, CA). Molecules were viewed with Pymol (Delano Scientific, California; http://www.pymol.org). NOEs were assigned manually as much as possible, with a starting set of 1506 unambiguously assigned NOEs. The structure was calculated in CNS 1.1 [36] using a final set of 1788 distance restraints obtained from NOESY spectra (approximately 13 restraints per residue), and 148 angle restraints from TALOS [37]. Hydrogen bond restraints were added at a later stage in the structure calculation, after the secondary structure was already clearly established, to avoid biasing the calculation. Analysis of the structures calculated using the final set of restraints showed that 50 out of 100 structures calculated had closely similar energies and structures. Thirty of these were refined in ARIA 1.2 using explicit water refinement [38], which resulted in slightly worse restraint violations and a greater difference from ideal values, but a better Ramachandran distribution.

2.4. Cell adherence assays

R. sphaeroides (∼9.6 x 10⁶ cells) were suspended at 34 °C in 10 ml M22+, 10 mM glucose, and 200 µl aliquots introduced into 96-well microtitre plates fitted with 96-peg lids (68.1 mm² submersed area). Adherent cells attached to pegs were counted after 5 days. Pegs were then removed and submersed in sterile distilled water to remove loosely-bound cells and transferred to 200 µl 1/2-strength Ringer’s diluent. Attached cells were removed into diluent using a sonicating water bath for 5 min. Suspensions of adherent cells were then spread-plated onto LBA agar for enumeration. Plates were incubated at 34 °C for 5 days. Additional assays were performed based on the crystal violet assay [39].

2.5. Isolation of the fdp gene and construction of insertionally-inactivated fdp mutants

The fdp gene is located on chromosome 1 (locus RSP1409; http://genome.ornl.gov), between two oppositely transcribed genes (one homologous to the endopeptidase Clp ATP-binding chain B of Mesorhizobium loti, and the other homologous to molybdoopterin binding domains of oxidoreductase enzymes). Therefore, fdp is not co-transcribed with any flanking genes and forms a single-gene operon, and inactivation of fdp is not expected to exert any polarity effects on flanking genes. Sequencing and restriction mapping of the fdp region in the NCIB 2853 strain confirmed that the arrangement is identical to that of the 2.4.1 sequenced strain. An fdp fragment possessing the entire gene was amplified by PCR using primers 5′-ATGCAATCGCTCTCGACGAC-3′ and 5′-CCGGCCATGTCATGTCGAC-3′. PCR was performed with 5% DMSO. The 1.9 kb product was purified and digested with BamHI. The 1.0 kb BamHI fdp fragment was then purified and labelled with digoxigenin using random priming. To isolate fdp-harbouring clones from a R. sphaeroides genomic DNA library by Southern hybridization, the labelled PCR product was used to screen a R. sphaeroides NCIB 2853 genomic library [40]. Hybridization was performed overnight at 65 °C. Membranes were washed in 0.2 × SSC, and detection of fdp-containing clones was by chemiluminescence. One positive clone (pSUP202/fdp-13) possessed fdp approximately centrally on a 4.0 kb HindIII fragment. This fragment was isolated and ligated into HindIII-digested pUC19 to give pUCfdpH-8 which has a unique SgrAI site which cleaves at base position 84 in the fdp gene. The 0.9 kb XmnI-ended Tn5 kanamycin resistance cassette of pUX-Km [41] was isolated and ligated into SgrAI-digested pUCfdpH-8. The orientation of the kanamycin cassette in the clones was checked by restriction digestion, and also by PCR using primers 5′-GTTGTTGCAGCATCCTCGC-3′ (in the fdp promoter region),
and 5′-TTGGTGCTCAGATGCCGAGCTAGCC-3′ (in the kanamycin resistance gene). The correct construct was called pUCfdpK4. The 4.9 kb HindIII fdp::kan fragment was isolated and cloned into HindIII-digested pSU202. The resulting plasmid pSU/pfdpKM was checked by restriction analysis and introduced into R. sphaeroides NCIB 8253. Kanamycin-resistant transconjugants were screened by Southern hybridization to check for loss of the suicide plasmid, and insertion of the kanamycin resistance cassette at the correct chromosomal position. Additional confirmation was obtained by PCR using the primers above with mutant genomic DNA as template.

