The EMBO Journal Vol.18 No.15 pp.4087–4095, 1999

Crystal structure of a tandem pair of fibronectin type III domains from the cytoplasmic tail of integrin α6β4

José María de Pereda, Gerhard Wiche[1] and Robert C.Liddington[2]

Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK and [1]Institute of Biochemistry and Molecular Cell Biology, Vienna Biocenter, 1030 Vienna, Austria

[2]Corresponding author
e-mail: RCL6@LE.AC.UK

The integrin α6β4 is an essential component of hemidesmosomes but it also plays a dynamic role in invasive carcinoma cells. The cytoplasmic tail of the β4 subunit is uniquely large among integrins and includes two pairs of fibronectin type III domains separated by a connecting segment. Here we describe the crystal structure of the first tandem domain pair, a module that is critical for α6β4 function. The structure reveals a novel interdomain interface and candidate protein-binding sites, including a large acidic cleft formed from the surfaces of both domains and a prominent loop that is reminiscent of the RGD integrin-binding loop of fibronectin. This is the first crystal structure of either a hemidesmosome component or an integrin cytoplasmic domain, and it will enable the intracellular functions of α6β4 to be dissected at the atomic level.
Keywords: carcinoma/crystal structure/fibronectin/hemidesmosome/integrin

Introduction

Integrins are heterodimeric transmembrane receptors that mediate cell–cell and cell–matrix adhesion. Ligand binding to their extracellular domains causes conformational changes that are transmitted through the plasma membrane to their cytoplasmic tails, triggering or augmenting a variety of intracellular signal transduction pathways (Hynes, 1992; Clark and Brugge, 1995; Schwartz *et al*., 1995). The integrin α6β4 is an essential component of hemidesmosomes, junctional complexes in the basal layer of stratified epithelia that connect the intermediate filament system to the extracellular matrix via binding to laminin (E.C.Lee *et al*., 1992; Niessen *et al*., 1994; Spinardi *et al*., 1995; Borradori and Sonnenberg, 1996; Dowling *et al*., 1996; van der Neut *et al*., 1996). Mutations in the α6β4 genes cause certain lethal types of epidermolysis bullosa, a disorder associated with severe skin blistering (Uitto *et al*., 1997). In keratinocytes, laminin binding to α6β4 leads to the activation of the Ras-ERK and Rac-JNK mitogen-activated protein kinase pathways (Mainiero *et al*., 1995, 1997). In carcinoma cells, expression of α6β4 promotes invasion by activation of the phosphatidylinositol-3 kinase pathway (Shaw *et al*., 1997), and stimulates lamellae formation and chemotactic migration by activation of a cAMP-specific phosphodiesterase (Rabinovitz and Mercurio, 1997; O'Connor *et al*., 1998). The switch to an invasive phenotype requires a change in the cellular localization of α6β4, from stable adhesive structures linked to intermediate filaments to dynamic motile structures linked to the actin cytoskeleton. Understanding the nature of this switch would represent a major achievement in cancer biology.

The large β4 cytoplasmic tail contains ~1000 amino acids, compared with ~50 amino acids for a typical integrin β subunit (Hogervorst *et al*., 1990; Suzuki and Naitoh, 1990; Tamura *et al*., 1990; Figure 1A). Sequence analysis revealed the presence of four regions with homology to fibronectin type III (FnIII) domains, arranged in two pairs separated by a connecting segment (CS; Suzuki and Naitoh, 1990), and a region with homology to Na^+/Ca^{2+} exchange proteins has also been described recently (May and Ponting, 1999). Most of the known functions of the cytoplasmic tail reside in the first pair of FnIII domains and the connecting segment (FN1:2:CS). This region is necessary for incorporation into hemidesmosomes in epithelial cells (Spinardi *et al*., 1993) and into vimentin-associated fibrillar structures in endothelial cells (Homan *et al*., 1998). In cells not naturally expressing α6β4, transfected FN1:2:CS controls the subcellular distribution of plectin (Niessen *et al*., 1997a; Homan *et al*., 1998; Rezniczek *et al*., 1998), a cytoskeletal linker protein present in hemidesmosomes which also interacts with the intermediate filament system (Wiche, 1998). Recombinant fragments of FN1:2:CS co-immunoprecipitate with plectin (Niessen *et al*., 1997b) and bind fragments of plectin *in vitro* (Rezniczek *et al*., 1998). In a yeast two-hybrid screen, FN1:2:CS interacts with $p27^{BBP/eIF6}$, a nuclear matrix protein that localizes into hemidesmosomes in epithelial cells expressing α6β4 (Biffo *et al*., 1997; Sanvito *et al*., 1999). A shorter fragment, containing the two FnIII domains and just the first 37 amino acids of the connecting segment, is necessary and sufficient for the localization of α6β4 into hemidesmosomes, and for controlling the cellular distribution of plectin (Niessen *et al*., 1997a; Nievers *et al*., 1998).

So far, structural data on integrins have been limited to the extracellular ligand-binding I-domain (Lee *et al*., 1995; Qu and Leahy, 1995; Emsley *et al*., 1997). Here we describe the crystal structure of the first tandem pair of FnIII domains from the cytoplasmic tail of α6β4.

Results and discussion

Overall structure

The crystal structure was solved using two heavy-atom derivatives and refined to a resolution of 2.15 Å (Table I). The crystals contain two copies of the two-domain module in the asymmetric unit, forming an intimate antiparallel embrace about a 2-fold axis close to the

