Molecular Cell
Article

The Structure of a Receptor with Two Associating Transmembrane Domains on the Cell Surface: Integrin $\alpha_{IIb}\beta_3$

Jieqing Zhu,[1,3] Bing-Hao Luo,[1,3,4] Patrick Barth,[2,3] Jack Schonbrun,[2,5] David Baker,[2] and Timothy A. Springer[1,*]
[1]The Immune Disease Institute, Children's Hospital Boston, and Department of Pathology, Harvard Medical School, 3 Blackfan Circle, Boston, MA 02115, USA
[2]Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA
[3]These authors contributed equally to this work
[4]Present address: Department of Biological Sciences, 202 Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA
[5]Present address: Spotfire, 212 Elm Street, Somerville, MA 02144, USA
*Correspondence: springeroffice@idi.harvard.edu
DOI 10.1016/j.molcel.2009.02.022

SUMMARY

Structures of intact receptors with single-pass transmembrane domains are essential to understand how extracellular and cytoplasmic domains regulate association and signaling through transmembrane domains. A chemical and computational method to determine structures of the membrane regions of such receptors on the cell surface is developed here and validated with glycophorin A. An integrin heterodimer structure reveals association over most of the lengths of the α and β transmembrane domains and shows that the principles governing association of hetero and homo transmembrane dimers differ. A turn at the Gly of the juxtamembrane GFFKR motif caps the α TM helix and brings the two Phe of GFFKR into the α/β interface. A juxtamembrane Lys residue in β also has an important role in the interface. The structure shows how transmembrane association/dissociation regulates integrin signaling. A joint ectodomain and membrane structure shows that substantial flexibility between the extracellular and TM domains is compatible with TM signaling.

INTRODUCTION

Membrane proteins with a single transmembrane (TM) domain per monomer are important in transmitting signals between the extracellular and cytoplasmic environments. For example, signal transmission in receptor kinases generally involves ligand-triggered association of monomeric single-TM proteins into homodimers or heterodimers (Zhang et al., 2006).

Integrins, adhesion receptors that transmit signals bidirectionally across membranes, are another heavily studied class of signaling receptors (Luo et al., 2007; Wegener and Campbell, 2008). Integrins have two subunits that constitutively heterodimerize through their large, ligand-binding extracellular domain. Reversible association through the single-pass α and β subunit TM domains mediates TM signaling. The cytoplasmic domains are short and can bind cytoskeletal proteins.

Structures are known in both active and inactive conformations for receptor kinase extracellular and cytoplasmic domain fragments (Zhang et al., 2006) and for integrin extracellular ligand-binding fragments (Luo et al., 2007) and cytoplasmic domains (Vinogradova et al., 2002). However, none of these structures include TM domains, and how signals are transmitted between these domains and across the membrane is largely unknown. In contrast, structural work on signaling proteins that span the membrane six or more times, such as channels and G protein-coupled receptors, is far more advanced.

Thus far, our understanding of how single-TM proteins associate in the membrane is confined to a limited number of experimental studies on isolated TM domains. Structures of the constitutively associating TM domains of the erythrocyte glycoprotein glycophorin A (GPA) have been solved in detergent and in pelleted vesicles (MacKenzie et al., 1997; Smith et al., 2001). A constitutively disulfide-linked TM dimer was also structurally defined in detergent by NMR (Call et al., 2006). Bicelles, made of a mixture of lipids and short-chain, detergent-like lipids, are closer mimics of bilayers than are detergents. Two further dimeric TM fragment structures (Bocharov et al., 2008a, 2008b) and monomeric α_{IIb} and β_3 TM domain structures (Lau et al., 2008a, 2008b) have recently been determined in bicelles. However, because regulated TM domain association is driven by domains outside of the membrane, structures of such associated TM domain fragments beg the question of whether they are physiologically relevant and, if so, whether they correspond to resting or active conformations.

The biological importance of integrins and the tractability of studying regulated TM and cytoplasmic domain association in the absence of ligand, in contrast to most other signaling receptors, have made them an attractive model system for studying association of their TM and cytoplasmic domains (Wegener and Campbell, 2008). In integrins, the α and β subunit TM domains associate in the resting state, driven by the close proximity of the C termini of α and β subunit ectodomains in the bent conformation. TM and cytoplasmic domain separation, induced by binding of the β subunit cytoplasmic domain through talin to

the actin cytoskeleton, drives integrin extension and shifts the ligand-binding α/β headpiece to an open, high-affinity conformation (Luo et al., 2007; Wegener and Campbell, 2008; Zhu et al., 2008). Disulfide crosslinking of portions of the TM domains in the outer membrane leaflet demonstrates helicity and the approximate orientations between the α_{IIb} and β_3 helices in the resting state, as well as an absence of association in the active state (Luo et al., 2004). Various TM domain orientations have been proposed based on these data or mutagenesis (Gottschalk, 2005; Li et al., 2005; Partridge et al., 2005). A comprehensive NMR complex structure showed how the integrin cytoplasmic domains associated and provided a model for integrin activation by showing that the cytoskeletal protein talin dissociated the complex (Vinogradova et al., 2002). However, different α/β complex structures or a lack of association have also been described (Ulmer et al., 2001; Vinogradova et al., 2004; Weljie et al., 2002). A unique feature of integrins is a highly conserved GFFKR motif following the α subunit TM domain. Mutation of any of the residues in this motif activates integrins by destabilizing association of α and β subunit TM domains (Hughes et al., 1996; O'Toole et al., 1994).

Disulfide crosslinking has been used to obtain structural information about membrane proteins, especially to probe TM domain organization and conformational change in bacterial chemoreceptors (Bass et al., 2007). However, development of this technique has not proceeded to the level of describing three-dimensional structures. Rosetta uses knowledge-based as well as physicochemical potentials and an efficient method of sampling conformational space to find low-energy structures. Furthermore, the potentials in Rosetta can be combined with experimental restraints (Das and Baker, 2008).

Here, we use disulfide-based distance restraints with Membrane Rosetta (Barth et al., 2007, 2009) to characterize the structure of the TM and juxtamembrane (JM) domains in intact integrins on the cell surface. The juxtamembrane GFFKR motif in α and a juxtamembrane Lys in β form an important part of the intersubunit interface. The conclusions are supported by mutational studies. Joining an $\alpha_{IIb}\beta_3$ ectodomain crystal structure to the membrane structure yields insights about integrin orientation on the cell surface and the finding that ectodomain flexibility is compatible with transmembrane signaling.

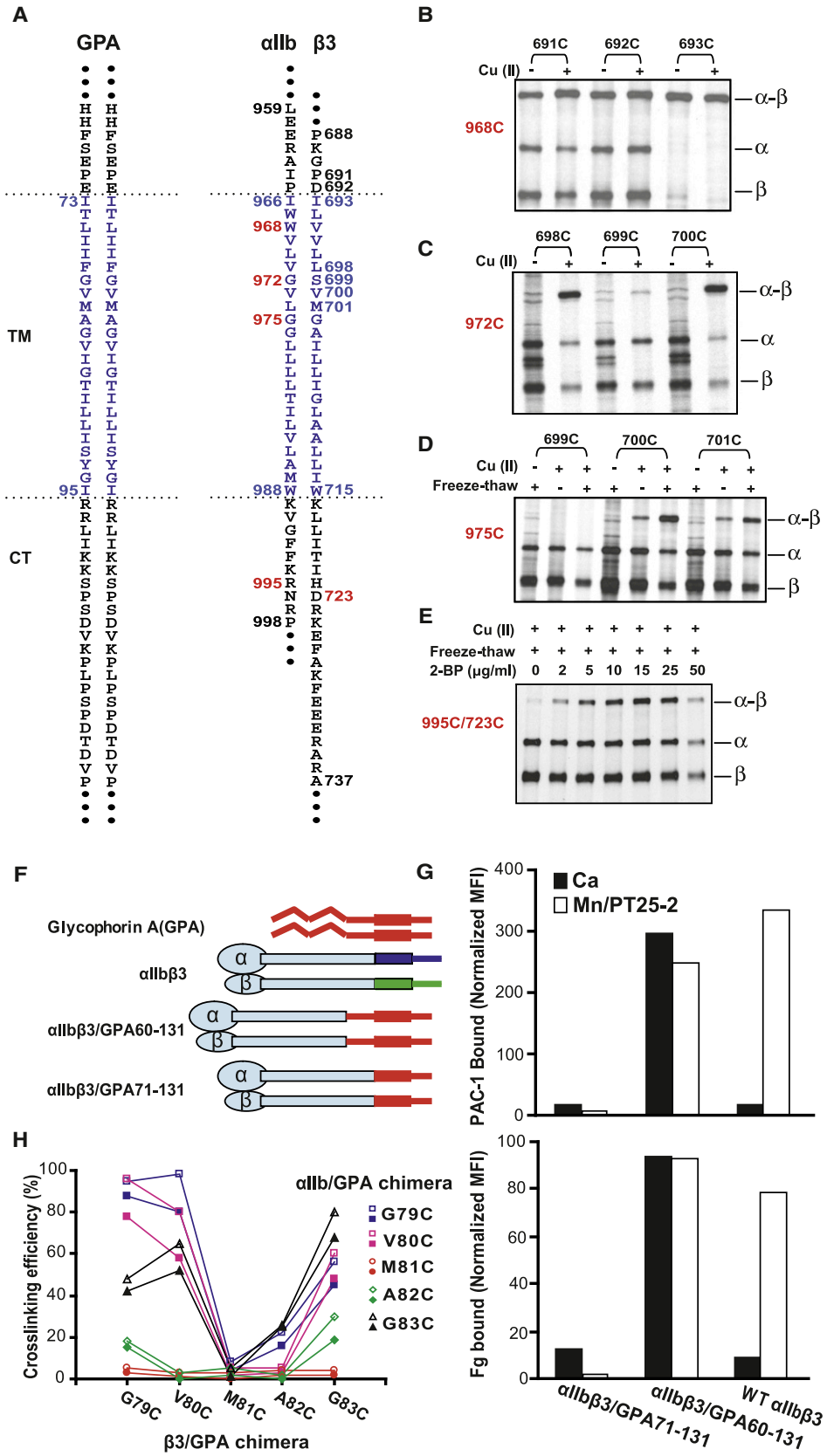
RESULTS

Method Development

We extended previous work on disulfide crosslinking of the first nine integrin TM residues (Luo et al., 2004) across the membrane and into the cytoplasm. Residues in the JM and TM segments of α_{IIb}, β_3, and GPA (Figure 1A) were individually mutated to cysteine. α_{IIb} and β_3 mutants were coexpressed in 293T transfectants. Cells were treated with or without Cu (II)-o-phenanthroline (Cu-phenanthroline) to catalyze disulfide bond formation, quenched with N-ethylmaleimide, and lysed with detergent. [^{35}S] integrins were immunoprecipitated and subjected to nonreducing SDS-PAGE (Figures 1B–1D), and disulfide-linked heterodimer was quantitated as percentage of total integrin. Disulfide bond formation within the first few TM residues was not increased further by Cu-phenanthroline (Figure 1B). Disulfide formation in this region and external to the membrane was also quantitated in redox buffers and after reduction with dithiothreitol followed by Cu-phenanthroline (Experimental Procedures and Figure S1 available online). By the seventh TM residue, disulfide formation was almost completely Cu-phenanthroline dependent (Figure 1C) (Luo et al., 2004), whereas by the tenth TM residue, Cu-phenanthroline alone had little effect (Figure 1D). However, disulfide crosslinking in the middle and cytoplasmic side of the membrane was made possible by freeze-thawing cells, which enables access of Cu-phenanthroline and oxygen to the otherwise reducing cytoplasmic environment (Figures 1D and 1E). Freeze-thawing had no effect on crosslinking of more exofacial residues (Figure S2). Cysteines near the membrane/cytoplasm interface may be palmitylated and, thus, unavailable for crosslinking (Kovalenko et al., 2004). Blocking palmitylation with 2-bromopalmitate (2-BP) markedly enhanced crosslinking efficiency (Figure 1E).

Standard crosslinking conditions thus employed 15 μg/ml 2-BP; a 1.5 hr pulse with [^{35}S] methionine and cysteine; a 17 hr chase; freeze-thaw; and treatment with Cu-phenanthroline for 10 min at 0°C. Kinetics showed that crosslinking was maximal at 10 min, even with cysteine pairs showing only partial crosslinking (Figure S3). In cysteine crosslinking, it is thought that a thiol peroxide (S-OH) equivalent to sulfenic acid is the first oxidation product and that nucleophilic attack on the oxidized cysteine by a second, unoxidized cysteine SH group results in disulfide formation, with release of water. We hypothesize that crosslinking by the second, unoxidized cysteine is in competition with oxidation of the second cysteine, which would prevent crosslinking to the first cysteine and would account for the lack of further crosslinking after 10 min. The dependence of disulfide formation on kinetics of oxidation of the first cysteine and its limitation by the kinetics of oxidation of the second cysteine may make the extent of crosslinking relatively insensitive to kinetic differences in SH oxidation at different depths within the membrane because the first and second cysteines must be nearby and at similar depths. This hypothesis agrees with the lack of correlation between depth in the membrane and extent of crosslinking (Figure 2).

Dimerization through the GPA TM domains is constitutive and is not dependent on the extracellular, O-glycosylated domain (MacKenzie et al., 1997). The α_{IIb} and β_3 ectodomains were fused to GPA residues 71–131, comprising the TM and cytoplasmic domains, or GPA residues 60–131, which include an additional 11 extracellular residues (Figure 1F). The use of α_{IIb} GPA /β_3GPA heterodimers enabled measurement of crosslinking not only between identical GPA TM residues (e.g., 80 with 80), but also between nonidentical residues (e.g., 80 with 81). Compared to wild-type $\alpha_{IIb}\beta_3$, the $\alpha_{IIb}\beta_3$/GPA 60–131 heterodimer was constitutively active, whereas $\alpha_{IIb}\beta_3$/GPA 71–131 was resistant to activation (Figure 1G). These results are consistent with the importance of close association between the C termini of integrin α and β subunit ectodomains in stabilizing the resting state (Luo et al., 2007) and stronger association between the GPA than the integrin TM domains. The αIIbβ3/GPA60-131 and αIIbβ3/GPA71-131 chimeras gave essentially identical disulfide crosslinking (Figure 1H), showing


A
GPA
αIIb β3
TM
CT
B
691C 692C 693C
Cu (II)
968C
α-β
α
β
C
698C 699C 700C
Cu (II)
972C
α-β
α
β
D
699C 700C 701C
Cu (II)
Freeze-thaw
975C
α-β
α
β
E
Cu (II)
Freeze-thaw
2-BP (μg/ml) 0 2 5 10 15 25 50
995C/723C
α-β
α
β
F
Glycophorin A(GPA)
αIIbβ3
αIIbβ3/GPA60-131
αIIbβ3/GPA71-131
G
Ca
Mn/PT25-2
PAC-1 Bound (Normalized MFI)
Fg bound (Normalized MFI)
αIIbβ3/GPA71-131
αIIbβ3/GPA60-131
WT αIIbβ3
H
Crosslinking efficiency (%)
αIIb/GPA chimera
G79C
V80C
M81C
A82C
G83C
β3/GPA chimera

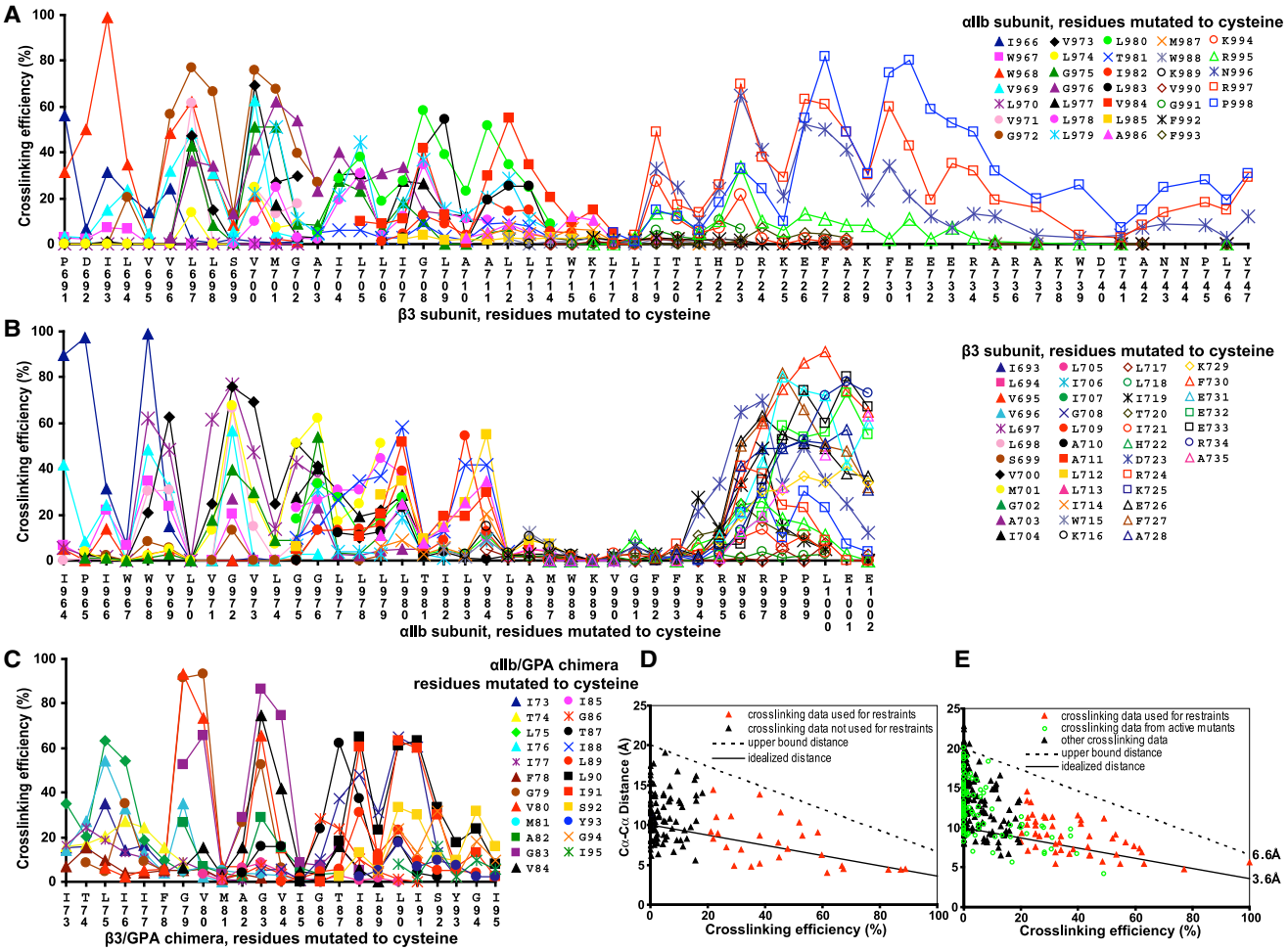


Figure 2. Disulfide Crosslinking Efficiency and Correlation with Cα-Cα Atom Distance

Disulfide crosslinking in transfectants between subunits with indicated cysteine mutations. As described in Experimental Procedures, data are with 2-BP, freeze-thaw, and Cu-phenanthroline, except that GPA residues 73–80 are with DTT and Cu-phenanthroline.

