Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation

Natalia Beglova[1,2], Stephen C. Blacklow[1,2], Junichi Takagi[1,3] and Timothy A. Springer[1,3]

[1]Department of Pathology, Harvard Medical School, [2]Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, Masschusetts 02115, USA. These authors contributed equally to the work. [3]The Center for Blood Research, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

Published online: 18 March 2002, DOI: 10.1038/nsb779

Cysteine-rich repeats in the integrin β subunit stalk region relay activation signals to the ligand-binding headpiece. The NMR solution structure and disulfide bond connectivity of Cys-rich module-3 of the integrin β2 subunit reveal a nosecone-shaped variant of the EGF fold, termed an integrin-EGF (I-EGF) domain. Interdomain contacts between I-EGF domains 2 and 3 observed by NMR support a model in which the modules are related by an approximate two-fold screw axis in an extended arrangement. Our findings complement a 3.1 Å crystal structure of the extracellular portion of integrin αVβ3, which lacks an atomic model for I-EGF2 and a portion of I-EGF3. The disulfide connectivity of I-EGF3 chemically assigned here differs from the pairings suggested in the αVβ3 structure. Epitopes that become exposed upon integrin activation and residues that restrain activation are defined in β2 I-EGF domains 2 and 3. Superposition on the αVβ3 structure reveals that they are buried. This observation suggests that the highly bent αVβ3 structure represents the inactive conformation and that release of contacts with I-EGF modules 2 and 3 triggers a switchblade-like opening motion extending the integrin into its active conformation.

Signaling cascades that impinge on integrin cytoplasmic domains, as well as truncation and mutation of these domains, activate ligand binding in the integrin headpiece[1]. Recent evidence shows that movement apart or unclasping of the juxtamembrane segments of the integrin α and β subunits is sufficient for activation[2]. Close association between membrane proximal segments restrains activation[3]. The Cys-rich modules of the C-terminal portion of the integrin β subunit extracellular domain are particularly important in signaling; most monoclonal antibodies (mAbs) to integrin activation epitopes map to this region, including mAbs that either directly activate ligand binding or selectively bind either when integrins are activated or have bound ligand[1]. In the integrin β2 subunit, activation epitopes have been mapped to specific mouse/human amino acid substitutions in Cys-rich repeats 2 and 3 (Fig. 1*b*)[4]. Xenogeneic amino acid substitutions that activate ligand binding by the integrin αXβ2 also map to these two Cys-rich repeats, suggesting that contacts between the second and third β subunit Cys-rich repeats and the α subunit stalk restrain integrins in the inactive conformation[5] (Fig. 1*b*). Consistent with opening of interfaces upon activation that are concealed in the resting state, activation epitopes are fully exposed on isolated β subunits[4,6].

In the integrin headpiece, the structure of the inserted (I) domain is known and has been shown to undergo significant conformational movement[7]. The recent structure of the αVβ3 integrin extracellular region[8] specifies the relative positions of 8 of the 12 integrin domains in a compact global architecture. However, the PSI (plexin-semaphorin-integrin) domain, Cys-rich modules 1 and 2, and the first 10 residues of Cys-rich module 3 could not be resolved in the reported αVβ3 structure (Fig. 1*a*). The second and third Cys-rich modules of the β subunit are wedged in the bend between the integrin stalk and headpiece domains (Fig. 1*a*), poised to communicate structural rearrangements between the cytoplasmic domains and the ligand-binding headpiece. To understand why exposure of antibody epitopes on these modules accompanies integrin activation, we have determined the structure of Cys-rich repeat 3 (module 3) of the integrin β2 subunit, identified sites of contact between modules 2 and 3, and modeled the mode of interaction of activation epitopes with other integrin domains.

NMR structure and disulfide bond determination

Recombinant Cys-rich repeats 2, 3 and 2+3 (the repeat two and three pair) from the integrin β2 subunit were constructed using

