Predicted and experimental structures of integrins and β-propellers

Timothy A Springer

Integrins and other cell surface receptors have been fertile grounds for structure prediction experiments. Recently determined structures show remarkable successes, especially with β-propeller domain predictions, and also reveal how ligand binding by integrins is conformationally regulated.

Addresses
Center for Blood Research and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA; e-mail: springeroffice@cbr.med.harvard.edu

Current Opinion in Structural Biology 2002, 12:802–813

0959-440X/02/$ – see front matter

Abbreviations

CASP	Critical Assessment of Structure Prediction
EGF	epidermal growth factor
EM	electron microscopy
Gβ	G protein β subunit
I	inserted
I-EGF	integrin EGF
LDLR	low-density lipoprotein receptor
LGA	local global alignment
MIDAS	metal ion dependent adhesion site
PDB	Protein Data Bank
rmsd	root mean square deviation
VWA	von Willebrand factor A

Introduction

Members of the integrin family of cell adhesion molecules undergo conformational changes that enable the affinity of their extracellular domain for ligand to be rapidly modulated by signals within the cell [1••]. Integrins are heterodimers formed by the noncovalent association of α and β subunits. Each subunit contains a large extracellular domain of >940 (α) and >640 (β) residues, a single transmembrane domain and a small, C-terminal cytoplasmic domain (Figure 1). As there are 18 integrin α subunits and 8 β subunits, each is designated with a letter or number; for example, the αIIb and αV integrin subunits associate with the β3 subunit to form the αIIbβ3 and αVβ3 integrin heterodimers.

Structural predictions of integrins were stimulated by their biological importance, the difficulty of determining their structure experimentally and the need to interpret a vast amount of data within a three-dimensional framework. Domain recognition was hampered by a lack of relation with exon structure, correlating with the evolution and fixation of integrin domain organization over a billion years ago [2••]. Initially, the only domain that could be recognized by sequence homology was the α subunit inserted or I domain [3,4] (Figure 1a). A similar, I-like domain was later recognized in the β subunit [5]. Folds were predicted for both [6–9]. Cysteine-rich repeats were recognized in the β-subunit sequence [10,11], but prediction of their correct boundaries, disulfide bond topology and relationship to epidermal growth factor (EGF) domains came much later [12•]. Four repeats with a Ca^{2+}-binding motif [13,14] and then three more without such a motif [3] were recognized in the α subunit, and later predicted to fold into a seven-bladed β-propeller [15] (Figure 1a).

Thus, the boundaries and structures of the I, I-like, β-propeller and integrin EGF (I-EGF) domains were predicted before their structures were solved. The αVβ3 heterodimer crystal structure and the β2 subunit EGF domain NMR structure [16••,17••] revealed most of these predicted domains in atomic detail and, for the first time, the thigh, calf-1 and calf-2 domains in the α subunit, the hybrid and β-tail domains in the β subunit, and a surprising overall bent conformation for integrins (Figure 1b). A β-subunit domain of ~50 residues with predicted α helices shared by plexins, semaphorins and integrins has been detected by sequence homology and termed the PSI domain [18] (Figure 1a). Although it was not resolved in the αVβ3 structure [16••], the predicted boundaries of the PSI domain correspond precisely to a disordered region, suggesting it is a discrete structural unit.

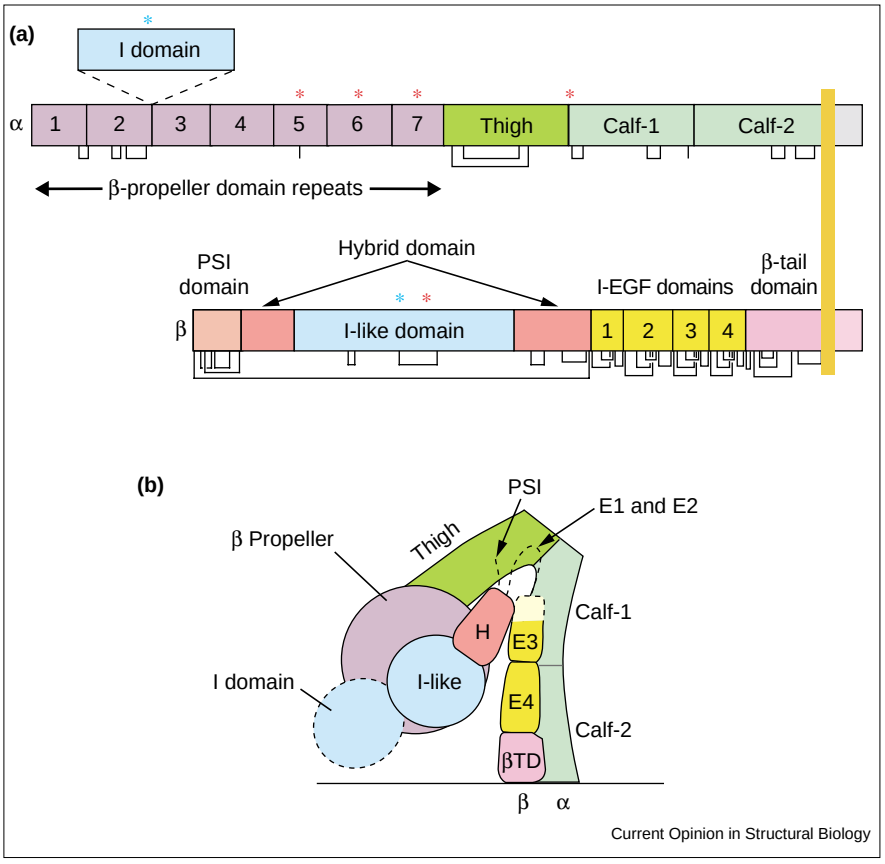
Structure prediction is now regularly evaluated in CASP (Critical Assessment of Structure Prediction) in a workshop setting, but how well does it work in 'real life', when the task at hand is advancing a biological problem? With experimental structures in hand, the predictions on integrins can now be critically evaluated. The prediction of a large family of extracellular six-bladed YWTD β-propeller domains [19] can also be evaluated based on a recently determined structure [20••]. The results show some remarkable successes. Furthermore, comparisons between crystal, NMR and EM structures reveal dramatic rearrangements that regulate ligand binding by integrins.

The inserted or von Willebrand factor A domain in the integrin α subunit

The inserted or I domain, found only in chordates and in half of vertebrate integrin α subunits (Figure 1), is a member of the von Willebrand factor A (VWA) family of domains, which are widely distributed in intracellular and extracellular proteins [21••]. In the integrins in which they are present, the I domains directly bind ligands. The VWA domain in general was predicted to have a fold similar to that of the small G protein ras-p21, with a doubly wound open twisted β-sheet structure flanked by α helices [6]. Shortly thereafter, the structure of the I domain of the integrin αM subunit confirmed this prediction [5,22]. The prediction of secondary structure was quite accurate (77%); however, no fold with the same number and order of α and β elements as

Figure 1

Integrin architecture. (a) Organization of domains within the primary structure. Some α subunits contain an I domain inserted in the position denoted by the dotted lines. Cysteines and disulfide bonds are shown as lines below the schematics. Red and blue asterisks denote Ca^{2+}- and Mg^{2+}-binding sites, respectively. (b) Arrangement of domains within the three-dimensional crystal structure of αVβ3 [16••], with an I domain added. Each domain is color coded as in (a).



predicted and found in the VWA domain was present in the Protein Data Bank (PDB). In ras-p21 and VWA domains, the positions in the fold of β-strands 2 and 3 are reversed, and the linked α helices differ in topology as well. This results in an estimated rmsd of the VWA model in templated regions of 12 Å. Nonetheless, the prediction was prescient. Two residues known to be important in metal binding [23] were correctly predicted to be close together in the three-dimensional structure. Furthermore, it is interesting that, in both ras-p21 and I domains, binding of ligand is coupled to conformational change elsewhere in the domain [1••].

The I-like or von Willebrand factor A domain in the integrin β subunit

At the same time that the crystal structure of the I domain in the integrin α subunit was determined, a similar, I-like domain was predicted to be present in the integrin β subunit [5]. This prediction was based on the presence of a conserved Asp-X-Ser-X-Ser (DXSXS) Mg^{2+}-binding motif, which was known to be important for ligand binding by both I and I-like domains, and on a similar hydropathy plot, which suggested similarly alternating surface-exposed α helices and buried β-strands.