3. Results

3.1. Structure of fdp

The fdp gene is designated ORF RSP1409 on chromosome 1 in the R. sphaeroides 2.4.1 database (http://genome.ornl.gov/microbial/rsp1). The protein is predicted to possess 155 residues (excluding initiating fMet), with residues 1–18 (RKTLLALSGLAAPPFA) constituting a signal peptide for translocation across the inner membrane. This results in a mature 137-residue protein, possessing the N-terminal sequence ETGDIVETATGA. Here, we number the protein in the full-length sequence, so that the first residue is residue 19. By PSI-BLAST, the closest sequence similarity (60% identical; 74% similar) is to S. meliloti Nextr18. Fdp is also related (39% identity; 55% similarity) to M. bovis MPB70 major secreted protein and Drosophila FAS1-4 (29% identity) (Fig. 1a). The sequence similarities are striking, since FAS1 domains generally exhibit low overall sequence conservation (<20%) [30]. The two regions of high conservation recognized for the FAS1 superfamily (H1 and H2) are also strongly conserved in Fdp. It is a single-domain protein, and is not co-transcribed with any other gene.

Fdp was expressed in E. coli with an N-terminal His6 tag for purification, as residues 19–155 of the full-length protein, which corresponds to the mature protein after cleavage of the N-terminal signal sequence. It constituted 12% of total soluble E. coli proteins and typical yields were 7.5 mg per litre of culture.

The NMR spectrum was sharp and well resolved and was assigned using standard triple resonance experiments on double labelled protein [42]. NMR spectra (particularly 15N relaxation experiments, not shown) indicate that the protein behaves as a monomer in solution, even at NMR concentrations. The structure was calculated using simulated annealing based on distance and angle restraints, and is shown in Fig. 2, with structural statistics in Table 1. It is an α + β structure, consisting of a wedge-shaped β-sandwich of approximately 30 Å diameter made up of two β-sheets, with six α-helices covering one face of the wedge. The structure is similar to those of other FAS1 domains whose structures have been determined: backbone RMSDs to TGFBIp, FAS1-4 and MBP70 are 2.4, 2.4 and 2.2 Å respectively (Fig. 2C and D). The structure of Fdp does not contain the helix d5 present in FAS1-4 (Fig. 2C), and has therefore a clearer split between the α-domain and the β-domain than does FAS1-4.

3.2. Fdp is involved in adherence properties of whole cells

Three independent insertionally inactivated fdp knockout mutants were constructed in R. sphaeroides and compared with wild type in adherence assays. Growth rates of mutant and wild type strains were similar under aerobic, semi-aerobic and anaerobic (photosynthetic) conditions, and there were no significant differences in levels of photosynthetic complexes as revealed by spectrophotometric analyses of dark/semi-aerobically cultured cells (data not shown). The assay measured the ability of stationary phase cells to clump together and thereby adhere to pegs in 96-well plates. Cell adherence was significantly reduced in the fdp mutants compared with the wild type strain (Fig. 3), from 8.8 × 10^2 cells mm^-2 in wild type to 0.87 ± 0.15 × 10^2 cells mm^-2 in the three mutants, confirming a clear role for Fdp in ability to adhere to external surfaces (Fig. 3A). This effect was confirmed to be specific for the fdp mutants and not attributable to the presence of the kanamycin resistance cassette present in these mutants by conducting experiments with other unrelated mutants containing this cassette, in which levels of adherent cells were comparable to wild type (data not shown).

An alternative adherence assay used crystal violet to measure adherence to the well [39]. In this assay, adherence was only reduced 2.5-fold (Fig. 3B). However, when the mutants were transformed with the complementation vector pRKfdp, almost full complementation (91%) by the fdp gene was achieved.