(A and B) The same $\alpha_{IIb}\beta_3$ disulfide crosslinking data, plotted against β_3 (A) or α_{IIb} sequence position (B).

(C) GPA TM domain disulfide crosslinking in $\alpha_{IIb}\beta_3$/GPA71–131 chimera.

(D and E) The same crosslinking data as in (A)–(C), plotted against Cα-Cα distance. (D) Crosslinking efficiency in $\alpha_{IIb}\beta_3$/GPA71–131 chimera is plotted against Cα-Cα distance in the GPA solid-state NMR structure (Smith et al., 2001). (E) Crosslinking efficiency in $\alpha_{IIb}\beta_3$ is plotted against Cα-Cα distance in the final integrin structure. Solid lines show the assumed relationship between crosslinking efficiency and Cα-Cα distance, and dashed lines show the distance above which restraint violation was penalized in Membrane Rosetta.

that GPA TM domain association was unaffected by the length of GPA extracellular domain included in the chimera. Crosslinking between nonidentical GPA transmembrane residues was also independent of the integrin subunit fusion partner (Figure S4).

Crosslinking and Restraints

The integrin and GPA TM domains show α-helical periodicity with crosslinking peaks every three to four residues, suggesting helix-against-helix packing (Figures 2A–2C). Similar periodic peaks in the β_3 cytoplasmic domain at residues I719, D723,

Figure 1. Disulfide Crosslinking in Native Cell Membrane

(A) Sequences of GPA and $\alpha_{IIb}\beta_3$ integrin TM and cytoplasmic domains. Numbers in red show positions tested in (B)–(E).

(B–E) Disulfide crosslinking of $\alpha_{IIb}\beta_3$ with indicated cysteine mutations in 293T transfectants, with or without Cu-phenanthroline, freeze-thaw, and 2-BP treatment as indicated. Immunoprecipitated ^{35}S-labeled material was subjected to nonreducing SDS-PAGE and autoradiography. Positions of αIIb (α), β3 (β), and αIIbβ3 heterodimer (α-β) are shown.

(F) Integrin αIIbβ3/GPA chimeras.

(G) Ligand binding by chimeras. Binding of ligand mimetic PAC-1 (IgM) (upper panel) or FITC-labeled fibrinogen (Fg) (lower panel) to 293T transfectants was measured in the presence of 1 mM Ca^{2+} or 1 mM Mn^{2+} plus 10 μg/ml activating mAb PT25-2. Binding is expressed as mean fluorescence intensity (MFI) of Fg or PAC-1 relative to MFI of Cy3-labled anti-β3 mAb AP3.

(H) αIIbβ3/GPA60-131 (closed symbols) and αIIbβ3/GPA71-131 (open symbols) chimeras show similar disulfide crosslinking.

Table 1. Experimental Restraints and Structure Statistics

Experimental Restraints		
	GPA	Integrin
Number of residues	46	90 (EC, 12; TM, 46; CT, 32)[a]
Number of restraints	54 (symmetrized)	48 (TM, 33; CT, 15)[a]
Structure Statistics		
	GPA (20 models)	Integrin (52 models)
	TM (46 residues)	TM + CT (78 residues)[b]
Cα-RMSD (Å) to cluster center model	0.68 ± 0.23	1.94 ± 0.42
Distance restraints RMS violation (Å)	0.01 ± 0.02 (0.00)[c]	0.06 ± 0.04 (0.00)[c]
Angle between TM helix axes (°)	−45.4 ± 1.8 (−43.1)[c]	−37.5 ± 3.1 (−36.9)[c]
Distance between TM helix axes (Å)	7.3 ± 0.49 (7.1)[c]	8.2 ± 0.77 (7.8)[c]
RMSDs from Idealized Geometry		
Bonds (Å)	0.0067 ± 0.008	0.012 ± 0.005
Angles (°)	0.28 ± 0.0015	0.57 ± 0.10
Ramachandran Statistics[d]		
Ramachandran favored	100.0%	99.8% ± 0.4%
Ramachandran outliers	0.0%	0.02% ± 0.15%

[a] EC, extracellular region; TM, transmembrane region; CT, juxtamembrane and cytoplasmic region.
[b] Only TM (I966–W988 of αIIb subunit and I693–W715 of β3 subunit) and CT (K989–P998 of αIIB subunit and K716–A737 of β3 subunit) residues are included.
[c] Values in parentheses are for the cluster center structure.
[d] Ramachandran statistics were measured by MolProbity (Davis et al., 2007).

F727, and F730 (Figure 2A) suggest that this region is α-helical as well and interacts with the Lys and Arg residues of the GFFKR motif and several following residues in the α subunit.

The GPA crosslinking data fit the GPA NMR structure well (Figure 2D), with the assumption that maximal crosslinking of 100% would be seen with maximal contact between Cys Cα atoms at the van der Waals distance of 3.6 Å and that crosslinking would linearly fall to 0% when this distance was increased to 10 Å (Figure 2D, solid line). Because distance is not the only factor that could affect crosslinking efficiency, it was important not to overrestrain the distance between crosslinked residues. Therefore, an upper distance bound was chosen, below which no penalty would be applied (Figures 2D and 2E, dashed lines). Restraints were used only for residue pairs with ≥ 20% crosslinking (Figures 2D and 2E) and only for residues that did not activate ligand binding by $\alpha_{IIb}\beta_3$ when individually mutated to cysteine (see below). Furthermore, to compensate for inherent flexibility or structural perturbations introduced by cysteine substitution, the lowest crosslinking value within two residues was subtracted prior to restraint calculation (Figure S5).

Structure Refinement

Rosetta has an efficient method for sampling structural diversity by assembling 3 and 9 residue fragments from the Protein Data Bank into protein structures. Fragments are selected that are similar in sequence to the target sequence and are enriched for (but do not all have) the predicted secondary structure of the target structure. Fragments are selected from high-resolution crystal structures irrespectively of whether they are membrane or water-soluble proteins. Structures are built using a Monte Carlo fragment insertion protocol, minimizing an energy function that favors hydrophobic burial, strand pairing, and other low-resolution properties of protein structures. A subsequent high-resolution refinement stage optimizes van der Waals packing, hydrogen bonding, and solvation interactions (Experimental Procedures).

Membrane Rosetta uses a model of a membrane with three slabs, a central hydrophobic slab corresponding to the lipid alkyl groups and two intermediate slabs corresponding to the interface regions between the alkyl groups and the polar head groups. Both the low-resolution (Barth et al., 2009) and high-resolution (Barth et al., 2007) potentials differ from water-soluble proteins in these slabs, with exposure of nonpolar groups favored rather than disfavored. The potentials for the head group regions and the external regions are the same as for water-soluble proteins. The embedding (depth and orientation) of proteins in the membrane slab is optimized to give the lowest energy. Disulfide distance restraints are incorporated as an additional potential term used to evaluate model energy; a distance greater than the upper bound shown in Figures 2D and 2E raises energy. This energy term is evaluated in every Monte Carlo step and thus affects the course of model building.

Building the integrin TM and juxtamembrane domains was done in several stages, as detailed in the Experimental Procedures and the Supplemental Data. In the final stage, all models with the 10% lowest energy were clustered by structural similarity (Figures S6 and S7). The vast majority of low-energy structures have similar structural features (Figure S7). Structure assessment below uses the structures at the center of the largest (top) GPA and integrin clusters (referred to as final structures), the structural ensembles within the top clusters (Tables 1 and 2), and the centers of the top five clusters (Figure S7).

Structure Assessment

Cross-validation with omitted restraints is used for the validation of crystal and NMR structures (Brunger et al., 1993). GPA and

Table 2. Comparison of Solution NMR Structure (1AFO) and Disulfide/Rosetta GPA Structures to Solid-State NMR Structure

	I73–I95 (46 Residues)			I76–I88 (26 Residues)		
	Cα	Backbone	All Heavy	Cα	Backbone	All Heavy
1AFO (20 models)	0.97 ± 0.10	1.14 ± 0.11	1.69 ± 0.08	0.67 ± 0.10	0.64 ± 0.09	1.09 ± 0.18
Disulfide/Rosetta (20 models)	0.96 ± 0.15	1.09 ± 0.13	1.59 ± 0.19	0.61 ± 0.12	0.59 ± 0.12	1.27 ± 0.20
Largest cluster center model	0.71	0.84	1.28	0.44	0.43	0.91

All values are RMSD for the indicated atoms (± SD for structural ensembles) in Å.

integrin models generated without restraints were far from the GPA NMR and integrin final models (Figures 3 and S8). When 12.5% restraints were used, models were much closer. When 25% and 50% restraints were used, the models approached a plateau in Cα RMSD for integrin of 1 Å for TM and 1.8 Å for TM + cytoplasmic domains and 0.8 Å for GPA (Figures 3A, 3C, and 3E). This provides one estimate of structure accuracy. Crystal structure accuracy is evaluated by omitting data from refinement and checking the agreement between the omitted data and the final structure (R_{free}). Similarly, we checked structures for violations of restraints that were not used in structure generation (Figures 3B, 3D, and 3E). When 50% of restraints were used in structure generation, there was little violation of the other 50% of omitted restraints (Figures 3B, 3D, and 3E). These results suggest that, at least for the regions of the structures that are well covered by restraints, i.e., the TM regions, the structures match the actual structure well.

The solid-state GPA NMR structure, which fits the solution-state restraints better than the solution NMR structure (Smith et al., 2001), fits the solution-state NMR and Disulfide/Rosetta structures, respectively, with Cα RMSD of 0.97 ± 0.10 and 0.96 ± 0.15 Å (see Table 2 for more details). These results show that the Disulfide/Rosetta structure is as similar to the solid-state NMR structure as the latter is to the solution-state NMR structure. The similarities between the Disulfide/Rosetta and NMR structures extend to details including side-chain rotamers (Figures 3G and 3H). For example, side chains in crosslinking peaks have identical rotamers, except for Leu-90 and Ile-91 in one of the two Disulfide/Rosetta TM helices (Figure 3H), which, in contrast to the solid-state NMR structure, were not constrained to be symmetric.

The final $\alpha_{IIb}\beta_3$ membrane structure showed no restraint violation (Figure 2E), and the ensemble showed little restraint violation (Table 1). Furthermore, crosslinking data from activating cysteine mutations that were omitted from structure calculation also did not violate restraints and trended well with the idealized crosslinking-distance relationship (Figure 2E and Table S1). The apparent lack of effect on crosslinking by these activating mutations may result from the use of 0°C and 37°C in crosslinking and activation assays, respectively.

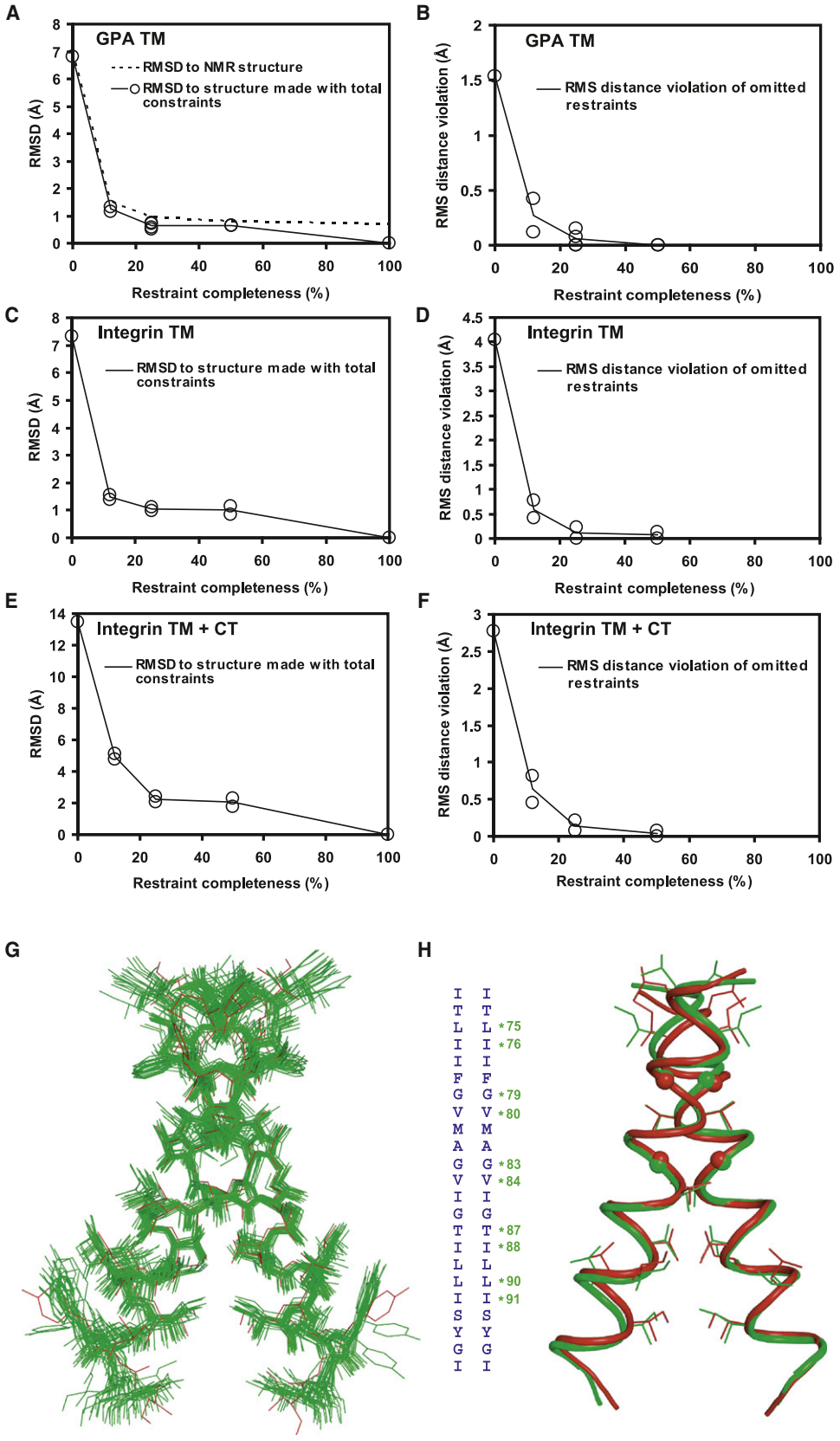
$\alpha_{IIb}\beta_3$ Membrane Structure and Comparison to Mutational Results

All low-energy $\alpha_{IIb}\beta_3$ structures have the same interhelical interface, and almost all have the Gly of GFFKR in a left-handed helix conformation that caps the α_{IIb} α helix and brings the two Phe of GFFKR into the interface with the β_3 α helix (Figures 4A–4E, S7A, and S7B). These key features are seen in all five top clusters with clustering at 2.0 Å (Figure S7A) and four out of five of the top clusters at 3.0 Å (Figure S7B). The sole exception, cluster 3 with clustering at 3.0 Å (Figure S7B), has the highest disulfide restraint violation (Figure S6B) and can be ruled out because it is incompatible with mutational data described below that show that the α_{IIb} GFFKR motif and β_3 Lys-716 have an important role in the α_{IIb}/β_3 interface. The β_3 TM α helix continues as α helix into the cytoplasm (Figure 4B), and in some clusters, the cytoplasmic portion of the β_3 TM α helix is divided into two segments separated by a turn near Glu-726 (Figure S7).

The integrin TM domains associate with a right-handed crossing angle of −37° compared to −41 to −45° in GPA (Table 1) (MacKenzie et al., 1997; Smith et al., 2001). The most prominent features of α-helical surfaces are ridges formed by the side chains of every fourth residue in sequence (i + 4n) and grooves in between (Figure 4F). In a common packing mode between adjacent α helices, observed for α helices that cross at angles of −50° ± 20°, a ridge formed by residues i + 4n on one helix fits into a groove on another helix between the ridges formed by residues j + 4n and j + 1 + 4n (Chothia et al., 1981). This has been termed "4-4 packing." The α_{IIb} and β_3 TM α helices show precisely this packing mode (Figures 4F and 4H). The ridge is on α_{IIb} and is formed by residues W968, G972, G976, and L980. The groove is on β_3 between residues V696 and L697, V700 and M701, I704 and L705, and G708 and L709 (Figure 4H). Residues in crosslinking peaks are found at the interface (Figures 4C, 4D, and 4H). Because the ridges and grooves spiral on the side of the cylindrical α helix, they are convex. The G972-XXX-G976 motif in α_{IIb} decreases the height of the ridge at the center of the packing interface, enabling the two helical cylinders to pack closer together.

In the NMR and cell-surface GPA structures, the TM helices pack most closely together at a GXXXG motif (MacKenzie et al., 1997) (Figures 3H and 4I). Two differences in interhelical orientation with integrins are apparent when the grooves in β_3 and one of the GPA monomers are superimposed (Figure 4G). First, although packing in GPA resembles ridge in groove (MacKenzie et al., 1997), the two low ridges bearing G79 and G80 move closer to one another than in integrin (horizontally in Figure 4I), providing more separation from the high ridges bearing V80 and V84. Second, the ridge in the integrin slides in its groove relative to GPA (vertically in Figure 4H) so that the spacing of residues in the interface differs (Figures 4H and 4I).

