

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 30, 1996.

Folding of the N-terminal, ligand-binding region of integrin α -subunits into a β -propeller domain

TIMOTHY A. SPRINGER

The Center for Blood Research and Harvard Medical School, Department of Pathology, 200 Longwood Avenue, Boston, MA 02115

Contributed by Timothy A. Springer, November 13, 1996

ABSTRACT The N-terminal ≈ 440 aa of integrin α subunits contain seven sequence repeats. These are predicted here to fold into a β -propeller domain. A homologous domain from the enzyme phosphatidylinositol phospholipase D is predicted to have the same fold. The domains contain seven four-stranded β -sheets arranged in a torus around a pseudosymmetry axis. The trimeric G-protein β subunit (G beta) appears to be the most closely related β -propeller. Integrin ligands and a putative Mg^{2+} ion are predicted to bind to the upper face of the β -propeller. This face binds substrates in β -propeller enzymes and is used by the G protein β subunit to bind the G protein α subunit. The integrin α subunit I domain, which is structurally homologous to the G protein α subunit, is tethered to the top of the β -propeller domain by a hinge that may allow movement of the domains relative to one another. The Ca^{2+} -binding motifs in integrin α subunits are on the lower face of the β -propeller.

Integrins are arguably the most structurally and functionally complex family of cell adhesion molecules (1–4). To date, 16 different integrin α subunits and 8 different integrin β subunits in mammals have been defined, which give rise to 21 different $\alpha\beta$ complexes. The extracellular domains are large, with up to 1114 aa for α and 678 aa for β . Diverse ligands on the cell surface and in the extracellular matrix are recognized. The adhesiveness of integrins can be rapidly modulated by signals from within the cell, a phenomenon termed “inside-out signaling” (5–7). Dynamic regulation of integrin adhesiveness is associated with alterations in affinity for ligand, expression of epitopes recognized by monoclonal antibodies (mAbs), clustering on the cell surface, and cytoskeletal association. The divalent cations Mg^{2+} or Ca^{2+} are required for ligand binding, and alteration of divalent cation composition in the medium can augment or inhibit adhesion. Regions in both the α and β subunit cytoplasmic domains have been defined that are important for regulation of adhesiveness, but how this is transduced and the nature of the alteration in the extracellular region are unclear (6, 7). Electron microscopy of integrins shows a globular ligand-binding domain, containing the N-terminal portions of the α and β subunits, that is connected to the membrane by two stalks corresponding to the C-terminal portions of the α and β subunit extracellular domains (8).

Regions important for ligand binding have been localized to the N-terminal portions of the α and β subunits (9). The N-terminal region of the α subunit is composed of seven repeats of about 60 aa each (10). These contain FG (phenylalanyl-glycyl) and GAP (glycyl-alanyl-prolyl) consensus sequences and will be termed FG-GAP repeats. A putative

Ca^{2+} -binding motif is present in FG-GAP repeat 4 in some integrins and in repeats 5 through 7 in all integrins (11). Ligand binding by certain $\beta 1$ and $\beta 3$ integrins has been mapped within the FG-GAP repeats of the α subunits by cross-linking to ligand and by site-directed mutagenesis (7, 9, 12–16). Within the third FG-GAP repeat in some integrins is an inserted (I) domain of 200 residues. The structure of the I domain of the integrins Mac-1 and LFA-1 has recently been solved and has a fold similar to that found in nucleotide-binding enzymes, small G proteins, and heterotrimeric G protein α subunits (17–20). The I domain contains a metal ion-dependent adhesion site (MIDAS). Two serines, a threonine, and two water molecules coordinate a Mg^{2+} , with a sixth coordination position available to bind an acidic residue in the ligand. Specificity for ligand is provided by residues that surround the MIDAS in the integrin LFA-1, suggesting that this is the ligand-binding site for the I domain-containing integrins (21). A similar MIDAS motif is present within a 250-aa region that is the most conserved portion of integrin β subunits, and this region has been predicted to adopt a fold similar to that of the I domain (17).

Insight into the fold of the seven α subunit FG-GAP repeats could substantially advance our ability to interpret current data and design further experiments on the structure and function of integrins. Here I predict the fold of this region of ≈ 440 amino acid residues. Conservation of the number of repeats as seven in evolution, symmetry arguments, and properties of the FG-GAP sequence repeats and of protein domains in general suggest that the only known fold that the seven repeats can adopt is the β -propeller. Strong support for this fold is provided by secondary structure and threading predictions, model building, and disulfide bond topology.

Methods

For prediction of secondary structure and disulfide-bonded cysteines, integrins present in Swiss-Prot (references ITA1_RAT, ITA2_HUMAN, ITA3_CRISP, ITA3_HUMAN, ITA4_HUMAN, ITA4_MOUSE, ITA5_HUMAN, ITA6_CHICK, ITA6_HUMAN, ITA8_CHICK, ITAB_HUMAN, ITAE_HUMAN, ITAL_HUMAN, ITAL_MOUSE, ITAM_HUMAN, ITAM_MOUSE, ITAP_DROME, ITAV_CHICK, ITAV_HUMAN, ITAX_HUMAN, YMA1_CAEEL, and YMB3_CAEEL) and selected integrins in GenBank (accession nos. A45226, ITA1HU; L25886, ITA2BO; S44142, ITA2MO; D13867, ITA3MO; U12683, ITA5XE; X69902, ITA6MO; L23423, ITA7MO; L36531, ITA8HU; A49459, ITA9HU; U37028, ITADHU; U12236, ITAEMO; S40311,

ITAPS1DR; U14135, ITAVMO; and X81108, ITAVPL) were used. For secondary structure prediction, the I domain was removed at structural boundaries (19), and integrins were truncated ≈ 12 residues after the end of the last FG-GAP repeat using SEQED in GCG (22). A PIR file was made with TOPIR of GCG, the format was edited with NISUS WRITER (Nisus, Solana Beach, CA), and sequences were submitted to PHD (23) as PIR lists. Insertions of >10 residues cannot be removed by PHD and are misaligned; therefore, ITAE_HUMAN and 1TAEMO were not used, and additionally ITA3_CRISP, ITA3_HUMAN, ITA3MO, and ITA6_CHICK were not used for α M prediction. For phosphatidylinositol phospholipase D (PIPLD), the Swiss-Prot PHLD_BOVIN and GenBank human L11701 and L11702 sequences were used. Because these sequences are highly identical, diversity was increased by separating the seven repeats and submitting PIR lists with $7 \times 3 = 21$ sequences. Lists with a different repeat of PHLD_BOVIN as the target sequence and with two different offsets at the ends were submitted, since predictions at each residue take into account the preceding and following six residues.

