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

A R T I C L E   I N F O

Article history:
Received 20 December 2012
Received in revised form 3 January 2013
Accepted 7 January 2013

Keywords:
NMR structure
Cell adhesion
Biofilm
Autoinhibition
Domain interface

A B S T R A C T

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 highly 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. Open access under CC BY-NC-ND license.

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 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 dystrophy, 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 believed to be a ligand for αvβ5 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 αβ 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 arabinoxylacan proteins required for plant growth and development [18,19], and as the Arabidopsis thaliana SOS5 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...
from only two sites, one in FAS1 domain 1 (FAS1-1) and one in do-
main 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 do-
mains 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
to TGFBIp [31] (Yoneyama et al., unpublished), and the single-
domain MP870 [32]. No clear binding site or mode of action has
been reported. Fasciclin I family, Fdp (Fasciclin I Domain Protein), a simple single-
component that is not expected to exert any polarity effects on
flanking genes. Sequencing and restriction mapping of the fdp region
in the NCIB 8253 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′-TCAGCCATTGGAACCCGAGACTGCGCA-3′ (Ndel site underlined)
and 5′-GCTAGGATCCATGCGCCGCCGATCGACC-3′ (BamHI site
underlined), using pSUP202-fdp-13 as template. The 413-bp fragment
was isolated and cloned into SmaI-digested pBluescript-SK to give
pBIFDPtr. The presence of inserts with correct sequence was veri-
fied by restriction digest analysis and sequencing. The fdp fragment
of BamHI Ndel-digested pBIFDPtr was cloned into pET14b (Novagen).
The final expression construct, pETfjetr, expresses a Fdp protein with
an N-terminal MGSS(H)6SSGLVPRGSHM sequence followed by Fdp
starting at E19. Fdp was expressed and purified as described [35]
and verified by N-terminal sequencing, electrospray mass spectrometry
and Western blotting.

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 con-
taining 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 cal-
ulation. 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 × 10^7 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 submerged in sterile distilled water to remove
loosely-bound cells and transferred to 200 μl ¼-strength Ringer’s
diluent. Attached cells were removed into diluent using a sonicitc
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.orgl.gov), between two oppositely transcribed genes (one homologous to the endopeptidase Clp ATP-binding chain B of
Mesorhizobium loti, and the other homologous to molybdopterin
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 po-
larity effects on flanking genes. Sequencing and restriction mapping of the fdp region in the NCIB 8253 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′-ATGACATGGCCTGATGCGACGA-3′ and 5′-
CCGGGATCCATGCGCCGCCGATCGACC-3′ (BamHI site
underlined), using pSUP202-fdp-13 as template. The 1.9 kb product was purified and digested with 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′-TCAGCCATTGGAACCCGAGACTGCGCA-3′ (Ndel site underlined)
and 5′-GCTAGGATCCATGCGCCGCCGATCGACC-3′ (BamHI site
underlined), using pSUP202-fdp-13 as template. The 413-bp fragment
was isolated and cloned into SmaI-digested pBluescript-SK to give
pBIFDPtr. The presence of inserts with correct sequence was veri-
fied by restriction digest analysis and sequencing. The fdp fragment
of BamHI Ndel-digested pBIFDPtr was cloned into pET14b (Novagen).
The final expression construct, pETfjetr, expresses a Fdp protein with
an N-terminal MGSS(H)6SSGLVPRGSHM sequence followed by Fdp
starting at E19. Fdp was expressed and purified as described [35]
and verified by N-terminal sequencing, electrospray mass spectrometry
and Western blotting.

2. Materials and methods

2.1. Bacterial strains and growth conditions

Escherichia coli strains were cultured aerobiically 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 conjugal transfer from E.
coli S17-1 [33]. R. sphaeroides NCIB 8253 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.
and 5’-TGTTGGTGTCAATGGCCAGTGGACC-3’ (in the kanamycin resistance gene). The correct construct was called pUCfpdK4-KM. The 4.9 kb HindIII fpd::kan fragment was isolated and cloned into HindIII-digested pSUP202. The resulting plasmid pSUPfpdKM 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://genom.ornl.gov/microbial/rspf). The protein is predicted to possess 155 residues (excluding initiating Met), with residues 1–18 (RTLLALSGLLAAPFA) 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 Nex18. 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 TGFβI, 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 β5 present in FAS1-4 whose structures have been determined: backbone RMSD to TGFβI, FAS1-4 and MBP70 are 2.4, 2.4 and 2.2 Å respectively (Fig. 2C and D).