Despite these similarities, it was very difficult to deduce the best alignment between the I and I-like domains. One group aligned the α I domain with a segment of similar length in the β subunit and produced a model of the β3 I-like domain [7]. However, two long insertions that are present in the I-like domain model compared to the I domain were not recognized, and thus only 3 of 13 secondary structure elements in the I-like domain were aligned with the corresponding elements in the I domain. This results in a rmsd of 17.5 Å between the model and the subsequent crystal structure [16••] (Table 1).

A longer I-like domain, incorporating 240 residues instead of 180 residues, was independently predicted by two groups [8,9]. This longer region could be identified because it was more evolutionarily conserved than the preceding and following regions, and because PHD secondary structure predictions [24] showed that it contained alternating α helices and β-strands, whereas surrounding regions were all-β. Furthermore, antibody epitopes were found that are composed of residues from both the first and last predicted α helices, which are adjacent in almost all doubly and singly wound α/β folds [8]. Both groups associated the predicted secondary structure elements in the I-like domain with the correct elements in the I domain and thus correctly placed the two long insertions. The strong sequence conservation of β-strand 4, as expected based on its central position in the hydrophobic β-sheet, was important in guiding alignment [9,25•]. One face of the domain contained hydrophobic elements and lacked

Table 1

Critical assessment of I-like domain models.

Theoretical model	Crystal structure	Templated residues		Model residues		Shift0 (%)[‡]	Secondary structure accuracy (%)[§]
		Number of residues	rmsd (Å)*	Number of residues	rmsd (Å)[†]		
β3 [7]	β3**	154	17.4	181	17.5	13	72
β2[#]	β3**	178	4.9	238	7.3	25	71
β3[¶]	β3**	175	4.6				65
β5[¥]	β3**	181	8.4				
βnu[¥]	β3**	185	11.4				

*Sequence-dependent structural superposition using all residues that were templated in the model. [†]Sequence-dependent structural superposition using all residues alignable between the model and structure. [‡]Model residues found by sequence-independent superposition to be within 3.8 Å of the correct residue in the crystal structure, with no other model residue closer [28••]. [§]The percentage accuracy of reported secondary structure prediction. [#]PDB code 1JX3 [25•]. [¶] Templated residues in the αM I domain structure (PDB code1jlm) were superimposed on the β3 I-like domain structure to estimate accuracy [9]. [¥]The alignments of human β5 (ITB5_HUMAN) and *Drosophila* βnu (Q27591) I-like domains with the mouse αM I domain (ITAM_MOUSE) in the SMART [27] alignment of VWA domains as of September 1st 2002 were separately used to guide alignment of the β3 I-like domain (PDB code 1JV2) and αM I domain (PDB code 1JLM). The rmsd of aligned residues was determined. **PDB code 1JV2 [16••].

N-linked glycosylation sites and antibody epitopes, and was predicted to be associated with other integrin domains [25•], as confirmed by the crystal structure [16••]. Although one group did not construct an I-like domain model, the reported alignment with the I domain template can be used to make a model for the templated residues (Table 1). Both models predict the position of templated residues reasonably well, with a rmsd of 4.6–4.9 Å. Even an ideal alignment with the template, determined by subsequent structure-based superposition, would only have yielded a rmsd of 3.3 Å. Inclusion of the two long, untemplated loops results in a rmsd of 7.3 Å.

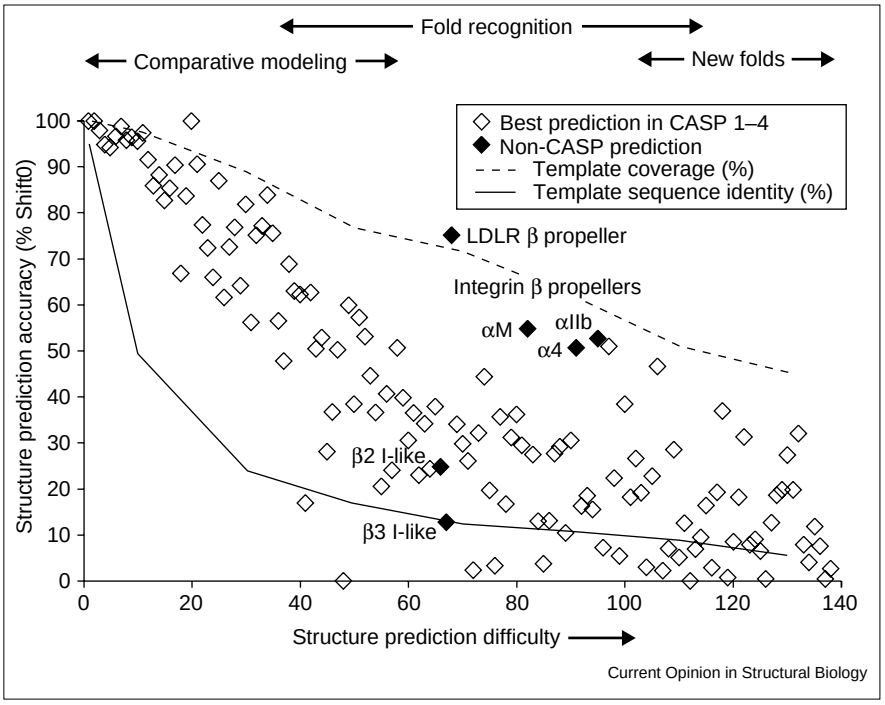
Improved methods have recently allowed statistically significant sequence similarity to be detected between the VWA domains present in integrin α subunits (I domains) and β subunits (I-like domains) [26•]. Threading earlier detected a significant structural relationship [9,25•], although threading gives notoriously unreliable sequence alignments. Does detection of statistically significant sequence similarity signify accurate alignments? The answer is emphatically 'no.' The two integrin I-like domains included in the SMART master VWA domain alignment [27] differ markedly in alignment and the automatic alignment places the two long insertions in erroneous locations. Use of these alignments results in far less accurate models (Table 1).

Comparison to CASP

How good are the predictions described here? To address this question, they have been benchmarked with data from the CASP workshops. In CASP experiments 1 to 4, sequences of proteins undergoing structure determination by NMR or crystallography were made available for prediction. After structure determination, the submitted models were compared. The benchmark used here is that which has been developed to compare predictions across all the CASP workshops [28••]. For each structure that is predicted (target), the best three-dimensional template available at the time of prediction is defined as that with the most residues superimposable within 5 Å of the target structure, using local global alignment (LGA) [28••]. All targets are then ranked based on the percentage of residues with Cα atom pairs within 5 Å in the superposition with the best template (percentage coverage) and the percentage sequence identity of these residues. These two ranks are then averaged to approximate the difficulty of prediction of ~140 different structures (x-axis, Figure 2). Models are assessed by sequence-independent superposition on the structure. Accuracy is defined as the percentage of correctly aligned residues with Cα atoms within 3.8 Å of one another and with no closer Cα atom ('Shift0,' y-axis, Figure 2). Superposition to this accuracy demonstrates correct alignment between sequence and structure, because residues adjacent in sequence have Cα atoms 3.8 Å apart.

The CASP experiments and biologically driven structure prediction differ in several important respects that affect their difficulty and assessment, but in the end these probably balance out. In CASP, the time from obtaining a sequence to submitting a structure is only about two months and a predictor may be working on multiple targets simultaneously. A biologist has more time and a better functional framework for understanding the sequence. On the other hand, all predictions assessed here were of multidomain proteins whose domain boundaries were difficult to detect, whereas in CASP most sequences were of a single domain. In CASP, the sequences of the structure and model are identical, whereas in real life the first example of a three-dimensional structure is often a homolog or ortholog of the model. The integrin β-propeller models evaluated here are orthologues, for example, an integrin α4 model compared to an integrin αV structure; sequence alignments were used to eliminate unpaired residues. Although some difficult to predict loop regions are thereby eliminated from analysis, this may be compensated by the lower structural similarity expected when a model and structure that share only 32 to 43% sequence identity are compared (Table 2). A final,

Figure 2



Accuracy of structure prediction as a function of target difficulty benchmarked with the best predictions for targets in CASP. Data on the best prediction for each of 132 targets in CASP 1–4 are from [28••] (http://PredictionCenter.llnl.gov/casp4/doc/supplement.html). Prediction accuracy and rank of prediction difficulty for non-CASP targets were calculated exactly as for the CASP targets using sequence-independent superposition with LGA [28••]. Because of gaps and insertions in the alignment of αV with the α4, αM and αIIb models, structure and model PDB files were edited to contain only the number of alignable residues shown in Table 2. These files were used both for comparisons to templates and for calculation of Shift0 values, as described in the text and Table 1 legend. The best templates available at the time of prediction were found using LGA to be the α2 I domain (PDB code 1AOX) for I-like domains, the Gβ β-propeller (PDB code 1TBG) for integrin β-propellers and the sialidase β-propeller (PDB code 1KIT) for the LDLR β-propeller.

important difference is that the benchmark is not the average prediction, but the best prediction submitted for each protein. In CASP4, so many groups made predictions (>18 for each target) that random variations in prediction accuracy will contribute to increasing the accuracy of the best prediction [28••]. In this context, it is not surprising that the better of the two I-like domain models is in the middle of the range of accuracy for a target of its difficulty (Figure 2).