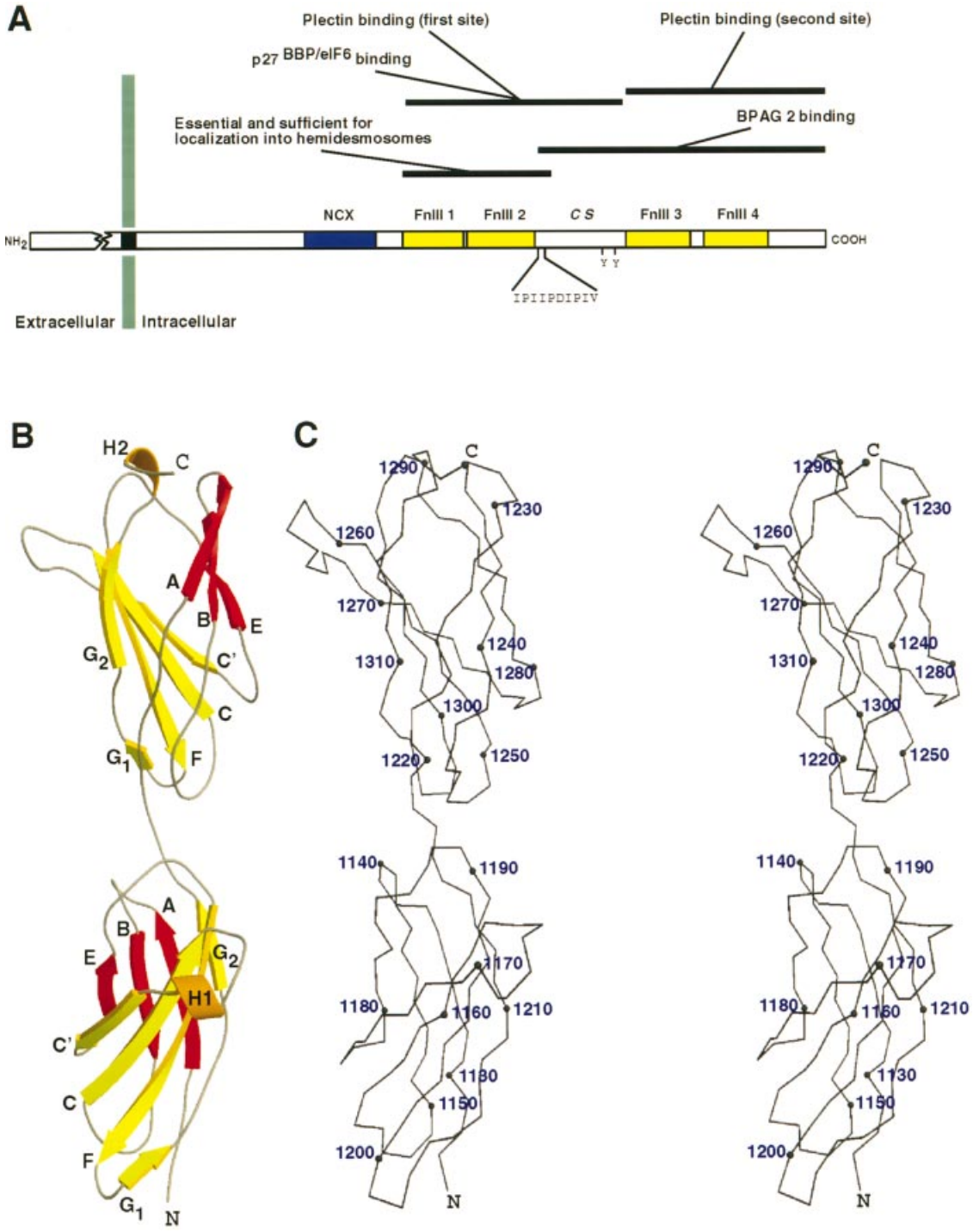

Fig. 1. (**A**) The intracellular organization and functional sites on integrin β4. The four FnIII domains are represented by yellow boxes, and the region with homology to Na^+/Ca^{2+} exchange proteins (NCX) is shown in blue. The proline-rich motif and the Tyr residues of the 'tyrosine activation motif' (Mainiero *et al.*, 1995) within the CS are indicated. Functional assignments are indicated by horizontal bars. (B and C) Structure of the first tandem pair of FnIII domains. (**B**) Ribbon diagram. The A–B–E β-sheet is shown in red, and the C–C′–F–G1–G2 β-sheet in yellow. (**C**) Stereo C_α plot. The molecule is in the same orientation as in (A); the C_α position of every tenth residue is labelled. All molecular figures were drawn with MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1997) or RENDER (Merritt and Bacon, 1997).

interdomain linker, which buries 2600 Å^2 of surface area. The protein is monomeric in solution up to a concentration of 0.75 mM (unpublished data), so that we cannot confidently ascribe functional significance to the crystallographic dimer. It remains possible that at the high effective concentrations observed when the integrin is clustered at hemidesmosomes, dimer formation could play a functional role. The two molecules were built independently, and there are no significant differences between the individual domains [root mean square deviation (r.m.s.d.) = 0.29 Å for all C_αs]. Small differences between the interdomain tilt and twist angles are noted below.

The molecule contains two domains with the typical

Table I. Summary of crystallographic analysis

Space group $P4_32_12$ a = 59.04 Å c = 246.1 Å	Two molecules in the asymmetric unit		41% solvent	
Data set	Native 1	Native 2	3 mM $HgCl_2$	3 mM EMTS
Wavelength/Å	1.037	1.007	1.007	1.007
Resolution/Å	2.15	2.80	2.80	2.80
Unique refections	22 724	9916	19 030[a]	19 195[a]
Overall completeness %	91.5	87.3	89.6	88.7
Anomalous completeness %			87.1	90.5
Redundancy	4.9	4.5	2.4[a]	2.2[a]
R_{sym} % (outer shell)	3.5 (6.8)	4.2 (4.9)	3.0[a] (6.6)	2.5[a] (4.4)
$<I/\sigma(I)>$ (outer shell)	34 (20)	36 (14)	25 (13)	26 (13)
R_{iso} (%)			28.5	30.0
Number of heavy atom sites			10	10
R_{cullis}: centric/acentric			0.53/0.57	0.66/0.67
Phasing power:centric/acentric			1.83/2.39	1.61/2.05
Figure of merit (outer shell)	0.75 (0.59)			

[a]Keeping Bijvoet pairs separate.
$R_{sym} = \Sigma|I_i - <I>|/\Sigma I_i$
$R_{iso} = \Sigma||F_{der}|-|F_{nat}||/\Sigma|F_{nat}|$, where F_{der} is the heavy-atom derivative structure factor and F_{nat} is the protein structure factor.
$R_{cullis} = \Sigma||F_{der} \pm F_{nat}| - F_{der}|/\Sigma|F_{der} \pm F_{nat}|$
Phasing power = $<|F_h|>/<E>$, where F_h is the heavy atom structure factor and E is the residual lack of closure error.

FnIII fold (Campbell and Spitzfaden, 1994) connected by a four-residue linker. Each domain is built up of two β-sheets, one with three (A, B and E) and one with four (C, C′, F and G) strands, packed in a β-sandwich enclosing a highly conserved hydrophobic core (Figure 1B and C). The C–C′ loops in both domains, and the loops at the bottom of the first domain, are more mobile than the core residues (as judged by their B values), but with the exception of a few surface side chains, there is unambiguous electron density throughout.

Domain structure

Domain 1 has several unusual features. At the top of strand A, an abrupt turn occurs at A1137, creating space for two buried water molecules which mediate hydrogen bonds between the main chain of the A–B loop and both main- and side chains of linker residues. The C–C′ loop includes a turn of 3_{10} helix which is stabilized in part by stacking against the tryptophan side chain of W1162 from strand C (see below). Strand G is divided into two segments separated by a structure that resembles a poly-proline II helix, extending from G1205 to L1210, which breaks the normal pattern of β-sheet hydrogen bonding. At the beginning of the helix, a water molecule is inserted between strands F and G and mediates main chain hydrogen bonds, whereas the side chain hydroxyl of S1208 makes hydrogen bonds to both its own carbonyl oxygen and the amide nitrogen of V1196 from strand F. A similar role for serine in stabilizing the poly-proline helix structure was observed in neuroglian (Huber *et al*., 1994).

