Definition of EGF-like, closely interacting modules that bear activation epitopes in integrin β subunits

Junichi Takagi*, Natalia Beglova†, Padmaja Yalamanchili*, Stephen C. Blacklow†, and Timothy A. Springer*‡

*Center for Blood Research and Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115; and †Brigham and Women's Hospital and Department of Pathology, Harvard Medical School, 75 Francis Street, Boston, MA 02115

Integrin β subunits contain four cysteine-rich repeats in a long extracellular stalk that connects the headpiece to the membrane. Most mAbs to integrin activation epitopes map to these repeats, and they are important in propagating conformational signals from the membrane/cytosol to the ligand-binding headpiece. Sequence analysis of a protein containing only 10 integrin-like, cysteine-rich repeats suggests that these repeats start one cysteine earlier than previously reported. By using the new repeat boundaries, statistically significant sequence homology to epidermal growth factor-like domains is found, and a disulfide bond connectivity of the eight cysteines is predicted that differs in three of four disulfides from a previous prediction of epidermal growth factor-like modules [Berg, R. W., Leung, E., Gough, S., Morris, C., Yao, W.-P., Wang, S.-x., Ni, J. & Krissansen, G. W. (1999) *Genomics* 56, 169–178]. N-terminally truncated β2 integrin stalk fragments were well expressed and secreted from 293 T cells when they began at repeat boundaries but not when they began one cysteine earlier or later. Furthermore, peptides that correspond to module 3 or modules 2 + 3 were expressed in bacteria and refolded. The module 2 + 3 fragment was as reactive with three mAbs to activation epitopes as a β2 fragment expressed in eukaryotic cells, indicating a native fold. Only one residue intervenes between the last cysteine of one module and the first cysteine of the next. This arrangement is consistent with a tight intermodule connection, a prerequisite for signal propagation from the membrane to the ligand binding headpiece.

Uniquely among adhesion molecules, integrins are found on all cells in metazoan organisms that are adherent or need to rapidly become adherent. They bind ligands on the surface of other cells and in the extracellular matrix, connect the extracellular environment to the actin and keratin cytoskeletons, regulate cell migration and growth, and communicate signals bidirectionally across the plasma membrane (1). Integrins contain two noncovalently associated, large glycoprotein α and β subunits with extracellular domains of >940 and >640 residues, respectively. A globular headpiece contains N-terminal domains of the α and β subunits, and two ≈16-nm-long stalk regions composed of more C-terminal segments from the α and β subunits connect the ligand-binding headpiece to the transmembrane and C-terminal cytoplasmic domains (2). In a process termed inside-out signaling, signals from the cytoplasm can rapidly activate ligand binding by promoting conformational reshaping of the headpiece (1, 3, 4). Four cysteine-rich repeats in the stalk region of the β subunit are an important link in inside-out signaling (3–7), and antibodies to this region can either directly activate ligand binding or act as probes that bind only to activated integrins (1, 8, 9).

Despite the importance of the integrin cysteine-rich repeats, they remain ill-defined. In many mammalian proteins with sequence repeats, the repeat boundaries define structural module boundaries (10). Recognition of the correct boundaries between these repeats is sometimes difficult; however, its accomplishment can lead to successful structure predictions and the design of fragments that are amenable to solution of atomic resolution structures. A recent example is the correct definition of the boundaries of the YWTD repeats, which led to the prediction that they fold into a six-bladed β-propeller domain with a specific predicted structure (11), and more recently, led to an atomic resolution crystal structure of this domain from the low density lipoprotein receptor (12).

Cysteine-rich repeats were identified in the first report of an integrin β-subunit sequence (13). Initially, three repeats each containing eight cysteines were recognized. Subsequently, a fourth more N-terminal repeat with a slightly different spacing of cysteines was identified (14). Many integrin β subunits have now been found and sequenced, including representatives from diverse metazoan phyla including Nematoda, Arthropoda, Cnidaria (corals), and Porifera (sponges) and eight different β subunits in mammals (15, 16). In each case, four cysteine-rich repeats are present, i.e., there has been no expansion or contraction of their number in the last 1.2 billion-1.5 billion years (17). Since their initial identification, the boundaries of the cysteine-rich repeats have never been questioned. However, definition of the boundaries is difficult because the adjacent N- and C-terminal segments are also cysteine rich, and aside from the cysteines and a few glycines, sequence conservation among the repeats is low. Furthermore, although the cysteines are disulfide-bonded, the disulfide connectivity has not been defined because of the close spacing of the cysteines and the resistance of the repeats to proteolysis; in contrast, disulfide bond connectivity has been defined for more widely separated cysteines present in other regions of the β subunit (18). The integrin β subunit cysteine-rich repeats have been reported to be homologous to laminin-epidermal growth factor (EGF) modules, a special type of EGF domain with eight cysteines (19). However, the amino acid sequence identity was low and was not statistically evaluated, i.e., an evolutionary relationship was not established. Moreover, our own unpublished analysis with BLAST and PSI-BLAST (20) shows that the reported relationship is not statistically significant.

Recently, a novel cDNA was reported that encodes a putative secreted protein termed TIED with 10 integrin-like cysteine-rich repeats (21). The amino acid sequence identity between the TIED and integrin cysteine-rich repeats is as high as 68% (21) and is highly statistically significant (our results show $E = 10^{-32}$ with BLAST), implying that the TIED and integrin cysteine-rich repeats have the same protein fold. Our examination of the TIED and integrin sequences demonstrates that the cysteine-rich repeats start one cysteine earlier than previously proposed. These repeats show statistically significant sequence homology with EGF-like domains. Compared to a previous alignment with EGF domains (19, 21), only three of eight cysteines and one of four disulfide bonds are equivalently assigned. Our module boundaries have been confirmed by truncation in both mammalian and bacterial cells and by expression of mAb-defined

---

BIOCHEMISTRY

activation epitopes. The integrin EGF domains have unusual features, including only one noncysteine residue between modules, suggesting that they are closely interconnected and well suited for conveying signals between the cytoplasmic domains and the ligand-binding integrin headpiece.