Integrin EGF-like domains

Despite the early identification of cysteine-rich repeats in integrin β subunits (Figure 1) [10,11] and the suggestion that they were EGF-like [29], establishing statistically significant sequence similarity to EGF domains and prediction of the correct boundaries and disulfide bond connectivities of these domains [12•] awaited the sequence of a protein composed solely of ten integrin EGF-like domains (TIED) [30]. Shifting the boundaries by one cysteine residue from those previously assigned in integrins gave eight cysteines in each repeat and also increased homology between the repeats [12•]. Correct prediction of the I-EGF domain boundaries was verified by autonomous refolding of I-EGF domains expressed in *Escherichia coli* [12•,17••] and truncation experiments [12•,31•]. A consensus sequence derived from 114 of these repeats in TIED and integrin β subunits was submitted to the Structure-3D server [32]. This initiated a PSI-BLAST search, which revealed statistically significant sequence homology to EGF-like domains in proteins with known three-dimensional structure, such as coagulation factor IX. Six of the eight cysteines were aligned with those in known EGF domains and predicted to form equivalent disulfide bonds.

Table 2

Critical assessment of β-propeller models.

Theoretical model	Residues	Crystal structure	Residues	Templated residues		Alignable residues		Shift0 (%)‡	Secondary structure accuracy (%)§
				Number of residues	rmsd (Å)*	Number of residues	rmsd (Å)†		
α4 [15]	431	αV**	438	310	3.1	412	4.7	51	72
αM#	403	αV**	438	300	3.1	385	5.0	55	70
αIIb¶	451	αV**	438	306	3.9	431	6.2	53	70
LDLR¥	262	LDLR††	254			254	2.5	75	72

*Sequence-dependent structural superposition using all residues that were templated in the model. †Sequence-dependent structural superposition using all residues alignable between the model and structure. ‡Model residues found by sequence-independent superposition to be within 3.8 Å of the correct residue in the crystal structure, with no other model residue closer [28••]. §The percentage accuracy of reported secondary structure prediction. #PDB code 1A8X [43]. ¶PDB code 1JX5 [38•]. ¥PDB code 1LRX [19]. **PDB code 1JV2 [16••]. ††PDB code 1IJQ [20••].

The 'extra' two cysteines were aligned with residues that were close to one another in EGF domains and were predicted to link up to form a fourth disulfide bond [12•].

In the αVβ3 crystal structure, good electron density was present for only a portion of the predicted four EGF-like domains [16••]. I-EGF domain 4 was well resolved and had exactly the predicted disulfide bond connectivity. I-EGF domains 1 and 2 could not be resolved. Although the authors reported the structure of I-EGF domain 3, it had high B factors and their definition of this domain lacked its N-terminal nine residues, including two cysteines defined by the prediction as belonging to I-EGF domain 3 [12•]. The six cysteines reported to be present in I-EGF domain 3 in the crystal structure formed three disulfide bonds [16••], only one of which had the predicted connectivity. Subsequently, an NMR structure was determined for I-EGF domain 3 of the integrin β2 subunit and the connectivity of the four disulfide bonds was directly determined by chemical methods [17••]. The disulfide bond pattern was exactly as predicted. Comparison to the crystal structure suggested that a chain trace mistake had probably occurred, which, together with the lack of density for the correct partners of two of the cysteines in the structure, led to the misassignment of two disulfide bonds. Although this is an example where the electron density was weak, it suggests that, at least in some cases, prediction can outperform or be used to complement experimental structure determination.

β-Propeller domains in integrins and the YWTD repeat family

The primary structure of integrin α subunits revealed three or four repeats with statistically significant sequence similarity [3,13,14]. These contained putative divalent cation binding sites that were similar to positions 1 to 9 of the EF-hand Ca^{2+}-binding site, but lacked a Ca^{2+}-coordinating residue at position 12. Furthermore, secondary structure prediction and the presence of a proline suggested that these sites were embedded in a β-pleated sheet, rather than between α helices, as in the EF-hand [3,13]. Manual sequence inspection recognized weaker, more N-terminal repeats that lacked Ca^{2+}-binding motifs, but shared consensus F-G-F/Y/A-S-V/L and L/V-V/L-V/I/L-G-A-P motifs [3]. A total of seven of these 'FG-GAP' repeats, containing ~65 amino acid residues each, were recognized to be present in the N-terminal portion of all integrin α subunits (Figure 1a) [3]. FG-GAP repeats 1 to 7 were aligned with one another to predict their secondary structure [33]. The results suggested that each repeat contained four segments, which, in turn, were 1, ambiguous in structure; 2, α helix followed by β-strand; 3, β-strand; and 4, β-strand. The putative cation-binding sites were in the loop between segments 2 and 3. A later secondary structure prediction used the PHD method [24] and a multiple alignment containing 30 different integrin α subunits in which repeats 1 to 7 were not aligned with one another and thus were each predicted independently [15]. With this method, segments 1, 2, 3 and 4 were predicted to be β-strands in 6/7, 6/7, 7/7 and 7/7 cases, respectively.

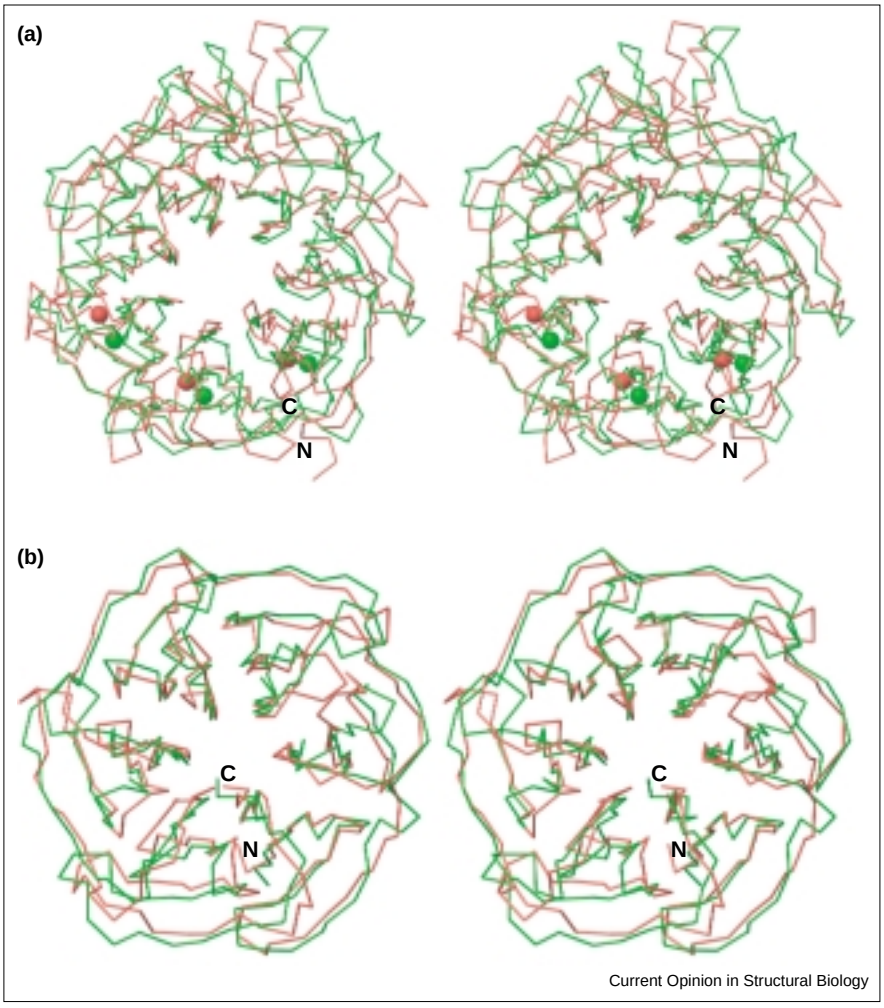
What fold did the repeats adopt? This question was of great importance to the integrin field because, in integrins that lack I domains, there was abundant evidence that a subset of the repeats bound to ligand. The repeats were being intensively studied by site-directed mutagenesis, swapping subsets of repeats between different integrins, expression of fragments containing only some of the seven repeats, and peptide synthesis of individual repeats. This work was going on without a structural framework for guiding the experiments or interpreting the results.