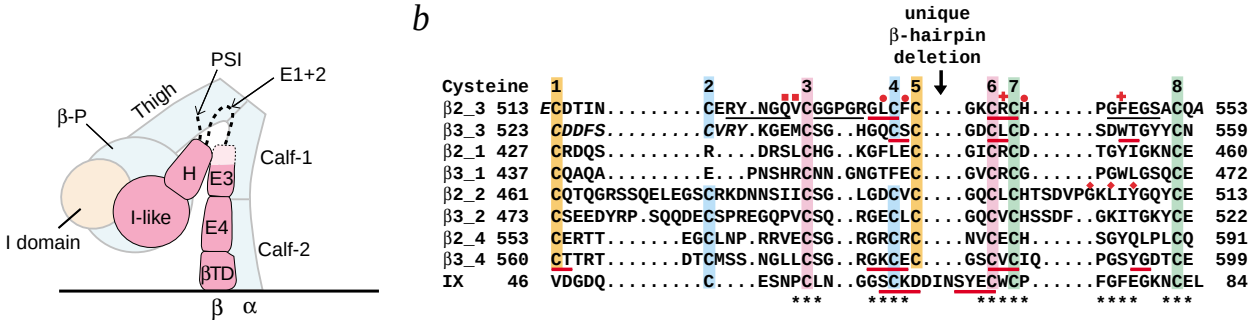

Fig. 1 Integrin domain organization and the Cys-rich repeats. ***a***, Diagram of integrin architecture based on the structure of αVβ3 (ref. 8). Domains are β-propeller (β-P), plexin-semaphorin-integrin (PSI), hybrid (H), EGF (E) and β-terminal domain (βTD). A probable position for the I domain is shown. Domains not resolved or only partially resolved in αVβ3 are shown with dashed lines. ***b***, Structural alignment of I-EGF3 to EGF module 1 of coagulation factor IX[12], I-EGF4 of β3 and alignment by sequence to other I-EGF modules. β-strands defined here in β2, β3 (ref. 8) and factor IX[12] are underlined in red; loops at the C-terminal end of the I-EGF3 module are underlined in black. Residues not defined in the β3 I-EGF3 (β3_3) structure are italicized, as are residues that were defined in the β2 I-EGF3 (β2_3) structure but not formally included within the module boundary defined here. Residues in activation epitopes are marked with red symbols above them: filled circles are residues common to the MEM48 and CBR LFA-1/2 epitopes; plus signs are additional residues in the MEM48 epitope and diamonds are residues in the KIM127 epitope[4]. Residues that restrain activation of αXβ2 are marked with red squares[5]. Asterisks below the sequences indicate the best conserved portions of the structures. Cys residues that are disulfide bonded to one another in β2 I-EGF3, β3 I-EGF4 (β3_4) and factor IX are highlighted in the same color; a different disulfide bond arrangement was reported for β3 I-EGF3, as discussed in the text.

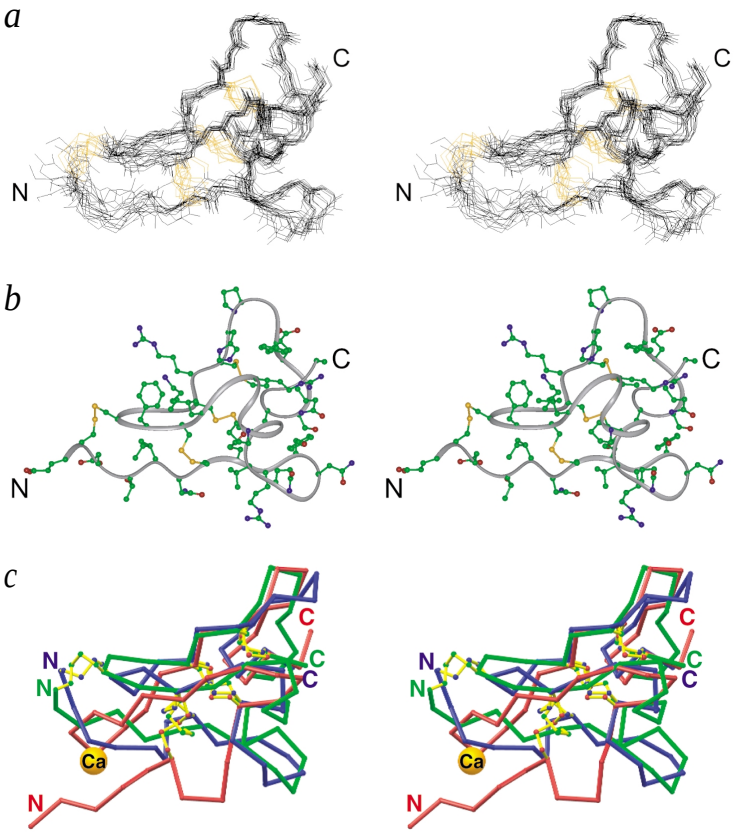

Fig. 2 Solution structure of I-EGF3. ***a***, Stereo view of the 15 lowest energy structures of I-EGF3. The best-fit superposition of the polypeptide backbone is shown. Side chains of the Cys residues engaged in the four disulfide bonds are illustrated in orange. ***b***, Stereo view of the minimized mean structure of I-EGF3. The side chains are in ball-and-stick representation on a ribbon trace of the module backbone. ***c***, Best-fit superposition of I-EGF3 with the crystal structures of EGF module 1 of coagulation factor IX[12] and I-EGF4 of αVβ3 (ref. 8). The Cα trace and Cys side chain atoms are green for I-EGF3, red for the factor IX EGF module and blue for I-EGF4 of αVβ3. Cys side chain bonds are yellow. The Ca^{2+} coordinated to the factor IX EGF module is depicted as a gold sphere. Figure generated using Insight 2000 (MSI) in (*a*, *b*) and Ribbons[39] in (*c*).

newly defined module boundaries[9,10]. The calculated NMR structure of repeat 3 (Table 1) reveals a variant of the EGF fold, termed the integrin- or I-EGF fold (Figs 1*b*, 2*a*–*c*). Disulfide bonds were independently determined by protease digestion and resolution of disulfide-bonded peptides by mass spectrometry (Table 1). In I-EGF3, the structure is stabilized by four disulfide bonds between the first and fifth, second and fourth, third and sixth, and seventh and eighth Cys residues (Fig. 1*b*). The structure-based sequence alignment (Fig. 1*b*) shows that three disulfide bonds are shared with classical EGF domains. The disulfide unique to I-EGF domains links the N-terminus to the turn between the two β-strands (Figs 1*b*, 2*c*).