Models of $\alpha 4$ were made with SEGMOD of LOOK (Molecular Applications Group, Palo Alto, CA). Alignments (see Fig. 2) used seven circularly permuted coordinate files for G beta (protomer B of the $\beta\gamma$ dimer; ref. 24) each beginning with strand 4 of a different sheet (W). To force *cis*-proline in the five $\alpha 4$ integrin Ws with GAP, which appeared required to obtain a strand 1 that extended through the FG sequence that is β -bridged to the GAP motif, the *cis*-proline in W4 of galactose oxidase (GO) in the same alignment position and one residue on each side were superimposed and used as the template. The best representative for each W was chosen from among multiple $\alpha 4$ models; criteria were a well developed β -sheet structure as determined by DSSP, and F and A residues from the FG and GAP motifs that were buried and in proximity of one another. These Ws were used as templates for a final model. The template selected for W2 contained the C81–C85 and C150–C165 disulfides that had spontaneously formed; the C58–C68 strand 2 to strand 3 and C111–C132 strand 3 to strand 4 disulfides were present in all models as a consequence of the alignment. Three copies of the Ca^{2+} -binding structure from galactose-binding protein (25) were aligned with W5, W6, and W7 (see Fig. 2), superimposed with LOOK using β -strand residues Q142–V144, and residues L135–Q142 and the Ca^{2+} were used as templates. Residues G577–D586 in GO were used to model a representative integrin $\beta 2$ subunit sequence, residues V238–F247, as a polypeptide finger with an aspartic acid at its tip. It was placed three residues lower in the central cavity with respect to strand 1 than in GO. The associated Cu^{2+} was modeled as Mg^{2+} . The Mg^{2+} was centered on the pseudosymmetry axis at the average position of $\text{C}\alpha$ carbons in bridge position 2 in strand 1. This gave a similar spatial relationship between metals and the $\text{C}\alpha$ carbons of metal-ligating histidine and aspartic acid residues in GO, the model, and the integrin I domain (17). The side chains of residues in position 2 of strand 1 were adjusted to an orientation similar to that of T102, C148, and S189 in G beta. Models were evaluated for β -sheet structure with DSSP (26) and for native-like packing environments of their residues with the QUACHK module of WHAT IF (27).

Deduction of a Protein Fold

Many surface proteins contain tandem, independently folded domains, which commonly are unpaired and extend linearly like beads on a string; however, a number of features suggest that the seven N-terminal repeats of integrin α subunits are not independent domains, but rather fold into a single globular domain. (i) The number of disulfide bonds in FG-GAP repeats is inadequate for small, autonomous domains. FG-GAP repeats are 62 ± 7 aa. Structurally characterized repeating

domains of similar size found on cell surfaces include complement control repeats of 61 ± 4 residues, fibronectin type II repeats of 60 ± 1 residues, and epidermal growth factor repeats of 40 ± 3 residues (28). Disulfide bonds are particularly important to stabilize small domains; all domains of ≈ 60 residues and smaller contain two intradomain disulfides. However, few FG-GAP repeats contain cysteines. Furthermore, some disulfides are interrepeat, rather than intrarepeat (29), which is unlikely for tandem independent domains. (ii) FG-GAP repeats contain two long consensus hydrophobic segments, VVVGAP and VYLF (see Fig. 2), that cannot be accommodated within a small domain. Small domains have two layers, whereas hydrophobic segments require a structure with three or more layers, so that they can be buried beneath amphipathic segments (30). (iii) Integrins always contain seven repeats, whereas for linear tandem repeats, the number of repeats is not constrained. Even within families of closely related adhesion molecules, repeat number varies. Among IgSF members that bind integrins, there are two or five IgSF domains in intercellular adhesion molecules (ICAMs) and three to seven in vascular cell adhesion molecule-1 (VCAM-1). Among selectins, there are two to nine complement control repeats (31). Cadherin superfamily members usually contain five domains, but one member contains 34 (32). The number of FG-GAP repeats is conserved at seven in all known integrin α subunits, including those in *Caenorhabditis elegans* and *Drosophila melanogaster* and the 16 different α subunits in mammals, despite insertion of the I domain in some of these. (iv) Exchange of the FG-GAP repeats before and after the I domain between the integrins Mac-1 ($\alpha\text{M}\beta 2$) and p150,95 ($\alpha\text{X}\beta 2$) reveals epitopes that require the presence of both regions (33). Furthermore, although not previously pointed out, mAb HP1/3 to VLA-4 interacts with both FG-GAP repeat 1 and repeats 5–7 (34).

Because this evidence strongly suggests that FG-GAP repeats cannot be autonomously folded units, I surveyed protein structures for those in which the FG-GAP repeats could be cooperatively folded. Proteins have recently been systematically classified according to their fold—i.e., the way in which the β -strand and α -helical secondary structure elements are arranged within the primary structure and folded in three dimensions to form domains (30, 35). Almost without exception, proteins with sequence homology adopt the same fold. The fold is more evolutionarily conserved than sequence homology and may also arise through convergent evolution, so that proteins with the same fold may not show statistically significant sequence homology. The similarities in sequence suggest that each FG-GAP repeat adopts a similar structure; therefore, for the seven units to fold cooperatively, the protein fold should contain seven pseudosymmetric units. Although several folds with pseudosymmetry exist, only one can have seven pseudosymmetric units, an all- β structure known as the β -propeller.

β -propellers contain four, six, seven, or eight β -sheets arranged radially and pseudosymmetrically around a central axis. Each sheet contains four antiparallel β -strands (see Figs. 1 and 3). Strand 1 is closest to and runs parallel to the central axis (36). The overall shape of the domain is cylindrical, with strand 4 on the outside of the cylinder. Because β -sheets have

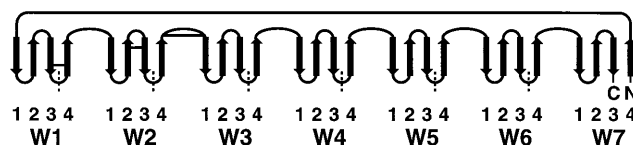


FIG. 1. Folding topology of the integrin β -propeller. The Ws are upright. β -strands are arrows, known disulfides in αIIb (29) are horizontal lines, and boundaries between FG-GAP repeats are marked with vertical dashes.

an inherent twist, the angle of successive strands increases with respect to the central axis (36), giving rise to a propeller pattern in which each sheet represents one blade. Strands 1, 2, 3, and 4 are connected by successive hairpin turns, and strand 4 of one sheet is connected to strand 1 of the next (Fig. 1). Known β -propellers with seven blades are the trimeric G protein beta subunit (G beta; refs. 24 and 37), GO (38), and methylamine dehydrogenase (39). Viral and bacterial neuraminidases contain six blades (40–42).

Secondary Structure and Fold Prediction

A previous prediction of secondary structure using a multiple sequence alignment suggested that the FG-GAP repeats each have four structural segments (43). One segment was of uncertain structure, and a second segment was predicted to contain an amphipathic helix followed by a partly buried β -strand. Two further segments were predicted to be buried β -strands. Since this secondary structure prediction, which was excellent for its time, further methods with improved accuracy have been developed, and a greater number of integrin α subunits are present in the data base, which allows further improvement in accuracy. Since secondary structure is highly conserved within a family, prediction accuracy is improved when information from the aligned sequences of the family is taken into account.