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 TGFβI, 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 β5 present in FAS1-4 whose structures have been determined: backbone RMSD to TGFβI, FAS1-4 and MBP70 are 2.4, 2.4 and 2.2 Å respectively (Fig. 2C and D).

### Table 1. Structural statistics for Fdp structure determination.

|                      | Unrefined ensemble* | Refined best† |
|----------------------|---------------------|---------------|
| Restraint violations |                     |               |
| NOE violations > 0.2 Å | 0                  | 1             |
| Dihedral violations > 5 | 0                  | 0             |
| RMSD from experimental restraints | 0.12 ± 0.0009 | 0.039         |
| Dihedral restraints (°) | 0.14 ± 0.02 | 0.75          |
| Coordinate precision (Å) |               |               |
| Backbone             | 0.87 ± 0.17         | NA            |
| Dihedral restraints (°) | 1.29 ± 0.55 | NA            |
| Ramachandran analysis |                     |               |
| Most favoured region (%) | 79.1             | 86.4          |
| Additionally allowed region (%) | 19.1            | 11.8          |
| Generously allowed region (%) | 0.9            | 0.9           |
| Disallowed region (%) | 0.9                | 0.9           |
| Energy (kcal/mol) |                     |               |
| Overall              | 165.4 ± 12.3        | -4126.8       |
| Bond                 | 7.4 ± 0.7           | 33.8          |
| Angle                | 74.3 ± 4.0          | 178.6         |
| Dihedral             | 9.8 ± 1.1           | 619.6         |
| VoW                  | 54.4 ± 6.4          | -369.4        |
| Electrostatics       | NA                  | -4863.9       |
| NOE                  | 19.4 ± 2.9          | 124.0         |
| Dihedral (TALOS)     | 0.2 ± 0.1           | 5.1           |
| Difference from ideal values |                   |               |
| Bonds (A)            | 0.012 ± 0.001       | 0.0042        |
| Angles (°)           | 0.37 ± 0.01         | 0.57          |
| Improper (°)         | 0.26 ± 0.02         | 2.12          |

* Ensemble of 30 lowest energy structures picked from 100 in CNS 1.1.
† Best structure after energy refinement in water using ARP (1.2).
‡ 1788 Restraints, consisting of 619 intra-residue, 494 sequential, 304 medium-range (2 ≤ i – j ≤ 4), 317 long-range (i – j > 4) and 27 pairs of hydrogen bond restraints.
§ 148 (74 f and 74 ϕ) obtained from TALOS.
¶ After alignment of backbone residues of mutants 22–154.
* Calculated using Procheck-NMR [54].
† Calculated using CNS 1.1 and ARP 1.2.

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. 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
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 (DFAS1: PDB 1o70, domain 4); the fourth FAS1 domain of human TGFBIp (TGFBI: PDB 2vxp); 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 DFAS1, yellow denotes residues described here as being interacting residues. In TGFBIp, blue denotes R555, one of the two major sites giving rise to corneal dystrophy. Other disease-causing sites are indicated in cyan. The sequence YH, suggested as a possible binding site [16], is shown in magenta. 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/V sequence is marked by asterisks. Domains 2 and 4 are more highly conserved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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, α5 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.)

Fig. 3. Adherence of *R. sphaeroides* NCIB 8253 strains. (A) Viable cells growing in biofilm attached to 68.1 mm² pegs following 5 days aerobic growth at 34°C. WT, wild type; Fdp1, Fdp2 and Fdp3, mutants possessing an insertionally-inactivated fdp gene; control, unincoculated medium (incorporating extremely low cell numbers that arise upon sonication). Standard deviation derived from nine replicate samples. (B) Biofilm formation during complementation studies of Fdp1 using pRK fdp determined by crystal violet staining, and expressed as a percentage of total culture growth in wells. Culturing was after 5 days at 34°C. Fdp1/pRKfdp is a complementation by Fdp1 harbouring fdp inserted on pRK415. Standard deviation derived from nine replicate samples. Similar results were obtained for the other mutants.

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.

oxygen-dependent regulatory system (Phillips-Jones et al., unpublished observations).