The key insight into fold prediction emerged from framing the question of whether the seven FG-GAP repeats folded autonomously into independent domains or cooperatively into a single domain [15]. Two general principles of protein structure argued that they had to fold cooperatively. Firstly, examination of all structurally known extracellular modules revealed a 'size-to-cysteine' rule, with the number of disulfide bonds per domain inversely related to the number of residues per domain. To maintain the folded state of small proteins with ≤70 residues, disulfide bonds appear to be required to offset the small size of the hydrophobic core. Integrin FG-GAP repeats contain ~65 residues and mostly lack cysteines, and therefore, according to the size-to-cysteine rule, could not fold autonomously. Secondly, the FG and GAP motifs are present within long hydrophobic sequences of 4 to 6 residues, which would have to be buried in a large structure with three or more layers. However, a small domain of ~65 residues could contain only two layers.

The only known fold in which seven similar structural units cooperatively fold into a single domain is the seven-bladed β-propeller [15]. These contain seven β-sheets that pack into a torus or propeller shape, with each β-sheet representing one blade of the propeller. In agreement with the secondary structure prediction of four β-strands per repeat, each blade contains four antiparallel β-strands connected by β-hairpin loops. β-sheets pack against one another side-by-side, with β-strand 1 lining a central, water-filled cavity containing the pseudo sevenfold axis and β-strand 4 forming the outer rim (see Figure 3 below). Threading programs [34] confirmed the prediction of the β-propeller fold.

For modeling, the first problem was to determine the offset between the four predicted β-strands in each sequence repeat and the four β-strands in each β-sheet [15]. The sequence threads through the β-sheets of a β-propeller circularly, so that typically the final sheet contains β-strands from both the first and last sequence repeats, which knit together the propeller. In β-propellers, β-strand 1 is closely packed and shows a preference for small amino acid sidechains, β-strands 2 and 3 are completely buried and thus the most hydrophobic, whereas β-strand 4 is an amphipathic edge strand [35]. The properties of segments

Figure 3



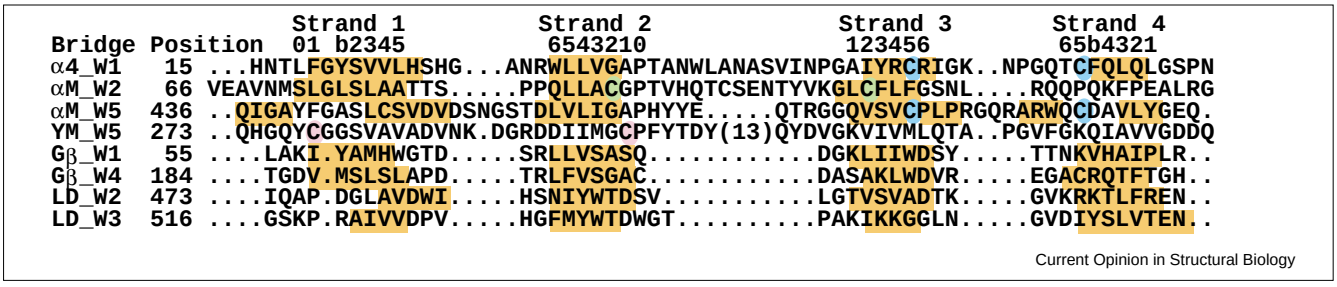
Stereo diagrams of superimposed β-propeller models and structures. (a) The integrin α4 model [15] and αV structure [16••] superimposed, showing all 412 alignable residues. The α4 and αV Cα traces are red and green, respectively. The three Ca^{2+} ions bound in propeller blades 5, 6 and 7 are shown as spheres with the same color code. C and N termini are marked. (b) The human LDLR model [19] and structure [20••] superimposed, showing all 254 common residues. The model and structure Cα traces are red and green, respectively.

1, 2, 3 and 4 in the sequence repeats allowed them to be associated with β-strands 4, 1, 2 and 3 in the β-sheets, respectively. This offset is shared with the WD40 repeats of the G protein β subunit (Gβ) β-propeller domain. Among seven-bladed propeller templates, this yielded the best-packed models. Although there is no statistically significant similarity between integrin FG and GAP motifs and Gβ WD40 repeats, once the offset was established, alignment was guided by the position of predicted β-strands and by similarities in the position of hydrophobic residues (Figure 4). Furthermore, a few cysteines were present that were predicted to form disulfide bonds between adjacent β-strands and these cysteines were aligned so they were in equivalent β-ladder positions (Figure 4). In turn, homology of strands containing cysteines to strands in other repeats aided their alignment. This worked for β-strands 1, 2 and 3, but not for β-strand 4, which shows no inter-repeat similarity. The sequence-to-structure alignment was tested by varying the position in the β-ladder of each β-strand by an off set of –2, –1, +1 or +2, making models and determining their packing quality with QUACHECK of WHATIF [36].

Comparison to the integrin αV crystal structure shows that the β-propeller models [15] were remarkably successful. The integrin α4, αM and αIIb β-propeller domain models, despite the large number of residues (385 to 431) that can be aligned by sequence to αV, show rmsds of only 4.7 to 6.2 Å (Table 2). Using the CASP benchmark, the models are better than or equal to all models of comparable difficulty (Figure 2).

The β-propeller models provided an important framework for interpreting and guiding structure/function work on integrins [15]. The models showed that the ligand-binding site was composed of the loops connecting β-strands 4 and 1, and the loops connecting β-strands 2 and 3, which are on the same face at the 'top' of the β-propeller. Only these loops in blades 2, 3 and 4, together with β-strand 4 of blade 3, contributed to ligand binding [37•,38•]. These results were confirmed by a crystal structure of a ligand mimetic peptide bound to αVβ3 [39] and were in agreement with the general finding that binding sites on β-propeller domains are on their 'top' and 'side' [40].

Figure 4



Representative alignments of integrin and LDLR receptor β-propeller sequences with the Gβ β-propeller template used in modeling. Alignments show one β-propeller β-sheet (W) on each line [15,19]. Predicted and known β-strands in the targets and Gβ template are shaded gold. The selected β-sheets in integrins contain pairs of cysteines predicted or known to form disulfide bonds, which are shaded the same color. Residues in the same bridge position in Gβ are at equivalent positions in the β-sheet ladder; residues in b positions are in β bulges. LD, LDLR; YM, *C. elegans* integrin YMA1 (no secondary structure was predicted).

The model predicted that specific residues preceding β-strand 1 (0 position, Figure 4) and in a bulge in β-strand 1 (b position, Figure 4) point 'up' and form an interaction surface surrounding the pseudosymmetry axis on the 'top' of the β-propeller [15]. This is the consequence of an unusual gap that was introduced into the bulge of Gβ β-strand 1 at the position of the G of the FG motif (Figure 4). This gap allowed the F of the FG motif to be aligned with buried residues in Gβ, and large and small residues to be properly aligned in β-strand 1 in positions b and 2, respectively. This resulted in a previously unprecedented shape for the loop connecting β-strands 4 and 1, which indeed is observed in the integrin β-propeller crystal structure and forms what is called the cup motif [16••]. The residues in the 0 and b positions form an upper and lower tier of residues surrounding the entrance to the central cavity of the β-propeller (Figure 3a). This configuration was likened to an 'amphitheatre' [15] and later to a 'cage' [16••]. The sidechains of these 'upwardly' pointing residues in the 0 and b positions were modeled in nearly the correct positions and were shown to contribute to the interaction between the Gβ β-propeller domain and its ras-like G protein α subunit [15]. Indeed, the αVβ3 crystal structure shows that these residues in the integrin β-propeller interact analogously with the ras-like I-like domain [16••].

Although its cogency and agreement with experimental data led to widespread acceptance in the integrin field, the β-propeller model was not universally accepted. The EF-hand-like motifs were adopted by some workers as the basis of an alternative model. The β-propeller model showed that the loops between β-strands 1 and 2 bearing these motifs were on the 'bottom' of the β-propeller, far from the ligand-binding site [15]. Indeed, the positions of Ca^{2+} ions were modeled to an accuracy of 3.1 ± 1.0 Å (Figure 3a). However, other work continued to be driven by the hypothesis that these Ca^{2+} ions were embedded in an α-helical EF-hand fold and were at the center of the ligand-binding site [41].