Mutational studies are in excellent agreement with the interhelical integrin interface identified here. Cysteine and leucine scanning show that the most activating and, hence, structurally disruptive mutations within the TM domains are of the three interfacial Gly residues: α_{IIb} Gly-972, α_{IIb} Gly-976, and β_3 Gly-708 (Figure 5A) (Luo et al., 2005). By contrast, mutations of α_{IIb}


A
GPA TM
RMSD to NMR structure
RMSD to structure made with total constraints
RMSD (Å)
Restraint completeness (%)
B
GPA TM
RMS distance violation of omitted restraints
RMS distance violation (Å)
Restraint completeness (%)
C
Integrin TM
RMSD to structure made with total constraints
RMSD (Å)
Restraint completeness (%)
D
Integrin TM
RMS distance violation of omitted restraints
RMS distance violation (Å)
Restraint completeness (%)
E
Integrin TM + CT
RMSD to structure made with total constraints
RMSD (Å)
Restraint completeness (%)
F
Integrin TM + CT
RMS distance violation of omitted restraints
RMS distance violation (Å)
Restraint completeness (%)
G
H
I I
T T
L L *75
I I *76
I I
F F
G G *79
V V *80
M M
A A
G G *83
V V *84
I I
G G
T T *87
I I *88
L L
L L *90
I I *91
S S
Y Y
G G
I I

Gly-975 and β_3 Gly-702, which are not interfacial (Figures 4D and 4H), are not activating (Figure 5A). These results correlate with the observations that the interfacial, but not the noninterfacial, Gly residues are invariantly small (Gly, Ala, or Ser) residues in integrin TM domains (Figure S10).

The GFFKR motif has a structural role in $\alpha_{IIb}\beta_3$ association consonant with its long-known functional importance. The α_{IIb} TM α helix extends beyond the 23 residue TM hydrophobic segment through Lys-989 and Val-990. However, at Gly-991 of the GFFKR motif, the backbone adopts a left-handed rather than right-handed α-helical conformation (Figure 4E). Gly in left-handed conformation at the C termini of α helices is the most common C-cap motif and is responsible for the marked overrepresentation of Gly at the +1 position following α helices (Richardson and Richardson, 1989). Almost all (91%) low-energy models and all but one outlier cluster (Figure S7) have a turn at the Gly (Figure S11). Many of these models, including the final structure, have Phi/Psi angles for this Gly that are disallowed for non-Gly residues (Figure S11), potentially explaining why mutation of this residue is so destabilizing. The turn at the Gly brings the two Phe residues of the GFFKR motif into the interface between the α_{IIb} and β_3 TM α helices (Figure 4E). The α_{IIb} and β_3 TM α helices are further apart near this interface than anywhere else in the membrane, and Phe-993 in particular occupies space between the two helices and is central in the interface. Phe-993 contacts on the β_3 subunit Leu-712 and is nearby the aliphatic portion of Lys-716. In the α_{IIb} subunit, Phe-993 abuts Met-987 and Phe-992. Both Phe side chains orient toward the hydrophobic core of the membrane. A meander in the α_{IIb} cytoplasmic domain containing the Lys and Arg of the GFFKR motif and three more C-terminal residues abuts the β_3 cytoplasmic α helix.

Mutational data support the juxtamembrane α_{IIb} structure. Cysteine substitutions in the GFFKR motif are activating (Figure 5A), emphasizing the important role of this motif in the juxtamembrane interface between the α_{IIb} and β_3 subunits. Mutation of FF of GFFKR to AA, YY, WW, or LL demonstrated that even the aromatic substitutions YY and WW are activating (Figure 5B). These results are consistent with the important packing role, particularly of Phe-993, in the $\alpha_{IIb}\beta_3$ interface and not with the general role that aromatic residues, particularly Trp, often play in the hydrophobic/polar membrane interface.

In the β_3 juxtamembrane domain, residues Lys-716, Ile-719, and Asp-723 are in the interface with the α_{IIb} GFFKR motif (Figures 2A and 4E). Remarkably, the Lys-716 ε-amino group hydrogen bonds to the α_{IIb} backbone (Figure 4E). Hydrogen bonds to either or both of the α_{IIb} Phe-992 and Lys-994 carbonyls, with Phe-992 predominating, are found in 31% of all low-energy models, 65% of the cluster 1 ensemble, and in five out of nine of the cluster center models shown in Figure S7 that have the turn at Gly-991.

Cysteine scanning of β_3 showed that the only activating mutation in the juxtamembrane/cytoplasmic domain is K716C (Figure 5A). Furthermore, this mutation is more activating than the single other activating Cys mutation identified, G708C in the TM domain. These results confirm the important structural role of Lys-716 in the $\alpha_{IIb}\beta_3$ interface. β_3 Lys-716 was also mutated to Ser, Arg, Glu, Leu, and Pro. All substitutions were activating except for Arg (Figure 5B), the only other side chain that could form a similar hydrogen bond to an α_{IIb} backbone carbonyl oxygen, and is found at this position in integrin β subunits (Figure S10). Mutation showed that, in contrast to Arg, polar (Ser, Glu) and aliphatic (Leu) side chains could not substitute for Lys-716. These findings lend credence to the hydrogen bond between the β_3 Lys-716 side chain and the α_{IIb} backbone.

Arg-995 of the α_{IIb} GFFKR motif is nearby both β_3 Asp-723 and Glu-726 (Figure 4E), consistent with the proposal based on mutational studies of a salt bridge between Arg-995 and Asp-723 (Hughes et al., 1996). However, the side chains of Arg-995 and Asp-723 sample a large conformational space and are within hydrogen-bonding distance in only 30% of the cluster 1 structural ensemble and 16% of low-energy structures (Table S2).

We have consistently observed little effect of mutation of β_3 Asp-723, the putative salt bridge partner of α_{IIb} Arg-995 in either 293T transfectants (Figures 5A and 5B) or transient CHO-K1 transfectants (Figure 5C). Because Asp-723 mutants have been found to be active in stable CHO transfectants (Hughes et al., 1996), this may reflect differences in levels of cell-surface density or in assay sensitivity.

Structure of the Full-Length αIIbβ3 Integrin

Models of the entire receptor were generated by assembling the structures of the TM + cytoplasmic domains with the ectodomain crystal structure (Zhu et al., 2008). Redox buffer disulfide restraints were used in the 6 residue α_{IIb} and 4 residue β_3 linkers between these segments (Figure S1B and Table S3B). Restraints were validated based on crosslinking between nearby residues defined in $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$ crystal structures; because of flexibility in crystal structures of the β_3 tail domain (Zhu et al., 2008), restraints were loosened relative to the TM region (Figure S12). The linkers were modeled while simultaneously optimizing the rigid body orientation of the ectodomain with respect to the membrane. Two long, disordered loops at the membrane-proximal base of the α_{IIb} calf-2 domain were included as flexible loops

Figure 3. Estimation of Structure Accuracy and Validation of the GPA Cell-Surface Structure

(A–F) Effect of restraint omission. Models were generated using different subsets of 12.5%, 25%, or 50% of the total restraints as described in the Supplemental Data. In (A), (C), and (E), models were compared over Cα atoms to the final structure made with all restraints, and, in the case of GPA, also to the GPA NMR structure. By definition, models made with 100% of restraints are identical to the final structure. In (B), (D), and (F), models generated with a subset of restraints were scored for violation of the omitted restraints, i.e., those not used in model generation. The per residue RMS distance violation, i.e., distance above the upper bound (see Figures 2D and 2E), is shown. How well the models satisfy omitted restraints is a measure of model accuracy similar to the R_{free} value in crystallography.

(G) The ensemble of 20 GPA Disulfide/Rosetta structures, showing all heavy atoms (green) superimposed on the solid-state NMR structure (red).

(H) Cartoon of the central GPA Disulfide/Rosetta structure (green) superimposed on the solid-state NMR structure (red) with side chains (or spheres for Gly) shown for residues in crosslinking peaks (asterisks in the sequence insert).

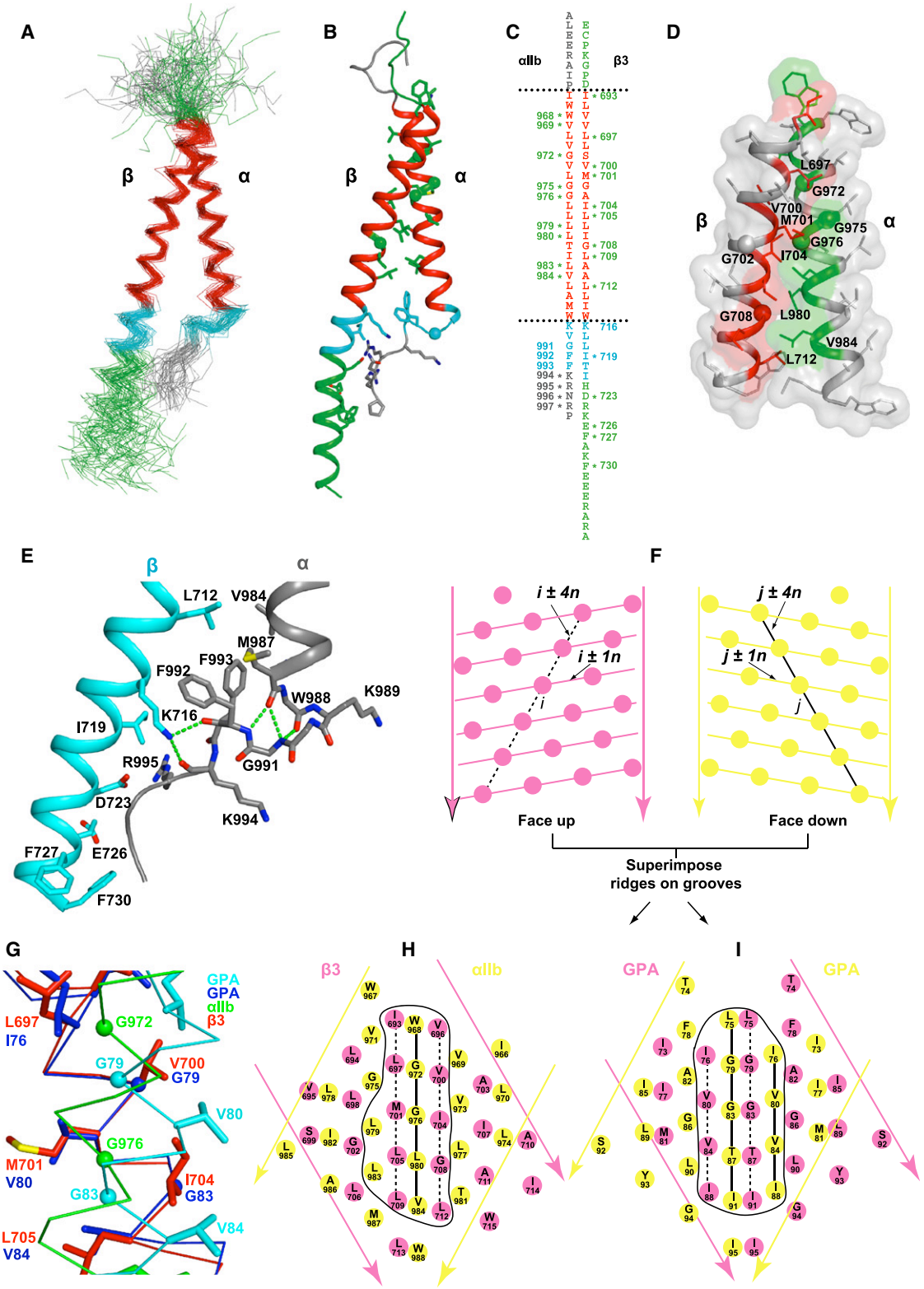


Figure 4. Structure of the Membrane Region of an Integrin

(A–D) Side chains and Gly Cα spheres in the crosslinking peaks in the 23 residue hydrophobic segment are shown in red or green. Residues in the cytoplasmic juxtamembrane interface region (18 to 12 Å from the membrane center) are cyan.

(A) The Disulfide/Rosetta structural ensemble superimposed on the 46 residue TM segments.

in models because they may impact orientation on the cell surface. The lowest-energy models exhibit a diversity of ectodomain orientations (Figure 6A). However, the range of orientations was more restricted than would be imposed solely by the physical barrier of the membrane bilayer. These results show that the TM and linker regions do not completely restrict the orientation of the ectodomain, which is still able to explore a substantial amount of conformational space. This is an important finding because it shows that models of signal transmission between the integrin ecto and TM domains have to be compatible with substantial flexibility at the interfaces between these domains.

DISCUSSION

Structures for Intact Receptors with Two Associating TM Domains

There are currently no crystal or NMR structures of intact signaling receptors with a single transmembrane domain per monomer. NMR studies to date have been on small fragments of subunits that either constitutively associate (Call et al., 2006; MacKenzie et al., 1997; Smith et al., 2001) or undergo regulated association but are devoid of the regulating extracellular domains (Bocharov et al., 2008a, 2008b). Even the two fragment structures that nominally represent regulated receptors may be special cases because the EphA1 receptor dimerizes in SDS, and ErbB2 physiologically only forms heterodimers, not homodimers. NMR experiments on fragments are capable of yielding more detailed structural information, but not of defining for regulated receptors whether the conformations that are experimentally captured are physiologically relevant and whether they correspond to "on" or "off" states. The method developed here is complementary because it characterizes structures in the presence of extracellular and cytoplasmic domains that regulate transmembrane association in intact receptors on the cell surface. The method is applicable to a large class of signaling receptors that have a single transmembrane domain per monomer or subunit and undergo regulated dimerization.

Our method combines sparse experimental data (disulfide restraints) with energy terms in Rosetta analogously to how experimental and energy terms are used in NMR and crystal model building, except that there are more energy terms in Rosetta. In crystal and NMR structures, energy terms for bond angles and distances and van der Waals repulsion are generally used; additional energy terms may be used in lower-resolution structures. In lower-resolution crystal and NMR structures, there is less experimental data, and the energy terms receive more weight. We may similarly think of the Membrane Rosetta models made here as low-resolution structures because the weight of the energy terms relative to the experimental data is high.

Our GPA ensemble is as close to the solid-state NMR structure as the solution-state NMR ensemble is to the solid-state structure. The GPA Disulfide/Rosetta structure suggests that the GPA TM domain NMR structures determined in dodecylphosphocholine detergent (MacKenzie et al., 1997) and dimyristoylphosphatidylcholine or 1-palmitoyl-2-oleoylphosphatidylcholine multilamellar dispersions (Smith et al., 2001) are similar to the GPA TM domain structure on the cell surface.

Methods for validating structures with sparse experimental data such as reported here are an important area for further development (Das and Baker, 2008). However, analogous to cross-validation in NMR or omission of data from refinement of crystal structures (R_{free}) (Brunger et al., 1993), we have used restraint omission to attempt to evaluate the accuracy of our integrin structure. The estimated RMSD Cα accuracy is 1 Å for the TM portion of the integrin and 0.8 Å for GPA, which is comparable to the RMSD between our GPA membrane structure and the NMR structure. The lack of restraint violation by mutationally important residues that were omitted from the structure refinement provides further support for our integrin structure.

Previous experience with Rosetta has shown that structural features are most accurately modeled when shared by the majority of low-energy structures and when alternative low-energy structures can also be ruled out by experimental data (Das and Baker, 2008). Consequently, we have focused not on a single structure but on the overall characteristics of the 10% lowest-energy structures. We have only emphasized conclusions when they were validated by a majority of low-energy structures and when alternatives could be ruled out. The most critical example of this is the GFFKR motif. We obtained restraints between the α_{IIb} and β_3 TM helices preceding GFFKR and between the helical JM portion of β_3 and the K and R residues of GFFKR and several following residues. Thus, low-energy structures of the GFF moiety had to be found by Rosetta that were consistent with these restraints. The structure of the GFF moiety illustrates one of the strengths of Rosetta, which uses a large library of fragments from high-resolution structures. The most common C-cap motif for α helices involves a Gly in left-handed helical conformation (Richardson and Richardson, 1989), and the backbone hydrogen bonds in this motif include the first F in GFFKR (Figure 4E). Thus, this conformation was readily found, and 91% of low-energy structures were built including this motif. The other 9% of structures continued in right-handed α-helical conformation through the G and do not have GFF in α_{IIb} and K716 in β_3 in the interface. These structures can be ruled out because mutational studies show that GFF in

(B) The cluster-center structure.
(C) The sequences, with asterisks designating residues in crosslinking peaks.
(D) Transparent molecular surfaces for α_{IIb} and β_3 TM segments, with all side chains shown.
(E) A blow-up of the lower TM, JM, and cytoplasmic segments. Hydrogen bonds are dashed green lines. Nitrogens, oxygens, and sulfurs are blue, red, and yellow, respectively.
(F, H, and I) Crick helical-net diagrams (Chothia et al., 1981). Cylindrical helical surfaces are cut at one position along the circumference, unrolled, and aligned at the helical interfaces.
(G) Superposition of the integrin and GPA TM helices on residues that form the two ridges forming the groove in β_3 and one GPA monomer in which the GXXXG motifs nestle.