We have thus shown that Fdp in R. sphaeroides has a similar function to that in other members of the FAS1 family, namely cell adhesion. In bacteria, cell adhesion plays many important roles, particularly in the formation of biofilms, which is an important feature of many colonizing bacteria [43]. R. sphaeroides is however not pathogenic and lives in aquatic environments. It can grow chemoheterotrophically in the dark or light, photosynthetically in anaerobic environments or by anaerobic respiration in the dark [44]. A regulatable ability to aggregate would give it much greater control over its location. The ability to adjust its depth in the water column in response to environmental signals is thus likely to be crucial to its ability to move to suitable locations. In this context, it is significant that the expression of Fdp is regulated by redox status, being downregulated by the Prr

<table>
<thead>
<tr>
<th>Table 1. Structural statistics for Fdp structure determination.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unrefined ensemble</strong></td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td><strong>Restraint violations</strong></td>
</tr>
<tr>
<td>NOE violations &gt; 0.2 Å</td>
</tr>
<tr>
<td>Dihedral violations &gt; 5</td>
</tr>
<tr>
<td><strong>RMSD from experimental restraints</strong></td>
</tr>
<tr>
<td>Distance restraints (Å)</td>
</tr>
<tr>
<td>Dihedral restraints (°)</td>
</tr>
<tr>
<td><strong>Coordinate precision (Å)</strong></td>
</tr>
<tr>
<td>Backbone</td>
</tr>
<tr>
<td>All heavy atoms</td>
</tr>
</tbody>
</table>

* Ensemble of 30 lowest energy structures picked from 100 in CNS 1.1.  
* Best structure after energy refinement in water using ARIA 1.2.  
* 1788 Restraints, consisting of 619 intra-residue, 404 sequential, 304 medium-range (2 ≤ i − j ≤ 4), 317 long-range (i − j > 4) and 27 pairs of hydrogen bond restraints.  
* Calculated using Procheck-NMR [54].  
* Calculated using CNS 1.1 and ARIA 1.2.
Fig. 1. Sequence alignments of FAS1 domains. (A) Alignments based on structural similarity. Sequences are those of the four FAS1 domains with three-dimensional structures: The Drosophila FAS1-3/4 pair (DFA1: PDB 1b70, domain 4); the fourth FAS1 domain of human TGFBip (TGFB1: PDB 2xpd); the M. bovis secreted protein MPB70 (MPB70: PDB 1nyo); and Fdp (Fdp: PDB 1w7d). The NMR structure of the fourth FAS1 domain of human TGFBip (PDB 1x3b) is very similar to the crystal structure and was not used as an independent structure. Colour code: in DFA1, yellow denotes residues described here as being interacting residues. In TGFBip, blue denotes RS55, one of the two major sites giving rise to corneal dystrophy. Other disease-causing sites are indicated in cyan. In MPB70, cyan indicates suggested interaction sites [32]. Highly conserved and completely conserved residues are indicated on Fdp in yellow and red respectively. Locations of regular secondary structure, and the conserved regions H1 and H2, are indicated below the sequences. (B) Domains 1 through 4 from Drosophila FAS1, TGFBip and periostin, each of which contains four tandem FAS1 domains. The alignments encompass the two regions (separated by a blue box) discussed here as being binding sites. Comparisons are more reliable in the second sequence, which is longer and better conserved. Conserved residues are highlighted in green; the important DI domains. The alignments encompass the two regions (separated by a blue box) discussed here as being binding sites. Comparisons are more reliable in the second sequence, which is longer and better conserved. Conserved residues are highlighted in green; the important DI domains. The alignments encompass the two regions (separated by a blue box) discussed here as being binding sites. Comparisons are more reliable in the second sequence, which is longer and better conserved. Conserved residues are highlighted in green; the important DI domains. The alignments encompass the two regions (separated by a blue box) discussed here as being binding sites. Comparisons are more reliable in the second sequence, which is longer and better conserved. Conserved residues are highlighted in green; the important DI domains.

Fig. 2. The solution structure of Fdp. (A) 10 overlaid structures, shown as a rainbow view, from blue at the N-terminus to red at the C-terminus. Only backbone atoms (Cα, C, N) are shown. (B) Best structure as a cartoon, same colour scheme and orientation. The α-helix and β-sheet numbering is indicated. Labelling of helices and sheets follows that in [30]. This means that the first helix is α1 rather than αL, α4 has a large bend in the middle, β5 is a helical turn rather than a full helix, and β6 is a short strand followed by a longer extended strand. (C) Best fit superposition to Drosophila FAS1-4 (Fdp red, FAS1 yellow). (D) Best fit superposition to M. tuberculosis MPB70 (Fdp red, MPB70 green). Superpositions and RMSD values in the text were based on the most highly conserved regions of secondary structure, corresponding to residues 45–50, 55–65, 82–94, 102–104, 114–122, 126–128 and 135–151 from Fdp. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Our results do not provide any information on the nature of the binding partner of Fdp, except that Fdp shows no indications of dimerizing, even at NMR concentrations, implying that homomeric interactions are unlikely. There are also no other identified fasciclin domains in the R. sphaeroides genome, further ruling out homomeric interactions. In eukaryotic homologs, the ligand is a cell-surface integrin glycoprotein, which is the most likely type of binding partner.