Predictions were made for well-characterized α subunits associated with three different β subunits; the β 3-associated α Ib (CD41) subunit, the β 1-associated α 4 (CD49d) subunit, and the β 2-associated α M (CD11b) subunit. The N-terminal segments containing the seven FG-GAP repeats, with the I domain removed where required, of each α subunit as the target sequence, together with corresponding regions from ≥ 30 different integrin α subunits were submitted to PHD (23). In contrast to the previous prediction in which all seven repeats from 16 α subunits were aligned together with one another to yield one consensus prediction for a repeat (43), the seven repeats were aligned along their entire length, and thus the prediction for each repeat is independent. Remarkably, out of the 28 positions in which β -strand is expected for a β -propeller (four strands \times seven sheets), β -strand was predicted in 25, 25, and 26 instances for α Ib, α 4, and α M, respectively (Fig. 2).

The congruence between a β -propeller fold and secondary structure prediction was further tested with PIPLD. The C-terminal portion of this enzyme contains seven FG-GAP repeats with Ca^{2+} -binding motifs in repeats 1, 2, 3, and 6 (refs. 43 and 46; Fig. 2). Indeed, the sequence homology between the seven repeats in integrins and PIPLD is highly significant ($P < 10^{-6}$ with BLASTP), showing duplication from a common ancestral gene and strongly suggesting adoption of the same protein fold. β -Strands were predicted by PHD in 27 of the 28 positions expected for a β -propeller in PIPLD (Fig. 2).

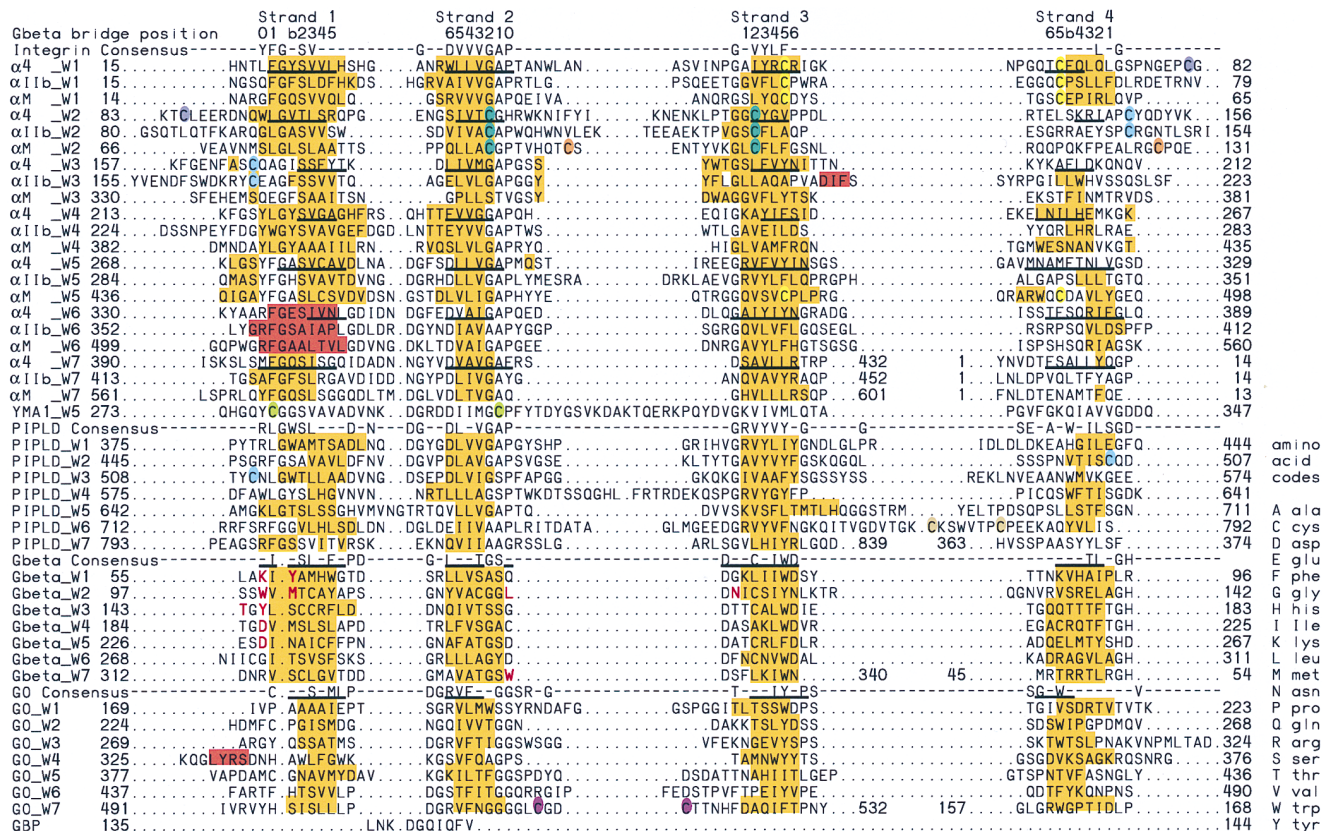


FIG. 2. Alignments, secondary structure, and disulfide bonds. Residues in human integrin α -subunits and bovine phosphatidylinositol phospholipase D (PIPLD) predicted by PHD to be β -strand and α -helix are highlighted in gold and magenta, respectively. The known secondary structure elements of G beta and GO β -propeller domains are highlighted with the same colors. Residues present in β -strands in the α 4 model as determined by DSSP (26) are underlined. Cysteines known in α Ib (29) or predicted in α 4 and α M to be disulfide bonded together are coded with the same color. Equivalent ladder positions in each strand are numbered at the top for G beta; G beta and GO are structurally aligned. The position of framework residues is marked with a line below the consensus sequence for G beta and GO. Residues in G beta that contact the switch II region of the G α -subunit (37, 44) are shown in red. The alignment unit is the sheet (W) rather than the sequence repeat; note that the N-terminal segment is in strand 4, W7. For structural alignments, the seven β -sheets (Ws) of GO (38) and G beta (24, 37, 44) were cut out and structurally superimposed on one another, and the intact β -propellers were also superimposed using the program 3DMALIGN (45). Sequence alignments were prepared with PILEUP in GCG (22) and manually adjusted with MEGALIGN (DNAstar, Madison, WI). All gaps except those between strands were removed. Consensus sequences were calculated with PRETTY in GCG using a plurality of 3/7 of the total number of W.

Automatic fold prediction algorithms confirmed a β -propeller. Input of the seven repeat region into the sequence threading program THREADER (47) showed that all the top, significant hits (Z scores < -3.0) were β -propeller structures: viral and bacterial neuraminidases and methylamine dehydrogenase. The secondary structure and accessibility prediction-based threading program TOPITS (48) yielded the β -propeller domain within GO as the top hit for all three integrins, with a significant score for α IIb ($Z = -3.68$).

The Relation Between β -Sheets and FG-GAP Repeats

Using a common convention for β -propellers, I will call each sheet a W, with each leg of the W corresponding to one β -strand. In the conventional view from the side with the Ws in the propeller upright, the strand 1 to 2 (1-2) loops and the strand 3 to 4 (3-4) loops are on the bottom, and the strand 2 to 3 (2-3) loops and the connections between strand 4 of one W and strand 1 of the next W (4-1 loop) are on the top (Fig. 1). For brevity, 4-1 loops are named after the following W—e.g., the loop between W1 and W2 is the 4-1 loop of W2. The FG-GAP repeats and the Ws could be offset relative to one another, since the amino acid sequence threads through the propeller in a circular fashion, with the N and C termini adjacent in the structure. In most β -propellers, the N- and C-terminal segments are tied together by having both contribute strands to W7 (49).