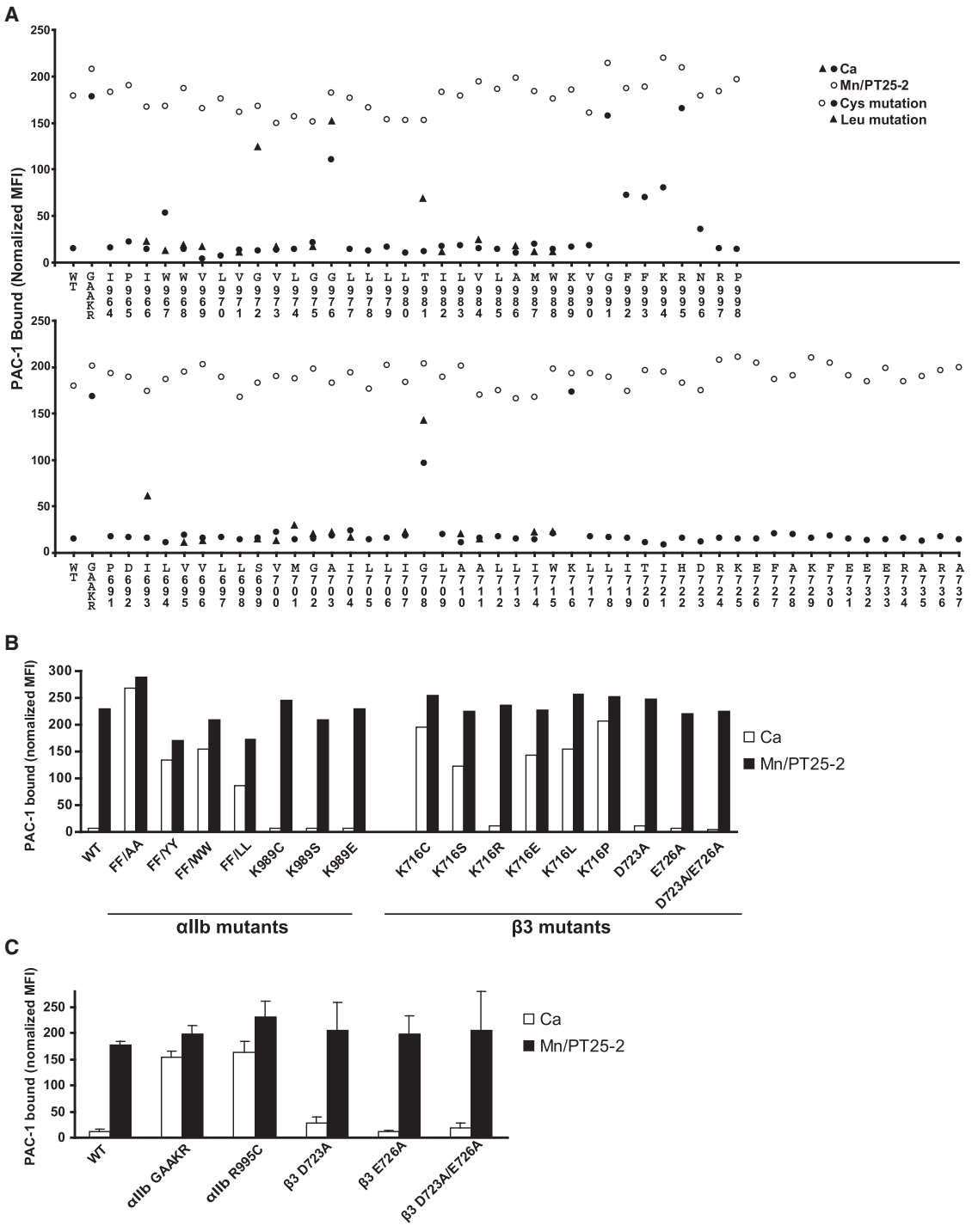


Figure 5. Ligand Binding by αIIbβ3 Integrin Mutants

(A) Binding of ligand-mimetic PAC-1 IgM by cysteine-scanning mutants in α_{IIb} (upper panel) and β_3 (lower panel). Results in absence (Ca) and presence of activation (Mn/PT25-2) are shown, along with wild-type (WT) and GFFKR/GAAKR mutant controls. Leucine-scanning results from a previous study (Luo et al., 2005) are also plotted.

(B and C) Effects of other mutations in 293T (B) and CHO-K1 (C) transient transfectants. Error bars represent the average of three independent experiments ± SD.

α_{IIb} and K716 in β_3 have a critical role in $\alpha_{IIb}\beta_3$ association. Interestingly, the only cluster lacking the turn at the G showed an unnatural, continuously curved TM β_3 α helix, as a consequence of satisfying disulfide restraints between α_{IIb} and β_3 residues on either side of the GFF motif (Figure S7B), presenting yet another criterion for discarding this model. Therefore, we conclude that

both the TM and JM portions of our $\alpha_{IIb}\beta_3$ membrane structure are strongly supported.

The Integrin TM Interface

Our study definitively identifies the interface between the α_{IIb} and β_3 TM α helices. Previous disulfide crosslinking of the first nine residues of the TM domains established only the approximate orientation between their helices and did not include the second G of the GXXXG motif (Luo et al., 2004). Based on that data, approximate orientations between the α_{IIb} and β_3 TM helices (Lau et al., 2008a) or detailed models (Gottschalk, 2005) have subsequently been presented. As an example of the lack of sufficiency of the previous data, the detailed model has an Cα RMSD of 2.7 Å over 44 TM residues and does not follow the ridge-in-groove packing we describe in the inner-half of the membrane (Gottschalk, 2005). Subsequent to Luo et al. (2004), another group used a different approach and produced two models that superimpose on our structure over 44 TM residues with Cα RMSD of 6.8 and 8.0 Å (Partridge et al., 2005). Thus, there has previously been no consensus on the orientation between the α_{IIb} and β_3 TM domains.

Compared to previous homodimer TM structures, the integrin TM heterodimer structure illustrates the important principle that ridge-in-groove packing in heterodimers is not constrained by symmetry, enabling packing of the ridge at intermediate positions in the groove not allowed in symmetric homodimers (vertically in Figures 4H and 4I). This finding is likely to reflect general differences between homodimeric and heterodimeric TM domain association. Many single-span receptors, such as the EGFR/ErbB family, can associate either as homo or heterodimers. ErbB family heterodimers have an important role in signaling, and their TM domains associate more stably than the corresponding homodimers (Duneau et al., 2007). It will be interesting to compare homo and heterodimers within a single receptor family to determine whether differences in packing modes between homo and heterodimers regulate the stability or signaling of TM domains.

Whereas association between GPA TM domains is constitutive and occurs even in SDS, association between the integrin TM domains is regulated and labile to detergents (Li et al., 2001; MacKenzie et al., 1997). The finding here that integrin ectodomain fusion directly to the GPA TM/cytoplasmic domains in the $\alpha_{IIb}\beta_3$ /GPA 71–131 chimera results in resistance to activation is compatible with the greater stability of the GPA dimer. This stability is not related to surface area buried in the interface, which with a 1.4 Å radius probe is 780 Å^2 for GPA, compared to 1050 Å^2 for the TM domain portion of integrin and 1300 Å^2 when the membrane-embedded JM portion is included. The β-branched residues, Ile, Thr, and Val, which only have one available rotamer in α helices, are present in the GPA TM domain interface. This minimizes loss of entropy when these residues are transferred from the lipid environment to a protein interface (Liu et al., 2003; MacKenzie et al., 1997). By contrast, integrins have a substantial number of interfacial Leu and Met residues, for which the conformational entropy loss upon TM helix-helix association will be much greater. Furthermore, hydrogen-bonded Thr residues stabilize the GPA, but not the integrin interface (MacKenzie et al., 1997; Smith et al., 2001).

The JM Interface and Comparison of Associated and Dissociated Integrin Subunits

The juxtamembrane $\alpha_{IIb}\beta_3$ structure is most unusual. The α_{IIb} α helix is terminated by a turn at the Gly of GFFKR motif that brings its Phe residues into the interface with the β_3 subunit and partially embeds them in the hydrophobic portion of the membrane bilayer. Furthermore, a Lys residue immediately following the β_3 TM hydrophobic segment also participates in the interface with α_{IIb}. These key structural findings are all supported by mutational studies. Our study demonstrates the importance of juxtamembrane segments in dimerization. Previous homodimeric NMR structures of associating TM domains (Bocharov et al., 2008a, 2008b; Call et al., 2006; MacKenzie et al., 1997; Smith et al., 2001) show different packing modes between two TM α helices but do not contain additional segments that contribute to packing within the membrane. JM amphipathic segments intervene between TM domains and intracellular signaling domains in many receptors, including the epidermal growth factor receptor, and thus may be of key importance in signal transduction (McLaughlin et al., 2005). The demonstration here that the integrin α subunit GFFKR motif is an amphipathic, JM segment that directly participates in the association between the α and β TM domains raises the intriguing possibility that amphipathic JM segments in other receptors may participate in dimerization interactions between their TM domains.

Careful studies by Ulmer et al. have described the NMR structures of isolated α_{IIb} and β_3 TM/cytoplasmic domain fragments in bicelles (Lau et al., 2008a, 2008b). In contrast to findings that similar α_{IIb} and β_3 constructs formed homomultimers in detergents (Li et al., 2001), Ulmer and coworkers found that these subunits were monomeric in the more bilayer-like bicelle environment. Comparisons between the isolated peptide and membrane complex structures show a remarkable similarity in the α_{IIb} structures in the GFFKR motif (Figures 6B–6D); it should be noted that no information from the monomer structures was used in the calculation of the $\alpha_{IIb}\beta_3$ complex structure. The backbone conformation is identical through the GFFKR motif to Lys-994, which is the last structured residue in the bicelle α_{IIb} structure (Lau et al., 2008a). The same hydrogen bonds that support the turn at Gly-991, between the backbone O atom of Met-987 and N atoms of Gly-991 and Phe-992, are found in both structures and, in general, in the C-cap motif with Gly in left-handed helical conformation (Richardson and Richardson, 1989). Moreover, Phe-993, which is central in the $\alpha_{IIb}\beta_3$ complex structure and has a well-defined rotamer in the α_{IIb} structure, has the same rotamer in the complex structure (Figures 6B–6D).

The isolated α_{IIb} and β_3 TM domain structures provide the basis for comparisons to our complex structure. There is no precedent for association between TM monomers, one of which has a significant juxtamembrane segment that associates with the α-helical TM segment. It was unknown whether the α_{IIb} juxtamembrane segment would alter conformation after association. Indeed, there was good reason to expect that the conformation of the α_{IIb} GFFKR segment would alter after association with β_3 because, as pointed out (Lau et al., 2008a), residues in this motif were found to be in α-helical conformation in an $\alpha_{IIb}\beta_3$ complex NMR structure containing the entire GFFKR

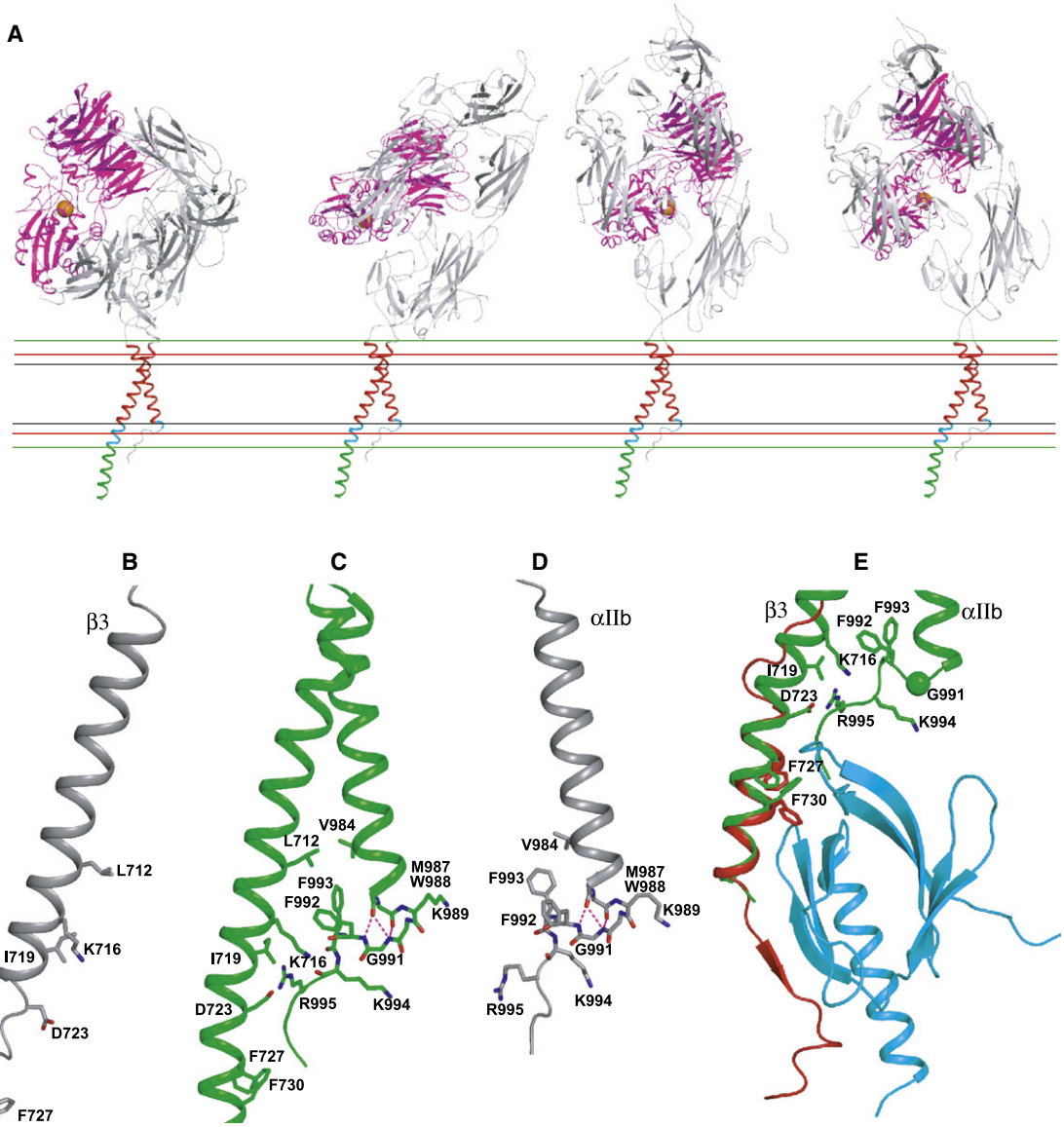


Figure 6. Structure in the Membrane and Comparison to Isolated α_{IIb} and β_3 TM Domains and β_3 Talin Complex

(A) Low-energy orientations of the $\alpha_{IIb}\beta_3$ ectodomain on the cell surface. Four favorable cell-surface orientations of intact $\alpha_{IIb}\beta_3$ are shown, superimposed on the TM domains. The ligand-binding α_{IIb} β propeller and β_3 I domains are shown in magenta, and ligand-binding I domain Mg^{2+} ion is shown as an orange sphere. The outer bounds of the hydrophobic, interface, and polar regions of the membrane are shown as black, red, and green lines, respectively.

(B–D) The NMR bicelle structures of β_3 (Lau et al., 2008b) (B) and α_{IIb} (Lau et al., 2008a) (D) are shown in the same orientation as the subunits of the $\alpha_{IIb}\beta_3$ complex (C).

(E) Superposition on the $\alpha_{IIb}\beta_3$ complex of the NMR structure of β3 integrin cytoplasmic tail fragment in complex with talin F3 domain (Wegener et al., 2007). The integrin heterodimer structure is in green. The talin F3 domain is in cyan, and the β3 cytoplasmic tail fragment of the NMR structure is in red. The side chains of the two phenylalanine residues (F727 and F730) are shown as sticks.

motif and two preceding residues (Vinogradova et al., 2002). Thus, it is surprising that the only significant difference between the structured portions of the peptide and complex structures is a difference in tilt between the TM and JM segments. This tilt brings Phe-993 in monomeric α_{IIb} 2 Å closer to Val-984 to pack against it in van der Waals contact (Figure 6D). In the $\alpha_{IIb}\beta_3$ complex, β_3 Leu-712 contacts α_{IIb} Val-984, and there is not room for Phe-993 to come as close (Figure 6C). Thus, some conformational adjustments must occur between the associated and dissociated states of the α_{IIb} subunit.

Another interesting juxtamembrane feature is the side chain-backbone intersubunit hydrogen bond donated by the β_3 Lys-716 side chain. A caveat is that this hydrogen bond appears in only 65% of structures in the ensemble. However, the presence of such a bond is strongly supported by the ability of Arg, but not other amino acids, to substitute for Lys. This bond is to the α_{IIb}

backbone near the end of the GFFKR motif and thus stabilizes both α β association and GFFKR conformation. In our structure, Lys-716 is by far the most important β_3 residue for interaction with α_{IIb} in the JM region, and it is fitting that β_3 Lys-716 and the α_{IIb} GFFKR motif, which interact with one another, are the sites where mutations are most disruptive to α β interaction. Arg and Lys are abundant at the boundary between TM and cytoplasmic domains, where they determine the polarity of membrane insertion by the positive-inside rule (von Heijne, 1992); thus, though Lys and Arg are of general importance, a specific role for β_3 Lys-716 was unanticipated.

Comparison to Previous Complex Structures

Surprisingly, our results disagree with a prominent NMR structure of the $\alpha_{IIb}\beta_3$ JM/cytoplasmic complex (Vinogradova et al., 2002). There is no similarity whatsoever in the interfaces between the NMR complex structure in aqueous media and our cell-surface complex structure (Figure S13A). The previous comprehensive study showed that mutations that activate integrins disrupted the $\alpha_{IIb}\beta_3$ complex and that the talin head domain bound to the β_3 JM/cytoplasmic fragment and disrupted its complex with the α_{IIb} JM/cytoplasmic fragment (Vinogradova et al., 2002). Other NMR studies have yielded conflicting results. Refinement by one group did not converge on a single NMR structure but on two different structures (Weljie et al., 2002), neither of which is in agreement with that of Vinogradova et al. (2002) or described here (Figures S13B and S13C). Furthermore, the same α_{IIb} JM/cytoplasmic fragments that associated in aqueous media were found not to associate in the presence of dodecylphosphocholine (Vinogradova et al., 2004). Moreover, a lack of complex formation has been reported even with proximity enforced by fusion of α_{IIb} and β_3 JM/cytoplasmic fragments to coiled coils (Ulmer et al., 2001).

In retrospect, perhaps it is not surprising that none of the NMR structures resemble the Disulfide/Rosetta structure (Figure S13). The two Phe residues of the GFFKR motif pack in the hydrophobic lipid environment and in an interface against the side chains of the α and β TM helices, all of which are missing in the discrepant structures. The lack of appropriate structure of the juxtamembrane/cytoplasmic segments when isolated from the TM segments and the lack of association between the α and β TM segments in intact cells when the GFFKR motif is deleted (Luo et al., 2004) demonstrate cooperativity between these segments for folding and assembly, consistent with their intimate interaction in the Disulfide/Rosetta structure.