3.3. Database

The atomic coordinates for FDP have been deposited with the Protein Data Bank; PDB: 1w7e (ensemble) and PDB: 1w7d (minimized best structure).

4. Discussion

4.1. Location of the protein interaction site

Here we report the structure of a new member of the FAS1 family, which unusually has only a single FAS1 domain. This new member is predicted to possess a signal peptide at the N-terminus, and the program PSORTb v3.0 [45] predicts a very high probability that it is attached to the cytoplasmic membrane, presumably via a C-terminal covalent anchor, consistent with its role in cell adhesion. Attempts to raise antibodies specific enough to identify the location of Fdp have proven unsuccessful.
The ubiquity of this domain across phyla suggests that it may represent an evolutionarily ancient cell adhesion domain, most likely functioning by binding to cell-surface proteins [1,18]. We therefore looked for residues that are conserved across a wide range of species, and are likely to be functionally important. To this end, we have prepared a new sequence alignment that is based on structural similarity rather than simply sequence similarity, using the existing structures as guides. Our structure of Fdp is important in guiding the alignment, because of the low level of sequence similarity between existing sequences, and the presence of insertions and deletions. The alignment is shown in Fig. 1, and identifies a number of highly conserved residues, in particular the H1 and H2 regions previously identified (Fig. 1a). These regions are adjacent in the structure, and form a large surface patch, followed by two β-strands that form the protein core and emerge on the opposite surface (Fig. 4A). Thus much of the conserved sequence appears to be essential because of its role in maintaining the structure, leaving the most likely binding site as the contiguous surface patch comprising residues 136–144 from H2 plus K50 and D52 from H1 (Fig. 4B).

There have previously been several attempts to identify binding sites on FAS1 domains. An analysis of *M. tuberculosis* MPB70, based on highly conserved residues and disease-inducing mutations, identified the same region as being important, together with other residues on the opposite face of the protein that were suggested to form a second interaction site [32]. The best-supported site is the two residues DI or DV (136–137 in Fdp), at the start of H2. These residues have been shown to be important in cell adhesion of TGFβlp via integrin α3β1, since mutations in these positions showed loss of function, and synthetic pentapeptides containing this sequence blocked cell adhesion [8,17,46].

Other binding sites have also been proposed. In particular, residues Tyr71-His72 were suggested to form an alternative binding site specific for αvβ5 integrin [16]. However, His72 is largely buried in both Fdp and FAS1, implying that it is unlikely to be involved in protein recognition [47]. In addition, several hydrophobic residues flanking Tyr71-His72 were identified as important for interaction with αvβ5 integrin [16]. In Fdp these are generally either absent or buried, again making it unlikely that this site is important for Fdp. We conclude that the most likely binding site for FAS1 domains is residues 136–144 plus 50 and 52 (Fig. 1a).

It is of course possible that the remarkable conservation of the H1 and H2 regions, from bacteria to plants and humans, is unrelated to function, and that our imputation of a binding region here is incorrect.

Against this we would argue that the cell adherence function of the fasciclin I domain is strongly conserved, and that as far as is known, the ligand type is also conserved [1]; that Fdp has a 29% sequence identity with *Drosophila* FAS1-4, this being a high enough similarity to make similarity of function very likely [48]; that the conserved residues identified here are surface-exposed and have no obvious structural role; and that most studies to date on a range of FAS domains have agreed in highlighting this region as the most likely binding site.