The disulfide bond connectivity of I-EGF3 assigned here using chemical methods and further verified by NMR distance restraint data corresponds to the arrangement predicted by structure-based alignment and tested by expression of truncated β2 subunits in mammalian cells[9,10]. However, this connectivity differs from the assignments reported in the recent 3.1 Å crystal structure of αVβ3 (ref. 8). The first 10 residues of I-EGF3 could not be resolved in the αVβ3 structure because of poor electron density; thus, Cys 1 and 2 are missing from the β3 I-EGF3 module. The remaining Cys residues were reported to be linked in a 3–5, 4–6 and 7–8 arrangement, in contrast to the 1–5, 2–4, 3–6 and 7–8 arrangement found here. The most plausible explanation for the discrepancy is that authors of the αVβ3 paper incorrectly traced the backbone in the N-terminal portion of I-EGF3 and could not properly assign the disulfides with Cys 4 and 5 because their disulfide-bonded partners were unresolved in the electron density. This misinterpretation is explicable given the 3.1 Å resolution of their native X-ray data and the lack of good density in the unresolved region of I-EGF3 and in the contiguous I-EGF modules 1 and 2. Sequence alignment predicts that in all integrin β subunits, the disulfide bond connectivity for I-EGF modules 2 and 4 will match that of I-EGF3 (Fig. 1*b*)[9,10]. Indeed, the disulfide connectivity of I-EGF4 reported in the αVβ3 crystal structure[8] is identical to the pattern reported here for I-EGF3 of the β2 subunit. Furthermore, despite three insertions and deletions, our β2 I-EGF3 structure superimposes well with the β3 I-EGF4 structure, with a root mean square (r.m.s.) deviation of 2.6 Å for all 37 alignable positions (Figs 1*b*, 2*c*).

Structure comparison

Compared to classical EGF modules with three disulfide bonds[11], the I-EGF3 module is less elongated, with a nosecone-like shape (Fig. 2*c*). The antiparallel sheet between the two β-strands is shortened because four highly conserved residues among classical EGF-like modules are deleted in I-EGF modules (Fig. 1*b*). The loop at the C-terminal end of the module between Cys 2 and 3 is unusually long compared to EGF-like domains and contributes to the flare at the C-terminus that helps dock the nose of the next module. The nose at the N-terminal end differs markedly in conformation from classical EGF modules (Figs 1*b*, 2*c*). The N-terminal tip is bent because of the additional disulfide bond in the hairpin turn between the shortened β-strands. This extra disulfide bond is structurally analogous to calcium coordination in the calcium-binding subset of EGF-like modules[11–13], in which calcium coordination also bridges the turn connecting the two β-strands to the N-terminal end of the module (Fig. 2*c*).

Relationship between modules 2 and 3

NMR chemical shift perturbation studies, in which heteronuclear fingerprint spectra of individual modules 2 and 3 are compared with the spectrum of the module 2+3 pair, demonstrate that intimate contacts are present between modules 2 and 3 in the module pair (Fig. 3*a*,*b*). Covalent connection of the modules leads to perturbations in chemical shift and linewidth for residues in the C-terminal half of module 2, at the N-terminal end of module 3 and in the turn between β-strands 1 and 2, which is connected to the N-terminal end of the module by the disulfide connecting the first and fifth Cys residues. Residues with the largest amount of disorder in the structure of module 3 (Fig. 2*a*) comprise most of those exhibiting perturbation upon covalent connection to module 2.

To build a model of the module 2+3 pair, we used three unambiguous NOEs between modules to define their relative orientation. The resulting model predicts that modules 2 and 3 are related by an approximate two-fold screw axis and are positioned in an extended arrangement with respect to each other