Buried in the otherwise hydrophobic core of domain 2 is a tyrosine residue (Y1296); the analogous position in other FnIII domains (except for the sixth FnIII domain of fibronectin, which also contains a tyrosine, but whose structure is unknown) is occupied by a pure hydrophobic side chain, typically F, M, V or I. Three water molecules are also buried, making a hydrogen-bonded chain between the tyrosine hydroxyl and the protein surface at the top of the domain. Domain 2 lacks the poly-proline helix observed in strand G of domain 1, but does contain a β-bulge at P1308, which creates a cavity filled by a water molecule in an environment similar to that in domain 1.

Comparisons between FnIII domains

The two FnIII domains of β4 integrin can be closely superimposed (r.m.s.d. of 1.3 Å for the C_α positions of 58 core residues). Domain 2 is 10 residues longer than domain 1, with insertions in the B–C, C–C′ and C′–E loops, where the major structural differences occur. Both β4 domains are as similar or more similar to other FnIII domains (r.m.s.d. in the range 0.7–1.2 Å for core residues). Other known structures lack the buried water molecules found in the β4 domains. An overlay of the two β4 domains with the tenth FnIII domain of fibronectin (Fn10) is shown in Figure 2. The main structural differences are as follows. The top of strand A takes a different course in the three structures; this region is generally variable in FnIII domains. The B–C loop in domain 2 of β4 is unusually long, three residues longer than in Fn10 or tenascin; it is involved in domain–domain interactions (see below). The C–C′ loop in domain 2 of β4 is also unusually long, four residues longer that in Fn10 or tenascin, and is a candidate protein-binding site (see below). The C′–E loop is longer in domain 2 and shorter in domain 1 of β4; this loop has a variable length in other FnIII domains. The F–G loop of both β4 domains is of standard length and lacks the insertion containing the RGD sequence in Fn10.

Domain–domain interactions

Analysis of known interdomain relationships in FnIII modules has revealed no consensus or general rules that would predict their relative orientations and packing (Leahy *et al*., 1996; Sharma *et al*., 1999), and our structure is different from any described previously. The module adopts an extended conformation. Domain 2 is related to domain 1 by a translation of 44 ± 1 Å and a 210 ± 2° rotation (average and variation between two copies of the

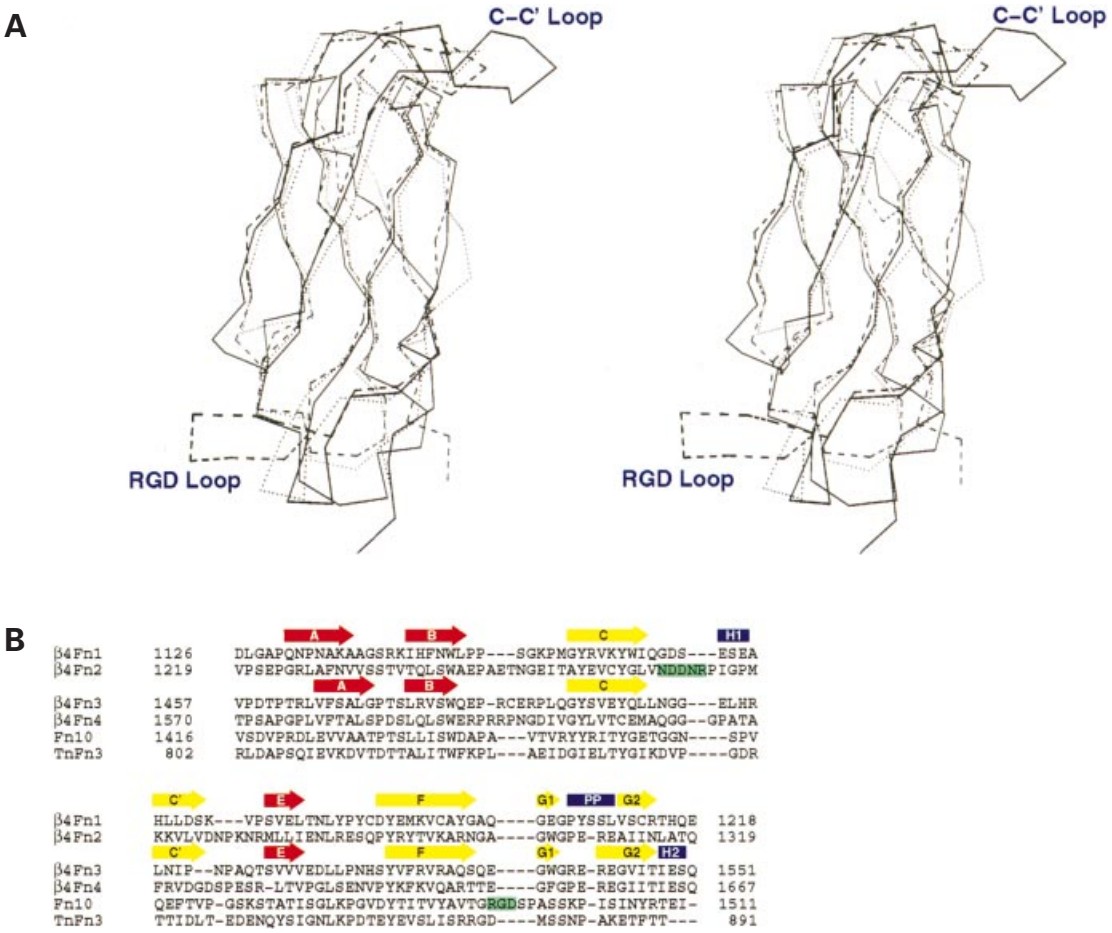

Fig. 2. Structure and sequence alignment of FnIII domains. (**A**) Stereo C_α plot comparing the second domain of integrin β4 (solid line) with the first domain (dotted line) and the tenth FnIII domain of fibronectin (dashed line). The prominent loop at lower left contains the integrin-binding RGD sequence of fibronectin. The C–C′ loop at top right contains the NDDNR motif in domain 2 of integrin β4. (**B**) Sequence alignment. The first and second domains of β4, the tenth FnIII domain of fibronectin (Fn10) and the third FnIII domain of tenascin (TnFn3) were aligned after superposition of their structures. Secondary structure elements of the β4 FnIII domains are indicated, including turns of 3_{10} helix (H1 and H2) and a segment of polyproline helix (PP). The RGD and NDDNR motifs are highlighted in green. The sequence of the third (β4Fn3) and fourth (β4Fn4) domains of β4 were aligned with those of the first and second modules using the program CLUSTALW (Thompson *et al*., 1994).

molecule in the crystal asymmetric unit) and only a small tilt angle: the angle between the major axes of the domains is 11 ± 2°. The surface area buried in the interface is 459 ± 3 Å^2, which is in the mid-range of values for other FnIII modules (330–770 Å^2). The interface is formed by contacts between the A–B and E–F loops of domain 1, the B–C and F–G loops of domain 2, and the four-residue linker (Figure 3).