Materials and Methods

Sequence Alignment and Homology. The 36 integrin β subunit sequences used here have been described (15). TIED repeats 1–10, all integrin cysteine-rich repeats from the 36 integrin subunits that contain eight cysteines each, i.e., repeats 2, 3, and 4 but not repeat 3 from $\beta4$ and $\beta8$ (104 sequences total), and EGF modules from tenascin X (repeats 3–17 from GenBank accession no. gi9087217) and reelin (repeats 1–8 from gi4826978) were aligned with PRRP (22) by using the Gonnet amino acid substitution matrix with a gap opening penalty of 2, and elongation penalties of 6, 7, 8, or 9. PAIRWISE was used to make multiple sequence alignment profiles and consensus sequences (23). The sequence of the EGF module of ADAM-11 is from gi3510511. The 10 TIED repeats and 104 integrin cysteine-rich repeats were aligned (multiple alignment with 114 sequences) to obtain a consensus sequence that was used to seed PSI-BLAST (20), 3D-PSSM (24), and STRUCTURE-3D (25) sequence homology searches. Statistical significance was determined from the expectation (E) value returned by STRUCTURE-3D, 3D-PSSM, PSI-BLAST, or BLAST.

mAbs. The murine mAbs CBR LFA-1/2 and CBR LFA-1/7 have been described (26). MEM48 (27) was a gift from V. Horejsî, Institute of Molecular Genetics, Czechoslovakia. KIM127 (28) and KIM185 (29) were gifts from M. Robinson, Imperial Cancer Research Fund, London. mAb 6.7 (30) was obtained from the Fifth International Leukocyte Workshop (45).

DNA Construction and Transfection. By using wild-type human $\beta2$ cDNA as a template, segments starting from residues 455, 460, 474, or 513 and ending at residue 678 were PCR-amplified (mature sequence numbering). The 5′ primer included a *Not*I site, and the 3′ primer included a stop codon followed by a *Xba*I site. PCR fragments were ligated into the pEF1/V5-HisA vector (Invitrogen) carrying a mouse entactin signal peptide (J.T., unpublished work) by using *Not*I and *Xba*I sites so that the $\beta2$ sequences begin immediately after the signal cleavage site. All fragments were verified by DNA sequencing. DNAs were transiently transfected into 293T cells by using calcium phosphate (31).

Metabolic Labeling and Immunoprecipitation. Transfected cells in 6-well plates ($\approx80\%$ confluency, 24 h after transfection) were incubated with 0.3 mCi [^{35}S]methionine and cysteine (NEN) in 1.5 ml of labeling medium (methionine and cysteine-free RPMI 1640 medium containing 10% dialyzed FCS) for 1 h and chased by adding the same volume of labeling medium containing 500 μg/ml cysteine and 100 μg/ml methionine for an additional 16 h. The culture supernatants were harvested and centrifuged to remove cell debris and subjected to immunoprecipitation by using $\beta2$ mAbs and Protein G agarose. Materials eluted from the beads were subjected to 10% SDS/PAGE and fluorography.

Preparation of a Native Monomeric $\beta2$ Mutant. The DNA fragment encoding Tyr-103–Tyr-338, corresponding to the I-like domain, was deleted from the full-length $\beta2$ cDNA in plasmid AprM8 (32) by overlap PCR. This construct preserves Cys-3 and Cys-425, which form a long-range disulfide bond (18). Termination codon and *Not*I restriction sites were introduced after Asn-678 by extension PCR, and the entire insert was excised by *Xho*I and *Not*I and cloned into the pBJ5/GS expression vector (33). $\beta2$ mutant DNA (2 μg) was cotransfected with 0.1 μg of pEF-puro

into Chinese hamster ovary Lec 3.2.8.1 cells and selected against 8 μg/ml puromycin. Stable clones secreting mutant $\beta2$ were screened by sandwich ELISA using CBR LFA-1/7 as capture antibody and biotinylated CBR LFA-1/2 as detection antibody. The clone with the highest expression was cultured in roller bottles (Corning no. 25291). Soluble monomeric mutant $\beta2$ ($\beta2\Delta I$) was purified from ≈4 liters of culture supernatant by immunoaffinity chromatography on CBR LFA-1/7-Sepharose followed by gel filtration on Superdex 200 HR (Amersham Pharmacia). The purified protein showed a single 52-kDa band on SDS/PAGE and eluted as an 84-kDa molecular species on a Superdex 200 column, suggesting an elongated shape.

Preparation of Refolded Peptides. Expression, refolding, and purification of individual and tandem cysteine-rich modules will be published elsewhere in detail. In brief, DNA fragments encoding module 3 (residues 513–552) or modules 2 + 3 (residues 460–552) of $\beta2$, each flanked by Ala residues at both ends, were PCR-amplified by using appropriate oligonucleotides and cloned into the vector pMM-LR5 (34). Recombinant modules were expressed as trpLE fusion peptides, cleaved from the fusion peptide with cyanogen bromide followed by RP-HPLC purification, and refolded in a redox buffer (34).