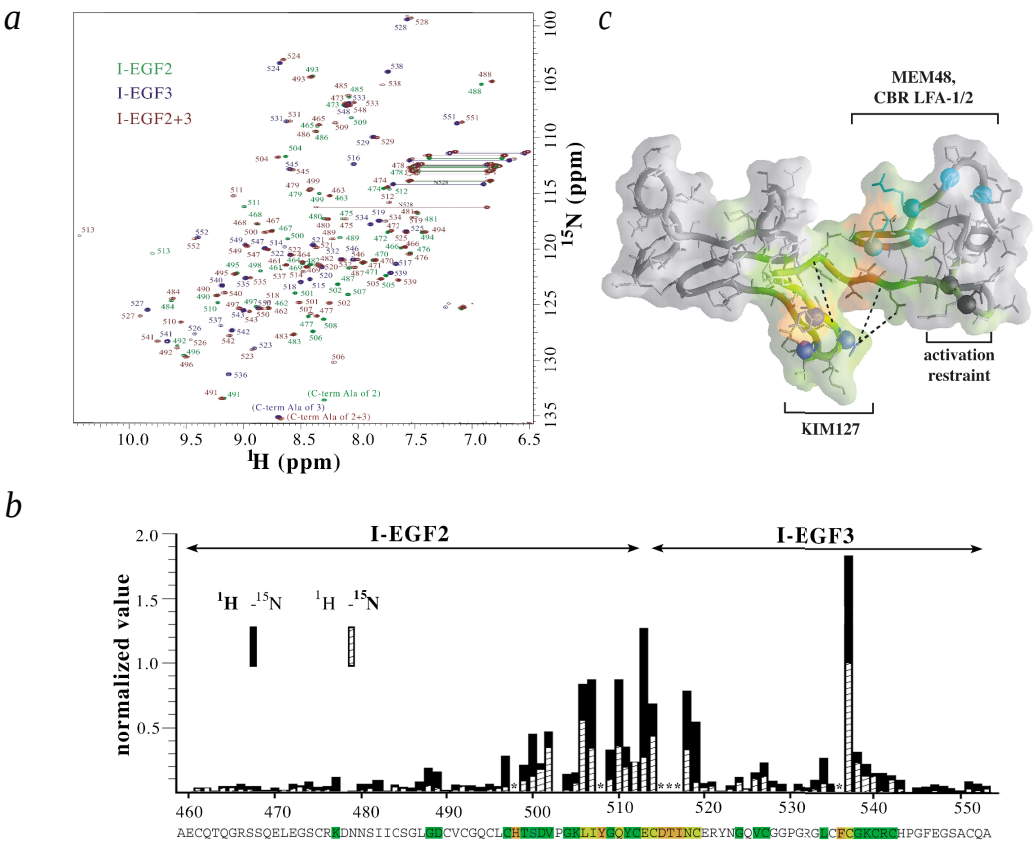

Fig. 3 Interdomain interactions between I-EGF2 and I-EGF3. ***a***, Comparison of assigned ^{15}N-^{1}H heteronuclear single quantum coherence spectra of I-EGF2 (green), I-EGF3 (blue) and the I-EGF2+3 domain pair (red) in an overlay plot. ***b***, Sum of the normalized HN- and ^{15}N-chemical shift differences between resonances of I-EGF2+3 and the corresponding resonances of I-EGF2 and I-EGF3. The maximum chemical shift perturbation in the proton dimension is 0.63 p.p.m., and the maximum chemical shift perturbation in the nitrogen dimension is 5.07 p.p.m. The amino acid sequence of the I-EGF2+3 domain pair is at the bottom, color-coded by the extent of chemical shift perturbation at each site. Resonances of residues with a white background are not significantly perturbed in the spectrum of the domain pair. Residues showing perturbations >2 standard deviations (s.d.) from the mean of the unperturbed residues are on a green background. A yellow background indicates a perturbation of >4 s.d. Resonances of residues displayed on an orange background undergo line broadening and are not detected in the spectrum of the module pair but are observed in the spectra of individual modules (asterisked in the plot). ***c***, Model for the structure of the I-EGF2+3 domain pair. A surface representation of the structure has been superimposed on a ribbon trace of the module backbone with side chains represented as sticks. The surface has been colored according to the perturbation scheme in (*b*). Intermodule NOEs used to estimate the approximate interdomain orientation are illustrated by dotted lines. Cα atoms of residues in the CBR LFA-1/2 (Leu 534, Phe 536 and His 543) and MEM48 (also includes Arg 541 and Phe 546) epitopes are light blue spheres; Cα atoms in the KIM127 (residues Gly 504, Leu 506 and Tyr 508) epitope are purple spheres; and Cα atoms of residues that restrain activation of human αXβ2 (Gln 525 and Val 526) are black spheres[5]. Figure was prepared with GRASP[40].

(Fig. 3*c*). This interdomain orientation is similar to the orientation between I-EGF modules 3 and 4 in αVβ3 (ref. 8) and that between tandem calcium-binding EGF-like modules in which coordination of calcium sets the interdomain orientation[13].

Structure and exposure of integrin activation epitopes

The structure of I-EGF3 and the model of the module 2+3 pair reveal for the first time the position of residues in integrin activation epitopes. In I-EGF3, residues Leu 534, Phe 536 and His 543 are shared in the epitopes for the activating antibodies CBR LFA-1/2 and MEM48 (Figs 1*b*, 3*c*). The epitope for MEM48 extends farther to include Arg 541 and Phe 546 (Figs 1*b*, 3*c*)[4]. Gln 525 and Val 526, located on the other side of the domain, participate in constraining αXβ2 in its resting conformation when paired with an αX subunit from the same species, suggesting that they lie at an important species-specific αβ contact site in the latent heterodimer[5]. In the epitope in I-EGF2 for the activating antibody KIM127, residues Gly 504, Leu 506 and Tyr 508 are located in the long loop protruding from the C-terminal end of the module (Figs 1*b*, 3*c*).

Several studies suggest that the mechanism of integrin activation is general. For example, activation epitopes also become exposed on I-EGF modules of β1- and β3-integrins[14–18], although they have not been mapped to specific residues as for β2-integrins. To understand the molecular rearrangements that occur in an intact integrin during activation, we superimposed the model of the β2 I-EGF2+3 domain pair on the partial I-EGF3 module of αVβ3 (see Methods). The αVβ3 extracellular domain unexpectedly exhibits an overall V-shape in which the ligand-binding headpiece bends back toward the base of the stalk region and is oriented toward the cell membrane[8] (Figs 1*a*, 4*a*). In the superposition, I-EGF modules 2 and 3 are wedged into the extreme bend or 'genu' between the integrin headpiece and stalk (Fig. 4*a*). I-EGF3 sits between the β subunit hybrid domain and the α subunit calf-1 domain, whereas the C-terminal portion of I-EGF2 nestles between the α subunit thigh and calf-1 domains, close to the fulcrum at the genu (Fig. 4*a*). In this bent integrin conformation, residues Gln 525 and Val 526 of I-EGF3, which are in an αβ interface that restrains integrins in the inactive state[5], are buried and face