The four segments in FG-GAP repeats can be identified with particular strands in the Ws by their hydrophobicity, which is well defined in β -propellers. β -Propellers contain a central cavity that is lined with strand 1 and that contains water molecules. Therefore, β -strand 1 is partially exposed to solvent. β -Strand 4 on the exterior of the propeller is fully exposed to solvent and is the least hydrophobic strand. β -Strands 2 and 3 are completely buried and therefore are the most hydrophobic. Hence, the first segment of each FG-GAP repeat, which is the least hydrophobic and least conserved between repeats, corresponds to the outer strand 4, the second segment, which is partially hydrophobic, corresponds to the inner strand 1, and the third and fourth segments, which are hydrophobic, correspond to the buried strands 2 and 3 (Figs. 1 and 2). This topology, with W7 containing strand 4 contributed by the N-terminus and strands 1-3 contributed by the C terminus, is shared by most β -propellers (24, 37, 38, 40–42, 49).

Disulfide Bonds

Three disulfide bonds have been chemically defined within the FG-GAP repeats of α IIb (29). The interrepeat C56–C65 disulfide links strand 3 and strand 4 in W1 (Figs. 1 and 2) and thus is intrasheet. The C107–C130 disulfide links strand 2 and strand 3 of W2. The C146–C167 disulfide is in the 4-1 loop of W3. Disulfide bonds between strands predict their ladder relationship. In β -sheets, each strand is bridged to the next by hydrogen bonds between the peptide carbonyl and amide groups; the ladder positions of residues in all strands of the sheet are thus defined. The residues in each strand at the same ladder or bridge position in the β -sheets of G beta are given the same number at the top of Fig. 2. Disulfide bonds within β -sheets occur between cysteines at the same ladder position on neighboring strands; the $C\alpha$ carbons of residues offset by one or more position are too far apart. The influenza neuraminidase β -propeller contains five such intra-sheet disulfides (40, 41). Integrins were aligned with G beta and GO, such that disulfide bonds would occur between residues in equivalent bridge positions. This defined the ladder relationships between integrin β -strands 2 and 3 in all W, since the alignment between the cysteine and non-cysteine-containing strands is excellent; and between strands 3 and 4 in only those W containing a disulfide, since strand 4 is poorly conserved.

The disulfide bridges in integrins other than α IIb can be predicted from homology to cysteines in α IIb and from multiple alignments. Cysteines at two different alignment positions that are shared by the same subset of integrins are predicted to be linked. All predicted disulfides are compatible with the β -propeller fold. For example, C478 and C490 in W5 of α M are predicted to form a disulfide between β -strands 3 and 4 at the same position as the disulfide in W1 (Fig. 2). C446 of α M and C278 of α 4 do not have a match, and are predicted not to form a disulfide. Indeed, C278 in α 4 has been demonstrated to be free (50). In W5 of the *C. elegans* integrin YMA1 (Fig. 2), a disulfide is predicted between cysteines that replace the F and the A of the FG-GAP motif. The F and the A are aligned in ladder position 1 of strands 1 and 2, providing further support for the alignment shown in Fig. 2.

Sequence Alignment

Among β -propellers, only G beta and GO resemble integrins and PIPLD in number of blades and topology in W7. A structure-based sequence alignment between the Ws of G beta and GO (Fig. 2) reveals the well known WD40 repeats in G beta (51) and sequence repeats within GO as previously mentioned (36). Integrins were aligned with G beta and GO according to the positions of the predicted and known β -strands, and the ladder positions defined by disulfides; sequence similarities were then noted that confirmed or fine-tuned the alignment, particularly with G beta (Fig. 2). For example, the residue before the beginning of β -strand 1 in G beta (“0” position at top of Fig. 2) aligns with similar hydrophilic and bulky residues in integrins and PIPLD. At position 1 of strand 1, the consensus F or L residue in integrins and PIPLD aligns with a buried, hydrophobic consensus I residue in G beta. A β -bulge follows, in which the G of the FG repeat is inserted. There is a precedent for insertion in evolution of a glycine into a β -bulge (52). The subsequent, β -bulged residue in G beta (“b” position, Fig. 2) is located where the polypeptide turns from running across the top of the propeller to down the inner lining of the central cavity. The residue in the b position is often too bulky to be accommodated at position 2, where the β -strands are very closely packed and small residues are found. At positions 2 and 3 in strand 1, consensus SV and SL residues in integrins and PIPLD align with consensus SL residues in G beta. Other similarities in strands 2, 3, and 4 are present (Fig. 2), including alignment in strand 2 of the consensus GA in integrins and PIPLD with the consensus GS in G beta.

Alternative Splicing

The β -propeller model is compatible with alternative splicing of the *Drosophila* PS2, and the mammalian α 6, α 7, and α 3 integrin subunits (53–55). In all cases, the alternatively spliced or deleted region corresponds to β -strands 3 and 4 of W3. Of any strands in a β -propeller, the loss or replacement of β -strands 3 and 4 would have the least deleterious effect, because they constitute the external half of the sheet and leave the packing in the interior of the domain intact.

Integrin β -Propeller Models

Models were built to further test the β -propeller fold, and to make predictions about integrin structure and function. Models made using GO or G beta as templates and the alignment in Fig. 2 yielded packing quality scores in the range -1.4 to -1.7 (27), which suggests that they were threaded correctly—i.e., had the proper fold. Scores were better with G beta than GO as template, in agreement with greater similarity to integrins in sequence and in length of the predicted β -strands (Fig. 2). β -strand 1 extended through the bulge position in the majority of Ws, and in the remaining Ws it had a similar