ELISA. Microtiter wells were coated with 5 μg/ml $\beta2\Delta I$ at 4°C overnight, and nonspecific binding sites were blocked by 1% BSA in 20 mM Tris and 150 mM NaCl, pH 7.5 (TBS). Varying concentrations of $\beta2\Delta I$ or cysteine-rich module peptides followed by $\beta2$ mAbs (0.5 μg/ml for CBR LFA-1/2, MEM48, KIM127, KIM185, and 1:30,000 dilution of ascites for 6.7) were added to the wells in 50 μl of TBS containing 0.1% BSA and incubated at room temperature for 1 h. After washing three times with TBS, wells were further incubated with a 1:3,000 dilution of anti-mouse IgG-peroxidase (ICN) for 30 min and washed four times with TBS, and color was developed with the substrate 2,2′-azinobis[3-ethylbenzothiazoline-6-sulfonic acid]-diammonium salt (Zymed).

Results

Repeat Assignment. Previous assignment of the boundaries of the cysteine-rich repeats in integrins was made difficult by the presence of adjacent N- and C-terminal cysteine-rich regions. In contrast, TIED contains only 10 integrin-like cysteine-rich repeats and a total of 80 cysteines. Parsimoniously, TIED can be divided into 10 modules, each containing eight cysteines capable of forming four intramodule disulfide bonds (Fig. 1*A*). However, the previously reported repeat boundaries for TIED were the same as previously assumed for integrins and resulted in one "extra" cysteine before repeat one, eight cysteines each in repeats 1–9, and only seven cysteines in repeat 10 (21). Our alternate repeat boundary assignment results in an alignment that is improved in several respects (Fig. 1*A*). (*i*) The previously omitted cysteine, residue 17 in the mature protein, readily aligns with the cysteines in other repeats because the spacing between it and the next cysteine is identical in repeats 1, 3, 5, 7, and 9 (Fig. 1*A*). (*ii*) The other residues added to the alignment in repeat 1, residues 18–27, show many identities and similarities with residues in repeats 2–10. Indeed, in this region, repeat 1 is at least as similar to repeats 2–10 as any of these latter repeats are among themselves (Fig. 1*A*). (*iii*) Finally and most important structurally, an even number of cysteines per repeat would allow all of the cysteines to form disulfide bonds within the repeats, and hence modules could be formed that would contain structurally equivalent disulfide bonds.

   To examine whether similar repeat boundaries could be identified in integrins, a multiple alignment sequence profile of the 10 TIED repeats was aligned with integrin β subunits by using PAIRWISE (23). Four high-scoring segments were identi-
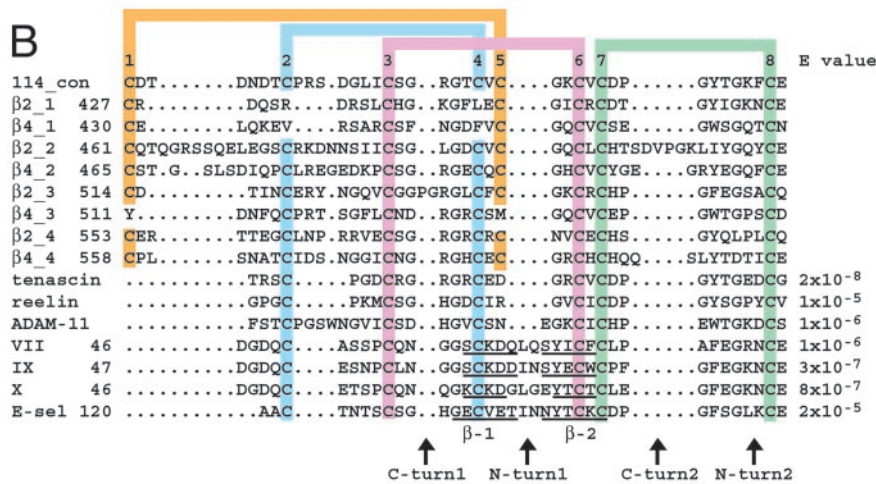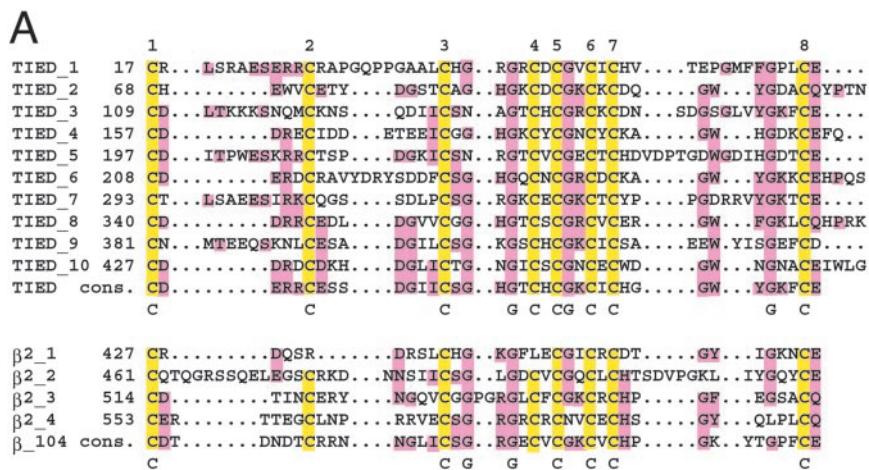
Fig. 1. (*A*) Alignment of the 10 repeats in the TIED protein and the four repeats in integrin β subunits. The most commonly occurring (consensus) residue at each sequence position is colored. Numbering is for mature sequences; in TIED the predicted signal peptide is 23 residues. Residues that are identical in all sequences are shown below each alignment. Alignments were made by using PRRP (22). (*B*) Sequence homology of integrin cysteine-rich repeats to structurally characterized EGF domains in coagulation factors and E-selectin, and annotated EGF domains in tenascin, reelin, and ADAM-11. The expectation (E) values are from STRUCTURE-3D (25), PSI-BLAST (20), or 3D-PSSM (24) searches that were seeded with the consensus sequence of 104 integrin and 10 TIED cysteine-rich repeats (114-con). The cysteines are numbered above the alignment and cysteines predicted or known to be disulfide bonded are connected by colored lines. β-strands 1 and 2 in the structurally characterized EGF domains are underlined and positions of turns indicated with arrows. The factor VII, IX, and X sequences are from the Protein Data Bank structures 1DAN (L chain), 1EDM (B chain), and 1XKA (L chain), respectively. The human tenascin X and human reelin sequences are consensus sequences prepared as described in *Materials and Methods*; the expectation values shown are for one specific repeat from these proteins.