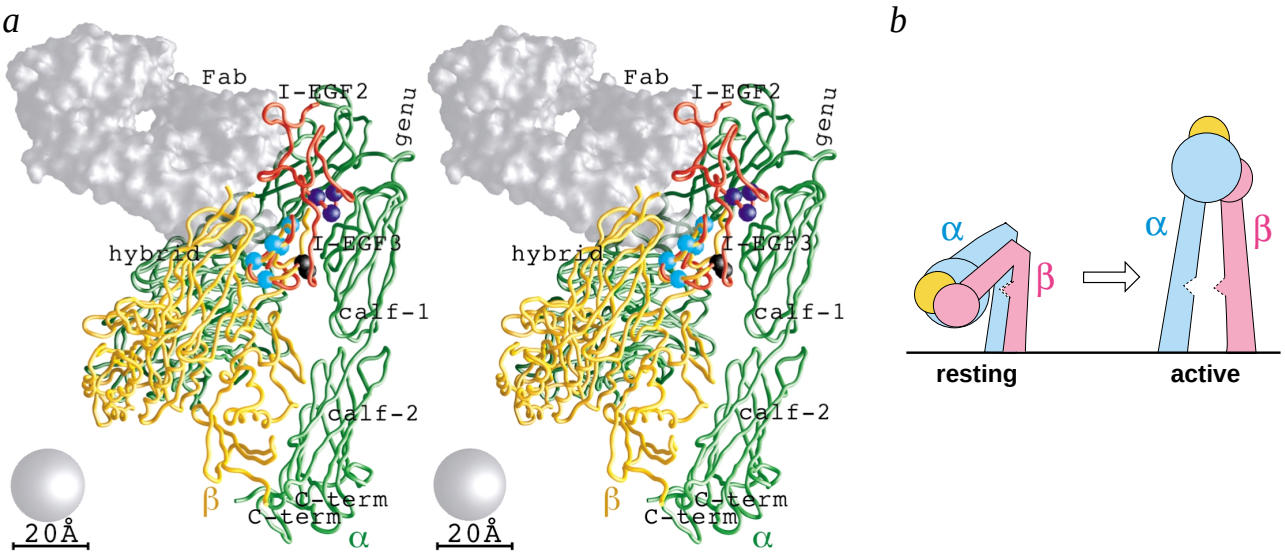

Fig. 4 Integrin activation. ***a***, Superposition of I-EGF2+3 from β2 upon the C-terminal end of αVβ3 (see Methods). A backbone ribbon trace of β2 I-EGF2+3 and αVβ3 is shown. Polypeptide backbones are red for β2 I-EGF2+3, green for αV and yellow for β3. Cα atoms of residues in the KIM127 epitope (Gly 504, Leu 506 and Tyr 508) are purple, and cyan in the CBR LFA-1/2 (Leu 534, Phe 536 and His 543) and MEM48 (also includes Arg 541 and Phe 546) epitopes[4]. Cα atoms of residues that restrain activation of human αXβ2 (Gln 525 and Val 526) are black[5]. The surface of a representative Fab antibody fragment (gray; PDB code 1A3R), with its antigen-binding site oriented toward and at the closest approach to the KIM127 and CBR LFA1/2 epitopes, is also shown to illustrate that the KIM127 and CBR LFA-1/2 epitopes are inaccessible to its antigen-binding site (see text). A gray sphere 20 Å in diameter is included at the bottom left as a size reference. Figure prepared with GRASP[40]. ***b***, Model for integrin activation. The model of the resting integrin resembles the form of αVβ3 observed in the crystal structure[8]. The active integrin is depicted in an extended conformation as seen in electron micrographs of αIIbβ3 bound to its ligand fibrinogen[18]. The triangular interaction site in I-EGF3 dovetails with the α subunit in the resting state and not in the active state.

calf-1 domain of the α subunit (Fig. 4*a*). Three other segments of I-EGF2 and I-EGF3 that restrain activation[5] are also masked. Furthermore, activation epitopes are buried in the bent integrin conformation. The α subunit thigh domain and, in particular, the β subunit hybrid domain mask the face of I-EGF3 presenting the CBR LFA-1/2 and MEM48 activating epitopes (Fig. 4*a*). Moreover, the KIM127 epitope of I-EGF2 is deeply buried in the crevice formed by the bend between the α subunit thigh and calf-1 domains. A surface representation of a manually docked antibody Fab fragment illustrates that the epitopes are inaccessible to its antigen-binding site (Fig. 4*a*). To objectively measure antigenic burial, we calculated the antibody-accessible surface using the approach to calculate solvent-accessible surface, except that we rolled over the surface a 20 Å diameter sphere to generously estimate the number of residues that might be accessible to a 40 Å diameter Fab fragment (Fig. 4*a*). All of the three identified residues in the KIM127 epitope are accessible to a 20 Å diameter sphere in an extended, active conformation (Fig. 4*b*), whereas none are accessible in the bent αVβ3 conformation. Furthermore, three of five residues in the CBR LFA-1/2 and MEM48 epitopes in I-EGF3 are accessible in the extended conformation, whereas one is accessible in the bent conformation and only 14% as much as in the extended conformation.

αVβ3/β2 models the inactive conformation

Given the strong conservation of structure–function relationships among integrins, the masking of activation epitopes and the positioning in αβ contacts for residues that restrain integrins in the inactive state, our composite αVβ3/β2 model forces us to suggest that the genuflected conformation represents a latent, inactive structure. The genuflected conformation has been suggested to be unlikely to occur on the cell surface[8]. However, extensive contacts between the headpiece and stalk regions stabilize the bent conformation in which 1,400 Å^2 of solvent-accessible surface area are buried[19,20]. This value is in the range of physiological interaction surfaces and out of the range of crystal packing artifacts ($P < 0.02$)[21]. The observation that antigenic and interaction sites that are inaccessible or buried in an inactive integrin are also buried in the composite model suggests that the bent conformation is physiologically relevant and represents the inactive state. Experience from previous crystal structures of integrin I domains, which adopt the inactive state when there is no ligand[7], also suggests that αVβ3 represents the latent conformation because it was crystallized in the absence of a physiological ligand. Although the headpiece in the αVβ3 structure was suggested to be in an active conformation[8], the lines of evidence in that paper actually support the opposite conclusion. First, αVβ3 was crystallized with Ca^{2+} and without exposure to Mg^{2+} or Mn^{2+} (ref. 8), yet Ca^{2+} is known to stabilize integrins in the inactive conformation, whereas Mg^{2+} and especially Mn^{2+} are activators. Although a similar αVβ3 preparation has been shown to be active in the presence of Ca^{2+} in a solid phase binding assay when pretreated during coating on plastic with 10 μM Mn^{2+} (ref. 22), αVβ3 does not bind ligand in the presence of Ca^{2+} in the absence of such pretreatment[23]. Therefore, the αVβ3 that was crystallized and the αVβ3 used in ligand binding assays seem to represent different conformational states. Integrins that are released from the cell surface usually are active in binding ligand, suggesting release of a restraint that keeps them inactive on the cell surface. However, studies of both cell surface and soluble integrins suggests that the latent conformation is maintained by complementary interactions holding the integrin α and β subunit juxtamembrane segments in close association[2,3,24]. The C-termini of the αV and β3 extracellular domains are very close together in the αVβ3 structure[8], mimicking the close association