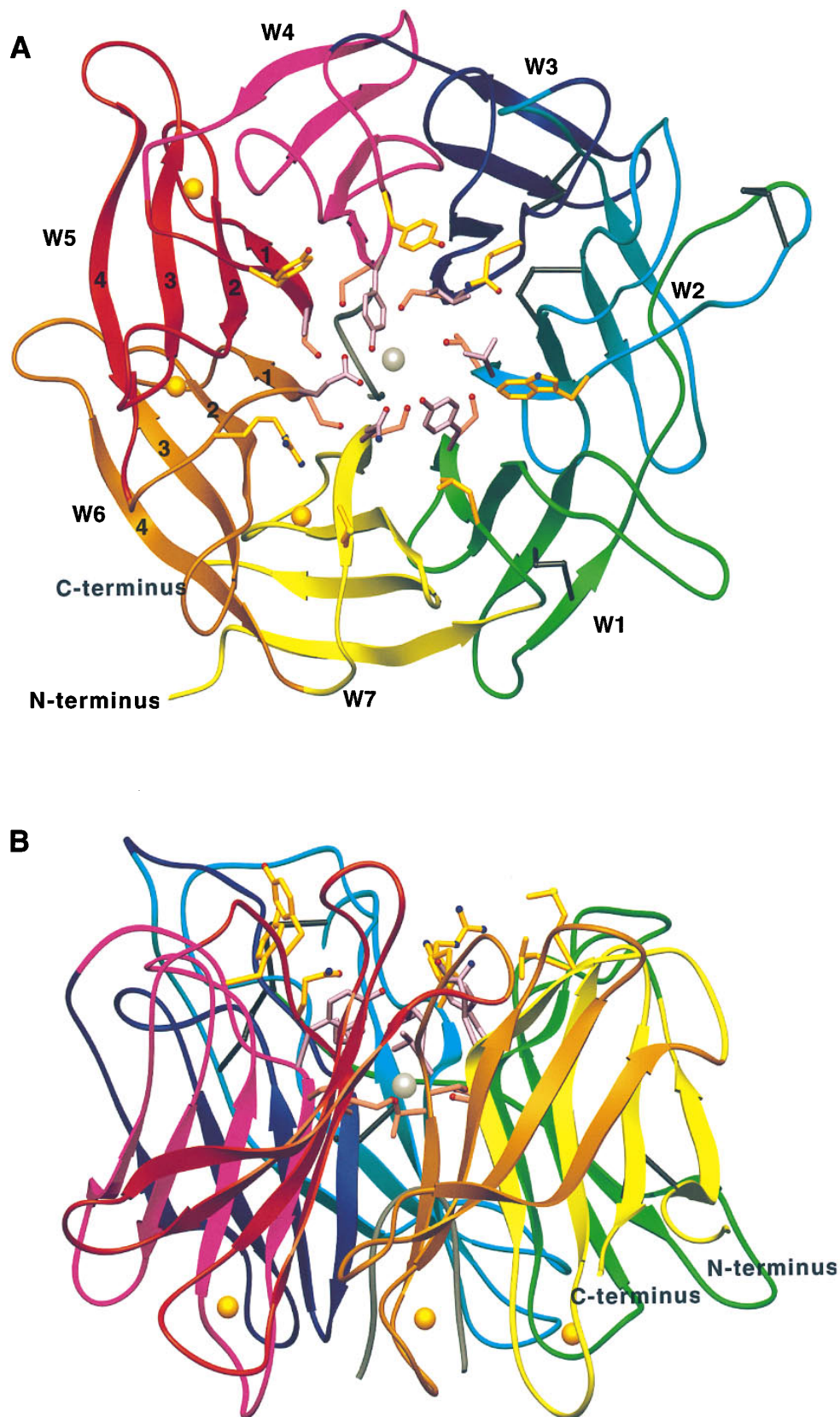


FIG. 3. Ribbon diagrams (65) of the model for the integrin α_4 -subunit β -propeller domain. Views are from the top (A) and side (B). Each W is shown in a different color. A hypothetical polypeptide finger in the central cavity is gray. Cysteines in disulfides are black. Side chains in β -strand 1 in positions 0, b, and 2 are shown in gold, lavender, and rose, respectively; their oxygens and nitrogens are red and blue, respectively. Ca^{2+} ions and a hypothetical Mg^{2+} ion are gold and silver spheres, respectively.

polypeptide chain path (Figs. 2 and 3). Furthermore, all disulfide bonds were properly formed (Fig. 3, black).

The upper surface of the integrin β -propeller models has a pocket resembling a slightly oval amphitheater, with tiers of

residues surrounding the central cavity. The residues in the last part of the 4-1 loop and in the bulge region of strand 1 are successively lower as they approach the turn into the central cavity; strand 2 exits further away from the cavity and parallels strand 1 as residues are successively higher and further from the cavity. Among the residues pointing up into a putative ligand-binding region at the center of the pocket, the residues at the b position in strand 1 (Fig. 3, lavender) line the upper rim of the central cavity and form the inner tier. The next tier is formed by the last residue of the 4-1 loop (0 position of strand 1; Fig. 3, gold). These residues line an inner pocket around the top of the central cavity. Tyrosine, which is frequently present in hapten-binding pockets in antibodies, is abundant at both the 0 and b positions in integrins (Figs. 2 and 3). In PIPLD, tryptophan occurs three times in these positions, and six of the seven residues in the 0 position are basic (Fig. 2), forming a ring of positive charge around a putative active site. The residue equivalent to the P of GAP, and the next usually hydrophilic residue in the 2-3 loop, form an overlapping and more distal tier. Residues in these positions in strand 2 and in the 0 and b positions in strand 1 in G beta (Fig. 2, shown in red) make contact with the switch II region of the G protein α subunit (37, 44).

The β -propeller enzymes all have their active sites at the top of the β -propeller, near the pseudosymmetry axis. It is generally true that enzymes with the same fold have their active sites in the same location and that these sites are present where adjacent loops run in opposite directions (30). The top of the β -propeller is richly endowed with such loops, the alternating 2-3 and 4-1 loops. These often vary in length, providing excursions for substrate and ligand binding and make extensive contacts with one another, particularly toward the axis. By contrast, the 1-2 and 3-4 loops on the bottom of the β -propeller are simple hairpins that are more uniform in length and make few contacts with one another. Fold is likely to be a reliable predictor of binding site for those adhesion molecules with divalent cations in their recognition sites, because similar specializations are required for binding divalent cations and substrates. Indeed, in the case of the nucleotide-binding or double-twisted fold that is shared between enzymes, small G proteins and regulatory G protein α subunits, and integrin I domains (17, 18), the same site that binds divalent cation and substrate in the enzymes and G proteins binds Mg^{2+} and ligand in integrin I domains.

The prediction based on the protein fold of a ligand-binding site on the upper surface of the integrin β -propeller domain is supported by mutational studies. Irie *et al.* (14) surveyed the region of the integrin $\alpha 4$ subunit to which function-blocking mAb mapped by mutating 50 residues in repeats 1-4. Mutation of three residues, Y187, W188, and G190, in loop 2-3 of W3 (Fig. 2), specifically reduced binding to VCAM-1 and fibronectin. Mutation of three residues in the same location in the integrin $\alpha 5$ subunit (14) and in the integrin αIIb subunit (16) similarly reduced or abolished ligand binding. Mutational studies on the $\alpha 3$ subunit have shown that substitution of residues 153-165 with the corresponding sequence from $\alpha 4$ abolishes binding to laminin (56). This region corresponds to the 4-1 loop of W3. In complete agreement with the β -propeller model, these regions in $\alpha 4$ and $\alpha 3$ integrins are not only present on the top of the domain, but the 4-1 loop of W3 and the 2-3 loop of W3 are adjacent in the structure.