### 4.2. Corneal dystrophy mutations affect structural integrity not binding

There have been detailed studies of mutations in TGFβlp, which lead to a range of corneal dystrophies, characterized by amyloid-like protein deposits in the eye. Over half of the cases studied are caused by two mutations, at R124 in FAS1-1 and R555 in FAS1-4. The equivalent position to R555 is not well conserved in Fdp (Fig. 1; in Fig. 2 it is residue 75, just above the text a4 in Fig. 2B). In the *Drosophila* FAS1 structure, the equivalent residue is in a turn, and it was concluded that it should also be exposed in TGFβlp, and consequently mutations here could affect interactions with other proteins [30]. It is however diametrically opposite to the interaction site suggested here, and in our structure corresponds to a partially buried valine. We therefore suggest that mutations of R555 may lead to restructuring of the loop, and thus perturbation to the adjacent H1/H2 strands. In support of this suggestion, we note that different mutations at R555 can have either stabilizing or destabilizing effects [49,50]. Almost all the other disease-causing mutations are at sites that are buried in Fdp, and are therefore likely to lead to instability and consequent amyloid formation, rather than loss of interactions, as also suggested by others [27,29,30,32].

### 4.3. The interaction site is at the dimer interface

The N-terminus of Fdp is immediately adjacent to the proposed binding site, while the C-terminus is on the opposite face of the protein (Fig. 4). Assuming that the membrane attachment site is in its usual location at the C-terminus, then the Fdp binding site is in the most exposed region of the protein, as expected.

There is an important difference for eukaryotic homologs. In these proteins, the FAS1 domains generally occur in pairs. Our most detailed understanding comes from the crystal structure of the FAS1-3/4 pair from *Drosophila*, in which there is a substantial domain interface of 1700 Å² [30]. Mutational studies of the homologous TGFβlp, discussed above, implicate the C-terminal domain as being by far the most important for function. The importance of the C-terminal domain can also be seen by sequence comparisons of *Drosophila* FAS1, TGFβlp and peristin (Fig. 1b), which show that the binding site residues are much more highly conserved in domains 2 and 4 (i.e., the C-terminal domain from each pair) than in the other two domains [51]. Studies using recombinant proteins and antagonist peptides identified domains 2 and 4 as both being important [8]. The clear implication is that the binding site in these proteins is located mainly or entirely on domains 2 and 4, which means that the binding site is more than 50% obscured by the domain/domain interaction (Fig. 4B), implying that binding must involve a competition between intramolecular and intermolecular binding (Fig. 5). The key residues D136 and V137 are almost completely buried in the interface (Fig. 4B). Further support for this hypothesis comes from the observation that one of the two mutations in TGFβlp that is not within domain 4 (P501T) is in the interface between domains 3 and 4, which could potentially disrupt the domain reorientation. Small-angle X-ray scattering (SANS) has suggested that TGFβlp has a ‘beads on a string’ structure, with the four domains roughly extended in solution: there is thus clearly some motional freedom between domains, allowing the C-terminal domain to open out and expose the binding surface when required [52]. Inspection of the *Drosophila* FAS1 structure shows that the interdomain loop
is long enough to allow considerable flexibility.

4.4. Implications of the binding site location

It is common to observe binding sites that are obscured by weak intramolecular binding. Such behaviour is often termed autoinhibition [53], and is used to regulate binding, such that the binding site is not available ‘accidentally’, only presenting when a genuine ligand binds. This reduces the probability of incorrect signal transmission. It can also be used to create further binding sites. Data presented here suggest that autoinhibition may be occurring in eukaryotic homologues of Fdp, with the binding site being the C-terminal domain, and its N-terminal partner serving as an inhibitor (Fig. 5). This may explain why the affinity of FAS1 proteins for their ligands is apparently weak; it also suggests that single C-terminal FAS1 constructs may bind more tightly. It is therefore likely that antagonists based on the C-terminal FAS1 domain would bind more tightly to their ligands than the full-length protein, and could form the basis for useful drug targets.

Acknowledgements

We thank Mary Phillips-Jones, Eun-Lee Jeong and Sam Broad (University of Leeds) for advice and plasmids, Peter J.F. Henderson (University of Leeds, UK) for helpful discussions and for reading an early version of the manuscript, Samuel Kaplan (University of Texas Health Science Center, USA) for discussions and provision of unpublished information, Richard Tunnicliffe for help with structure calculation and Arthur Moir (University of Sheffield) for protein N-terminal sequencing. This work was funded through a White Rose Studentship (RGM).

References