fied, corresponding to the previously identified cysteine-rich repeats, except for the change in repeat boundaries (Fig. 1*A*). Repeat 1 in integrin β subunits differs from the others because it contains six rather than eight cysteines. Repeat 1 was defined as beginning with Cys-427 because the preceding cysteine, Cys-425, is predicted to form a long-range disulfide bond to Cys-3 (human integrin $\beta2$ numbering), based on determination of disulfide bonds in the $\beta3$ subunit (18).

Sequence Homology to EGF-Like Domains. A consensus sequence for integrin and TIED cysteine-rich repeats was prepared from an alignment of the 10 TIED repeats and 104 repeats from 36 integrin β subunits. The consensus sequence was submitted to a PSI-BLAST search (20) as implemented by the STRUCTURE-3D server, to find amino acid sequence homologies to proteins with known structures (25). Very significant sequence homology (expectation values from 2×10^{-5} to 3×10^{-7}) was found to EGF modules of known three-dimensional structures in coagulation factors VII, IX, and X, and E-selectin (Fig. 1*B*). Furthermore, statistically significant homology was found with PSI-BLAST 2.2, which corrects for compositional bias because of high cysteine content (20) to EGF-like sequence repeats in tenascin X, reelin, and their relatives (35) and with 3D-PSSM (24) to a segment in the transmembrane disintegrin ADAM-11 (Fig. 1*B*). After finding this relationship, our study of the reelin literature revealed that a similarity of repeats in reelin to those in tenascins and integrins had previously been suggested (35), although statistical significance had not been demonstrated, and the partial alignment with five cysteines could be consistent with

several alternative repeat boundaries. All six cysteines in the structurally known EGF domains as well as in the annotated EGF-like repeats align with cysteines in the integrin repeats (Fig. 1*B*). Other important residues besides cysteines are aligned, including residues important in making turns such as glycines. The region from β-strand 1 to the C terminus is particularly well conserved (Fig. 1*B*). The statistical significance of the sequence homology demonstrates that these modules are evolutionarily related and thus have the same fold. Therefore, we refer to the cysteine-rich repeats in integrins as integrin EGF-like modules.

EGF modules have two canonical antiparallel β-strands (β-1 and β-2, Fig. 1*B*) and variably may have additional β-strands or α-helices (10). The EGF domain is elongated along the axis between its N- and C-terminal ends, and the structural elements run parallel to this axis. The turns between these elements are marked in Fig. 1*B* according to whether the turn is at the N-terminal end (N-turn) or C-terminal end (C-turn) of the elements they connect.

A number of structural features confirm the conclusion that integrin modules are EGF-like. (*i*) Insertions and deletions occur in turns (C-turn1, N-turn1, and C-turn2, Fig. 1*B*) and thus can be readily accommodated structurally. Strands 1 and 2 are shortened an equal amount by the deletion of four residues from the turn between them, allowing the remaining portions of these antiparallel β-strands to hydrogen bond in a similar manner. (*ii*) The three disulfide bonds shared with EGF domains are appropriate. Six of the eight cysteines in the integrin modules align to the six cysteines in the structurally defined EGF domains and are predicted to form disulfides between cysteines 2–4, 3–6, and 7–8
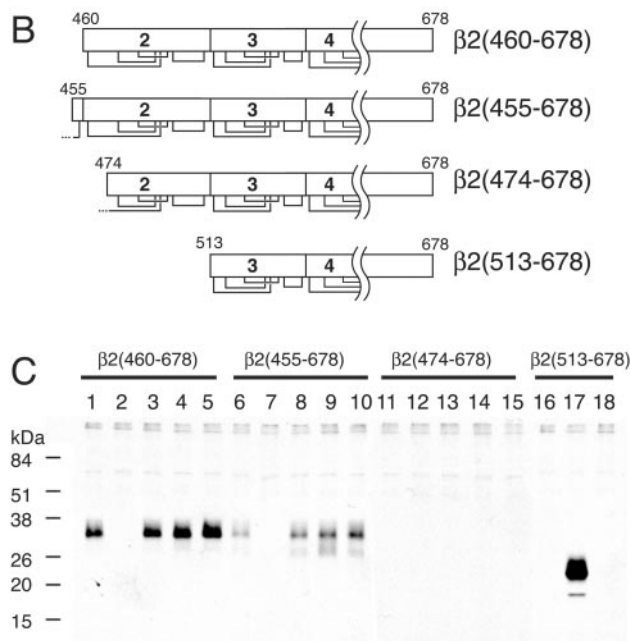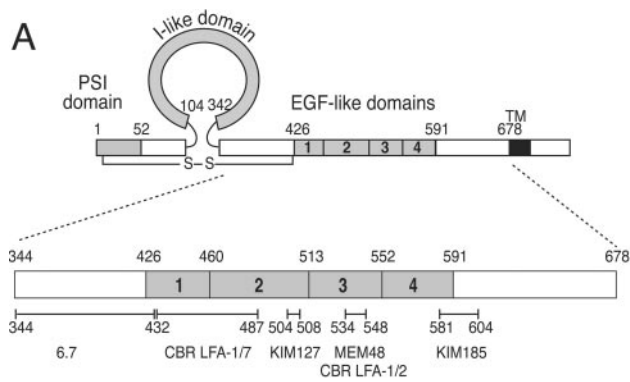
BIOCHEMISTRY