Integrin Structure on the Cell Surface

Using a crystal structure of the $\alpha_{IIb}\beta_3$ extracellular domain (Zhu et al., 2008), the $\alpha_{IIb}\beta_3$ membrane structure, and disulfide restraints on the short extracellular linkers, we characterized the structure of intact $\alpha_{IIb}\beta_3$ on the cell surface. Multiple low-energy orientations between the extracellular and transmembrane domains were found. This finding suggests that the orientation of integrins on the cell surface is dynamic rather than fixed and that the 6 residue α_{IIb} and 4 residue β_3 linkers between the extracellular and TM domains are flexible. The lack of a fixed structure for this region is consistent with marked variation in the sequence and length of these linkers, particularly among integrin α subunits. For example, the α_V subunit, which also associates with β_3, has an 8 residue QPAPMPVP linker compared to the 6 residue EERAIP linker in α_{IIb}.

In contrast to our results, a cryo-electron microscopy (cryoEM) structure of native $\alpha_{IIb}\beta_3$ in octylglucoside detergent suggested a defined, extended orientation between the extracellular and TM domains (Adair and Yeager, 2002). However, the cryoEM extracellular domain structure does not agree with that determined in crystals or negative stain (Zhu et al., 2008), and it has been suggested that the cryoEM structure may have averaged over bent and extended integrin conformations (Luo et al., 2007).

Our finding that limited, but not extensive, flexibility between the extracellular and TM domain is compatible with signal transduction has important implications. Flexibility and variation in the ectodomain-TM domain linkers are compatible with activation by large-scale movements such as separation in the plane of the membrane between the TM domains (Luo et al., 2004), but not with small-scale twisting or pistoning motions with which the linkers could comply. On the other hand, extreme flexibility in the linkers decouples extracellular domain activation from the TM domains, as shown here by the $\alpha_{IIb}\beta_3$ /GPA 60–131 chimera and previously by insertion of artificial linkers (Xiong et al., 2003). Decoupling likely arises because long, flexible linkers enable integrin ectodomain extension and activation even when the TM domains remain associated.

Inside-Out Activation of Integrins

Ligand binding by integrins can be activated by binding of the actin cytoskeleton to the β cytoplasmic domain. Two classes of cytoskeleton-associated proteins mediate this linkage, talin and kindlins (Moser et al., 2008; Wegener and Campbell, 2008). Talin and kindlin bind through their FERM domains to NPXY motifs at β_3 residues 744–747 and 756–759, respectively. Talin also binds to more membrane-proximal β_3 regions, including Phe-727 and Phe-730 (Wegener et al., 2007). The Disulfide/Rosetta structure shows that these residues are in the β_3 α helix and face the α_{IIb} cytoplasmic tail. Superposition of the talin β_3 and $\alpha_{IIb}\beta_3$ complexes demonstrates that talin clashes with α_{IIb} residues that interact with the β_3 α helix and immediately follow the GFFKR motif, i.e., α_{IIb} residues N996–P998 (Figure 6E). However, the position of the segment with talin-binding residues Phe-727 and Phe-730 is not well defined, given that some clusters show two helices in the β_3 cytoplasmic domain separated by a turn near Glu-726 (Figures S7A and S7B). Furthermore, the talin complex structure shows a loss of helicity in β_3 N terminal to His-722, and isolated β_3 in bicelles loses α helicity after residue His-722 (Lau et al., 2008b). Thus, it is plausible either that talin binding to β_3 would perturb its association with α_{IIb} or that flexibility in the β_3 α helix between its sites of association with α_{IIb} and talin would eliminate clashes in a talin/$\alpha_{IIb}\beta_3$ complex. Because the evidence that talin interferes with association between α_{IIb} and β_3 comes from an NMR study on a complex that our results suggest is nonphysiologic, the mechanism by which talin activates integrins requires further structural study.

Once integrins bind through talins or kindlins to actin filaments, lateral force exerted by the cytoskeleton will favor an extended integrin conformation with the hybrid domain in β swung out in

the direction of the pulling force in the open conformation with high affinity for ligand (Zhu et al., 2008). This mechanism can explain activation by talin as well as by kindlins, which bind to an NPXY motif that is distal from the α_{IIb}/β_3 interface (Moser et al., 2008). Lateral force would tilt the dissociated β_3 TM domain in the plane of the membrane. The long continuous α helix seen in the dissociated β_3 TM domain is well suited to tilting (Lau et al., 2008b).

We have described the structure of an intact receptor containing two TM domains on the cell surface. The results shed important light on the structural basis for transmembrane signaling through integrins, demonstrate that the principles differ for association between heterodimeric and homodimeric TM domains, and demonstrate that JM domains can have an important role in TM domain association. The method can be extended to many other important classes of cell-surface receptors with two associating TM domains and can yield information on how regulated changes in the association state of the TM domains transmit signals between the extracellular and intracellular environments.

EXPERIMENTAL PROCEDURES

Disulfide Crosslinking and Immunoprecipitation

The 293T transfectants were pretreated with 15 μg/ml of 2-BP for 1 hr, labeled with [^{35}S] cysteine/methionine (10 mCi/ml, PerkinElmer Life Science) for 1.5 hr, and chased for 17 hr. Cells were detached and kept intact or broken by freeze-thaw, treated on ice with 200 μM $CuSO_4$/1000 μM o-phenanthroline for 10 min, treated with 10 mM N-ethylmaleimide for 10 min, and lysed in 2% Triton X-100. Lysate supernatants were immunoprecipitated with anti-β3 mAb AP3 and protein G agarose and subjected to nonreducing 7.5% SDS-PAGE. Radioactivity was quantitated with a PhosphorImager. Crosslinking efficiency was quantitated as disulfide-linked heterodimer as percentage of heterodimer plus α_{IIb} and β_3 monomers. For constitutively crosslinked residues, crosslinking was also measured after 1 hr in 5 mM cysteamine/1 mM cystamine redox buffer and after DTT treatment followed by Cu-phenanthroline. Details are in the Supplemental Data.

Membrane Structure Generation

The energy terms used in Membrane Rosetta (Barth et al., 2007, 2009) include a Lennard-Jones potential, a backbone torsional term, a knowledge-based pairwise interaction term between amino acid residues, a penalty term for the disulfide restraints that is proportional to distance beyond the upper bound, an implicit Lazaridis-Karplus solvation term for the aqueous and hydrophobic environments (with the hydrophobic potential based on experimental transfer-free energies from water to cyclohexane), and an orientation-dependent hydrogen-bonding term. Models are first built with centroids for side chains and are then refined in all-atom mode. The all-atom mode membrane environment has an inner, slab-shaped, 24 Å thick, isotropic, hydrophobic phase representing the lipid alkyl groups, outer isotropic phases representing the head group and membrane-external regions, and, in between, two 10 Å thick, anisotropic phases representing the interface region where alkyl, acyl, and glycerol groups are located. The centroid membrane environment is similar but replaces each anisotropic slab with two isotropic slabs. The solvation and hydrogen bond potentials differ in the hydrophobic and aqueous environments and are interpolated in the interface region. The hydrogen bond potential includes weak CαH-O bonds and bifurcated hydrogen bonds. The position of models in the membrane slab is varied to find the lowest-energy embedding.

Structures were built in several stages. First, 12 residue ideal helices were docked. Then, all but two residues of each helix were discarded, and chain growing with fragment assembly was used to build longer structures that could, in principle, have any secondary structure. A total of two (GPA) or three (integrin) chain-growing stages were used, alternating in N-to-C and C-to-N direction, and previously built residues were discarded in each stage so that no remnants from the first one (GPA) or two stages (integrin) were present in the final structures (Figure S6). Approximately 10,000 models were generated in each stage. Disulfide restraints were used as part of the total energy function in each Monte Carlo step within each chain-growing stage. Furthermore, between each stage, the 100 models used as seeds for the next stage were selected by the independent criteria of low restraint violation and low energy (in this case, energy did not include the restraint violation penalty). The stages are summarized in more detail in Figure S6 and Supplemental Experimental Procedures. Structures are provided in the Supplemental Data.

Modeling the Intact Integrin in the Membrane Bilayer

To provide starting TM structures, we performed one additional stage of chain growing from the TM regions to grow the linker regions, using redox buffer disulfide restraints for the linker region and distance restraints for the two C-terminal residues present in α_{IIb} and β_3 in the crystal structure (Zhu et al., 2008). After joining these TM structures to the crystal structure, the two linkers were simultaneously remodeled with Membrane Rosetta using the linker disulfide restraints to find low-energy orientations with respect to the membrane (Supplemental Data). Two long, membrane-proximal loops in the calf-2 domain that are disordered in the crystal structure were included in the models.

Ligand-Binding Assay

Ligand mimetic IgM PAC-1 (Becton Dickinson, San Jose, CA) and FITC-labeled human fibrinogen (Enzyme Research Laboratories, South Bend, IN) binding to transfected cells was as described (Luo et al., 2004).

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, 13 figures, 5 tables, and 3 sets of structure coordinates and can be found with this article online at http://www.cell.com/molecular-cell/supplemental/S1097-2765(09)00138-5.

ACKNOWLEDGMENTS

We thank Junichi Takagi and Guo-hui Li for early help with the project. This work was supported by NIH grant HL-48675 and HHMI.

Received: October 28, 2008
Revised: December 23, 2008
Accepted: February 18, 2009
Published: April 23, 2009

REFERENCES

Adair, B.D., and Yeager, M. (2002). Three-dimensional model of the human platelet integrin αIIbβ3 based on electron cryomicroscopy and x-ray crystallography. Proc. Natl. Acad. Sci. USA *99*, 14059–14064.

Barth, P., Schonbrun, J., and Baker, D. (2007). Toward high-resolution prediction and design of transmembrane helical protein structures. Proc. Natl. Acad. Sci. USA *104*, 15682–15687.

Barth, P., Wallner, B., and Baker, D. (2009). Prediction of membrane protein structures with complex topologies using limited constraints. Proc. Natl. Acad. Sci. USA *106*, 1409–1414.

Bass, R.B., Butler, S.L., Chervitz, S.A., Gloor, S.L., and Falke, J.J. (2007). Use of site-directed cysteine and disulfide chemistry to probe protein structure and dynamics: applications to soluble and transmembrane receptors of bacterial chemotaxis. Methods Enzymol. *423*, 25–51.

Bocharov, E.V., Mayzel, M.L., Volynsky, P.E., Goncharuk, M.V., Ermolyuk, Y.S., Schulga, A.A., Artemenko, E.O., Efremov, R.G., and Arseniev, A.S. (2008a). Spatial structure and pH-dependent conformational diversity of dimeric transmembrane domain of the receptor tyrosine kinase EphA1. J. Biol. Chem. *283*, 29385–29395.

Bocharov, E.V., Mineev, K.S., Volynsky, P.E., Ermolyuk, Y.S., Tkach, E.N., Sobol, A.G., Chupin, V.V., Kirpichnikov, M.P., Efremov, R.G., and Arseniev, A.S. (2008b). Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. J. Biol. Chem. *283*, 6950–6956.

Brunger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R., and Nilges, M. (1993). Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. Science *261*, 328–331.

Call, M.E., Schnell, J.R., Xu, C., Lutz, R.A., Chou, J.J., and Wucherpfennig, K.W. (2006). The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor. Cell *127*, 355–368.

Chothia, C., Levitt, M., and Richardson, D. (1981). Helix to helix packing in proteins. J. Mol. Biol. *145*, 215–250.

Das, R., and Baker, D. (2008). Macromolecular modeling with rosetta. Annu. Rev. Biochem. *77*, 363–382.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., III, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, W375-W383.

Duneau, J.P., Vegh, A.P., and Sturgis, J.N. (2007). A dimerization hierarchy in the transmembrane domains of the HER receptor family. Biochemistry *46*, 2010–2019.

Gottschalk, K.E. (2005). A coiled-coil structure of the αIIbß3 integrin transmembrane and cytoplasmic domains in its resting state. Structure *13*, 703–712.

Hughes, P.E., Diaz-Gonzalez, F., Leong, L., Wu, C., McDonald, J.A., Shattil, S.J., and Ginsberg, M.H. (1996). Breaking the integrin hinge. J. Biol. Chem. *271*, 6571–6574.

Kovalenko, O.V., Yang, X., Kolesnikova, T.V., and Hemler, M.E. (2004). Evidence for specific tetraspanin homodimers: Inhibition of palmitoylation makes cysteine residues available for cross-linking. Biochem. J. *377*, 407–417.

Lau, T.L., Dua, V., and Ulmer, T.S. (2008a). Structure of the integrin alphaIIb transmembrane segment. J. Biol. Chem. *283*, 16162–16168.

Lau, T.L., Partridge, A.W., Ginsberg, M.H., and Ulmer, T.S. (2008b). Structure of the integrin beta3 transmembrane segment in phospholipid bicelles and detergent micelles. Biochemistry *47*, 4008–4016.

Li, R., Babu, C.R., Lear, J.D., Wand, A.J., Bennett, J.S., and Degrado, W.F. (2001). Oligomerization of the integrin aIIbb3: Roles of the transmembrane and cytoplasmic domains. Proc. Natl. Acad. Sci. USA *98*, 12462–12467.

Li, W., Metcalf, D.G., Gorelik, R., Li, R., Mitra, N., Nanda, V., Law, P.B., Lear, J.D., Degrado, W.F., and Bennett, J.S. (2005). A push-pull mechanism for regulating integrin function. Proc. Natl. Acad. Sci. USA *102*, 1424–1429.

Liu, W., Crocker, E., Siminovitch, D.J., and Smith, S.O. (2003). Role of side-chain conformational entropy in transmembrane helix dimerization of glycophorin A. Biophys. J. *84*, 1263–1271.

Luo, B.H., Springer, T.A., and Takagi, J. (2004). A specific interface between integrin transmembrane helices and affinity for ligand. PLoS Biol. *2*, e153.

Luo, B.H., Carman, C.V., Takagi, J., and Springer, T.A. (2005). Disrupting integrin transmembrane domain heterodimerization increases ligand binding affinity, not valency or clustering. Proc. Natl. Acad. Sci. USA *102*, 3679–3684.

Luo, B.H., Carman, C.V., and Springer, T.A. (2007). Structural basis of integrin regulation and signaling. Annu. Rev. Immunol. *25*, 619–647.

MacKenzie, K.R., Prestegard, J.H., and Engelman, D.M. (1997). A transmembrane helix dimer: Structure and implications. Science *276*, 131–133.

McLaughlin, S., Smith, S.O., Hayman, M.J., and Murray, D. (2005). An electrostatic engine model for autoinhibition and activation of the epidermal growth factor receptor (EGFR/ErbB) family. J. Gen. Physiol. *126*, 41–53.

Moser, M., Nieswandt, B., Ussar, S., Pozgajova, M., and Fassler, R. (2008). Kindlin-3 is essential for integrin activation and platelet aggregation. Nat. Med. *14*, 325–330.

O'Toole, T.E., Katagiri, Y., Faull, R.J., Peter, K., Tamura, R., Quaranta, V., Loftus, J.C., Shattil, S.J., and Ginsberg, M.H. (1994). Integrin cytoplasmic domains mediate inside-out signal transduction. J. Cell Biol. *124*, 1047–1059.

Partridge, A.W., Liu, S., Kim, S., Bowie, J.U., and Ginsberg, M.H. (2005). Transmembrane domain helix packing stabilizes integrin aIIb3 in the low affinity state. J. Biol. Chem. *280*, 7294–7300.

Richardson, J.S., and Richardson, D.C. (1989). Principles and Patterns of Protein Conformation. In Prediction of Protein Structure and the Principles of Protein Conformation, G.D. Fasman, ed. (New York: Plenum Press), pp. 1–98.

Smith, S.O., Song, D., Shekar, S., Groesbeek, M., Ziliox, M., and Aimoto, S. (2001). Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. Biochemistry *40*, 6553–6558.

Ulmer, T.S., Yaspan, B., Ginsberg, M.H., and Campbell, I.D. (2001). NMR analysis of structure and dynamics of the cytosolic tails of integrin alpha IIb beta 3 in aqueous solution. Biochemistry *40*, 7498–7508.

Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T.A., Plow, E.F., and Qin, J. (2002). A structural mechanism of integrin $a_{IIb}b_3$ "inside-out" activation as regulated by its cytoplasmic face. Cell *110*, 587–597.

Vinogradova, O., Vaynberg, J., Kong, X., Haas, T.A., Plow, E.F., and Qin, J. (2004). Membrane-mediated structural transitions at the cytoplasmic face during integrin activation. Proc. Natl. Acad. Sci. USA *101*, 4094–4099.

von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. *225*, 487–494.

Wegener, K.L., Partridge, A.W., Han, J., Pickford, A.R., Liddington, R.C., Ginsberg, M.H., and Campbell, I.D. (2007). Structural basis of integrin activation by talin. Cell *128*, 171–182.

Wegener, K.L., and Campbell, I.D. (2008). Transmembrane and cytoplasmic domains in integrin activation and protein-protein interactions. Mol. Membr. Biol. *25*, 376–387.

Weljie, A.M., Hwang, P.M., and Vogel, H.J. (2002). Solution structures of the cytoplasmic tail complex from platelet αIIb- and β3-subunits. Proc. Natl. Acad. Sci. USA *99*, 5878–5883.

Xiong, Y.M., Chen, J., and Zhang, L. (2003). Modulation of CD11b/CD18 adhesive activity by its extracellular, membrane-proximal regions. J. Immunol. *171*, 1042–1050.

Zhang, X., Gureasko, J., Shen, K., Cole, P.A., and Kuriyan, J. (2006). An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell *125*, 1137–1149.

Zhu, J., Luo, B.H., Xiao, T., Zhang, C., Nishida, N., and Springer, T.A. (2008). Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. Mol. Cell *32*, 849–861.

Note Added in Proof

A structure similar to that reported here has been obtained for the complex between α_{IIb} and β_3 transmembrane/cytoplasmic domain peptides in bicelles: Lau, T.L., Kim, C., Ginsberg, M.H., and Ulmer, T.S. (2009). The structure of the integrin $\alpha_{IIb}\beta_3$ transmembrane complex explains integrin transmembrane signalling. EMBO J. Published online March 12, 2009. 10.1038/emboj.2009.