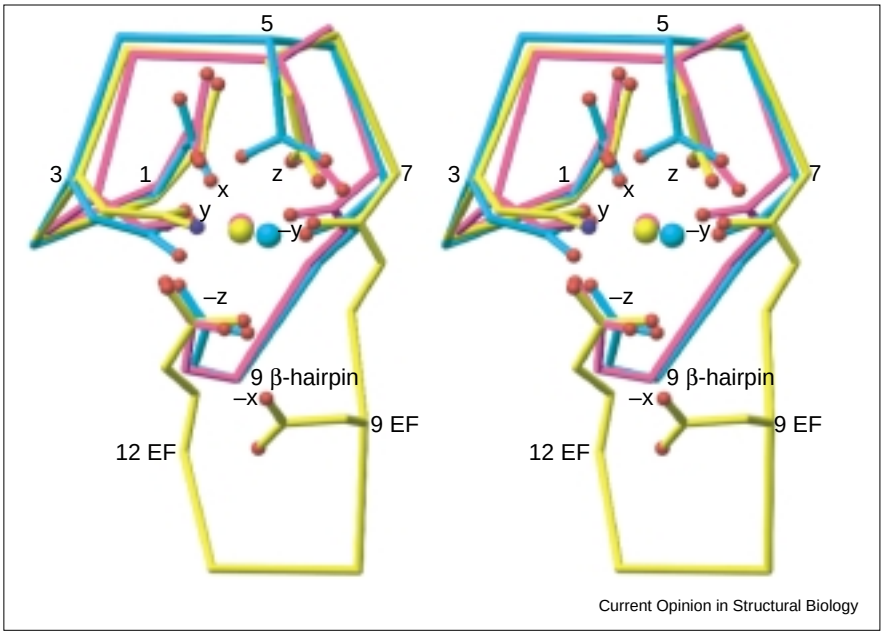
It was therefore important to understand how an EF-hand-like Ca^{2+}-binding motif could be located in a loop between two β-strands. Ca^{2+} was suggested to bind to these sites in integrins, despite the lack of one of the Ca^{2+}-coordinating residues found in EF-hands [42]. Based on the constraints of modeling the Ca^{2+}-binding site in a hairpin loop between β-strands and on sequence differences between integrin and EF-hand Ca^{2+}-binding sequences, a variant mode of coordination was suggested [43]. The residues in positions 1, 3, 5 and 7 were predicted to contribute equivalent oxygens to equivalent octahedral coordination positions. By contrast, the sidechain Oδ atom of the aspartic acid residue in position 9 in integrins was predicted to occupy a site more similar to that occupied by the residue in position 12 in EF-hands, which donates the -z coordination (Figure 5).

A survey of the structure database revealed a previously unannotated Ca^{2+}-binding site in a β-hairpin loop in a bacterial alkaline protease [44,45]. The sequence of this Ca^{2+}-binding motif was found to match that of integrins better than that of the EF-hand. Furthermore, residue 9 in the Ca^{2+}-binding β hairpin occupied the same octahedral coordination position (-z) as position 12 in the EF-hand, as previously predicted [43] (Figure 5). The structure of this β-hairpin Ca^{2+}-binding loop was predicted to closely match that in integrins [45], as indeed verified by the αV integrin crystal structure [16••] (Figure 5).

YWTD β-propeller domains

Could the size-to-cysteine rule be extended to predict β-propeller domains in other extracellular proteins [19]? A comprehensive review contained a table with all known extracellular sequence repeats, both with and without solved three-dimensional structures, listed in order of domain size and with the number of cysteines [46]. Scanning down the list showed that all sequence repeats with 70 or fewer residues contained two or more cysteines per repeat, with three notable exceptions. Two of the exceptions were known to have structures in which the repeats folded cooperatively into large domains.

Figure 5



Comparisons in stereo of Ca^{2+}-binding β-hairpin motifs in integrins and a bacterial protease, and the Ca^{2+}-binding helix-turn-helix motif in the EF-hand. Superimposed structures are alkaline protease (PDB code 1KAP [44]), residues 446 to 454 (magenta); human integrin αV (PDB code 1JV2 [16••]), residues 284 to 292 (cyan); and the EF-hand motif of troponin C (PDB code 5TNC), residues 142 to 153 (yellow). Numbers show the Cα positions of coordinating residues in positions 1, 3, 5, 7, 9 and 12. Letters show octahedral coordination positions. Coordinating sidechains (and carbonyl at position 7) are shown, together with a Cα trace from positions 1–7 and 8–9 or 8–12, and all backbone atoms between positions 7 and 8.

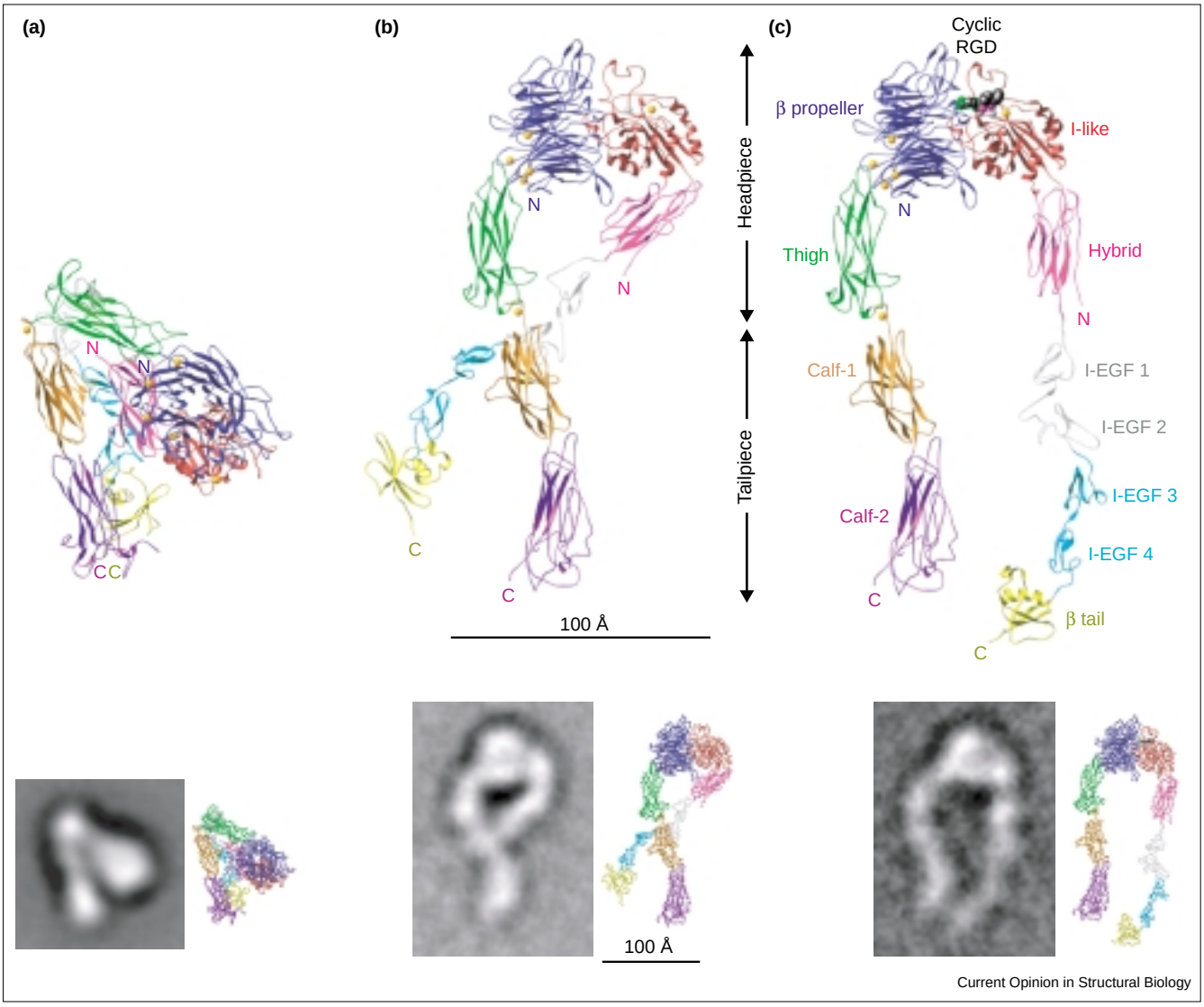
The third exception was a repeat of ~40 residues with a Tyr-Trp-Thr-Asp (YWTD) consensus sequence. Using neighboring EGF modules to define their boundaries showed that YWTD repeats were always present in tandem groups of six, rather than in groups of five, as had been previously thought [19]. Four β-strands were predicted in each repeat. Each tandem group of six YWTD repeats was therefore predicted to fold into a six-bladed β-propeller domain, as strongly supported by threading. The offset was predicted to be different to that in integrins and Gβ, with the four predicted β-strands in each repeat corresponding to β-strands 2, 3, 4 and 1 in the propeller β-sheets. The CASP criteria show that the closest template at the time was a six-bladed bacterial sialidase (PDB code 1KIT) [47]. However, it was recognized that a β-propeller with greater pseudosymmetry would be a better template. Therefore, a six-bladed β-propeller template was modeled from the seven-bladed Gβ β-propeller and permuted to begin with β-strand 2. Retrospective comparison with the low-density lipoprotein receptor (LDLR) β-propeller structure [20••] using LGA shows the artificial template has 233 aligned residues with a rmsd of 2.2 Å, compared to only 210 aligned residues with a rmsd of 2.9 Å with the sialidase. Thus, use of the artificial template was a key step in obtaining a high quality model. Alignment with the template followed the same principles as for the integrin β-propeller model and was tested by model building. There was little apparent sequence similarity; however, the similar lengths of the YWTD and WD40 repeats, and the predicted similar lengths of the 1–2, 3–4 and 4–1 loops aided alignment (Figure 4).