Fig. 2. β2 integrin domains and effect of truncation position on expression of fragments. (*A*) Putative domains are shown together with residue numbers of boundaries. The long-range disulfide is shown. Bars with residue numbers below denote epitope locations of β2 mAbs (7). (*B*) Design of C-terminal fragments. Cysteines and their putative disulfide bonds are shown below each fragment. (*C*) Expression of fragments. 293T cells were transiently transfected with each plasmid followed by metabolic labeling. Culture supernatants were immunoprecipitated with β2 antibodies CBR LFA-1/2 (lanes 1, 6, 11, and 16), CBR LFA-1/7 (lanes 2, 7, and 12), KIM127 (lanes 3, 8, and 13), KIM185 (lanes 4, 9, 14, and 17), and MEM48 (lanes 5, 10, 15, and 18).

(Fig. 1*B*). In integrin EGF module 1 (β2_1 and β4_1, Fig. 1*B*), only six cysteines are present. The alignment predicts that the pair of cysteines that is missing in module 1 is a disulfide-bonded pair, cysteines 2 and 4. (*iii*) Cysteines 1 and 5, which form the "extra" disulfide compared to EGF domains, are specifically missing in module 3 of all integrin β4 and β8 subunits sequenced to date (β4_3, Fig. 1*B*). This provides additional evidence for the disulfide assignment. (*iv*) The extra 1–5 disulfide is structurally sound. Cysteine 1 of the module is at its N terminus, whereas cysteine 5 that it bonds to aligns with a residue that is nearby in EGF domain structures. In EGF domains, the Cα distances between the residues similar to cysteine 1 (A120 in E-selectin and D46 or D47 in factors VII, IX, or X), and those that align with cysteine 5 are 5.9–7.4 Å, compatible with disulfide bond formation.
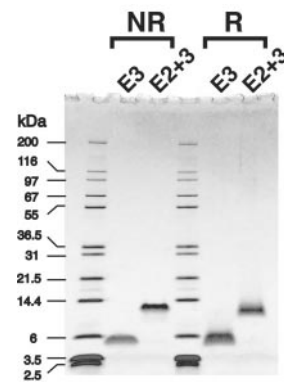

Fig. 3. SDS/PAGE of refolded integrin EGF modules. Bacterially expressed peptides (2 μg) corresponding to module 2 (E3) or modules 2 + 3 (E2 + 3) were refolded in redox buffer to allow disulfide bond formation, purified by RP-HPLC, subjected to SDS/PAGE on a 10–20% gradient Tricine gel, and stained with Coomassie brilliant blue. M_r of standard proteins are shown on the left.

To test the modular character of the integrin EGF-like domains, and verify their boundaries, different C-terminal fragments of the integrin β2 subunit were expressed in mammalian cells (Fig. 2*B*). The fragment beginning one residue before the first predicted cysteine of module 2, β2(460–678), was well expressed in 293T cells and was immunoprecipitated by mAbs CBR LFA-1/2, KIM127, KIM185, and MEM48 (Fig. 2*C*, lanes 1 and 3–5), which map to species-specific residues C-terminal to residue 460 (7) (Fig. 2*A*), and not by mAb CBR LFA-1/7, which maps to the 432–487 interval (Fig. 2*C*, lane 2). The fragment including the last cysteine of repeat 1, β2(455–678), was recognized by the same mAbs (Fig. 2*C*, lanes 6 and 8–10); however, it was expressed only 30% as well as β2(460–478). Fragment β2(474–678), which lacks the first predicted cysteine of module 2, was not precipitated by any of the mAb. By contrast, the fragment beginning one residue before the first cysteine of repeat 3, β2(513–678), was well expressed as shown by precipitation with mAb KIM185 (Fig. 2*C*, lane 17). It was not recognized by mAbs CBR LFA-1/2 and MEM48 (Fig. 2*C*, lanes 16 and 18), in agreement with the requirement for module 2 for recognition by these mAb (see below). Because residues 513–678 are sufficient for recognition by mAb KIM185, the lack of reaction with β2(474–678) suggests that this fragment was misfolded.

To further test our predictions of integrin EGF-like domains and to test their modularity, they were expressed in the absence of other N-terminal or C-terminal domains. Module 3 (residues
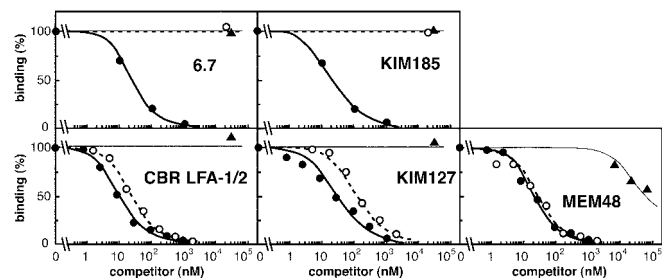

Fig. 4. Reactivity of refolded modules with β2 mAbs. Binding of antibodies to β2ΔI was determined in the presence of increasing concentrations of module 3 (▲, thin line), modules 2 + 3 (○, dashed line), or β2ΔI (●, thick line) as competitors. Values are the percentage of the binding in absence of any competitor. Data are from one representative of three independent experiments.