A key element of the interface is the unusually long B–C loop in domain 2, which makes the only loop–loop contacts between domains, as well as hydrogen bonds to the linker and β-sheet interactions with residues immediately downstream. The B–C loop presents a 10 Å long flat hydrophobic surface comprising the main- and side-chains of N1247, G1248 and E1249, which pack against the side chains of P1188 and Y1189 from the A–B loop of domain 1, making extensive van der Waals contacts.

The linker residues (HQEV 1216–1219) make a variety of bonds which stabilize the conformation. In the middle of the linker, E1218 makes a full complement of main- and side chain hydrogen bonds: its main chain nitrogen hydrogen bonds to a main chain carbonyl in the B–C loop; its carbonyl oxygen hydrogen bonds to the buried side chain of N1302 in the F–G loop of domain 2; and its carboxylate side chain hydrogen bonds to the side chain of S1140 and a buried water molecule at the top of domain 1. In an extended conformation, the carboxylate of E1218 would make a salt bridge to R1141, but in the crystal a side chain from another molecule inserts between them. H1216 makes further stabilizing bonds: its main chain nitrogen makes a water-mediated bond to the A–B turn of domain 1; its side chain was modelled to make a hydrogen bond with a sulfate ion in the solvent channel, but the electron density is also consistent with a 180° flip about its Cβ–Cγ bond, which would allow the formation of a hydrogen bond with the carbonyl oxygen of G1139 in the A–B loop of domain 1. The side-chain Oε of Q1217 nestles between the B–C and F–G loops of domain 2, making hydrogen bonds with the main chain nitrogens in both loops. Of the linker residues, only V1219 makes no hydrogen bonds, but it is not flexible because its peptide planes are fixed by the downstream proline (P1220),

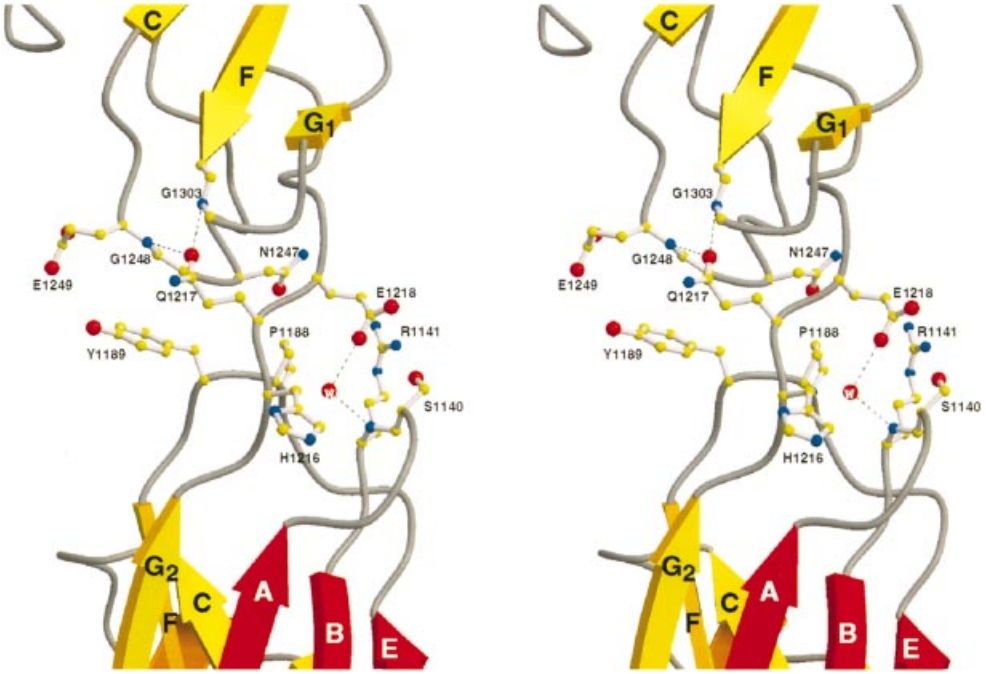

Fig. 3. Stereo close-up of the interdomain interface. Selected residues are shown in ball-and-stick form; hydrogen bonds between linker side chains and domain loops are indicated by dashed lines. The view is rotated by 180° about a vertical axis compared with Figure 1B.

which is packed into the hydrophobic core, and the upstream E1218, which is fully hydrogen bonded, as noted previously.

How well does this conformation represent the two-domain module in solution? The extent and nature of the interface, with its network of hydrogen bonds and van der Waals contacts, suggest that it does represent a major conformation, and independent evidence from small angle solution X-ray scattering also indicates an extended shape (P.Chacón and J.M.Andreu, personal communication). It should be noted that the similarity of the tilt angles and interface structures for the two molecules in the asymmetric unit does not provide additional evidence for rigidity, because the crystal contacts between them are extensive and quasi-symmetrical.

Potential binding motifs

The minimal functional unit for targeting into hemidesmosomes and controlling the cellular distribution of plectin is the first two FnIII domains plus 37 residues into the connecting segment (CS) (Niessen *et al*., 1997a), therefore these three regions must all contribute either directly or indirectly to binding its target proteins.

In other systems, FnIII domains display a variety of binding modes in their interactions with other proteins, involving individual loops [e.g. the RGD loops of fibronectin (Main *et al*., 1992; Dickinson *et al*., 1994; Leahy *et al*., 1996) and tenascin (Leahy *et al*., 1992)] as well as binding surfaces created by domain–domain organization [e.g. the human tissue factor (Banner *et al*., 1996) and the human growth hormone receptor (de Vos *et al*., 1992)]. Our structure contains two striking features that might constitute protein-binding motifs. The first is an acidic V-shaped groove 20 Å wide and 10 Å deep created by the upper surface of domain 1 (the C–C′ and E–F loops) and the lower surface (B–C loop) of domain 2 (Figure 4). The C–C′ loop in domain 1 extends outwards from the body of the domain as a distorted type II β-turn and a turn of 3_{10}-helix which together with the E–F loop form the lower half of the V-shaped groove. The upper half of the groove is formed by the B–C loop of domain 2. Six acidic residues line the perimeter of this surface (D1166 and E1170 from the C–C′ loop; D1191 from the E–F loop; and E1242, E1245 and E1249 from the B–C loop), which is relatively hydrophobic at its centre.