Previous studies have emphasized the importance of the FG-GAP repeats, but they have conceptualized them as separate domains and mapped ligand binding to different locations (9). Ligand cross-linking sites correspond to W5 (12) and to the region from W3 to W6 (13). The ligand specificity of αIIb can be transferred with a region spanning from strand 4 of W7 to strand 3 of W5, but not with a region spanning from strand 4 of W3 to strand 3 of W7 (15). mAbs to $\alpha 4$ that block ligand binding map to W3 (34, 57). Ligand-mimetic antibodies

to the fibrinogen-binding site on αIIb map to the 2-3 loop of W3 (16). By contrast, eight different mAb to the $\alpha 3$ subunit that block adhesion to laminin and homotypic aggregation map to a region extending from strand 4 of W7 to the 1-2 loop of W2 (56). These results are difficult to interpret with a model in which the FG-GAP repeats represent tandem, independent domains, but are highly consistent with a β -propeller model in which all FG-GAP repeats contribute to a common surface.

Since almost all integrins require Mg^{2+} for ligand binding and their ligands contain functionally important glutamic or aspartic acid residues that may coordinate a divalent cation (1-4, 9), it is reasonable to inspect models to determine whether a Mg^{2+} -binding site might be present in the putative ligand-binding site on the upper surface. A substantial proportion of β -propellers bind cations, and the sites are located on the upper surface in the 1-4 and 2-3 loops or in the central cavity (38, 41, 58-61). In integrin β -propeller models, residues with side chains containing oxygen are abundant near the upper end of the central cavity, although few of these residues are acidic (Fig. 3). The only type of divalent cation-binding region that is apparent would be similar to the MIDAS in I domains (17). The coordination chemistry in the MIDAS is unusual for Mg^{2+} and never seen for Ca^{2+} . Three oxygens from serine and threonine side chains and one water molecule donate coordinations to a Mg^{2+} ion in an xy plane. The lack of charged residues in the primary coordination sphere in the MIDAS has significance for the strength of binding to an acidic residue in the ligand. Serines and threonines are abundant in an xy plane at the top of the central cavity in integrin β -propeller models (Fig. 3, rose residues), in position 2 of strand 1 (Fig. 2). This is at the narrowest aperture of the central channel. Moving around the circumference of this aperture, each group of three adjacent serine and threonine residues—i.e., those in adjacent β -strands—is in a remarkably similar disposition to the serines and threonine in the MIDAS. The C α carbons of the adjacent residues are 5.21 ± 0.23 Å apart, compared with 5.55 Å in the MIDAS, and the next most adjacent are 9.37 ± 0.37 Å apart, compared with 8.4 Å in the MIDAS. Only one Mg^{2+} could be bound at a time, but there are as many as seven different possible binding sites distributed around the pseudosymmetry axis, which could have important implications for regulation of ligand affinity and specificity. The fourth position in the xy plane might be occupied by a water molecule; the residues are too far apart to form four coordinations to an ion, as occurs in four-bladed β -propellers (59-61). An acidic residue to hold a water in position to coordinate the Mg^{2+} in the $-z$ position as found in the I domain MIDAS appears to be missing. However, one might imagine a polypeptide finger from another domain—e.g., the β subunit—protruding up into the central cavity. Such a finger could have a residue at its tip that coordinates with the cation from below, as occurs in the central cavity of GO. Indeed, the cavities of all seven- and eight-bladed propellers except G beta are filled with a polypeptide finger or prosthetic group, and G beta may not be a real exception, because it might bind to loops from seven transmembrane receptors (37). There is no evidence for such a finger in integrins; however, it was included in models to demonstrate that there would be room for one, particularly when placed lower in the central cavity than in GO, as appropriate for the predicted lower position of the cation. The central cavity of G beta is narrower, but more conical, than that of GO. The remaining $+z$ coordination position for a hypothetical Mg^{2+} would be above it in the center of the upper surface of the β -propeller, in an ideal position for coordination to a bound ligand.

The putative Ca^{2+} -binding motifs in integrins located in the 1-2 loops were modeled on the Ca^{2+} -binding motif in galactose-binding protein (25), since, in contrast to the EF-hand motif, in galactose-binding protein one of the coordinating residues is located in a β -strand, as predicted in integrins (Fig.

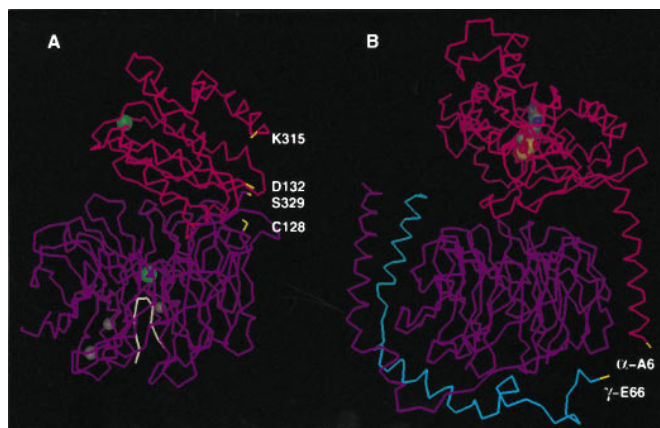


FIG. 4. Superimposition of the α M integrin subunit I domain and β -propeller domain on a trimeric G protein. (A) The α M I domain (magenta; ref. 17), β -propeller domain (purple), and polypeptide finger (white) are shown with Ca^{2+} traces. Mg^{2+} and Ca^{2+} ions are green and grey spheres, respectively. Residues at the ends of the domains that may be connected by linker segments are shown in yellow. According to the superimposition, the α carbons of C128 and D132 are 7.4 Å apart, and K315 and S329 are 11 Å apart. (B) The trimeric G protein α -subunit (magenta) and bound GDP (space-filling), β -subunit (purple), and γ -subunit (cyan; ref. 44) are shown in the same orientation. The residues in yellow are separated by 5–7 residues from lipid-modified termini that bind to the inner face of the plasma membrane (37, 44). These interactions with the plasma membrane increase association between the α -subunit and the $\beta\gamma$ -subunit dimer and to some extent may act like the covalent linkage between the I domain and β -propeller domain in integrin α -subunits. The α M β -propeller domain was modeled on α 4, and superimposed on G beta with the alignment in Fig. 2 and a 2/7 rotation on the pseudosymmetry axis—i.e. with α M W1 on G beta W6. A total of 49 residues in α M I domain elements β A, α 1, β B, β D, β E, β F, and α 7 (17) was superimposed on G α β 1, α 1, β 3, β 4, β 5, β 6, and α 5 (44), respectively, with an RMS of 1.56 Å using LOOK. The figure was made with LOOK.

2). The Ca^{2+} -binding sites are predicted to be near one another on the lower surface of the β -propeller domain (Fig. 3). The N terminus of integrins is also predicted to be on the lower surface and has highly conserved Asn and Asp residues at positions 2 and 4 that might ligate a cation. Since removal of Ca^{2+} activates ligand binding by many integrins and destabilizes association with the β subunit in several integrins (4, 6, 9), the lower surface of the β -propeller domain in the α subunit may be involved in interactions with the β subunit and in conformational changes that regulate ligand binding.

The I domain is inserted into the β -propeller domain in the 4-1 loop between W2 and W3. Integrins with I domains uniquely contain cysteines in the 2-3 loop of W2 and in the 4-1 loop, just before the beginning of the I domain, that are predicted to be disulfide-bonded (α M C97 and C128; Fig. 2). The two polypeptide chain connections and the disulfide bond are all predicted to be close together on the upper rim of the β -propeller domain. The connections to the β -propeller are also close together in the I domain structure (17), and a hydrophilic linker segment is present on the C terminal side of the I domain. These features suggest a hinge between the I domain and the β -propeller domain.