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
the conserved sequence appears to be essential because of its role in protein core and emerge on the opposite surface (Fig. 4A). Thus much of a large surface patch, followed by two conserved residues, in particular the H1 and H2 regions previously identified. (B) Suggested binding surface of Fdp; residues 50, 52 and 136–144, all other residues being green. Acids residues are shown in red, basic in blue, hydrophobic in cyan, and hydrophilic in magenta. The key binding residues D136 and V137 are indicated. The orientation is the same as in (A). The orange surface is domain 3 of Drosophila FAS1, from the crystal structure of domains 3 and 4 [30], oriented so that domain 4 aligns with Fdp. For clarity, domain 4 is not shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 domains 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 TGFBIp via integrin αvβ3[17], since mutations in these positions showed loss of function, and synthetic pentapeptides containing this sequence blocked cell adhesion [8,17,46].

Another 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 TGBF1p, 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 TGBF1p, and consequently mutations here could affect interactions with other proteins [30]. It is however diagnostically 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 TGBF1p, 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, TGBF1p and periostin (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 TGBF1p 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 TGBF1p 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 homologs of Fd, 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

[1] Burroughs A.M., Balaji S., Iyer L.M., Aravind L. (2007) Small but versatile: the expanded domain of metazoan Ig-like superfamilies. PLoS Comput. Biol. 3, e196.
[2] Elkins T., Zinn K., McAllister L., Hoffmann F.M., Goodman C.S. (1990) Genetic analysis of the Drosophila neural cell-adhesion molecule: interaction of fasciclin-I and AbdB tyrosine kinase mutations. Cell. 60, 565–575.
[3] McAllister L., Goodman C.S., Zinn K. (1992) Dynamic expression of the cell adhesion molecule fasciclin-I during embryonic development in Drosophila. Development. 115, 267–276.
[4] McAllister L., Rehm E.J., Goodman C.S., Zinn K. (1992) Alternative splicing of micro-exons creates multiple forms of the insect cell-adhesion molecule fasciclin-I. Neurosci. 12, 895–905.
[5] Zinn K., McAllister L., Goodman C.S. (1988) Sequence analysis and neuronal expression of fasciclin-I in grasshopper and Drosophila. Cell. 53, 577–587.
[6] Hortsch M., Goodman C.S. (1990) Drosophila fasciclin-I, a neural cell adhesion molecule, has a phosphatidylinositol lipid membrane anchor that is developmentally regulated. J. Biol. Chem. 265, 15104–15109.
[7] Horisuchi K., Amizuka N., Takeshita S., Takamatsu H., Katsura M., Ozawa H. et al. (1999) Identification and characterization of a novel protein, periostin, with restricted expression to periostium and periodontal ligament and increased expression by transforming growth factor beta. J. Bone Miner. Res. 14, 1239–1249.
[8] Kim J.E., Kim S.J., Lee B.H., Park R.W., Kim K.S., Kim I.S. (2000) Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced protein, h3. J. Biol. Chem. 275, 30907–30915.
[9] Sasaki H., Auclair D., Kaji M., Fukai I., Kiyama H., Yamakawa Y. et al. (2001) Serum level of the periostin, a homologue of an insect cell adhesion molecule, in thymoma patients. Cancer Lett. 172, 37–42.
[10] Gilliott L., Matei D., Fishman D.A., Gerbis C.S., Karlan R.Y., Chang D.D. (2002) Periostin secreted by epithelial ovarian carcinoma is a ligand for a b1 and a b6 integrins and promotes cell motility. Cancer Res. 62, 5358–5364.
[11] Takeshita S., Kikuno K., Tezuka K., Amann E. (1993) Osteoblast-specific factor II: cloning of a putative bone adhesion protein with homology with the insect protein fasciclin-I. Biochem. J. 294, 271–278.
[12] Adachi H., Tsujimoto M. (2002) FEEL-1, a novel scavenger receptor with in vitro bacteria-binding and angiogenesis-modulating activities. J. Biol. Chem. 277, 34264–34270.
[13] Ruan K., Rao S., Ouyang G. (2009) The multifaceted role of periostin in tumorigenesis. Cell. Mol. Life Sci. 66, 2219–2230.