The second feature is the large C–C′ loop in domain 2. The loop is longer than in other FnIII domains, and protrudes from the body of the domain in a manner reminiscent of the integrin-binding RGD loop of the tenth FnIII domain of fibronectin, which, however protrudes from the F–G loop at the opposite end of the domain (Figures 2 and 4C). The C–C′ loop consists of residues 1260–1266, which adopt a β-hairpin with a modified type I β-turn. The turn is stabilized by the side-chain of N1261, which folds back towards the centre of the turn, with its Oε making hydrogen bonds with four peptide nitrogens: D1262 (3.2 Å), D1263 (2.6 Å), N1264 (3.1 Å) and R1265 (3.0 Å). The turn wraps around the body of the domain, such that the side chain of N1264 makes a close contact with I1313 in the C-terminal strand near the beginning of the connecting segment. The side chains of D1262, D1263 and R1265 do not form salt bridges; they are fully exposed to solvent and extend >10 Å from the body of the domain. A mutation in the β4 integrin gene in a patient suffering from junctional epidermolysis bullosa produced a deletion in residues 1249–1265, which encompass the C–C′ loop (Vidal *et al*., 1995). However, the deletion possibly leads to a general destabilization of domain 2.

The CS was not included in the construct used for crystallization, but it would emanate from the upper end of domain 2, surrounded by the upper loops (A–B, E–F

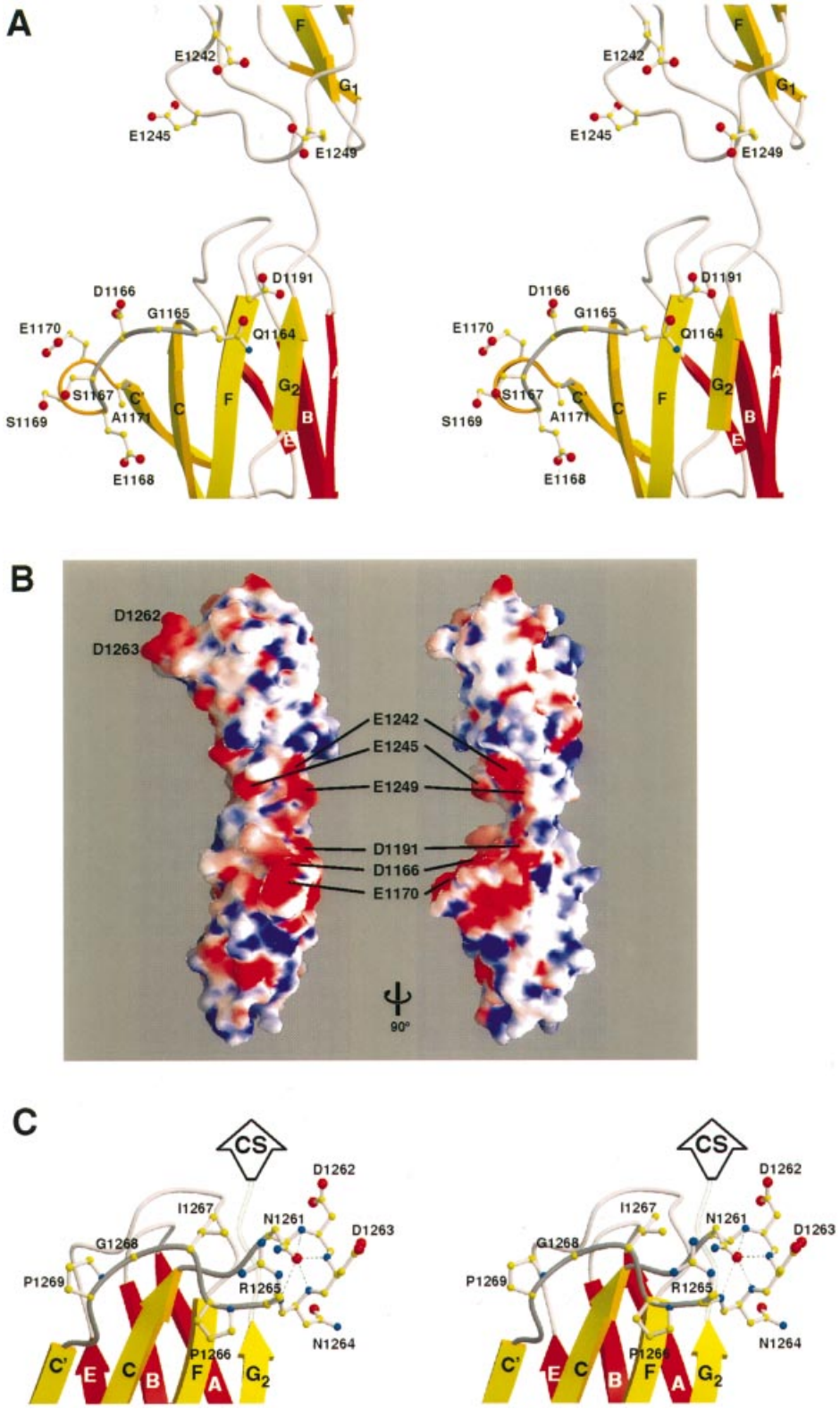

Fig. 4. Potential binding motifs. (A) and (B) Close-up of the V-shaped groove created by the surfaces of both domains. (**A**) The side-chains of the C–C′ loop of domain 1 are shown in ball-and-stick form. The loop ends in a turn of 3_{10} helix, which is viewed down its axis. Acidic residues in the B–C loop of domain 2, which line the upper perimeter of the groove, are also shown. The view is the rotated by 90° about a vertical axis compared with Figure 1B. (**B**) Molecular surface coloured according to surface charge distribution (red negative, white neutral, and blue positive) calculated with the program GRASP (Nicholls, 1992), showing the acidic groove. In the right panel the molecule is in the same orientation as in (A). An orthogonal view is shown on the left. (**C**) Stereo close-up of the C–C′ loop in domain 2, a hairpin with a β-turn at its end. Hydrogen bonds between the side chain of N1261 and the four main chain nitrogens of the β-turn are shown as dashed lines. The beginning of the CS is indicated by an arrow.