Table 1 Summary of NMR constraints, structural statistics and disulfide assignments

NOE restraints	
Long-range (|i – j| > 4)	145
Medium-range (1 < |i – j| < 4)	64
Sequential (|i – j| = 1)	144
Intraresidue	163
Total	516
Disulfide bond restraints	12
Dihedral angle restraints (Φ)	22
Hydrogen bonds	3
Pairwise r.m.s. deviation statistics[1]	
Residues superimposed	14–41
Backbone (Å)	0.8 ± 0.2
Non-hydrogen (Å)	1.6 ± 0.2

Ramachandran analysis (%)[2]	15 lowest energy	Minimized average
Most favorable region	47.1	37.9
Additional allowed region	36.3	51.7
Generously allowed region	14.3	10.3
Disallowed region	2.3	0.0
Average r.m.s. deviation[1]		
From experimental restraints		
Distance (Å)	0.056 ± 0.004	
Dihedral (°)	0.60 ± 0.15	
From idealized covalent geometry		
Bonds (Å)	0.0044 ± 0.0002	
Angles (°)	0.65 ± 0.03	
Impropers (°)	0.49 ± 0.06	

Assignment of I-EGF3 disulfide connectivity by mass spectrometry

Peptides identified (1 2 3 4 5 6 7 8[4])	Observed mass (Da)	Expected mass (Da)	Assigned S-S bond(s)
AE**C**DTIN**C**ERYNGQV**C**GGPGRGL**C**F**C**GK**C**R**C**HPGFEGSA**C**QA	4,391.5	4,391.0 (ox)	
CHPGFEGSA**C**QA	1,204.6	1,204.5	7–8
DTIN**C**ERY GL**C**F	1,449.4	1,449.9	2–4
DTIN**C**ER GL**C**F	1,286.5	1,286.8	2–4
YNGQV**C**GGPGR **C**R	1,381.2	1,382.8	3–6
NGQV**C**GGPGR **C**R**C**HPGFEGSA**C**QA	2,405.6	2,406.2	3–6; 7–8
AE**C**DTIN**C**ERY **C**F**C**GK	1,887.0	1,887.2[5]	1–5; 2–4

[1]R.m.s. deviations are calculated for the 15 lowest energy structures.
[2]Calculated energies (kcal mol^{-1}) from CNS[31] are E_{tot} = 240.85 ± 12.40, E_{bond} = 11.08 ± 1.11, E_{angle} = 66.03 ± 6.12, E_{cdih} = 0.51 ± 0.25 and E_{NOE} = 82.58 ± 11.07.
[3]Ramachandran analysis was performed using PROCHECK[38].
[4]The cysteines are in bold type and denoted by the numbers 1–8.
[5]This expected mass includes 18 Da from hydrolysis of the peptide bond between Phe 536 and Cys 537.

in the juxtamembrane region that maintains integrins in the inactive state. Finally, the I-like domain of αVβ3 was suggested to resemble the open, ligand-binding conformation of integrin I domains rather than the closed, inactive conformation[8]. However, automatic superposition[25] with either the closed or open conformations of the integrin α2 and αM I domains shows a significantly better fit of the I-like domain to the closed conformation than the open conformation, with 11 more residues within 3 Å of the common framework. These residues include five in the β6–α7 loop, which dramatically rearranges between the two conformations.

A general model for integrin activation

The flexibility at the genu "may be linked to integrin regulation"[8]. To accommodate the model of the I-EGF2+3 structure into a mechanism for integrin activation, we propose that activation triggers a switchblade-like opening of the interface between the headpiece and stalk, which extends the ligand-binding headpiece of the integrin heterodimer away from the plasma membrane (Fig. 4*b*). This extended form, which is commonly seen in electron micrographs of integrins, including integrins simultaneously bound to ligand and activating antibody[18], exposes not only the epitopes studied here, but also a much broader surface. This correlates with (i) selective exposure of epitopes on active and ligand-bound integrins, but not on inactive integrins, and (ii) the predominant localization of activation epitopes to the β subunit, because 70% of the surface buried in the headpiece–stalk interface is on the β subunit[19,20]. Conformational changes of this magnitude are well precedented; influenza hemagglutinin, which undergoes a conformational change to induce fusion of viral and cellular membranes, rearranges to reposition its hydrophobic fusion peptide by 100 Å toward the target cell membrane[26,27].

We propose that in resting integrins, some hinging motion occurs about the fulcrum at the genu, and the accessibility of antibodies increases as the distance of the epitope from the fulcrum becomes greater. This would explain why antibodies that bind in I-EGF2, such as KIM127, bind only to pre-activated integrins, whereas antibodies that bind further away from the fulcrum, such as CBR LFA-1/2 in I-EGF3 and KIM185 in I-EGF4 (refs 4,28), can bind to integrins in the inactive state and activate them by acting like a wedge in the headpiece–stalk interface. We envision that in integrin activation by biological inside-out signaling, movement apart of the juxtamembrane domains[2,3] results in separation of the α and β subunit stalks and destabilizes their interface with the headpiece, thereby triggering the switchblade-like opening and linked conformational changes within ligand-binding domains. Our data suggest that I-EGF modules 2 and 3, which reveal antibody-reactive epitopes during integrin activation, play an important role in these rearrangements.