Supplementary Information

Plasmid construction and transient transfection

Plasmids encoding full-length human α_{IIb} and β_3 were subcloned into pEF/V5-HisA and pcDNA3.1/Myc-His(+), respectively (Takagi et al., 2002). To make the integrin/glycophorin A chimeras, the transmembrane and cytoplasmic domains of the α_{IIb} and β_3 subunit were replaced with human glycophorin A sequences, E60-Q131 or P71-Q131. Single amino acid substitutions to cysteine were made using site-directed mutagenesis with the QuikChange Kit (Stratagene, La Jolla, CA). The wild type or mutant constructs were transfected into 293T cells using Fugene (Roche Diagnostics, Indianapolis, Indiana).

Ligand binding assay

Ligand mimetic IgM PAC-1 (Becton Dickinson, San Jose, CA) and FITC-labeled human fibrinogen (Enzyme Research Laboratories, South Bend, IN) binding to transfected cells was determined as described (Luo et al., 2004). In brief, 293T cell transfectants were incubated with 10 μg/ml PAC-1 or 50 μg/ml FITC-labeled fibrinogen at room temperature for 30 minutes in Hepes-buffered saline (20 mM Hepes, pH 7.4, 150 mM NaCl, 5.5 mM glucose, and 1% bovine serum albumin) containing 5 mM EDTA, or 1 mM Ca^{2+}/Mg^{2+}, or 1mM/Mn^{2+} plus 10 μg/ml αIIbβ3 specific activating mAb PT25-2. Then cells were incubated with 10 μg/ml of Cy3-labeled mAb AP3 on ice for 30 minutes before subjected to flow cytometry. For PAC-1 binding, cells were incubated with 10 μg/ml of Cy3-labled AP3 and 10 μg/ml of FITC-labeled goat anti-mouse IgM at the same

time. Ligand binding ability was expressed as the percentage of mean fluorescence intensity (MFI) of fibrinogen or PAC-1 binding after subtracting the MFI of fibrinogen or PAC-1 binding in EDTA condition relative to the MFI of AP3 binding.

Disulfide crosslinking and immunoprecipitation

Twenty-four hours after transfection, 293T cells in 12-well plates with 1.5 ml DMEM medium containing 10% FCS were pre-treated with 15 μg/ml of 2-BP for 1 hour, the medium was replaced with 0.75 ml Met, Cys-free RPMI1640 (Sigma R-7513), supplemented with 10% dialyzed FCS, 10 μl [^{35}S] cysteine/methionine (10mCi/ml, PerkinElmer Life Science), 15 μg/ml 2-BP. After 1.5 h at 37°C, 0.75 ml of RPMI1640 containing 10% FCS, 500 μg/ml cysteine, 100 μg/ml methionine, and 15 μg/ml 2-BP was added, and cells chased for at least 17 hours. Cells were detached by vigorous pipetting, washed, and suspended (10^6 cells in 100 μl) in Tris-buffered saline (TBS, 20 mM Tris-HCl, pH 7.5, 150 mM NaCl) containing 1 mM Ca^{2+}/1 mM Mg^{2+} and proteinase inhibitors (1 μg/ml each aprotinin, leupeptin, and pepstatin). The cells were kept intact or broken by 3 cycles of freezing on dry ice and thawing. Saponin (40 μg/ml) gave results identical to freeze-thawing, but freeze-thawing was adapted as the least membrane-perturbing. After chilling on ice for 5 minutes, 200 μM $CuSO_4$/1000 μM o-phenanthroline was added by 10 fold dilution from stock solution, and cells were incubated on ice for another 10 minutes. N-ethylmaleimide (10 mM) was added and after 10 minutes on ice, cells were lysed with an equal volume of TBS containing 2% Triton X-100 and 0.1% NP-40 for 10 minutes on ice. Cell lysates were cleared by centrifugation at 14000 RPM for 10 minutes and immunoprecipitated with anti-β3 mAb AP3 and protein G agarose at 4°C for 1 hour (Luo et al., 2004). The precipitated proteins were subjected to nonreducing 7.5% SDS-

PAGE. The SDS-PAGE gel was dried and exposed for 3 h to storage phosphor screens, which were measured with a Storm PhosphorImager (Molecular Dynamics, Sunnyvale, California, United States). Disulfide bond formation was quantitated as the intensity of the disulfide-bonded heterodimer band divided by the sum of the intensity of α_{IIb}, β_3, and heterodimer bands. Specific intensity of each band was determined by subtraction of background intensity.

For constitutively crosslinked extracellular and exofacial residues, crosslinking was also measured in redox buffer and after DTT treatment followed by Cu-phenanthroline. For redox buffer treatment, cells were suspended in pH 8.2 TBS containing 1 mM Ca^{2+}/1 mM Mg^{2+} and 5 mM cysteamine/1 mM cystamine, and incubated at 37 °C for 1 hour. Following addition of 10 mM N-ethylmaleimide, cells were lysed and immunoprecipitated as described above. For DTT treatment, cells were incubated in pH 8.2 TBS containing 1 mM Ca^{2+}/1 mM Mg^{2+} and 10 mM DTT for 10 min at 37 °C, washed three times with TBS containing 1 mM Ca^{2+}/1 mM Mg^{2+}, and broken by freeze/thaw and treated with Cu-phenanthroline as described above.

Constitutive crosslinking for GPA residues 73-78 showed only slight periodicity, whereas redox buffer and DTT/Cu-phenanthroline gave similar, periodic peaks (Fig. S1A). Furthermore, the crosslinking results for residues that were dependent on Cu-phenanthroline, such as 79-80, were similar with or without DTT pretreatment (Fig. S1A). Therefore, results with DTT/Cu-phenanthroline were used for restraint calculation for all crosslinks involving GPA residues 73-78 and are shown in Fig. 2. In contrast, exofacial residues in integrins showed good periodicity under constitutive conditions

(Fig. 2), and we did not use data with reducing agents (Fig. S1B) in restraint calculation because DTT can activate integrins (Peerschke, 1995).

Disulfide-based distance constraints

Crosslinking data and restraints are in Supplementary Materials. To minimize effect of flexibility or cysteine substitution on crosslinking, all values were subjected to near-valley correction before restraint calculation. The % crosslinking of residues i with j was subtracted by the higher of ((the lowest % crosslinking found for residue i with residues j ±2) or (the lowest % crosslinking found for residue j with residues i ±2)). This reduced broadening of crosslinking patterns (Fig. S5). Only crosslinking efficiencies equal to or larger than 20% were used for the calculation of upper boundary C_α-C_α distances. The distance constraints for GPA were symmetrized (Table S4).

Membrane Structure Generation

Structures were generated in three or four stages as shown in Fig. S6. In stage 1, 23-residue TM sequences were divided into three 12-residue, overlapping segments, denoted N-terminal (N) (α_{IIb} I966-L977, β_3 I693-I704, GPA I73-V84), middle (M) (α_{IIb} G972-L983, β_3 S699-A710, GPA G79-L90) and C-terminal (C) (α_{IIb} L977-W988, β_3 I704-W715, GPA V84-I95) (Fig. S6). The 12-residue N α_{IIb} with N β_3, M α_{IIb} with M β_3, etc. ideal helices were docked with an all-atom membrane force field (Barth et al., 2007) with the following command line: *Rosetta.gcc gd 1pdb chain i.d. –spanners 1pdb.span –fake_native –membrane –dock –dock_mcm –randomize1 –randomize2 –ex1 –ex2 –s*

1pdb.pdb –nstruct 10000 –paths paths.txt. This docking protocol samples backbone rigid-body and all side-chain conformational degrees of freedom. The 1,000 docked models with lowest Rosetta all-atom energy were selected from 10,000 total decoys and pairwise disulfide-based constraint violations were calculated. 100 models with low pairwise constraints violation (in practice all 100 had no violations) were selected as the starting structures for stage 2. For integrins, both N and M helices yielded >100 docked structures with no restraint violations. Very similar 46-residue integrin TM segments (1.2 Å RMSD) were obtained after stage 3 starting with either the N or M 12-residue helices. The C 12-residue helix was not tested because of weaker disulfide restraints in this region. For GPA, 12-residue helix docking with the N, M, and C segments gave for the 100 models with lowest restraint violation, violation scores of 0, 3.96±2.21, and 3.45±0.71, respectively. Therefore, only the N segment was used for GPA structure generation.

In subsequent "chain growing" stages, the backbone coordinates of two or more residues were kept from the previous stage and remaining residues were rebuilt from N-terminal to C-terminal or from C-terminal to N-terminal at low-resolution by Monte-Carlo-based peptide fragment insertion and then refined at all-atom using disulfide distance constraints in both low resolution and all atom potentials (Barth et al., 2007). After use of the integrin M-helices in stage 1, residues L974-W988, M701-W715 (stage 2), L978-L959 and L705-P688 (stage 3), and L970-P998 and L697-A737 (stage 4) were grown in α_{IIb} and β_3, respectively. GPA 12-residue N-helices were I73-V84 in stage 1, and in stages 2 and 3, residues A82-I73 and G79-I95 were grown, respectively.

Fragments used in fragment assembly were generated as described (Rohl et al., 2004), except that only SAM-T02 (Karplus et al., 2003) secondary structure prediction method was used during fragment selection. The "chain growing" stage was performed with the following command line: *Rosetta.gcc lg 1pdb chain i.d. –fake_native –membrane –no_filters –loops –fold –l pdblist –nstruct 10000 –paths paths.txt –hb_srbb_reweight 0.0 –hb_lrbb_reweight 0.0 –pc_reweight 1.0 –FA_pc_weight 1.0 –short_rang_hb_weight 0.5 –use_fold_constrints_no_minimize –minimize_exclude_helix –fa_refine –ex1 –ex2 –exlaro –extrachi_cutoff 0 –thickness 15 –steepness 10 –mem_solv –memb_hb –mem_env –Wmbenv 0.462289.* In each stage, 10,000 models were generated (100 decoys for each of the 100 starting structures) and 100 low energy, low constraint violation models were selected as starting structures for the following stage.

The final 10% lowest energy models were clustered based on Cα-RMSD (Bonneau et al., 2002). Clustering was robust to the RMSD cutoff used. In clustering with 1.5, 1.8, 2.0, 2.5, and 3.0 Å cutoffs, similar top (most populous) clusters were found, and the structures at the center of the 1.8, 2.0, 2.5 and 3.0 Å top clusters were identical (Fig. S7). Clustering used TM+JM+CT α_{IIb} I966-P998 and β_3 I693-E731 residues or GPA TM residues 73-95, and results are summarized in Fig. S6B,D.

For comparison, the coordinates for the GPA solid state NMR structure (Smith et al., 2001) were kindly provided by Dr. S. O. Smith, SUNY Stony Brook, Stony Brook NY. Coordinates for previous integrin models were provided by Dr. K. Gottschalk, Ludwig-Maximilians U., München, Germany (Gottschalk, 2005), and Dr. J. Bowie, UCLA, Los Angeles, CA (Partridge et al., 2005).

Cross-validation by restraint omission

Restraints were randomly partioned into approximately equal subsets (before symmetrization for GPA) of 1/8, 1/4, or 1/2 of restraints. These smaller subsets of 12.5%, 25%, or 50% of restraints, or no restraints, were used in all stages of structure generation, exactly as when all restraints were used. At least two groups of each size of restraint set were used in structure generation. In each run of structure generation, the final 10% lowest energy structures were clustered at a 3.0 Å cutoff. The central structure in the largest cluster was examined by superposition to the GPA NMR model or GPA or integrin models made with complete restraints to calculate RMSD. Models were also scored for violation of the omitted restraints. RMSD and restraint violations were calculated over the TM segment (46 residues) or TM+cytoplasmic segments (78 residues). The root mean square distance violation (above the upper bound) per restraint was calculated for all omitted restraints (whether violated or not).

Modeling of the full-length integrin receptor in the context of the membrane bilayer

A modeling protocol was developed to assemble different domains in a membrane environment, to find low-energy conformations for the linkers consistent with disulfide crosslinking data, and to optimize the rigid-body orientations of each domain with respect to the membrane. The full-length integrin receptor structure was modeled in four stages.

In the first stage, the structures of A958-W968 of α_{IIb} and E686-V695 of β_3 were rebuilt by chain growing protocol from C to N with the structures of the TM+cytoplasmic domains in the largest cluster (52 structures, Figure 4A and Table S6B) as initial stubs.

Redox buffer disulfide restraints were used for the residues L959-P695 of α_{IIb} and P688-D692 of β_3 (Table S3B). These restraints were validated based on crosslinking between nearby residues defined in $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$ crystal structures; because of flexibility in crystal structures of the β_3 tail domain (Zhu et al., 2008), restraints were loosened relative to the TM region (Fig. S12). Additional restraints were added from the $\alpha_{IIb}\beta_3$ crystal structure. Cα-CαCβ-Cβ ˜ ..C̃Cand O-O atom distances between residues α_{IIb} A958 and β_3 E686 and between residues α_{IIb} L959 and β_3 C687 were used. The upper and lower bound distance restraints were obtained by adding 2 Å and subtracting 1 Å, respectively, from distances in the $\alpha_{IIb}\beta_3$ crystal structure (Zhu et al., 2008). All the restraints for TM+cytoplasmic residues were also kept. 100 models were generated for each starting structure. The 10% lowest energy models were selected and clustered (RMSD cutoff of 2 Å) based on the structures of the rebuilt residues. The center models of top 10 clusters were used for the next stage in full length model generation.

In a second stage, the crystal structure was joined to the linker segments. The lowest energy cluster center models of the TM and cytoplasmic (TM+CT) regions were assembled to the crystal structure of the ectodomains (chains A and B) with one extended linker connecting the domains in chain A (command line used: *rosetta.intel -assemble -regions -extend -s A2B3_xtal_tmhAB.pdb -nstruct 1*). This step involved selecting residues on chain A (residues 958 to 959) connecting the TM+CT and extracellular regions and modeling these residues as an extended polypeptide chain connecting the two regions. The other chain was left unclosed.

In a third stage, the linkers were remodeled simultaneously to optimize the rigid-body orientation of the ectodomain with respect to the membrane bilayer and the TM

domains, and the remaining cut was closed. The redox buffer disulfide crosslinking restraints for the linker segments (Table S3B) were used as one of the energy terms (command line used: *rosetta_assemble_new.intel CE A2B3 -pose1 -pose_memb -jumping -pairing_file pairings.dat -membrane_cst_reweight 0 -membrane_exposure_reweight 0 -s A2B3.pdb -regions -regionfile region -spanfile A2B3.span -increase_cycles 0.5 -cut_for_assemble -memb_assemble -n A2B3.pdb -num_TMH_pairings 2 -skip_closure -skip_cut 953 -chain_break_reweight 50.0 -nstruct 20 -dump_pdb -paths paths.txt.)*. To keep the relative backbone orientations of each chain in both TM and ectodomains fixed, a "fold tree" was constructed for the polypeptide chain where two Cα positions in each domain were connected and fixed in space during folding . The fold tree was designed so that the TM regions stayed fixed and embedded in the membrane while the ectodomain region was allowed to adopt different conformations with respect to the membrane by sampling the space as a rigid-body. The conformational space sampled by the ectodomain was dictated by the conformations of the loops connecting the TM+CT to the ectodomain. The connecting loops were modeled by Monte-Carlo-based insertion of 3-mer peptide fragments into α_{IIb} and β_3 residues 958-965 and 688-692, respectively, while simultaneously scoring the entire receptor structure. This procedure allowed discarding loop conformations that would embed the ectodomain deeply in the membrane. A total of 6000 coarse-grained models were generated by modeling the loops connecting the TM+CT to the ectodomain. During this procedure, the original orientations of the linker fragments were altered. The models were clustered and the lowest energy cluster centers were refined at all-atom by sampling backbone / side-chain and side-chain

conformational degrees of freedom for the remodeled loops and the remaining regions of the entire integrin structure, respectively.

In the fourth and final stage, two membrane proximal loops that are missing in density in calf-2 domain were remodeled in the context of the entire receptor embedded in the membrane. Residues 838-855 and 858-874 in a loop cleaved by furin, and 762-777 in another loop were added, and modeled as flexible loops that could conform to the orientation with respect to the membrane.

Supplementary figure legends

Figure S1. Disulfide crosslinking with redox buffer and DTT treatment. For constitutively crosslinked extracellular and exofacial residues, crosslinking was measured in redox buffer and after DTT treatment followed by Cu-phenanthroline as described in methods. The crosslinking patterns were plotted for GPA (A) and integrin (B).

Figure S2. Crosslinking in the outer membrane leaflet is indistinguishable in intact and broken cells. 293T cells were transiently transfected with the indicated integrin cysteine mutants and metabolically labeled. Disulfide crosslinking was induced with Cu-phenanthroline in intact (-) or freeze-thawed broken cells (+). Integrin heterodimers were immunoprecipitated from the cell extract with mAb AP3 and subjected to nonreducing SDS-PAGE and autoradiography (upper panel) and phosphorimaging for quantitation of crosslinking efficiency (lower panel).

Figure S3. Kinetics of crosslinking. Crosslinking was measured as described in Methods, except treatment with Cu-phenanthroline was for 0, 10, 30 or 60 mins.

Figure S4. Crosslinking between non-identical GPA TM residues with different integrin subunit fusion partners (symmetry test). 293T cells were co-transfected with αIIb/GPA71-131 G79 or V80 cysteine mutants and the indicated β3/GPA71-131 cysteine mutants. Alternatively, 293T cells were co-transfected with β3/GPA71-131 G79 or V80 cysteine mutant and the indicated αIIb/GPA71-131 cysteine

mutants. Crosslinking efficiency was measured as described in Methods. The plots show that GPA TM crosslinking was independent of the integrin subunit fusion partner, demonstrating a lack of effect of the integrin fusion partner on the GPA structure, and suggesting that symmetry was maintained in the GPA TM domain despite coupling to the asymmetric integrin ectodomain.

Figure S5. **"Near valley correction" of disulfide crosslinking** efficiency. To compensate for inherent flexibility or structural perturbations introduced by cysteine mutation, the lowest crosslinking efficiency within 2 residues of crosslinked residues i or j was subtracted from the crosslinking efficiency of residue pair i and j as described in Methods, before restraint calculation.