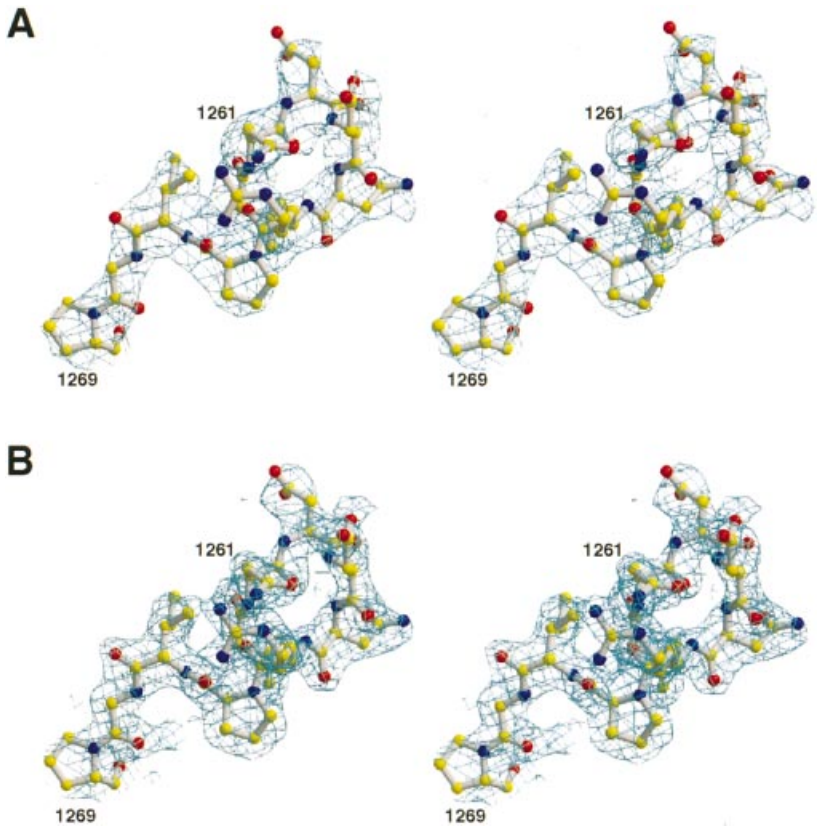

Fig. 5. Electron density maps of the C–C′ loop of domain 2. The refined structure of the loop is shown in ball-and-stick form. (**A**) Experimental solvent flattened map at 2.8 Å resolution, contoured at 1 σ. (**B**) $2F_o$–F_c map with phases calculated from the final model at 2.15 Å, contoured at 1 σ.

and C–C′; Figure 4C). The beginning of the CS contains a highly hydrophobic proline-rich motif, IPIIPDIPIV (1326–1335), which is reminiscent of, but distinct from, known SH3-binding motifs (Feng *et al*., 1994). These motifs are typically unstructured peptides in the absence of their binding partner, and indeed, based on circular dichroism (CD) spectra of recombinant proteins containing the CS, we find no evidence for significant secondary structure (unpublished data).

In conclusion, our crystal structure reveals two potential protein-binding motifs: an acidic groove that could, for example, act as a docking site for a positively charged protein domain; and a prominent loop with an NDDNR motif that could bind to a pocket in another domain. Just as importantly, the organization of the two-domain module shows how these motifs are organized in space and could cooperate with each other and with the connecting segment to create one or more binding surfaces. This information can be used to probe the atomic-level determinants of α6β4 function in both normal and invasive carcinoma cells.

Materials and methods

Cloning and purification

The DNA sequence corresponding to the first two FnIII domains of the human integrin β4, residues 1126–1322, was amplified by PCR using the vector pGR36 (Rezniczek *et al*., 1998) as a template. The upper primer (5′-TGA GAA TTC CAT ATG GAC CTG GGC GCC CCG CAG AAC-3′) contained an *Eco*RI site and a *Nde*I site; the lower primer (5′-GCC GAA TTC GGA TCC CTA CCT CTT GGG CTG GGT GGC-3′) contained an *Eco*RI site and a *Bam*HI site. The 624 bp product was digested with *Eco*R I and cloned into pBluescript SK (–) (Stratagene). The *Nde*I–*Bam*HI fragment was then cloned into pET15b (Novagen). The recombinant fragment includes the extra sequence MGSSHH-HHHHSSGLVPRGSHM (His-tag) at the N-terminus, which after thrombin digestion is reduced to GSHM. The PCR-generated fragment was sequenced in its entirety to verify the absence of mutations.

Escherichia coli BL21 (DE3) cells carrying the expression vector were grown at 37°C in M9ZB medium (Studier *et al*., 1990) containing 100 mg/l ampicillin, until absorbance at 600 nm reached 0.7–0.9 (saturated cultures were avoided). Cultures were induced with 0.25 mM 1-thio-β-D-galactopyranoside (Sigma) at 37°C for 3 h. Cells harvested by centrifugation were resuspended in 20 mM Tris–HCl pH 7.9, 500 mM NaCl, 5 mM imidazole, 0.1% Triton X-100, lysed by sonication and centrifuged. The expressed fragment was in the soluble fraction and was purified by metal chelate affinity chromatography on a 5 ml HiTrap chelating column (Amersham Pharmacia Biotech). After elution with 20 mM Tris–HCl pH 7.9, 500 mM NaCl, 200 mM imidazole, the β4 fragment was extensively dialysed against 20 mM Tris–HCl pH 7.5, 150 mM NaCl, digested with thrombin (Boehringer Mannheim) in the presence of 1 mM $CaCl_2$ at 30°C for 1.5 h, dialysed against 20 mM Tris–HCl pH 7.9, 500 mM NaCl, 5 mM imidazole and reloaded in the HiTrap chelating column. Protein lacking the His-tag was collected in the flow through, dialysed against 20 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 M sucrose, and finally flash-frozen and stored at –80°C. Typically 60 mg of pure β4 protein were obtained from each litre of bacterial culture. The molecular mass of the protein, as measured by mass spectrometry, was 22 342 ± 3 Da, in agreement with the theoretical mass (22 342 Da).

Crystallization and structure determination

Initial crystals were obtained at 4°C by vapour diffusion using 100 mM cacodylic acid/NaOH pH 6.2, 200 mM Li_2SO_4, 30% (v/v) polyethylene glycol 400. Subsequent crystals were obtained by microseeding of non-equilibrated drops containing equal volumes of 20 mg/ml protein in 10 mM Tris–HCl pH 7.5 and mother liquor. Crystals belong to the space group $P4_32_12$ with cell dimensions $a = 59.04$ Å, $c = 246.1$ Å. The best crystals had dimensions 0.25 mm × 0.25 mm × 0.05 mm, with the long cell axis parallel to the short crystal dimension.

Single crystals were frozen by immersion in liquid nitrogen. The mother liquor acted as cryoprotectant. Crystal quality was assessed by

taking test exposures with a Rigaku RU-200 X-ray generator and an R-Axis IV image plate. Good quality crystals were stored and transported frozen to the Advance Photon Source at the Argonne National Laboratory, IL, USA. Data were collected at 100 K at the beam lines BM-14C and BM-14D using Quantum 1 or Quantum 4 CCD detectors. Data were indexed with DENZO and reduced with SCALEPACK (Otwinowski and Minor, 1997). Two mercury derivatives were prepared by soaking crystals in mother liquor containing either 3 mM $HgCl_2$ or 3 mM EMTS at 4°C for 26 h. Data were collected for both derivatives to 2.8 Å resolution using a wavelength of 1.0071 Å to maximize the anomalous signal of mercury. The positions of five mercury atoms of the $HgCl_2$ derivative were found using SOLVE (Terwilliger and Berendzen, 1999). Another five mercury atom sites in the $HgCl_2$ derivative and 10 sites in the EMTS derivative were found by inspection of difference Fourier maps. Sites were refined and phases were calculated using MLPHARE (Collaborative Computational Project No. 4, 1994). An experimental map, obtained after solvent flattening with DM (Cowtan, 1994), was readily interpretable. The asymmetric unit contains two molecules and 41% solvent.