513–552) and modules 2 + 3 (residues 415–552) were expressed in *Escherichia coli*. After refolding in redox buffer a single peptide peak was obtained on RP-HPLC from each construct, whereas multiple peaks were obtained when 6 M guandine HCl was additionally present (N.B., J.T., T.A.S., and S.C.B., unpublished data). Module 3 and modules 2 + 3 migrated at 6 kDa and 12 kDa, respectively, in both reducing and nonreducing SDS/PAGE (Fig. 3). The presence of a single disulfide-linked isomer in HPLC and lack of disulfide-linked multimers in SDS/PAGE strongly suggest correct folding and disulfide formation, and hence correct assignment of module boundaries.

To further verify the native fold of these bacterially expressed fragments, their reactivity with mAbs was investigated. Several of these mAbs bind to epitopes that are fully exposed in isolated $\beta2$ subunits but are activation-dependent in $\alpha\beta$ complexes (7). To obtain a $\beta2$ fragment that was well expressed in mammalian cells in the absence of an associating α subunit, and could be used as a standard in ELISA assays, we deleted the $\beta2$ I-like domain ($\beta2\Delta I$, *Materials and Methods*). When adsorbed to microtiter wells, the purified soluble $\beta2\Delta I$ fragment was recognized by all $\beta2$ mAbs that have previously been mapped to epitopes outside the I-like domain (7) (Fig. 2*A*), including 6.7, CBR LFA-1/2, KIM127, KIM185, and MEM48 (Fig. 4). When added as a competitor in solution phase, $\beta2\Delta I$ inhibited binding of all these mAbs at IC_{50} values of 10–30 nM (Fig. 4). The refolded, bacterially expressed, module 2 + 3 fragment inhibited binding of the KIM127 mAb that maps to module 2, and the CBR LFA-1/2 and MEM48 mAbs that map to module 3, but not mAbs 6.7 and KIM185, which map more N and C terminal, respectively. Remarkably, in molar terms the bacterial module 2 + 3 fragment was essentially as effective as the eukaryotic $\beta2\Delta I$ fragment, suggesting a fully native structure. The module 3 fragment bound 5,000-fold more weakly to MEM48 mAb than the module 2 + 3 fragment and did not detectably bind to CBR LFA-1/2 mAb. These two mAbs recognize species-specific residues in module 3 (7). Our results suggest that interaction between modules 2 and 3 is required for expression of these epitopes, either because of an influence of module 2 on the structure of the epitope in module 3, or because the epitope includes species-common residues in module 2.

Discussion

Questions of importance in integrins are the structural basis of conformational changes that regulate ligand binding in the headpiece and signaling in the cytoplasm and the linkages between domains that permit signals to be relayed over distances that are quite large relative to other signaling proteins. Answers to these questions depend on defining the structure and function of the smallest structural units, or modules of integrins, and how these modules work together. The cysteine-rich region of the β subunit is of major importance in signal transmission, as evidenced by the large number of activation-associated epitopes and mutations that map to this region.

We significantly advance the understanding of the β subunit cysteine-rich region by demonstrating the presence of modules, and that these modules differ in their boundaries, and hence in which cysteine residues they contain, from the repeat units that have been accepted since integrin β subunits were first sequenced in 1986 (13). Although the boundaries are difficult to discern in integrins, we were able to predict them because they divide a highly homologous protein termed TIED into 10 repeats each containing eight cysteines. Truncation of the integrin $\beta2$ subunit confirmed the boundaries. Truncation at the predicted N-terminal boundary of module 2 gave optimal expression, whereas addition of a small portion of module 1 containing one cysteine reduced expression. Furthermore, removal of a small portion of module 2 containing its first cysteine completely abolished expression. Moreover, further truncation to the N-

terminal boundary of module 3 again gave optimal expression. Modularity was further confirmed by expression in *E. coli* and refolding of module 3, and tandem modules 2 + 3. The refolded proteins were monomeric and well behaved biochemically, suggesting a native-like structure. Furthermore, tandem modules 2 + 3 had a native structure as shown by reactivity with mAbs that had previously been mapped to residues in module 2 (one mAb) or module 3 (two mAbs).

Although integrin cysteine-rich repeats have previously been suggested to be EGF-like (19, 21), our definition of the EGF modules in integrins differs markedly from these previous reports. We demonstrate statistical significance of the homologies to EGF modules and find them to EGF domains that contain six cysteines, rather than to laminin EGF domains that contain eight cysteines. The alignment to the six cysteines in these EGF domains predicts their disulfide connectivity. The two additional cysteine residues are thus predicted to bond to one another, as additionally supported by their specific absence in certain integrin EGF modules and the spatial proximity of the residues to which they align in EGF structures. These "extra" cysteines are unique to integrin EGF domains and differ markedly in location from the "extra" pair in laminin EGF modules. Compared to a previous alignment (21), in our alignment only three of the eight cysteines in the integrin EGF modules are aligned equivalently to EGF domains, and only one of four disulfide bonds is structurally equivalent. Our alignment to EGF domains is similar to that suggested for three integrin EGF repeats in a "seed" alignment for EGF-like domains in Pfam release 6.5 (36); however, the difficulty in detecting homology is emphasized by the absence of any integrin or TIED sequences in the 4,293 EGF-like modules that were found to be homologous to the seed alignment. Furthermore, the Pfam EGF-like alignment also lacks the tenascin EGF-like repeats shown here to be closely related to those in integrins, although it contains the similar repeats in reelin. The sequences aligned to EGF domains in SMART as of July 31, 2001 (37) do not contain integrin, TIED, tenascin, or reelin EGF-like repeats.