Models were built of important YWTD β-propellers, including those in LDLR and the extracellular matrix protein nidogen [19]. The most frequent human genetic disease, familial hypercholesterolaemeia, is caused by mutations in LDLR. Many of these are located in the β-propeller domain and the model enabled their molecular basis to be defined. Guided by the prediction, a fragment of LDLR was expressed that contained six YWTD repeats and two flanking EGF domains. The 1.5 Å crystal structure [20••] showed that the fold prediction was remarkably accurate, with a perfect alignment of the sequence with the fold template and a rmsd from the structure of only 2.5 Å for all 254 residues (Table 2). This is an unprecedented level of accuracy for a fold prediction. Using the CASP benchmark, the YWTD β-propeller prediction stands out as the best in the fold recognition class (Figure 2). Superposition of the LDLR β-propeller model and structure shows the backbone trace is highly similar (Figure 3b). Detailed inspection shows that many sidechains were modeled with great accuracy, including the tyrosine or phenylalanine of the YWTD motif in the six different blades, for which the C-zeta atoms show a rmsd of only 1.7 Å.

Seven-bladed β-propellers with a YVTN motif and significant sequence similarity to the six-bladed YWTD propellers have been predicted and structurally defined [48•]. These are present in archaebacterial surface layer proteins, together with another module, the PKD domain, also found in metazoan cell surface proteins. These findings suggest an ancient archaeal origin for metazoan cell surface receptors.

Because β-propeller domains are unusually large, another outstanding feature of the integrin and YWTD β-propeller models is the extraordinarily large number of residues that can be placed in three-dimensional relationship to one another. The β-propeller models would look even more

Figure 6



Conformational states of integrin αVβ3 extracellular domains defined by atomic structures [16••,17••,39] and EM [50••]. **(a)** Bent conformation (low affinity). **(b)** Extended conformation with closed headpiece (intermediate affinity). **(c)** Extended conformation with open headpiece (high affinity). In (a–c), the upper panel is a ribbon diagram, the panel on the lower left is an EM image average and the panel on the lower right is a Cα-atom representation. Scale bars shown are to the same scale. (a) The αVβ3 crystal structure [16••] in the same orientation as EM averages of the bent αVβ3 conformation [50••]. (b) Representative EM average of the extended conformation with the closed headpiece seen in Mn^{2+} [50••]. (c) Representative EM average of the extended conformation with the open headpiece seen with cyclic RGD peptide ligand [50••]. To obtain the ribbon and Cα-atom representations, I-EGF domains 1–3 were modeled and added to the αVβ3 crystal structure [1••]. For (b,c), the orientation of the αVβ3 crystal structure headpiece to the EM averages was determined by cross-correlation [50••]. The αV leg containing the calf-1 and calf-2 domains, and the β3 leg containing the I-EGF1 to β-tail domains were added in orientations suggested by EM averages. In (c), a downward movement of the C-terminal helix of the I-like domain was modeled to give the change in orientation of the hybrid domain [50••]. In panels on the lower right, all Cα atoms are represented by spheres, giving an impression of density similar to that in the EM averages in panels on the lower left. Panels at the top and lower right are shown in the same orientation. Ca^{2+} ions are gold and Mg^{2+} ions are silver. The cyclic RGD peptide is shown in CPK representation, with carbon atoms black, nitrogen green and oxygen magenta.

outstanding if the results in Figure 2 were plotted as the total number rather than the percentage of correctly aligned (Shift0) residues.

No doubt the outstanding success of the β-propeller models is, in a large measure, a consequence of the efficient use of special advantages afforded by the pseudosymmetry and highly conserved topology of this fold. This might be seen as a criticism of comparing β-propeller models to results in CASP, where no β-propeller folds have yet been included. On the other hand, it can be seen as justifying the investment of intellectual resources on problems whereby both important new structural insights can arise and accurate models can be produced.

The structural basis for regulation of ligand binding by integrins

The αVβ3 crystal structure revealed most integrin extracellular domains in atomic detail and the organization of these domains within an intact integrin in a surprising bent conformation (Figures 1b and 6a) [16••]. Subsequently, a ligand mimetic peptide was soaked into the crystals, revealing that it bound to an interface between the β-propeller domain and the I-like domain [39], exactly as predicted by domain models and mutagenesis data [37•]. The cyclic Arg-Gly-Asp (RGD) peptide bound with its acidic aspartic acid sidechain coordinating the metal ion of the metal ion dependent adhesion site (MIDAS) of the I-like domain and its arginine sidechain salt bridged to aspartic acid residues in the 4-1 loops of β-propeller blades 2 and 3 [39].

Despite these important advances, the crystal structures did not answer the most interesting question concerning integrins: how ligand binding is regulated. Studies with antibodies had shown that epitopes become exposed on integrin extracellular domains upon activation by signals from inside the cell and upon ligand binding. To define these conformational changes, epitopes were mapped to specific amino acid residues in β2 I-EGF modules 2 and 3 [49], and then structures were determined by NMR [17••]. Superposition on αVβ3 showed that, in the bent conformation, the activation epitopes were buried in the bend between the integrin headpiece and tailpiece, hidden from recognition by antibodies (Figure 6). This suggested that the bent conformation corresponded to the latent, low affinity conformation rather than the active conformation of integrins, as had previously been suggested. Furthermore, it was suggested that integrin activation resulted in a switchblade-like opening, making the epitopes accessible for recognition and placing the ligand-binding domains in the headpiece in a more favorable disposition for binding ligands on the surface of other cells or in the extracellular matrix [17••] (Figure 6c).

Negative-stain EM with image averaging, together with hydrodynamic studies and surface plasmon resonance ligand-binding measurements on the same preparations, has revealed the conformation free in solution, that is, unconstrained by a crystal lattice, of αVβ3 in different activation states [50••]. The resting conformation in Ca^{2+}/Mg^{2+} has low affinity for ligands and corresponds exactly to the bent conformation seen in crystals (Figure 6a). By contrast, activation of integrins by Mn^{2+} or ligand binding induces switchblade opening (Figure 6b,c). Two types of extended conformers are revealed by EM that differ in the angle between the β subunit I-like and hybrid domains. The extended conformer with the closed headpiece has the same angle between the I-like and hybrid domains as seen in the crystal structure (Figure 6b). An extended conformer with an open headpiece shows a marked change in the angle between the I-like and hybrid domains (Figure 6c). This suggests a dramatic conformational movement in the I-like domain in which a downward shift in its C-terminal helix is linked to a change in metal ion coordination in the MIDAS, as occurs in I domains [50••]. Mn^{2+} induces the extended conformers with both closed and open headpieces (Figure 6b,c), whereas binding of a cyclic RGD peptide induces exclusively the extended conformer with the open headpiece (Figure 6c). The extended conformers with closed and open headpieces appear to have intermediate and high affinity for ligand, respectively. Thus, RGD peptide can bind to the low affinity, bent conformation (Figure 6a), as revealed by the complex crystal structure [39]. Then, in the absence of crystal lattice restraints, rearrangement occurs to the extended conformer with high affinity for ligand (Figure 6c) [50••].