An initial model containing both molecules in the asymmetric unit was built using the program FRODO (Jones *et al*., 1991) and the tenth FnIII domain of fibronectin as a guide (PDB code: 1FNF; Leahy *et al*., 1996). After simulated annealing refinement using CNS (Brünger *et al*., 1998) the *R* factor (R_{work}) was 35.5% and the R_{free} (calculated on 5% of reflections omitted from refinement) was 42.5%. Further refinement included conjugate gradient minimization, B-factor refinement and manual model building. Two strong positive peaks (10 σ) in the F_{obs}–F_{calc} map lying on a 2-fold axis were modelled as cacodylate ions; another three strong peaks were modelled as sulfate ions. Water molecules were included in peaks of the F_{obs}–F_{calc} map higher than 3.5 σ when plausible hydrogen bonds were observed. The final *R* factor was 20.1% for data between 20 and 2.15 Å (bulk solvent corrected) and the R_{free} was 26.2%. The refined model includes residues 1126–1320 and 1127–1319 in the first and second molecule, respectively, 381 water molecules, two cacodylate ions and three sulfate ions. Main-chain torsion angles of the non-glycine residues lie in the most favoured (87.6%) or additional allowed regions (12.4%) of the Ramachandran plot. There are no *cis*-prolines. The r.m.s.d. from ideal bond length is 0.0056 Å and from ideal bond angles is 1.32°. The average temperature factors are 24.8 $Å^2$ and 25.8 $Å^2$, for the first and second molecule, respectively, 36.7 $Å^2$ for the water molecules, 66.2 $Å^2$ for the cacodylate ions and 43.0 $Å^2$ for the sulfate ions. A portion of the initial experimental map and the final refined $2F_o$–F_c map is shown in Figure 5. The crystals contain a non-crystallographic dyad parallel to *x*, at $y = 0$, $z = 1/8$. The nearby crystallographic 2-fold screw parallel to *x*, at $y = 1/4$, $z = 1/8$, gives rise to a non-crystallographic translation close to $x = 1/2$, $y = 1/2$, $z = 0$, which is evident as a peak in the native Patterson.

Accession numbers

PDB ID codes: 1qg3 (coordinates); r1qg3sf (structure factors). Coordinates will be released immediately on publication.

Acknowledgements

We thank Wilfried Schildkamp and Vukica Srajer at the Argonne National Laboratory (BioCARS) for user support and synchrotron access; Laurie Bankston and Günther Rezniczek for critical reading of the manuscript; Pablo Chacón and José Manuel Andreu for communication of unpublished results; and Tina Izard for advice on crystallization. This work was supported by the United Kingdom BBSRC and MRC (to R.C.L.) and by project grant 12389 from the Austrian Science Research Found (to G.W.). J.M.d.P. received a postdoctoral fellowship from the Spanish Ministry of Education and Culture.

References

Banner,D.W., D'Arcy,A., Chene,C., Winkler,F.K., Guha,A., Konigsberg, W.H., Nemerson,Y. and Kirchhofer,D. (1996) The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*, **380**, 41–46.

Biffo,S., Sanvito,F., Costa,S., Preve,L., Pignatelli,R., Spinardi,L. and Marchisio,P.C. (1997) Isolation of a novel β4 integrin-binding protein (p27 (BBP)) highly expressed in epithelial cells. *J. Biol. Chem.*, **272**, 30314–30321.

Borradori,L. and Sonnenberg,A. (1996) Hemidesmosomes: roles in adhesion, signaling and human diseases. *Curr. Opin. Cell Biol.*, **8**, 647–656.

Brünger,A.T., *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.

Campbell,I.D. and Spitzfaden,C. (1994) Building proteins with fibronectin type III modules. *Structure*, **2**, 333–337.

Clark,E.A. and Brugge,J.S. (1995) Integrins and signal transduction pathways: the road taken. *Science*, **268**, 233–239.

Collaborative Computational Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.

Cowtan,H.D. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.

de Vos,A.M., Ultsch,M. and Kossiakoff,A.A. (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science*, **255**, 306–312.

Dickinson,C.D., Veerapandian,B., Dai,X.P., Hamlin,R.C., Xuong,N.H., Ruoslahti,E. and Ely,K.R. (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J. Mol. Biol.*, **236**, 1079–1092.

Dowling,J., Yu,Q.C. and Fuchs,E. (1996) B4 integrin is required for hemidesmosome formation, cell adhesion and cell survival. *J. Cell. Biol.*, **134**, 559–572.

Emsley,J., King,S., Bergelson,J. and Liddington,R. (1997) Crystal structure of the I domain from integrin α2β1. *J. Biol. Chem.*, **272**, 28512–28517.

Esnouf,R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.*, **15**, 132–136.

Feng,S., Chen,J.K., Yu,H., Simon,J.A. and Schreiber,S.L. (1994) Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3–ligand interactions. *Science*, **266**, 1241–1246.

Hogervorst,F., Kuikman,I., von dem Borne,A.E. and Sonnenberg,A. (1990) Cloning and sequence analysis of β-4 cDNA: an integrin subunit that contains a unique 118 kd cytoplasmic domain. *EMBO J.*, **9**, 765–770.

Homan,S.M., Mercurio,A.M. and LaFlamme,S.E. (1998) Endothelial cells assemble two distinct α6β4-containing vimentin-associated structures: roles for ligand binding and the β4 cytoplasmic tail. *J. Cell Sci.*, **111**, 2717–2728.

Huber,A.H., Wang,Y.M., Bieber,A.J. and Bjorkman,P.J. (1994) Crystal structure of tandem type III fibronectin domains from *Drosophila* neuroglian at 2.0 A. *Neuron*, **12**, 717–731.

Hynes,R.O. (1992) Integrins: versatility, modulation and signaling in cell adhesion. *Cell*, **69**, 11–25.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjelgaard,M. (1991) Improved methods for building protein models into electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kraulis,P.J. (1991) Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Leahy,D.J., Hendrickson,W.A., Aukhil,I. and Erickson,H.P. (1992) Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science*, **258**, 987–991.

Leahy,D.J., Aukhil,I. and Erickson,H.P. (1996) 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell*, **84**, 155–164.

Lee,E.C., Lotz,M.M., Steel,G.D.J. and Mercurio,A.M. (1992) The integrin α6β4 is a laminin receptor. *J. Cell. Biol.*, **117**, 671–678.