The I domain is structurally homologous to the G protein α -subunit (17, 19), and when the integrin I domain and β -propeller domain are superimposed on the trimeric G protein structure (37, 44), the N and C termini of the I domain are very close to a 4-1 loop on the rim of the β -propeller domain, as proposed for the hinge connection (Fig. 4). The superimposition illustrates the relative sizes of the β -propeller and I domains, and one possible orientation between them. Since the I domain can be expressed independently of the integrin β -propeller domain (17), in an intact integrin, move-

ments at the hinge might allow reversible binding of the I domain to the upper surface of the β -propeller domain, as in interactions between G protein α and β subunits. Conformational changes in the β -propeller domain similar to those that regulate ligand binding might regulate I domain binding to the β -propeller domain and, in turn, regulate ligand binding by the I domain. Specifically, a conformational change in the I domain MIDAS that correlates with binding of a surrogate ligand and that is linked to a substantial change in position of the C-terminal α -helix (17–20) would be expected to be linked to movements of the I domain relative to the β -propeller domain.

Conclusion

The convergence of multiple, independent lines of evidence provides strong support for the conclusion that the FG-GAP repeats of integrins fold into a β -propeller domain. Indeed, GO, G beta, integrins, and PIPLD appear so similar that it is likely that they evolved from a common ancestral β -propeller domain, as previously suggested for GO and G beta (24). It is interesting that integrins have statistically significant sequence homology to the enzyme PIPLD, which, like integrins, is predicted here to have a β -propeller domain. This demonstrates an explicit evolutionary link between enzymatic and nonenzymatic proteins with β -propeller folds. It is also interesting that like integrins and PIPLD, most β -propeller enzymes are extracellular—i.e., membrane-bound (40, 41), secreted (38, 42, 60, 61), or periplasmic (39, 49, 58). The β -propeller domain appears to be well suited to many different types of functions. Indeed, because there are so many different members of the integrin (1–4) and G protein β -subunit or WD40 families (51), members of the β -propeller family involved in ligand binding, regulation, and signaling may be more numerous than the enzymes.

Previous fold predictions have been reviewed (62, 63). The prediction that the integrin I domain had the same fold as the small G protein ras—i.e., the double-twisted, nucleotide-binding fold (64) is most relevant. This was an outstanding and correct prediction (62), although the subsequent structure revealed a different variant of this fold, in which the order of two of the β -strands in the sheet is reversed (17). Among fold predictions, a β -propeller is auspicious. These are among the largest protein domains known, and therefore the relative disposition of a large number of amino acid residues can be predicted. Furthermore, the β -propeller fold is highly stereotyped (36). In most protein folds, variation can occur in the number of secondary structure elements and in the order in which the elements occur in the amino acid sequence. By contrast, all known β -propellers contain four β -strands per sheet and an invariant pattern for threading the sequence through the three-dimensional structure.

Three-dimensional models of the integrin β -propeller domain are therefore likely to be more accurate than usually expected for models based on templates that lack statistically significant sequence homology. Needless to say, there will be discrepancies with the actual structures, and accuracy will vary as function of position in the fold. The strands and nearby positions in loops with clear correspondence to the template will be predicted the most accurately. It is propitious that these positions include residues thought to be important in binding ligand and possibly a Mg^{2+} ion, on the upper surface of the β -propeller domain. Some 4-1 and 2-3 loops in integrins are particularly long and are not predictable; in β -propellers these may go up in the general direction of the pseudosymmetry axis, extend in a radial direction, or even drape over the outer cylindrical surface. However, the adjacency of loops is well defined in β -propellers as shown by inspection of G beta and GO. On the upper surface, most loops contact other loops in the same and two adjacent W; there are few more distant

contacts. On the lower surface, loops are more regular and the interactions more circumscribed. The 1-2 loops only interact with 1-2 loops in neighboring W and with 3-4 loops in the same and in the following W; 3-4 loops do not interact with neighboring 3-4 loops. One prediction of adjacencies by the β -propeller model has already been confirmed; the 1-2 loop of W5 is in close proximity to the 3-4 loop of W6 in the α M integrin subunit (C. Oxvig and T.A.S., unpublished).

I thank members of my family and laboratory for their understanding during this project. Biocomputing resources on the internet, particularly at European Molecular Biology Laboratory (EMBL), were essential for this work. I am grateful to the C. Sander group and especially to B. Rost for his support of the PIR list submission format for PHD; G. Vriend for WHAT IF; D. Jones for THREADER; A. Sáli for MODELLER and advice; J.-H. Wang for advice; S. Sprang and J. Sondek for G protein coordinates; R. Liddington for I domain coordinates; Y. Takata for a preprint; T. Huynh and C. Huang for installation and help with programs; C. Oxvig for help with editing sequences; J. Casanovas for help with RIBBONS; and S. Chen for help with spreadsheets for coordinate transformations and analysis of β -propellers. This work was supported by National Institutes of Health Grants CA31798 and CA31799.