Figure S6. Integrin and GPA structure generation.. Procedure for structure generation of αIIbβ3 integrin TM, JM and cytoplasmic domains (A and B) and GPA TM domain (C and D). Residues of TM domains are in blue and marked by dashed lines. TM regions were divided into three overlapping, 12-residue segments (blue lines labeled N, M, or C for N-terminal, middle, C-terminal, respectively). The 12-residue segments were fixed as ideal α-helices with rigid sidechains and docked with the Rosetta global docking protocol in a modeled membrane slab (stage 1). The color codes in the schematic show in subsequent stages which residues were kept from the previous stage and which were grown. Arrows show the direction of chain growing. The 1,000 models in the first stage, which had already been selected to be among the lowest 10% in energy, were scored with the disulfide-based distance constraints, and 100 models with the lowest (or no) pairwise

disulfide constraints violations were selected as the starting structures for chain growing. In chain growing stages 2-4, backbones of the first two or more residues of the starting structures were fixed, and structures of the remaining residues were built with the Rosetta fragment assembly protocol with energy scoring function for membrane protein and disulfide crosslinking-based distance constraints. Between each stage of building 5,000 to 10,000 models, 100 models with low disulfide restraint violation and low energy (not including disulfide restraint violation penalties) were selected for the next stage. At the final stage, the 10% low energy models were clustered based on root-mean-square deviation (RMSD) of C_α atoms for residues in the range, α_{IIb} I966-P998 and β_3 I693-E731 or GPA 73-95. Plots show the correlation of Rosetta all-atom energy (without disulfide restraint penalties) versus disulfide constraints violation scores in stages 1-4. The 100 models selected in successive stages are shown in magenta, and those in the largest cluster in the final stages are shown in red. The five most populous integrin (B) and GPA (D) clusters are summarized. Integrin results are shown both for clustering with 2.0 and 3.0 Å cutoffs, with the latter results in parentheses. The energies shown in A-D do not include disulfide restraint violation penalties.

Figure S7. Center models of the 5 largest clusters. Models and sequence are presented and colored as described in Figure 4. A and B. Central integrin models (from a total of 500 low energy structures) are shown both after clustering at 2.0 Å (A) and 3.0 Å (B) as described in the Fig. S6 legend. Clusters are after 4 stages, whereas the cluster structures shown in Fig. 4A are after a 5^{th} stage of growing extracellular linkers with

redox buffer disulfide restraints. C. Central GPA models (from a total of 1000 low energy structures) are after clustering at 1.0 Å.

Figure S8. Comparison of TM structures generated with and without disulfide-based distance constraints. A. Superposition of GPA NMR structure (red) and GPA models generated without disulfide-based distance constraints (green, center models of top 5 clusters). B. Superposition of integrin TM structures generated with (red, center model of the largest cluster) or without disulfide-based distance constraints (green, center models of top 5 clusters). Structures without distance restraints were generated as described in Figure S6, except structures to seed the next stage were selected based only on Rosetta energy.

Figure S9. Superposition of the 500 low energy $\alpha_{IIb}\beta_3$ models on the TM domains. This represents the ensemble after 4 stages, whereas the cluster structures shown in Fig. 4A are after a 5^{th} stage of growing the extracellular linkers with redox buffer disulfide restraints.

Figure S10. Sequence alignment of integrin TM and juxtamembrane regions. All human integrin α and β subunit sequences are aligned. The boundary of TM and juxtamembrane regions is marked with a dashed line. Green stars indicate residues showing disulfide crosslinking peaks in αIIb and β3. The conserved small amino acids (G, A or S) in both α and β TM domains are marked in yellow and correspond to the Gly

in α_{IIb} and β_3 TM interfaces. The Lys or Arg conserved at the boundary of TM and juxtamembrane region for β integrins are marked in cyan.

Figure S11. Ramachandran plots for residue α_{IIb} Gly-991. A. All 500 low energy models. B. The cluster 1 structural ensemble. The figure was generated by the RAMPAGE server (Lovell et al., 2003). In B, the position of the final cluster center structure is shown in red.

Figure S12. Disulfide crosslinking efficiency within regions defined in integrin ectodomain crystal structures near the linker to the TM domains and correlation with distance. Disulfide crosslinking efficiency in redox buffer is plotted against distance in $\alpha_{IIb}\beta_3$ (Zhu et al., 2008) and $\alpha_V\beta_3$ (Xiong et al., 2002) crystal structures. The specific residues tested in $\alpha_{IIb}\beta_3$ or homologous residues in $\alpha_V\beta_3$ are indicated. Note that a disulfide was introduced by mutation to cysteine of L959 and P688 in the $\alpha_{IIb}\beta_3$ crystal structure and these residues must by definition be close in the $\alpha_{IIb}\beta_3$ structure; however, the homologous I955 and P688 residues in $\alpha_V\beta_3$ are similarly close. To calculate distances in $\alpha_V\beta_3$, the α_{IIb} and α_V sequences were aligned; the sequence in the α_V structure extends more C-terminally than in α_{IIb}. The upper bound distance used in restraints for building ectodomain models is shown as a dashed line.

Figure S13. Lack of relationship among three NMR $\alpha_{IIb}\beta_3$ cytoplasmic domain structures and the Disulfide/Rosetta structure. NMR structures with the indicated PDB ID codes (blue) were superimposed on the juxtamembrane/cytoplasmic portion of the Disulfide/Rosetta structure (red) using Cα atoms of residues α_{IIb} 989-997

and β_3 716-737. Structures are from (Vinogradova et al., 2002) (A) and (Weljie et al., 2002) (B and C).

Spreadsheet. An excel worksheet containing all individual crosslinking results, their averages, and calculation of the disulfide restraints will be submitted. Data from this spreadsheet will be deposited at the PDB in a format analogous to that used for deposition of NMR restraints.

References

Barth, P., Schonbrun, J., and Baker, D. (2007). Toward high-resolution prediction and design of transmembrane helical protein structures. Proc. Natl. Acad. Sci. U.S.A *104*, 15682-15687.

Barth, P., Wallner, B., and D., B. (2009). Prediction of membrane protein structures with complex topologies using limited constraints. Proc. Natl. Acad. Sci. U.S.A *106*, 1409-1414.

Bonneau, R., Strauss, C.E., Rohl, C.A., Chivian, D., Bradley, P., Malmstrom, L., Robertson, T., and Baker, D. (2002). De novo prediction of three-dimensional structures for major protein families. J Mol Biol *322*, 65-78.

Bradley, P., and Baker, D. (2006). Improved beta-protein structure prediction by multilevel optimization of nonlocal strand pairings and local backbone conformation. Proteins *65*, 922-929.

Gottschalk, K.E. (2005). A coiled-coil structure of the αIIbβ3 integrin transmembrane and cytoplasmic domains in its resting state. Structure *13*, 703-712.

Karplus, K., Karchin, R., Draper, J., Casper, J., Mandel-Gutfreund, Y., Diekhans, M., and Hughey, R. (2003). Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. Proteins *53 Suppl 6*, 491-496.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Cα geometry: phi, psi and Cβ deviation. Proteins *50*, 437-450.

Luo, B.-H., Springer, T.A., and Takagi, J. (2004). A specific interface between integrin transmembrane helices and affinity for ligand. PLoS Biol. *2*, 776-786.

Partridge, A.W., Liu, S., Kim, S., Bowie, J.U., and Ginsberg, M.H. (2005). Transmembrane domain packing stabilizes integrin aIIbb3 in the low affinity state. J. Biol. Chem. *280*, 7294-7200.

Peerschke, E.I. (1995). Regulation of platelet aggregation by post-fibrinogen binding events. Thromb. Haemostas. *73*, 862-867.

Rohl, C.A., Strauss, C.E., Misura, K.M., and Baker, D. (2004). Protein structure prediction using Rosetta. Methods in enzymology *383*, 66-93.

Smith, S.O., Song, D., Shekar, S., Groesbeek, M., Ziliox, M., and Aimoto, S. (2001). Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. Biochemistry *40*, 6553-6558.

Takagi, J., Petre, B.M., Walz, T., and Springer, T.A. (2002). Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. Cell *110*, 599-611.

Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T.A., Plow, E.F., and Qin, J. (2002). A structural mechanism of integrin $a_{IIb}b_3$ "inside-out" activation as regulated by its cytoplasmic face. Cell *110*, 587-597.

Weljie, A.M., Hwang, P.M., and Vogel, H.J. (2002). Solution structures of the cytoplasmic tail complex from platelet αIIb- and β3-subunits. Proc. Natl. Acad. Sci. USA *99*, 5878-5883.

Xiong, J.P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., and Arnaout, M.A. (2002). Crystal structure of the extracellular segment of integrin αVβ3 in complex with an Arg-Gly-Asp ligand. Science (New York, N.Y *296*, 151-155.

Zhu, J., Luo, B.H., Xiao, T., Zhang, C., Nishida, N., and Springer, T.A. (2008). Structure of a Complete Integrin Ectodomain in a Physiologic Resting State and Activation and Deactivation by Applied Forces. Mol. cell *32*, 849-861.

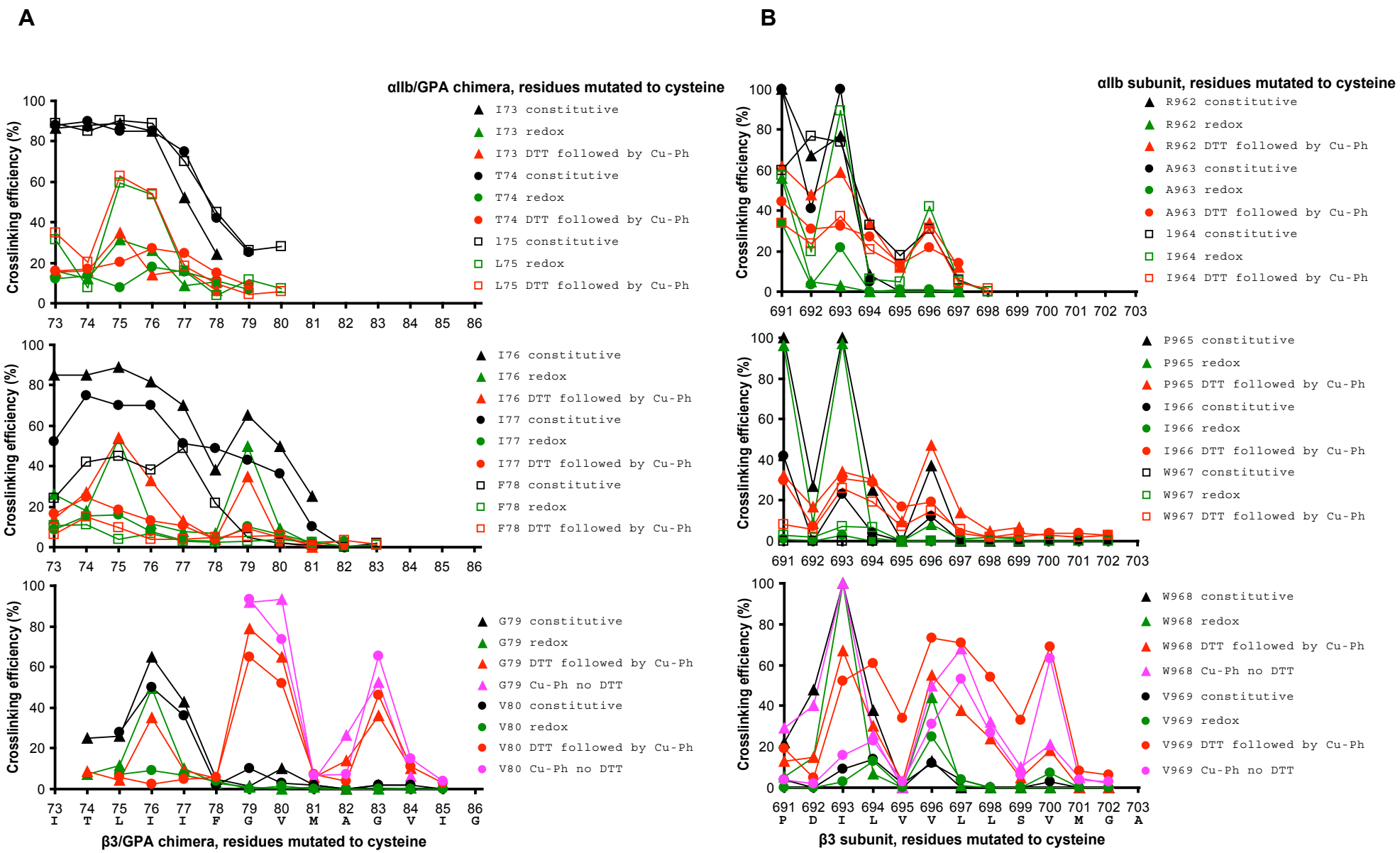

A
αIIb/GPA chimera, residues mutated to cysteine
Crosslinking efficiency (%)
I73 constitutive
I73 redox
I73 DTT followed by Cu-Ph
T74 constitutive
T74 redox
T74 DTT followed by Cu-Ph
l75 constitutive
L75 redox
L75 DTT followed by Cu-Ph
I76 constitutive
I76 redox
I76 DTT followed by Cu-Ph
I77 constitutive
I77 redox
I77 DTT followed by Cu-Ph
F78 constitutive
F78 redox
F78 DTT followed by Cu-Ph
G79 constitutive
G79 redox
G79 DTT followed by Cu-Ph
G79 Cu-Ph no DTT
V80 constitutive
V80 redox
V80 DTT followed by Cu-Ph
V80 Cu-Ph no DTT
73 74 75 76 77 78 79 80 81 82 83 84 85 86
I T L I I F G V M A G V I G
β3/GPA chimera, residues mutated to cysteine
B
αIIb subunit, residues mutated to cysteine
R962 constitutive
R962 redox
R962 DTT followed by Cu-Ph
A963 constitutive
A963 redox
A963 DTT followed by Cu-Ph
l964 constitutive
I964 redox
I964 DTT followed by Cu-Ph
P965 constitutive
P965 redox
P965 DTT followed by Cu-Ph
I966 constitutive
I966 redox
I966 DTT followed by Cu-Ph
W967 constitutive
W967 redox
W967 DTT followed by Cu-Ph
W968 constitutive
W968 redox
W968 DTT followed by Cu-Ph
W968 Cu-Ph no DTT
V969 constitutive
V969 redox
V969 DTT followed by Cu-Ph
V969 Cu-Ph no DTT
691 692 693 694 695 696 697 698 699 700 701 702 703
P D I L V V L L S V M G A
β3 subunit, residues mutated to cysteine