[14] Kudo Y., Srinivadsena B.M.S.M., Hatano H., Ogawa L., Takata T. (2007) Periostin: Novel diagnostic and therapeutic target for cancer. Histol. Histopathol. 22, 1167–1174.
[15] Nummela P., Lammi J., Soikkeli J., Saksela O., Laakkonen P., Holta E. (2012) Transforming growth factor beta-induced (TGBI) is an anti-adhesive protein regulating the invasive growth of melanoma cells. Am. J. Pathol. 180, 1663–1674.
[16] Kim J.E., Jeong H.W., Nam J.O., Lee B.H., Choi J.Y., Park R.W. et al. (2002) Identification of motifs in the fasciclin domains of the transforming growth factor-beta-induced matrix protein jig-h3 that interact with the a b1 integrin. J. Biol. Chem. 277, 46159–46165.
[17] Park S.W., Bae J.S., Kim K.S., Park S.H., Lee B.H. et al. (2004) Jig-h3 promotes renal proximal tubular epithelial cell adhesion, migration and proliferation through the interaction with a b2 integrin. Exp. Mol. Med. 36, 211–219.
[18] Huber O., Sumpner M. (1994) Algal-cams – isomers of a cell adhesion molecule in embryos of the alga Volvox with homology to Drosophila fasciclin-I. EMBO J. 13, 4212–4222.
[19] Johnson K.L., Jones B.J., Bacic A., Schultz C.J. (2003) The fasciclin-like arabinogalactan proteins of arabidopsis. A multigene family of putative cell adhesion molecules. Plant Physiol. 133, 1911–1925.
[20] Borner G.H.H., Scherrer D.J., Stevens T.J., Arken I.T., Dupree P. (2002) Prediction of glycosylphosphatidylinositol-anchored proteins in arabidopsis. A genomic analysis. Plant Physiol. 129, 486–499.
[21] Shi H.Z., Kim Y., Guo Y., Stevenson B., Zhu J.K. (2003) The Arabidopsis SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. Plant Cell. 15, 19–32.
[22] Nagai S., Matsumoto J., Nagasuga T. (1981) Specific skin-reactive protein from culture filtrate of Mycobacterium bovis BCG. Infect. Immun. 31, 1152–1160.
[23] Paulsrud P., Lindblad P. (2002) Fasciclin domain proteins are present in Nostoc symbionts of lichens. Appl. Environ. Microbiol. 68, 2036–2039.
[24] Reynolds W.S., Schwarz J.A., Weis V.M. (2000) Symbiosis-enhanced gene expression in cyanobacterial–algae associations: cloning and characterization of cDNA sim2, encoding a possible cell adhesion protein. Comp. Biochem. Physiol. A. Mol. Integr. Physiol. 126, 33–44.
[25] Ulstrup J.C., Jeansson S., Wikre H.G., Harboe M. (1995) Relationship of secretion pattern with osteoblast–sclerostin factor–2 to osteoblast h3 following Mycobacterium bovis BCG vaccination. Infect. Immun. 63, 672–675.
[26] Oke V., Long S.R. (1999) Bacterial genes induced within the nodule during the Rhizobium–legume symbiosis. Mol. Microbiol. 32, 837–849.
[27] Thapa N., Lee B.-H., Kim I.-S. (2007) TGBIP (Jig-h3 protein: a versatile matrix molecule induced by TGBF). Int. J. Biochem. Cell. Biol. 39, 2183–2194.
[28] Munier F.L., Frueh B.E., Othenin-Girard P., Uffer S., Cousin P., Wang M.X. et al. (2002) BIGH3 mutation spectrum in corneal dystrophies. Invest. Ophthalmol. Vis. Sci. 43, 949–954.
[29] Kannabiran C., Klintworth G.K. (2006) TGBF gene mutations in corneal dystrophies. Invest. Ophthalmol. Vis. Sci. 47, 3918–3923.
[30] Clont N.J., Tris D., Hohenester E. (2003) Novel fold revealed by the structure of a FAS1 domain pair from the insect cell adhesion molecule Fascin-I. Structure. 11, 197–203.
[31] Yoo J.-H., Kim E., Kim J., Cho H.-S. (2007) Crystallization and preliminary crystallographic analysis of the fourth FAS1 domain of human Ig-h3. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 63, 893–895.
[32] Carr M.D., Bloemink M.J., Dentten E., Whelan A.O., Gordon S.V., Kelly G. et al. (2003) Solution structure of the Mycobacterium tuberculosis complex protein murA – from tuberculosis pathogenesis to inherited human disease. J. Biol. Chem. 278, 43736–43743.
[33] Hunter C.N., Turner G. (1988) Transfer of genes coding for apoproteins of reaction center and light-harvesting LH1 complexes to Rhodobacter sphaeroides. J. Bacteriol. 170, 1471–1480.
[34] Porter C.A., Ward A., Laguri C., Williamson M.P., Henderson P.J.F., Phillips-Jones M.K. (2002) Expression, purification and characterization of full-length histidine protein kinase RegB from Rhodobacter sphaeroides. J. Biol. Mol. Biol. 320, 201–213.
[36] Brunger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W. et al. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921.