1. Springer, T. A. (1990) *Nature (London)* **346**, 425–433.
2. Hemler, M. E. (1990) *Annu. Rev. Immunol.* **8**, 365–400.
3. Hynes, R. O. (1992) *Cell* **69**, 11–25.
4. Smyth, S. S., Joneckis, C. C. & Parise, L. V. (1993) *Blood* **81**, 2827–2843.
5. Dustin, M. L. & Springer, T. A. (1989) *Nature (London)* **341**, 619–624.
6. Diamond, M. S. & Springer, T. A. (1994) *Curr. Biol.* **4**, 506–517.
7. Ginsberg, M. H. (1995) *Biochem. Soc. Trans.* **23**, 439–446.
8. Weisel, J. W., Nagaswami, C., Vilaire, G. & Bennett, J. S. (1992) *J. Biol. Chem.* **267**, 16637–16643.
9. Loftus, J. C., Smith, J. W. & Ginsberg, M. H. (1994) *J. Biol. Chem.* **269**, 25235–25238.
10. Corbi, A. L., Miller, L. J., O'Connor, K., Larson, R. S. & Springer, T. A. (1987) *EMBO J.* **6**, 4023–4028.
11. Tuckwell, D. S., Brass, A. & Humphries, M. J. (1992) *Biochem. J.* **285**, 325–331.
12. D'Souza, S. E., Ginsberg, M. H., Burke, T. A. & Plow, E. F. (1990) *J. Biol. Chem.* **265**, 3440–3446.
13. Smith, J. W. & Cheresch, D. A. (1990) *J. Biol. Chem.* **265**, 2168–2172.
14. Irie, A., Kamata, T., Puzon-McLaughlin, W. & Takada, Y. (1995) *EMBO J.* **14**, 5550–5556.
15. Loftus, J. C., Halloran, C. E., Ginsberg, M. H., Feigen, L. P., Zablocki, J. A. & Smith, J. W. (1996) *J. Biol. Chem.* **271**, 2033–2039.
16. Kamata, T., Irie, A., Tokuhira, M. & Takada, Y. (1996) *J. Biol. Chem.* **271**, 18610–18615.
17. Lee, J.-O., Rieu, P., Arnaout, M. A. & Liddington, R. (1995) *Cell* **80**, 631–638.
18. Qu, A. & Leahy, D. J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 10277–10281.
19. Lee, J.-O., Bankston, L. A., Arnaout, M. A. & Liddington, R. C. (1995) *Structure (London)* **3**, 1333–1340.
20. Qu, A. & Leahy, D. J. (1996) *Structure (London)* **4**, 931–942.
21. Huang, C. & Springer, T. A. (1995) *J. Biol. Chem.* **270**, 19008–19016.
22. Devereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
23. Rost, B. (1996) *Methods Enzymol.* **266**, 525–539.
24. Sondek, J., Bohm, A., Lambricht, D. G., Hamm, H. E. & Sigler, P. B. (1996) *Nature (London)* **379**, 369–374.
25. Vyas, N. K., Vyas, M. N. & Quijcho, F. A. (1987) *Nature (London)* **327**, 635–638.
26. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
27. Vriend, G. (1990) *J. Mol. Graphics* **8**, 52–56.
28. Barclay, A. N., Birkeland, M. L., Brown, M. H., Beyers, A. D., Davis, S. J., Somoza, C. & Williams, A. F. (1993) *The Leucocyte Antigen Facts Book* (Academic, London).
29. Calvete, J. J., Henschen, A. & Gonzalez-Rodriguez, J. (1989) *Biochem. J.* **261**, 561–568.
30. Branden, C. & Tooze, J. (1991) *Introduction to Protein Structure* (Garland, New York), pp. 1–302.
31. Springer, T. A. (1994) *Cell* **76**, 301–314.
32. Mahoney, P. A., Weber, U., Onofrechuk, P., Beissmann, H., Bryant, P. J. & Goodman, C. S. (1991) *Cell* **67**, 853–868.
33. Diamond, M. S., Garcia-Aguilar, J., Bickford, J. K., Corbi, A. L. & Springer, T. A. (1993) *J. Cell Biol.* **120**, 1031–1043.
34. Schiffer, S. G., Hemler, M. E., Lobb, R. R., Tizard, R. & Osborn, L. (1995) *J. Biol. Chem.* **270**, 14270–14273.
35. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995) *J. Mol. Biol.* **247**, 536–540.
36. Murzin, A. G. (1992) *Proteins* **14**, 191–201.
37. Wall, M. A., Coleman, D. E., Lee, E., Iniguez-Lluhi, J. A., Posner, B. A., Gilman, A. G. & Sprang, S. R. (1995) *Cell* **83**, 1047–1058.
38. Ito, N., Phillips, S. E. V., Stevens, C., Ogel, Z. B., McPherson, M. J., Keen, J. N., Yadav, K. D. S. & Knowles, P. F. (1991) *Nature (London)* **350**, 87–90.
39. Vellieux, F. M. D., Huitema, F., Groendijk, H., Kalk, K. H., Jzn, J. F., Jongejan, J. A., Duine, J. A., Petratos, K., Drenth, J. & Hol, W. G. J. (1989) *EMBO J.* **8**, 2171–2178.
40. Colman, P. M., Varghese, J. N. & Laver, W. G. (1983) *Nature (London)* **303**, 41–44.
41. Burmeister, W. P., Ruigrok, R. W. H. & Cusack, S. (1992) *EMBO J.* **11**, 49–56.
42. Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R. & Taylor, G. L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9852–9856.
43. Tuckwell, D. S., Humphries, M. J. & Brass, A. (1994) *Cell Adhes. Commun.* **2**, 385–402.
44. Lambright, D. G., Sondek, J., Bohm, A., Skiba, N. P., Hamm, H. E. & Sigler, P. B. (1996) *Nature (London)* **379**, 311–319.
45. Sáli, A. & Blundell, T. L. (1993) *J. Mol. Biol.* **234**, 779–815.
46. Scallon, B. J., Fung, W.-J. C., Tsang, T. C., Li, S., Kado-Fong, H., Huang, K.-S. & Kochan, J. P. (1991) *Science* **252**, 446–448.
47. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992) *Nature (London)* **358**, 86–89.
48. Rost, B. (1995) in *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology: Cambridge, U.K.*, ed. Rawlings, C. (AAAI, Menlo Park, CA), pp. 314–321.
49. Xia, Z.-x., Dai, W.-w., Xiong, J.-p., Hao, Z.-p., Davidson, V. L., White, S. & Mathews, F. S. (1992) *J. Biol. Chem.* **267**, 22289–22297.
50. Pujades, C., Teixido, J., Bazzoni, G. & Hemler, M. E. (1996) *Biochem. J.* **313**, 899–908.
51. Neer, E. J., Schmidt, C. J., Nambudripad, R. & Smith, T. F. (1994) *Nature (London)* **371**, 297–300.
52. Chan, A. W. E., Hutchinson, E. G., Harris, D. & Thornton, J. M. (1993) *Protein Sci.* **2**, 1574–1590.
53. Brown, N. H., King, D. L., Wilcox, M. & Kafatos, F. C. (1989) *Cell* **59**, 185–195.
54. Ziober, B. L., Vu, M. P., Waleh, N., Crawford, J., Lin, C.-S. & Kramer, R. H. (1993) *J. Biol. Chem.* **268**, 26773–26783.
55. Delwel, G. O., Kuikman, I. & Sonnenberg, A. (1995) *Cell Adhes. Commun.* **3**, 143–161.
56. Zhang, X.-P., Puzon-McLaughlin, W., Irie, A., Kovach, N., Prokopishyn, N. L., Laferte, S., Takeuchi, K. I., Tsuji, T. & Takada, Y. (1997) *J. Biol. Chem.*, in press.
57. Kamata, T., Puzon, W. & Takada, Y. (1995) *Biochem. J.* **305**, 945–951.
58. Xia, Z.-x., Dai, W.-w., Zhang, Y.-f., White, S. A., Boyd, G. D. & Mathews, F. S. (1996) *J. Mol. Biol.* **259**, 480–501.
59. Faber, H. R., Groom, C. R., Baker, H. M., Morgan, W. T., Smith, A. & Baker, E. N. (1995) *Structure (London)* **3**, 551–559.
60. Li, J., Brick, P., O'Hare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E. & Blow, D. M. (1995) *Structure (London)* **3**, 541–549.
61. Libson, A. M., Gittis, A. G., Collier, I. E., Marmer, B. L., Goldberg, G. I. & Lattman, E. E. (1995) *Nat. Struct. Biol.* **2**, 938–942.
62. Russell, R. B. & Sternberg, M. J. E. (1995) *Curr. Biol.* **5**, 488–490.
63. Sippl, M. J. & Flockner, H. (1996) *Structure (London)* **4**, 15–19.
64. Edwards, Y. J. K. & Perkins, S. J. (1995) *FEBS Lett.* **358**, 283–286.
65. Carson, M. (1991) *J. Appl. Crystallogr.* **24**, 958–961.