The domain organization of integrins has been preserved since the divergence of sponges, corals, and nematodes. Given the selection against shuffling of these domains for the last 1.2 billion-1.5 billion years, it should not be surprising that exon boundaries in integrins provide few clues to domain organization. The only exception is the I domain, which is a recent evolutionary addition because it is found only in some vertebrate α subunits and is not present in *Caenorhabditis elegans* or *Drosophila*. The intron boundaries in the integrin EGF repeats of the $\beta1$, $\beta2$, and $\beta7$ subunits occur after the fifth Cys in repeat 1, after the third Cys in repeat 3, and three cysteines after the end of repeat 4 (38); furthermore, intron locations differ markedly in $\beta4$ (39). The lack of correlation between exon and module boundaries in integrins contrasts with the situation for most extracellular proteins, which have evolved more recently. Commonly, EGF-like modules are well demarcated by phase 1 introns (10, 40). Therefore, it is interesting that the EGF-like domains that are most similar to those in integrins, i.e., those in reelin, tenascin, and their relatives, are also not demarcated with introns. The nearest introns in reelin occur well before or within the EGF-like repeats (41) and in tenascin 19 EGF-like domains occur within a single exon. These facts suggest a common function, such as important interactions with neighboring domains, that disfavors exon shuffling.

Integrin EGF-like modules appear to closely interact with one another. "Typical" EGF modules contain approximately five noncysteine interdomain residues. Often, tandem EGF modules have been difficult to crystallize, and marked intermodule flexibility leading to disorder has been observed in crystal structures (42, 43). However, a subset of EGF domains contains a Ca^{2+}-binding site in the interdomain linker, which rigidifies the

linkage (44). The presence of only a single noncysteine residue between domains in integrin EGF-like domains is unique not only among EGF-like domains, but among extracellular domains in general (10), and no doubt contributed to the long delay in identifying the domain boundary. However, the spacings between the cysteines that are shared with other EGF-like domains are conserved; in integrins, the "extra" disulfide joins the interdomain linker to the turn between the shortened β-strands (Fig. 1B). The shortening of the interdomain linker from approximately five residues to one residue greatly limits the potential for interdomain flexibility and suggests a rigid connection between domains. Our finding that two mAbs, MEM48 and KIM127, recognize an epitope that depends on both integrin EGF modules 2 and 3 also emphasizes the close interactions between neighboring integrin EGF domains.

The close interconnection between integrin EGF domains has important implications for signal transmission. The portion of the β subunit stalk region that contains the EGF-like domains is known to play an important structural role in regulating the affinity of the ligand-binding site in the integrin headpiece. Integrin extracellular domains undergo extensive conformational change during affinity up-regulation (inside-out signaling) as well as after ligand is bound (outside-in signaling). Many mAbs including KIM127 that map to the β subunit EGF-like domains can report this change; i.e., their epitopes become exposed after activation or ligand binding. Moreover, binding of other mAbs including CBR LFA-1/2 to this region can induce conformational shifts in integrin heterodimers and activate receptor function. As a consequence of these data, and the finding that close association between the membrane proximal regions of the α and β subunits restrains integrins in the inactive state (3, 4), it is postulated that close association of α and β subunits at the stalk region is correlated with the low affinity state of the receptor. Indeed, species-specific residues in the third EGF-like module of $\beta2$ contribute to maintaining the αX$\beta2$ integrin in the low affinity state and are presumably located in an interface with the αX stalk (6). Release of the membrane-proximal restraint results in a movement apart of 14 nm at the C-terminal portion of the α and β subunit extracellular domains and activates ligand binding (4). There is no precedent among surface receptors for understanding how signals can be conveyed over such large distances as present in the integrin stalk regions of 16-nm length. If the linkages between domains in the stalk regions were flexible, conformational signals such as separation in the membrane of the transmembrane domains would be absorbed by interdomain movement before reaching the ligand-binding headpiece. Therefore, the tight interdomain linkage identified here between integrin EGF-like domains appears to be a structural adaptation that has evolved to function in signal transduction between the headpiece and transmembrane domains. Because identification of modular fragments that retain native structure is currently the rate-limiting step in structural studies on integrins, this work should advance atomic resolution studies of the integrin EGF-like domains, which in turn will help unravel their physiological relevance in integrin function.