Lee,J.-O., Rieu,P., Arnaout,M.A. and Liddington,R.C. (1995) Crystal structure of the A-domain from the the α subunit of integrin CR3 (CD11b/CD18). *Cell*, **80**, 631–635.

Main,A.L., Harvey,T.S., Baron,M., Boyd,J. and Campbell,I.D. (1992) The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell*, **71**, 671–678.

Mainiero,F., Pepe,A., Wary,K.K., Spinardi,L., Mohammadi,M., Schlessinger,J. and Giancotti,F.G. (1995) Signal transduction by the α6β4 integrin: distinct β4 subunit sites mediate recruitment of Shc/Grb2 and association with the cytoskeleton of hemidesmosomes. *EMBO J.*, **14**, 4470–4481.

Mainiero,F., Murgia,C., Wary,K.K., Curatola,A.M., Pepe,A., Blumemberg,M., Westwick,J.K., Der,C.J. and Giancotti,F.G. (1997) The coupling of α6β4 integrin to Ras-MAP kinase pathways mediated by Shc controls keratinocyte proliferation. *EMBO J.*, **16**, 2365–2375.

May,A.P. and Ponting,C.P. (1999) Integrin α- and β4-subunit-domain homologues in cyanobacterial proteins. *Trends Biochem. Sci.*, **277**, 12–13.
Merritt,E.A. and Bacon,D.J. (1997) Raster 3D photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
Nicholls,A. (1992) *GRASP: Graphical Representation and Analysis of Surface Properties.* Columbia University, New York.
Niessen,C.M., Hogervorst,F., Jaspars,L.H., de Melker,A.A., Delwel,G.O., Hulsman,E.H., Kuikman,I. and Sonnenberg,A. (1994) The α6β4 integrin is a receptor for both laminin and kalinin. *Exp. Cell Res.*, **211**, 360–367.
Niessen,C.M., Hulsman,E.H., Oomen,L.C., Kuikman,I. and Sonnenberg,A. (1997a) A minimal region on the integrin β4 subunit that is critical to its localization in hemidesmosomes regulates the distribution of HD1/plectin in COS-7 cells. *J. Cell Sci.*, **110**, 1705–1716.
Niessen,C.M., Hulsman,E.H., Rots,E.S., Sanchez-Aparicio,P. and Sonnenberg,A. (1997b) Integrin α6β4 forms a complex with the cytoskeletal protein HD1 and induces its redistribution in transfected COS-7 cells. *Mol. Biol. Cell*, **8**, 555–566.
Nievers,M.G., Schaapveld,R.Q., Oomen,L.C., Fontao,L., Geerts,D. and Sonnenberg,A. (1998) Ligand-independent role of the β4 integrin subunit in the formation of hemidesmosomes. *J. Cell Sci.*, **111**, 1659–1672.
O'Connor,K.L., Shaw,L.M. and Mercurio,A.M. (1998) Release of cAMP gating by the α6β4 integrin stimulates lamellae formation and the chemotactic migration of invasive carcinoma cells. *J. Cell. Biol.*, **143**, 1749–1760.
Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
Qu,A. and Leahy,D.J. (1995) Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, αLβ2) integrin. *Proc. Natl Acad. Sci. USA*, **92**, 10277–10281.
Rabinovitz,I. and Mercurio,A.M. (1997) The integrin α6β4 functions in carcinoma cell migration on laminin-1 by mediating the formation and stabilization of actin-containing motility structures. *J. Cell. Biol.*, **139**, 1873–1884.
Rezniczek,G.A., de Pereda,J.M., Reipert,S. and Wiche,G. (1998) Linking integrin α6β4-based cell adhesion to the intermediate filament cytoskeleton: direct interaction between the β4 subunit and plectin at multiple molecular sites. *J. Cell. Biol.*, **141**, 209–225.
Sanvito,F., Piatti,S., Villa,A., Bossi,M., Lucchini,G., Marchisio,P.C. and Biffo,S. (1999) The β4 integrin interactor p27 (BBP/eIF6) is an essential nuclear matrix protein involved in 60S ribosomal subunit assembly. *J. Cell. Biol.*, **144**, 823–838.
Schwartz,M.A., Schaller,M.D. and Ginsberg,M.H. (1995) Integrins: emerging paradigms of signal transduction. *Annu. Rev. Cell. Dev. Biol.*, **11**, 549–599.
Sharma,A., Askari,J.A., Humphries,M.J., Jones,E.Y. and Stuart,D.I. (1999) Crystal structure of a heparin- and integrin-binding segment of human fibronectin. *EMBO J.*, **18**, 1468–1479.
Shaw,L.M., Rabinovitz,I., Wang,H.H., Toker,A. and Mercurio,A.M. (1997) Activation of phosphoinositide 3-OH kinase by the α6β4 integrin promotes carcinoma invasion. *Cell*, **91**, 949–960.
Spinardi,L., Ren,Y.L., Sanders,R. and Giancotti,F.G. (1993) The β 4 subunit cytoplasmic domain mediates the interaction of α6β4 integrin with the cytoskeleton of hemidesmosomes. *Mol. Biol. Cell*, **4**, 871–884.
Spinardi,L., Einheber,S., Cullen,T., Milner,T.A. and Giancotti,F.G. (1995) A recombinant tail-less integrin β4 subunit disrupts hemidesmosomes, but does not suppress α6β4-mediated cell adhesion to laminins. *J. Cell. Biol.*, **129**, 473–487.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
Suzuki,S. and Naitoh,Y. (1990) Amino acid sequence of a novel integrin β4 subunit and primary expression of the mRNA in epithelial cells. *EMBO J.*, **9**, 757–763.
Tamura,R.N., Rozzo,C., Starr,L., Chambers,J., Reichardt,L.F., Cooper,H.M. and Quaranta,V. (1990) Epithelial integrin α6β4: complete primary structure of α6 and variant forms of β4. *J. Cell Biol.*, **111**, 1593–1604.
Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 849–861.
Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
Uitto,J., Pulkkinen,L. and McLean,W.H. (1997) Epidermolysis bullosa: a spectrum of clinical phenotypes explained by molecular heterogeneity. *Mol. Med. Today*, **3**, 457–465.
van der Neut,R., Krimpenfort,P., Calafat,J., Niessen,C.M. and Sonnenberg,A. (1996) Epithelial detachment due to absence of hemidesmosomes in integrin β4 null mice. *Nature Genet.*, **13**, 366–369.
Vidal,F., Aberdam,D., Miquel,C., Christiano,A.M., Pulkkinen,L., Uitto,J., Ortonne,J.-P. and Meneguzzi,G. (1995) Integrin β4 mutations associated with junctional epidermolysis bullosa with pyloric atresia. *Nature Genet.*, **10**, 229–234.
Wiche,G. (1998) Role of plectin in cytoskeleton organization and dynamics. *J. Cell Sci.*, **111**, 2477–2486.

Received April 21, 1999; revised and accepted June 16, 1999