[37] Cornilescu G., Delaglio F., Bax A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR. 13, 289–302.

[38] Linge J.P., Williams M.A., Spronk C., Bonvin A., Nilges M. (2003) Refinement of protein structures in explicit solvent. Proteins: Struct. Funct. Bioinf. 50, 496–506.

[39] Djordjevic D., Wiedmann M., McLandsborough L.A. (2002) Microtiter plate assay for assessment of Listeria monocytogenes biofilm formation. Appl. Environ. Microbiol. 68, 2950–2958.

[40] Hunter C.N., Coomber S.A. (1988) Cloning and oxygen-regulated expression of the bacteriochlorophyll biosynthesis genes bchE, bchB, bchA and bchC of Rhodobacter sphaeroides. J. Gen. Microbiol. 134, 1491–1497.

[41] Duggan P.S., Parker S.D., Phillips-Jones M.K. (2000) Characterisation of a Rhodobacter sphaeroides gene that encodes a product resembling Escherichia coli cytochrome b562 and R. sphaeroides cytochrome b562. FEMS Microbiol. Lett. 189, 239–246.

[42] Moody R.G., Phillips-Jones M.K., Williamson M.P. (2007) NMR assignment of the Rhodobacter sphaeroides fasciclin-1 domain protein (Fdp). Biomol. NMR Assign. 1, 11–12.

[43] Jefferson K.K. (2004) What drives bacteria to produce a biofilm? FEMS Microbiol. Lett. 236, 163–173.

[44] Roh J.H., Smith W.E., Kaplan S. (2004) Effects of oxygen and light intensity on transcriptome expression in Rhodobacter sphaeroides 2.4.1 – Redox active gene expression profile. J. Biol. Chem. 279, 9146–9155.

[45] Yu N.Y., Wagner J.R., Laird M.R., Meli G., Rey S., Lo R. et al. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 26, 1608–1615.

[46] Park S.J., Park S., Ahn H.C., Kim I.S., Lee B.J. (2004) Conformational resemblance between the structures of integrin-activating pentapeptides derived from jg-h3 and RGD peptide analogues in a membrane environment. Peptides. 25, 199–205.

[47] Clout N., Hohenester E. (2003) A model of FAS1 domain 4 of the corneal protein jg-h3 gives a clearer view on corneal dystrophies. Mol. Vis. 9, 440–448.

[48] Feng D.F., Doolittle R.F. (1996) Progressive alignment of amino acid sequences and construction of phylogenetic trees from them. Methods Enzymol. 266, 368–382.

[49] Elavazhagan M., Lakshminarayanan R., Zhou L., Ting L.W., Tong L., Beuerman R.W. et al. (2012) Expression, purification and characterization of fourth FAS1 domain of TGFβlp-associated corneal dystrophic mutants. Protein Expr. Purif. 84, 108–115.

[50] Runager K., Basaiawmoit R.V., Deva T., Andreasen M., Valnickova Z., Sørensen C.S. et al. (2011) Human phenotypically distinct TGFβlp corneal dystrophies are linked to the stability of the fourth FAS1 domain of TGFβlp. J. Biol. Chem. 286, 4951–4958.

[51] Orecchia P., Conte R., Balza E., Castellani P., Borsi L., Zardi L. et al. (2011) Identification of a novel cell binding site of peristin involved in tumour growth. Eur. J. Cancer. 47, 2221–2229.

[52] Basaiawmoit R.V., Oliveira C.L.P., Runager K., Sørensen C.S., Behrens M.A., Jonsson B.H. et al. (2011) SAXS models of TGFβlp reveal a trimeric structure and show that the overall shape is not affected by the Arg124His mutation. J. Mol. Biol. 408, 503–513.

[53] Williamson M.P. (2011) How Proteins Work. New York: Garland Science.

[54] Laskowski R.A., Rullmann J.A.C., MacArthur M.W., Kaptein R., Thornton J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. 8, 477–486.