Thus, we are beginning to understand how signals that regulate ligand binding are transmitted in integrins, with large-scale rearrangements in the quaternary organization of extracellular domains linked to specific conformational changes within domains. Furthermore, a recent NMR structure of an integrin α and β subunit cytoplasmic domain complex [51••] shows that a noncovalent clasp between juxtamembrane segments maintains integrins in the latent, low affinity state. This is consistent with the close proximity of the extracellular, juxtamembrane segments of the α and β subunits in the bent αVβ3 crystal structure (Figure 6a) [16••], but not in the extended αVβ3 structure (Figure 6c) [50••]. Furthermore, the effect of clasping the legs artificially suggests that transmission of the activating signal across the cell membrane is accomplished by movement apart of the juxtamembrane segments [52,53]. In the bent conformation, substantial interfaces of >2000 $Å^2$ are present between the headpiece and tailpiece, and between the α and β subunits legs in the tailpiece [17••,50••]. Therefore, it appears that separation of the integrin transmembrane domains strains these interfaces, resulting in a switchblade-like opening that is linked to conformational movements at the ligand-binding site straddling the β-propeller and I-like domains, and increasing integrin affinity for ligand.

Prediction in studies of conformational regulation in integrins

Both prediction and structure determination have made important contributions to understanding how integrins work. Interest now focuses on using modeling and prediction to guide experimentation on how conformational change in integrins is regulated. Mutations have been designed that stabilize integrin I domains in low affinity, intermediate affinity and high affinity conformations [54•,55,56••–58••], and stabilize integrin heterodimers in the bent, low affinity conformation [50••]. The αL I domain has been crystallized in three different conformational states and in complex with its Ig superfamily ligand, ICAM-1 [58••]. Current efforts are addressing how conformational change is relayed from one integrin domain

to another. Signaling from the cytoplasm appears to be relayed from integrin cytoplasmic and transmembrane domains to the interface between the α and β subunits in the tailpiece, and the interface between the tailpiece and the headpiece. Signals also appear to be relayed within the integrin headpiece from the hybrid domain to the I-like domain, and from the I-like domain to the I domain. Binding to ligand appears to relay the same signals in the reverse direction [1••,58••]. Recent successes with mutations predicted to stabilize distinct conformational states in integrins suggest that prediction has an important role to play in future studies on these sophisticated, bi-directional signaling machines

Acknowledgements

Supported by National Institutes of Health grants CA31799 and HL48675. Adam Zemla at Lawrence Livermore National Laboratory is gratefully acknowledged for the LGA server and for calculating Shift0 values. S Blacklow, R Hynes, J Takagi, C Ponting, S Perkins and M Humphries are thanked for suggestions on the manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

••1. Shimaoka M, Takagi J, Springer TA: Conformational regulation of integrin structure and function. *Annu Rev Biophys Biomol Struct* 2002, 31:485-516.

A thorough and provocative review of conformational change in integrins, particularly in I domains, and how it regulates affinity for ligand.

2. Hynes RO: Integrins: bi-directional, allosteric, signalling machines. *Cell* 2002, 110:673-687.

•• An excellent recent overview on integrins.

3. Corbi AL, Miller LJ, O'Connor K, Larson RS, Springer TA: cDNA cloning and complete primary structure of the α subunit of a leukocyte adhesion glycoprotein, p150,95. *EMBO J* 1987, 6:4023-4028.

4. Pytela R: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the integrin family and an additional domain related to von Willebrand factor. *EMBO J* 1988, 7:1371-1378.

5. Lee J-O, Rieu P, Arnaout MA, Liddington R: Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell* 1995, 80:631-638.

6. Edwards YJK, Perkins SJ: The protein fold of the von Willebrand factor type A domain is predicted to be similar to the open twisted β-sheet flanked by α-helices found in human ras-p21. *FEBS Lett* 1995, 358:283-286.

7. Tozer EC, Liddington RC, Sutcliffe MJ, Smeeton AH, Loftus JC: Ligand binding to integrin $\alpha_{IIb}\beta 3$ is dependent on a MIDAS-like domain in the β3 subunit. *J Biol Chem* 1996, 271:21978-21984.

8. Huang C, Lu C, Springer TA: Folding of the conserved domain but not of flanking regions in the integrin β_2 subunit requires association with the α subunit. *Proc Natl Acad Sci USA* 1997, 94:3156-3161.

9. Tuckwell DS, Humphries MJ: A structure prediction for the ligand-binding region of the integrin β subunit: evidence for the presence of a von Willebrand factor A domain. *FEBS Lett* 1997, 400:297-303.

10. Tamkun JW, DeSimone DW, Fonda D, Patel RS, Buck C, Horwitz AF, Hynes RO: Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin. *Cell* 1986, 46:271-282.

11. Kishimoto TK, O'Connor K, Lee A, Roberts TM, Springer TA: Cloning of the β subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family. *Cell* 1987, 48:681-690.

•12. Takagi J, Beglova N, Yalamanchili P, Blacklow SC, Springer TA: Definition of EGF-like, closely interacting modules that bear activation epitopes in integrin β subunits. *Proc Natl Acad Sci USA* 2001, 98:11175-11180.

Prediction of integrin EGF-like domains and their disulfide bond connectivity.

13. Argraves WS, Suzuki S, Arai H, Thompson K, Pierschbacher MD, Ruoslahti E: Amino acid sequence of the human fibronectin receptor. *J Cell Biol* 1987, 105:1183-1190.

14. Poncz M, Eisman R, Heidenreich R, Silver SM, Vilaire G, Surrey S, Schwartz E, Bennett JS: Structure of the platelet membrane glycoprotein IIb: homology to the alpha subunits of the vitronectin and fibronectin membrane receptors. *J Biol Chem* 1987, 262:8476-8482.

15. Springer TA: Folding of the N-terminal, ligand-binding region of integrin α-subunits into a β-propeller domain. *Proc Natl Acad Sci USA* 1997, 94:65-72.

16. Xiong J-P, Stehle T, Diefenbach B, Zhang R, Dunker R, Scott DL,
•• Joachimiak A, Goodman SL, Arnaout MA: Crystal structure of the extracellular segment of integrin αVβ3. *Science* 2001, 294:339-345.

The important determination of the crystal structure of integrin αVβ3.

17. Beglova N, Blacklow SC, Takagi J, Springer TA: Cysteine-rich
•• module structure reveals a fulcrum for integrin rearrangement upon activation. *Nat Struct Biol* 2002, 9:282-287.

Based on a prediction, integrin EGF-like domains were expressed in *E. coli*, refolded and used for NMR structure determination. The structure localizes integrin activation epitopes to buried positions in the bent αVβ3 structure, suggesting that it represents the low affinity conformation.

18. Bork P, Doerks T, Springer TA, Snel B: Domains in plexins: links to integrins and transcription factors. *Trends Biochem Sci* 1999, 24:261-263.

19. Springer TA: An extracellular β-propeller module predicted in lipoprotein and scavenger receptors, tyrosine kinases, epidermal growth factor precursor, and extracellular matrix components. *J Mol Biol* 1998, 283:837-862.

20. Jeon H, Meng W, Takagi J, Eck MJ, Springer TA, Blacklow SC:
•• Implications for familial hypercholesterolemia from structure of the LDL receptor YWTD-EGF domain pair. *Nat Struct Biol* 2001, 8:499-504.

The β-propeller prediction was used to guide the expression of a fragment containing the β-propeller domain and flanking EGF-like domains. The 1.5 Å structure yields important insights into the YWTD β-propeller fold and the molecular basis for a large number of LDLR mutations.

21. Whittaker CA, Hynes RO: Essay from the genome annotation
•• series: distribution and evolution of the von Willebrand/integrin A domain: a widely dispersed domain with roles in cell adhesion and elsewhere. *Mol Biol Cell* 2002, 13:3369-3387.

An excellent review of VWA domains from a phylogenetic perspective, with reference to integrin I and I-like domains.

22. Edwards YJK, Perkins SJ: Assessment of protein fold predictions from sequence information: the predicted α/β doubly wound fold of the von Willebrand factor type A domain is similar to its crystal structure. *J Mol Biol* 1996, 260:277-285.

23. Michishita M, Videm V, Arnaout MA: A novel divalent cation-binding site in the A domain of the β2 integrin CR3 (CD11b/CD18) is essential for ligand binding. *Cell* 1993, 72:857-867.

24. Rost B: PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol* 1996, 266:525-539.

25. Huang C, Zang Q, Takagi J, Springer TA: Structural and functional
• studies with antibodies to the integrin β2 subunit: a model for the I-like domain. *J Biol Chem* 2000, 275:21514-21524.

Out of twelve secondary structure elements shared by I and I-like domains, seven were aligned perfectly. Whereas adjusting the sequence alignment to the template using packing quality scores worked well for highly constrained β-propeller domains, it did not work well for this α/β fold. Although all five incorrectly aligned secondary structure elements were only misaligned by 1 to 3 residues, in each case the alignment would have been improved by centering the predicted secondary structure on the template secondary structure.

26. Ponting CP, Schultz J, Copley RR, Andrade MA, Bork P: Evolution of
• domain families. *Adv Protein Chem* 2000, 54:185-244.
A phylogenetic distribution of domain families. The integrin I-like domain is shown to be a VWA family member using PSI-BLAST.