Methods

Sample preparation. Integrin β2 subunit repeat two (I-EGF2, residues 460–513), repeat three (I-EGF3, residues 513–552) and the repeat 2+3 pair (I-EGF2+3, residues 460–552), all containing an extra C-terminal Ala residue, were cloned into the T7-based[29] pMM expression vector[30]. The I-EGF2, I-EGF3 and I-EGF2+3 polypeptides were expressed in *Escherichia coli* and purified in reduced form from inclusion bodies as described[30]. For NMR spectroscopy, ^{15}N-labeled protein was produced by growth in M9 minimal media supplemented with $^{15}NH_4Cl$ as the sole nitrogen source. Refolding of each polypeptide was performed by dialysis for 48 h

at 4 °C in a redox buffer containing reduced (2.5 mM) and oxidized (0.5 mM) glutathione (10 mM Tris, pH 8.2) to permit disulfide bond rearrangement, essentially as described[31]. The refolded modules were then purified by reversed-phase HPLC, and their identities were verified by MALDI-TOF mass spectrometry. The presence of a unique disulfide-bonded species among many possibilities, as judged by HPLC (not shown), provided convincing evidence that each individual repeat and the repeat pair are properly folded.

NMR spectroscopy. NMR spectra were recorded at 298 K, unless otherwise specified, on a Bruker 600 MHz spectrometer using standard procedures[32], and processed with GIFA v4.3 (ref. 33). I-EGF2, I-EGF3 and I-EGF2+3 were at 1 mM in 5 mM $NaPO_4$, pH 6.2. Sequential resonance assignments were obtained primarily from a combination of 3D-NOESY and 3D-TOCSY experiments: ^{15}N-HSQC-NOESY (τ_{mix} = 150 ms), ^{15}N-HSQC-TOCSY (τ_{mix} = 60 ms), ^{15}N-NOESY-HSQC (τ_{mix} = 150 ms) and ^{15}N-TOCSY-HSQC (τ_{mix} = 60ms). These data were supplemented with information from homonuclear ^{1}H–^{1}H 2D experiments, including NOESY (τ_{mix} = 100 ms and τ_{mix} = 150 ms), TOCSY (τ_{mix} = 60 ms) and DQF-COSY performed on unlabeled I-EGF3 samples in H_2O and D_2O. Coupling constants for I-EGF3 were derived from an ^{15}N-HMQC-J spectrum. Temperature coefficients for the backbone amide protons of I-EGF3 were determined from ^{15}N-HSQC spectra taken at 284 K, 293 K, 298 K, 303 K and 313 K. To identify slowly exchanging amide protons, an ^{15}N-labeled I-EGF3 sample was lyophilized and redissolved in D_2O. The decay of amide proton signal intensity was monitored for 6 h in a series of ^{15}N-HSQC experiments at 298 K. The first spectrum was taken 5 min after D_2O was added to the sample. Amide proton exchange rates for random coil peptides were calculated with SPHERE[34]. In the chemical shift perturbation experiment (Fig. 3), each single module ^{15}N-HSQC spectrum was compared with the ^{15}N-HSQC spectrum of I-EGF2+3.

Structure determination. NOE-based distance constraints derived from the spectra listed above were calibrated to intraresidue Hβ–Hβ crosspeaks and divided into three categories: strong (<3.0 Å), medium (<4.0 Å) and weak (<6.0 Å), with appropriate pseudo atom corrections introduced as necessary. Φ-torsion angle restraints of –120° ± 35° and –60° ± 35° were used for residues with measured $^{3}J_{HNH\alpha}$ coupling constants of >8 Hz and <4 Hz, respectively. Three distance restraints were incorporated to describe each disulfide bond. The distance between the two sulfur atoms was set to 2.02 ± 0.05 Å, and the Sγ–Cβ distances were constrained to 2.99 ± 0.05 Å. Structures were calculated by the standard hybrid distance geometry simulated annealing protocol as implemented in CNS[35]. Inclusion of hydrogen bond constraints for three hydrogen bonds between Arg 541 O and Leu 534 N, Lys 539 O and Phe 536 N, and Arg 532 O and His 543 N at the final stage of structure calculations is based on the following observations. (i) The vast majority (74%) of initial calculated structures with appropriate covalent geometry satisfied accepted hydrogen bond criteria (O···N distance ≤ 3.6 Å and O···H-N angle ≥ 120°). (ii) Amide protons of residues Leu 534, Phe 536 and His 543 show protection factors of 620, 700 and 1,630, respectively, on the basis of hydrogen exchange measurements. (iii) The temperature coefficient for these residues, determined by measuring the change in amide proton chemical shift with temperature, is <5 (–p.p.b. K^{-1}). (iv) The Hα chemical shift index predicts two β-strands spanning residues 533–536 and 539–543, encompassing all three identified hydrogen bonds. Hydrogen bond distance restraints, introduced during the last stage of calculations, were $r_{HN...O}$ = 1.5–2.3 Å and $r_{N...O}$ = 2.4–3.3 Å (ref. 36).

Modeling and superposition. The minimized average structure of I-EGF3 was used as a template to model I-EGF2 in the I-EGF2+3 pair using MODELLER-4 (ref. 37). The interface between the two repeats in the pair was modeled by including distance restraints corresponding to NOEs observed between the Leu 506 side chain and the backbone and side chain of Asn 518, and between the Ile 507 side chain and the backbone of Cys 514. Best-fit superposition of the factor IX EGF module[12] and β3 integrin I-EGF4 (ref. 8) upon β2 I-EGF3 was performed using MALIGN-3D of MODELLER-4. To create the superposition of β2 I-EGF2+3 on β3, we used the Cα coordinates of β2-residues Lys 539–Glu 547 and Ala 550–Gln 552 at the C-terminal end of I-EGF3 where the two structures are in good agreement, with an r.m.s. deviation of 1.3 Å.