1. Humphries, M. J. (2000) *Biochem. Soc. Trans.* **28,** 311–339.
2. Weisel, J. W., Nagaswami, C., Vilaire, G. & Bennett, J. S. (1992) *J. Biol. Chem.* **267,** 16637–16643.
3. Lu, C., Takagi, J. & Springer, T. A. (2001) *J. Biol. Chem.* **276,** 14642–14648.
4. Takagi, J., Erickson, H. P. & Springer, T. A. (2001) *Nat. Struct. Biol.* **8,** 412–416.
5. Du, X., Gu, M., Weisel, J. W., Nagaswami, C., Bennett, J. S., Bowditch, R. & Ginsberg, M. H. (1993) *J. Biol. Chem.* **268,** 23087–23092.
6. Zang, Q. & Springer, T. A. (2001) *J. Biol. Chem.* **276,** 6922–6929.
7. Lu, C., Ferzly, M., Takagi, J. & Springer, T. A. (2001) *J. Immunol.* **166,** 5629–5637.
8. Bazzoni, G. & Hemler, M. E. (1998) *Trends Biochem. Sci.* **23,** 30–34.
9. Takagi, J., Isobe, T., Takada, Y. & Saito, Y. (1997) *J. Biochem. (Tokyo)* **121,** 914–921.
10. Bork, P., Downing, A. K., Kieffer, B. & Campbell, I. D. (1996) *Q. Rev. Biophys.* **29,** 119–167.
11. Springer, T. A. (1998) *J. Mol. Biol.* **283,** 837–862.
12. Jeon, H., Meng, W., Takagi, J., Eck, M. J., Springer, T. A. & Blacklow, S. C. (2001) *Nat. Struct. Biol.* **8,** 499–504.
13. Tamkun, J. W., DeSimone, D. W., Fonda, D., Patel, R. S., Buck, C., Horwitz, A. F. & Hynes, R. O. (1986) *Cell* **46,** 271–282.
14. Kishimoto, T. K., O'Connor, K., Lee, A., Roberts, T. M. & Springer, T. A. (1987) *Cell* **48,** 681–690.
15. Huang, C., Zang, Q., Takagi, J. & Springer, T. A. (2000) *J. Biol. Chem.* **275,** 21514–21524.
16. Brower, D. L., Brower, S. M., Hayward, D. C. & Ball, E. E. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9182–9187.
17. Wang, D. Y., Kumar, S. & Hedges, S. B. (1999) *Proc. R. Soc. London Ser. B.* **266,** 163–171.
18. Calvete, J. J., Henschen, A. & González-Rodríguez, J. (1991) *Biochem. J.* **274,** 63–71.
19. Yuan, Q., Jiang, W.-M., Krissansen, G. W. & Watson, J. D. (1990) *Int. Immunol.* **2,** 1097–1108.
20. Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V. & Altschul, S. F. (2001) *Nucleic Acids Res.* **29,** 2994–3005.
21. Berg, R. W., Leung, E., Gough, S., Morris, C., Yao, W.-P., Wang, S.-x., Ni, J. & Krissansen, G. W. (1999) *Genomics* **56,** 169–178.
22. Gotoh, O. (1996) *J. Mol. Biol.* **264,** 823–838.
23. Birney, E., Thompson, J. D. & Gibson, T. J. (1996) *Nucleic Acids Res.* **24,** 2730–2739.
24. Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. (2000) *J. Mol. Biol.* **299,** 499–520.
25. Huynen, M., Doerks, T., Eisenhaber, F., Orengo, C., Sunyaev, S., Yuan, Y. & Bork, P. (1998) *J. Mol. Biol.* **280,** 323–326.
26. Petruzzelli, L., Maduzia, L. & Springer, T. A. (1995) *J. Immunol.* **155,** 854–866.
27. Bazil, V., Stefanova, I., Hilgert, I., Kristofova, H., Vanek, S. & Horejsi, V. (1990) *Folia Biol. (Prague)* **36,** 41–50.
28. Robinson, M. K., Andrew, D., Rosen, H., Brown, D., Ortlepp, S., Stephens, P. & Butcher, E. C. (1992) *J. Immunol.* **148,** 1080–1085.
29. Andrew, D., Shock, A., Ball, E., Ortlepp, S., Bell, J. & Robinson, M. (1993) *Eur. J. Immunol.* **23,** 2217–2222.
30. David, V., Leca, G., Corvaia, N., Le Deist, F., Boumsell, L. & Bensussan, A. (1991) *Cell. Immunol.* **136,** 519–524.
31. Oxvig, C., Lu, C. & Springer, T. A. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2215–2220.
32. Huang, C. & Springer, T. A. (1995) *J. Biol. Chem.* **270,** 19008–19016.
33. Casasnovas, J. M., Stehle, T., Liu, J.-h., Wang, J.-h. & Springer, T. A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 4134–4139.
34. North, C. L. & Blacklow, S. C. (1999) *Biochemistry* **38,** 3926–3935.
35. D'Arcangelo, G., Miao, G. G., Chen, S. C., Soares, H. D., Morgan, J. I. & Curran, T. (1995) *Nature (London)* **374,** 719–723.
36. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D. & Sonnhammer, E. L. (1999) *Nucleic Acids Res.* **27,** 260–262.
37. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5857–5864.
38. Jiang, W.-M., Jenkins, D., Yuan, Q., Leung, E., Choo, K. H. A., Watson, J. D. & Krissansen, G. W. (1992) *Int. Immunol.* **4,** 1031–1040.
39. Pulkkinen, L., Kurtz, K., Xu, Y., Bruckner-Tuderman, L. & Uitto, J. (1997) *Lab. Invest.* **76,** 823–833.
40. Doolittle, R. F. (1995) *Annu. Rev. Biochem.* **64,** 287–314.
41. Royaux, I., Lambert de Rouvroit, C., D'Arcangelo, G., Demirov, D. & Goffinet, A. M. (1997) *Genomics* **46,** 240–250.
42. Campbell, I. D. & Bork, P. (1993) *Curr. Opin. Struct. Biol.* **3,** 385–392.
43. Padmanabhan, K., Padmanabhan, K. P. & Tulinsky, A. (1993) *J. Mol. Biol.* **232,** 947–966.
44. Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D. & Handford, P. A. (1996) *Cell* **85,** 597–605.
45. Schlossman, S. F., Boumsell, L., Gilks, W., Harlan, J., Kishimoto, T., Morimoto, T., Ritz, J., Shaw, S., Silverstein, R., Springer, T., *et al.*, eds. (1995) *Leukocyte Typing V: White Cell Differentiation Antigens* (Oxford Univ. Press, New York).