27. Schultz J, Milpetz F, Bork P, Ponting CP: SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* 1998, 95:5857-5864.

28. Venclovas C, Zemla A, Fidelis K, Moult J: Comparison of performance
•• in successive CASP experiments. *Proteins Suppl* 2001:163-170.
Every two years, the latest prediction techniques are tested at CASP (http://PredictionCenter.llnl.gov/). A supplement in *Proteins* analyzes the results and is the place to find out how well these techniques work. For fold recognition, automatic structure prediction is currently no match for computation augmented by expert human knowledge and is being separately evaluated in Critical Assessment of Fully Automated Structure Prediction (CAFASP).

29. Yuan Q, Jiang W-M, Krissansen GW, Watson JD: Cloning and sequence analysis of a novel beta 2-related integrin transcript from T lymphocytes: homology of integrin cysteine-rich repeats to domain III of laminin B chains. *Int Immunol* 1990, 2:1097-1108.

30. Berg RW, Leung E, Gough S, Morris C, Yao W-P, Wang S, Ni J, Krissansen GW: Cloning and characterization of a novel β integrin-related cDNA coding for the protein TIED ("ten β integrin EGF-like repeat domains") that maps to chromosome band 13q33: a divergent stand-alone integrin stalk structure. *Genomics* 1999, 56:169-178.

31. Tan S-M, Walters SE, Mathew EC, Robinson MK, Drbal K, Shaw JM,
• Law SK: Defining the repeating elements in the cysteine-rich region (CRR) of the CD18 integrin β2 subunit. *FEBS Lett* 2001, 505:27-30.
Excellent experimental work defining the boundaries of integrin EGF-like domains.

Huynen M, Doerks T, Eisenhaber F, Orengo C, Sunyaev S, Yuan Y, Bork P: Homology-based fold predictions for ***Mycoplasma genitalium*** proteins. *J Mol Biol* 1998, 280:323-326.

33. Tuckwell DS, Humphries MJ, Brass A: A secondary structure model of the integrin α subunit N-terminal domain based on analysis of multiple alignments. *Cell Adhes Commun* 1994, 2:385-402.

34. Jones DT, Miller RT, Thornton JM: Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing. *Proteins* 1995, 23:387-397.

35. Murzin AG: Structural principles for the propeller assembly of β-sheets: the preference for seven-fold symmetry. *Proteins* 1992, 14:191-201.

36. Vriend G: WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 1990, 8:52-56.

37. Humphries MJ: Integrin structure. *Biochem Soc Trans* 2000,
• 28:311-339.
An excellent review of structural knowledge about integrins and their ligand-binding sites before the αVβ3 crystal structure was determined.

38. Kamata T, Tieu KK, Springer TA, Takada Y: Amino acid residues in
• the αIIb subunit that are critical for ligand binding to integrin αIIbβ3 are clustered in the β-propeller model. *J Biol Chem* 2001, 276:44275-44283.
Mutagenesis combined with the β-propeller model goes a long way toward defining the ligand-binding site on integrin αIIbβ3.

39. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL, Arnaout MA: Crystal structure of the extracellular segment of integrin αVβ3 in complex with an Arg-Gly-Asp ligand. *Science* 2002, 295:151-155.

40. Russell RB, Sasieni PD, Sternberg MJ: Supersites within superfolds. Binding site similarity in the absence of homology. *J Mol Biol* 1998, 282:903-918.

41. Baneres JL, Roquet F, Green M, LeCalvez H, Parello J: The cation-binding domain from the α subunit of integrin α5β1 is a minimal domain for fibronectin recognition. *J Biol Chem* 1998, 273:24744-24753.

42. Tuckwell DS, Brass A, Humphries MJ: Homology modelling of integrin EF-hands. *Biochem J* 1992, 285:325-331.

43. Oxvig C, Springer TA: Experimental support for a β-propeller domain in integrin α-subunits and a calcium binding site on its lower surface. *Proc Natl Acad Sci USA* 1998, 95:4870-4875.

44. Baumann U, Wu S, Flaherty KM, McKay DB: Three-dimensional structure of the alkaline protease of ***Pseudomonas aeruginosa***: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J* 1993, 12:3357-3364.

45. Springer TA, Jing H, Takagi J: A novel Ca^{2+}-binding β-hairpin loop better resembles integrin sequence motifs than the EF-hand. *Cell* 2000, 102:275-277.

46. Bork P, Downing AK, Kieffer B, Campbell ID: Structure and distribution of modules in extracellular proteins. *Q Rev Biophys* 1996, 29:119-167.

47. Crennell S, Garman E, Laver G, Vimr E, Taylor G: Crystal structure of ***Vibrio cholerae*** neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* 1994, 2:535-544.

48. Jing H, Takagi J, Liu J, Lindgren S, Zhang R-G, Joachimiak A, Wang J,
• Springer TA: Archaeal surface layer proteins contain β-propeller, polycystic kidney disease, and β-helix domains, and are related to metazoan cell surface proteins. *Structure* 2002, 10:1453-1464.
A combined prediction and crystal structure study.

49. Lu C, Ferzly M, Takagi J, Springer TA: Epitope mapping of antibodies to the C-terminal region of the integrin β2 subunit reveals regions that become exposed upon receptor activation. *J Immunol* 2001, 166:5629-5637.

50. Takagi J, Petre BM, Walz T, Springer TA: Global conformational
•• rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* 2002, 110:599-611.
EM image reconstruction demonstrates interconversion between one bent and two extended integrin conformations. Mutagenic introduction of a disulfide bond to stabilize the bent conformation abolishes activation of cell surface integrins.

51. Vinogradova O, Velyvis A, Velyviene A, Hu B, Haas TA, Plow EF, Qin J:
•• A structural mechanism of integrin $\alpha_{IIb}\beta_3$ "inside-out" activation as regulated by its cytoplasmic face. *Cell* 2002, 110:587-597.
A unique orientation between integrin α and β subunit cytoplasmic domains is found that is consistent with previous mutagenesis studies.

52. Lu C, Takagi J, Springer TA: Association of the membrane-proximal regions of the α and β subunit cytoplasmic domains constrains an integrin in the inactive state. *J Biol Chem* 2001, 276:14642-14648.

53. Takagi J, Erickson HP, Springer TA: C-terminal opening mimics "inside-out" activation of integrin α5β1. *Nat Struct Biol* 2001, 8:412-416.

54. Shimaoka M, Shifman JM, Takagi J, Mayo SL, Springer TA:
• Computational design of an integrin I domain stabilized in the open, high affinity conformation. *Nat Struct Biol* 2000, 7:674-678.
Repacking the hydrophobic core of the αM integrin I domain to stabilize the open, high affinity conformation, or the closed, low affinity conformation.

55. Xiong J-P, Li R, Essafi M, Stehle T, Arnaout MA: An isoleucine-based allosteric switch controls affinity and shape shifting in integrin CD11b A-domain. *J Biol Chem* 2000, 275:38762-38767.

56. Lu C, Shimaoka M, Ferzly M, Oxvig C, Takagi J, Springer TA: An
•• isolated, surface-expressed I domain of the integrin αLβ2 is sufficient for strong adhesive function when locked in the open conformation with a disulfide. *Proc Natl Acad Sci USA* 2001, 98:2387-2392.
Designed disulfide bonds can lock the αL integrin I domain in high or low affinity conformations.

57. Shimaoka M, Lu C, Palframan R, von Andrian UH, Takagi J, Springer
•• TA: Reversibly locking a protein fold in an active conformation with a disulfide bond: integrin α L I domains with high affinity and antagonist activity ***in vivo***. *Proc Natl Acad Sci USA* 2001, 98:6009-6014.
Surface plasmon resonance demonstrates a 10 000-fold increase in affinity of the αL integrin for ICAM-1 when a disulfide is introduced that stabilizes the open conformation.

58. Shimaoka M, Xiao T, Liu J-H, Yang Y, Dong Y, Jun C-D, McCormack A,
•• Zhang R, Joachimiak A, Takagi J *et al.*: Structures of the αL I domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell* 2002, in press.
Seven distinct disulfide bonds are introduced to stabilize alternative conformations of the αL I domain. Two have high affinity for the ligand ICAM-1, three intermediate affinity and two low affinity. High and intermediate affinity I domains are each crystallized in two distinct forms. One structure is a complex with ICAM-1. A novel intermediate conformation of I domains is revealed and the shape-shifting pathways induced by ligand binding and by pull down of the C-terminal α-helix are shown to be equivalent.