Coordinates. Coordinates have been deposited in the Protein Data Bank (accession code 1L3Y).

Acknowledgments
Supported by grants from the NIH, Pew Charitable Trust and the American Heart Association.

Competing interests statement
The authors declare that they have no competing financial interests.

Correspondence should be addressed to T.A.S. *email: SpringerOffice@cbr.med.harvard.edu* and S.C.B. email: sblacklow@rics.bwh.harvard.edu.

Received 8 February, 2002; accepted 14 February, 2002.

1. Humphries, M.J. *Biochem. Soc. Trans.* **28**, 311–339 (2000).
2. Takagi, J., Erickson, H.P. & Springer, T.A. *Nature Struct. Biol.* **8**, 412–416 (2001).
3. Lu, C., Takagi, J. & Springer, T.A. *J. Biol. Chem.* **276**, 14642–14648 (2001).
4. Lu, C., Ferzly, M., Takagi, J. & Springer, T.A. *J. Immunol.* **166**, 5629–5637 (2001).
5. Zang, Q. & Springer, T.A. *J. Biol. Chem.* **276**, 6922–6929 (2001).
6. Huang, C., Lu, C. & Springer, T.A. *Proc. Natl. Acad. Sci. USA* **94**, 3156–3161 (1997).
7. Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J. & Liddington, R.C. *Cell* **101**, 47–56 (2000).
8. Xiong, J.-P. *et al. Science* **294**, 339–345 (2001).
9. Takagi, J., Beglova, N., Yalamanchili, P., Blacklow, S.C. & Springer, T.A. *Proc. Natl. Acad. Sci. USA* **98**, 11175–11180 (2001).
10. Tan, S.-M. *et al. FEBS Lett.* **505**, 27–30 (2001).
11. Bork, P., Downing, A.K., Kieffer, B. & Campbell, I.D. *Q. Rev. Biophys.* **29**, 119–167 (1996).
12. Rao, Z. *et al. Cell* **82**, 131–141 (1995).
13. Downing, A.K. *et al. Cell* **85**, 597–605 (1996).
14. Shih, D.T., Edelman, J.M., Horwitz, A.F., Grunwald, G.B. & Buck, C.A. *J. Cell Biol.* **122**, 1361–1371 (1993).
15. Faull, R.J. *et al. J. Biol. Chem.* **271**, 25099–25106 (1996).
16. Bazzoni, G., Shih, D.-T., Buck, C.A. & Hemler, M.A. *J. Biol. Chem.* **270**, 25570–25577 (1995).
17. Takagi, J., Isobe, T., Takada, Y. & Saito, Y. *J. Biochem. (Tokyo)* **121**, 914–921 (1997).
18. Du, X. *et al. J. Biol. Chem.* **268**, 23087–23092 (1993).
19. Lee, B. & Richards, F.M. *J. Mol. Biol.* **55**, 379–400 (1971).
20. Bailey, S. *Acta Crystallogr. D* **50**, 760–763 (1994).
21. Janin, J. *Nature. Struct. Biol.* **4**, 973–974 (1997).
22. Mehta, R.J. *et al. Biochem. J.* **330**, 861–869 (1998).
23. Smith, J.W., Piotrowicz, R.S. & Mathis, D. *J. Biol. Chem.* **269**, 960–967 (1994).
24. Hughes, P.E., O'Toole, T.E., Ylanne, J., Shattil, S.J. & Ginsberg, M.H. *J. Biol. Chem.* **270**, 12411–12417 (1995).
25. Huang, C., Zang, Q., Takagi, J. & Springer, T.A. *J. Biol. Chem.* **275**, 21514–21524 (2000).
26. Carr, C.M. & Kim, P.S. *Cell* **73**, 823–832 (1993).
27. Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. *Nature* **371**, 37–43 (1994).
28. Andrew, D. *et al. Eur. J. Immunol.* **23**, 2217–2222 (1993).
29. Studier, F.W. & Moffatt, B.A. *J. Mol. Biol.* **189**, 113–130 (1986).
30. Blacklow, S.C. & Kim, P.S. *Nature Struct. Biol.* **3**, 758–762 (1996).
31. North, C.L. & Blacklow, S.C. *Biochemistry* **38**, 3926–3935 (1999).
32. Cavanagh, J., Palmer, I.A.G., Fairbrother, W. & Skelton, N.J. Protein NMR Spectroscopy: Principles and Practice. (Academic Press, San Diego; 1996).
33. Pons, J.-L. Malliavin, T.E. & Delsuc, M.A. *J. Biomol. NMR* **8**, 445–452 (1996).
34. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. *Proteins* **17**, 75–86 (1993).
35. Brunger, A.T. *et al. Acta Crystallgr. D* **54**, 905–921 (1998).
36. Stickle, D.F., Presta, L.G., Dill, K.A. & Rose, G.D. *J. Mol. Biol.* **226**, 1143–1159 (1992).
37. Sali, A. & Blundell, T.L. *J. Mol. Biol.* **234**, 779–815 (1993).
38. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. *J. Biomol. NMR* **8**, 477–486 (1996).
39. Carson, M. *Methods Enzymol.* **277**, 493–505 (1997).
40. Nicholls, A., Sharp, K.A. & Honig, B. *Proteins* **11**, 281–296 (1991).