Figure S1

F) Supplemental Text and Figures

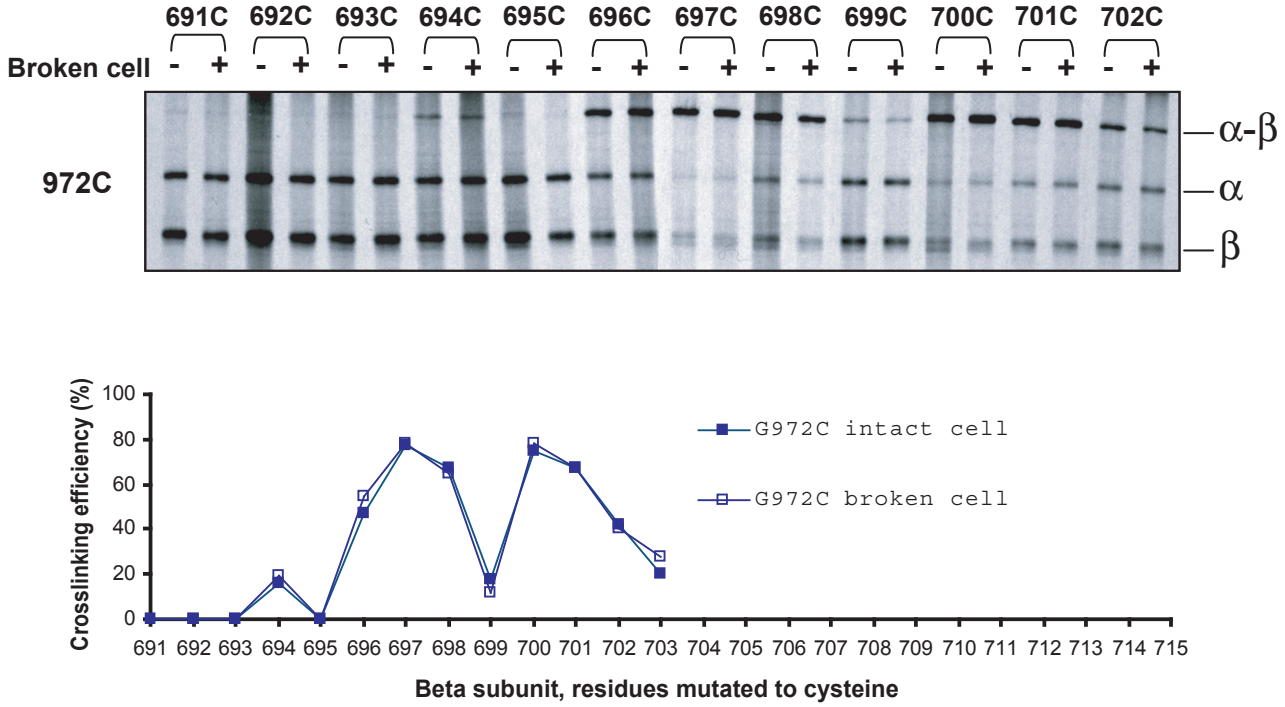


Figure S2

F) Supplemental Text and Figures

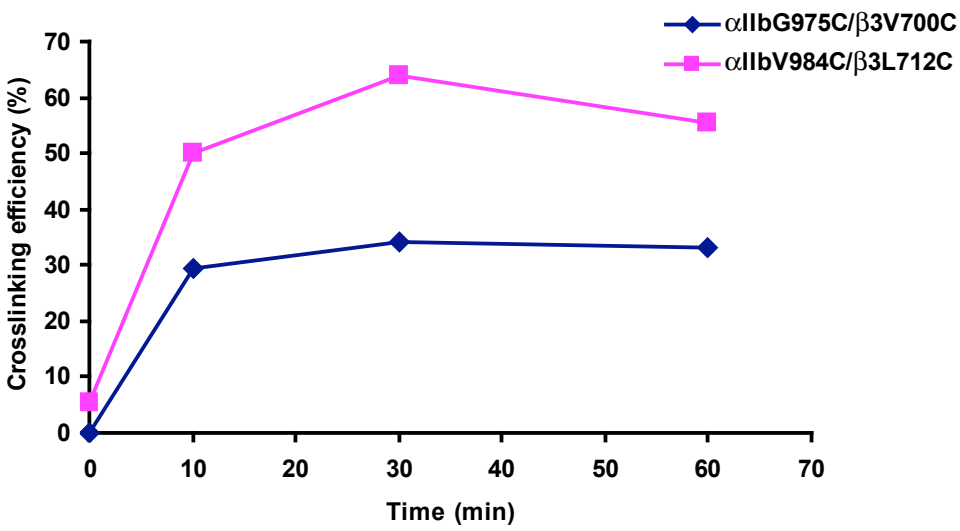


Figure S3

F) Supplemental Text and Figures

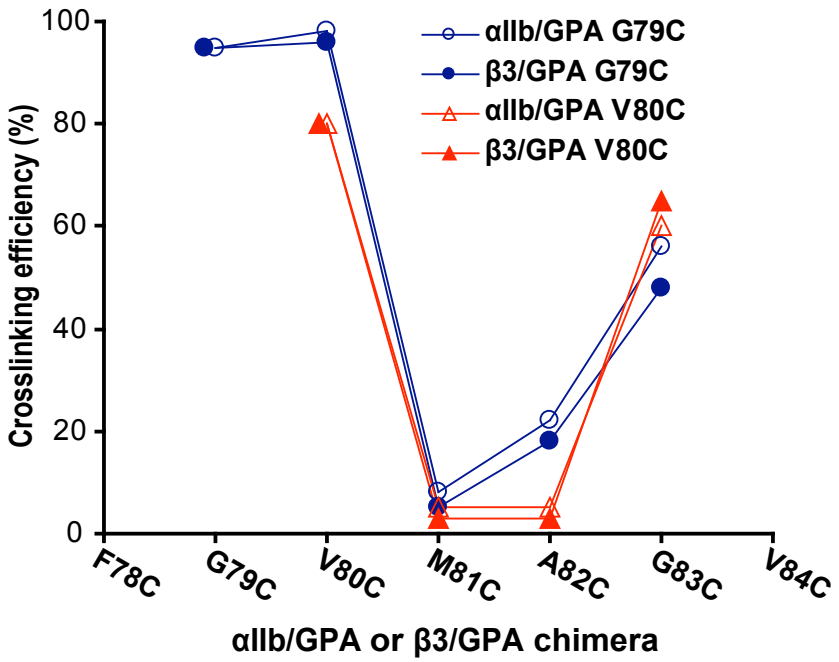


Figure S4

F) Supplemental Text and Figures

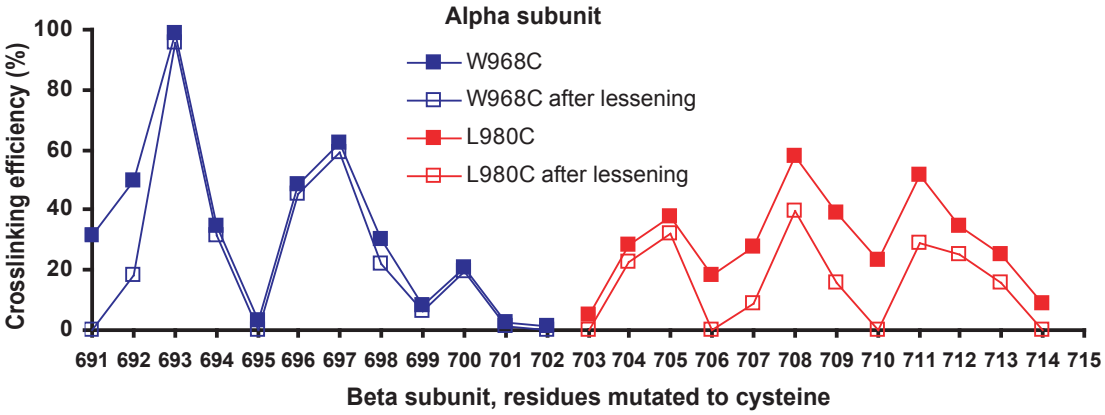
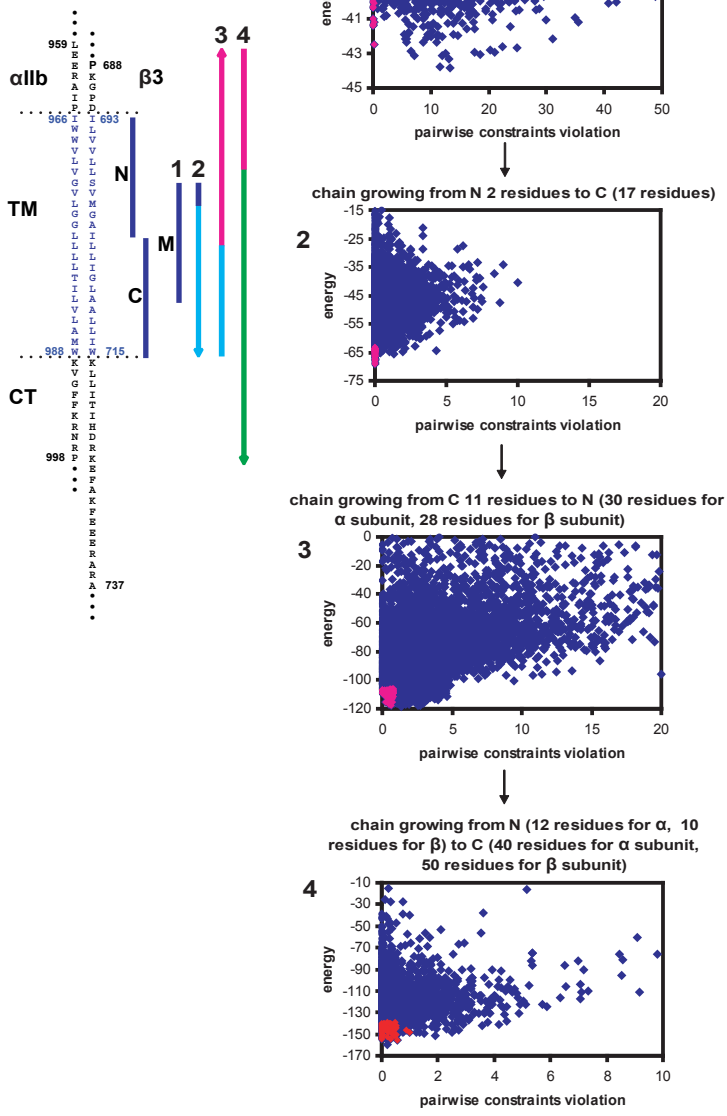


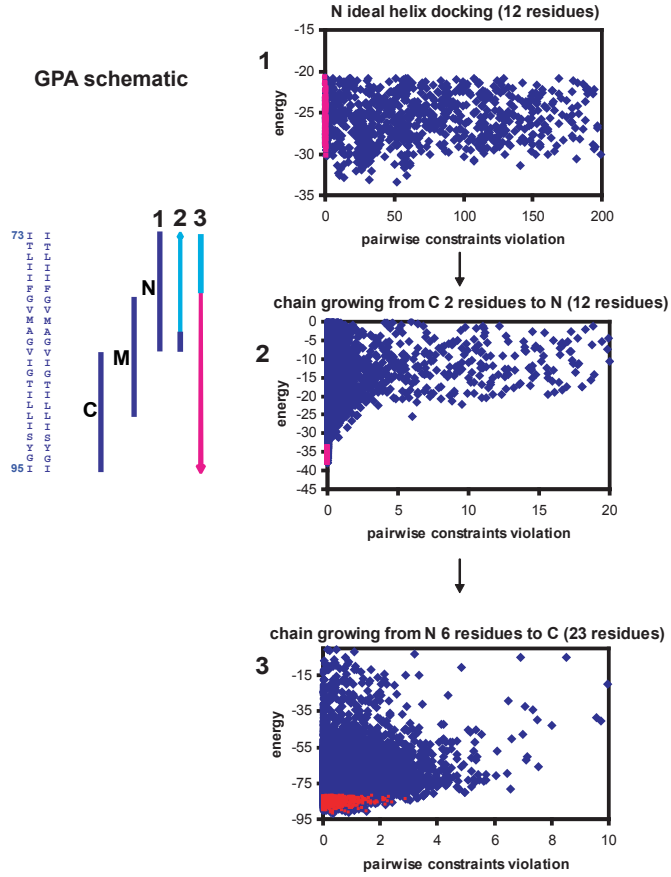
Figure S5

F) Supplemental Text and Figures

A



C



B

Select 10% lowest energy models (500) for clustering, rmsd cutoff: 2.0 Å (3.0 Å)

Cluster No.	Cluster size	Average energy	Average pairwise constraints violation
1	52 (224)	-145.82±4.61 (-144.26±4.11)	0.26±0.23 (0.26±0.27)
2	17 (79)	-144.93±3.82 (-143.64±3.51)	0.13±0.18 (0.27±0.29)
3	13 (23)	-142.74±3.99 (-143.68±3.68)	0.30±0.17 (0.44±0.56)
4	12 (18)	-143.83±2.78 (-144.34±3.93)	0.25±0.17 (0.25±0.21)
5	9 (17)	-142.53±2.75 (-144.46±3.21)	0.36±0.44 (0.31±0.24)

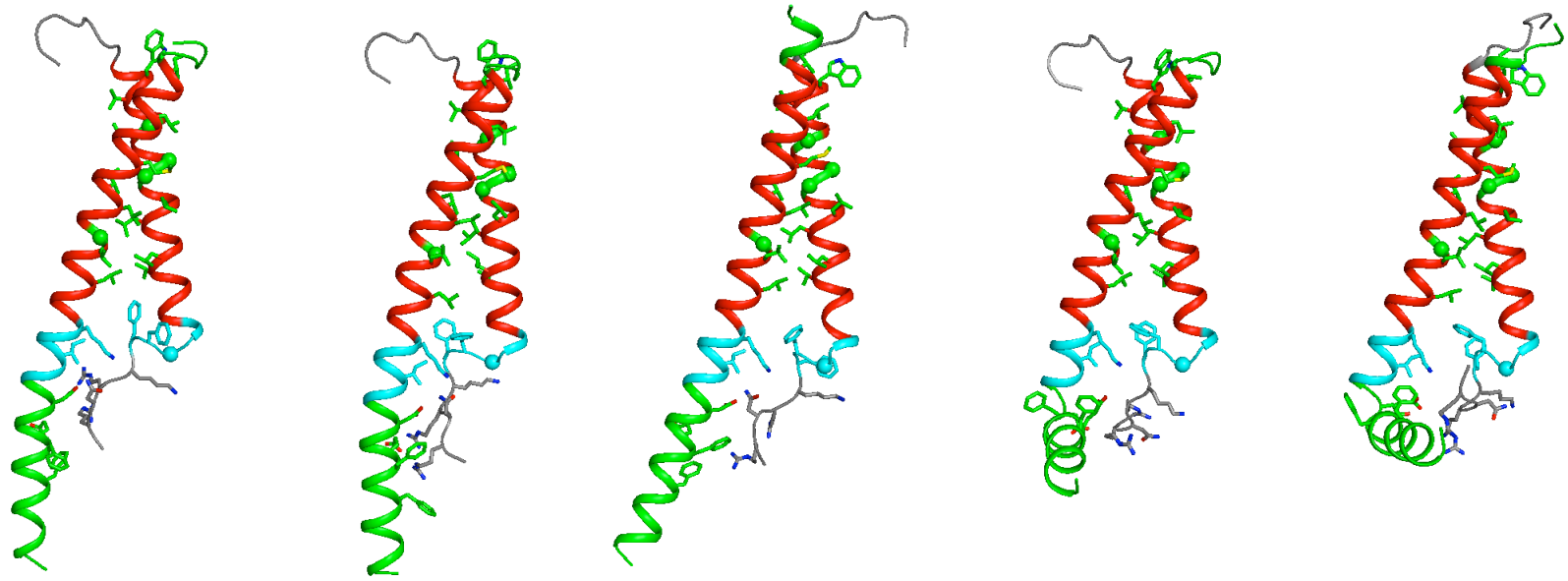
D

Select 10% lowest energy models (1000) for clustering, rmsd cutoff: 1.0 Å

Cluster No.	Cluster size	Average energy	Average pairwise constraints violation	rmsd of center model to NMR structure (Å)
1	652	-84.70±1.91	0.46±0.45	0.71
2	78	-84.66±1.73	0.28±0.47	0.99
3	48	-83.81±1.16	0.79±0.79	1.43
4	38	-84.37±1.95	0.65±0.65	1.55
5	34	-84.04±1.45	0.20±0.28	0.92

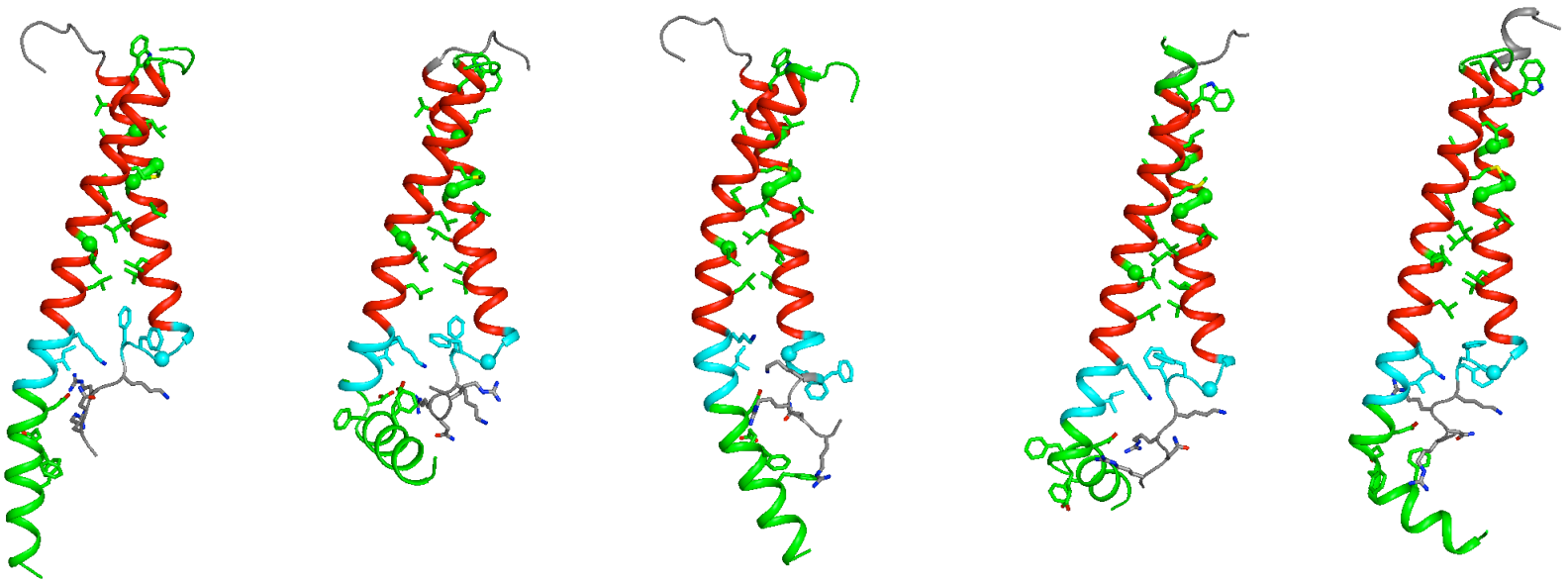
Figure S6

A. Clustered with 2.0 Å cutoff



Cluster 1 (52) Cluster 2 (17) Cluster 3 (13) Cluster 4 (12) Cluster 5 (9)

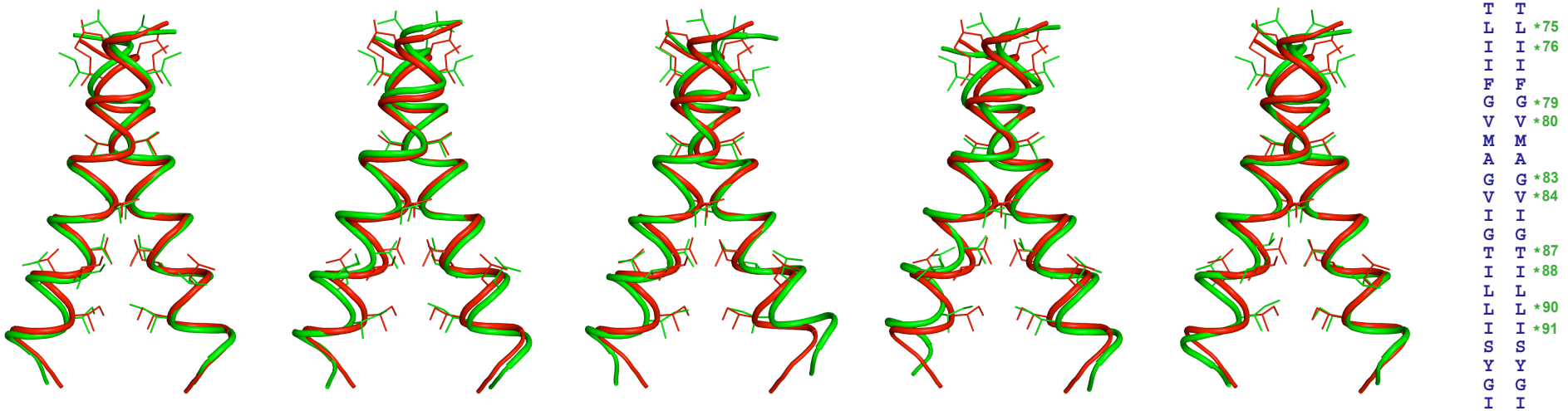
B. Clustered with 3.0 Å cutoff



Cluster 1 (224) Cluster 2 (79) Cluster 3 (23) Cluster 4 (18) Cluster 5 (17)

Figure S7

C. Clustered with 1.0 Å cutoff



Cluster 1 (652) Cluster 2 (78) Cluster 3 (48) Cluster 4 (38) Cluster 5 (34)

Figure S7

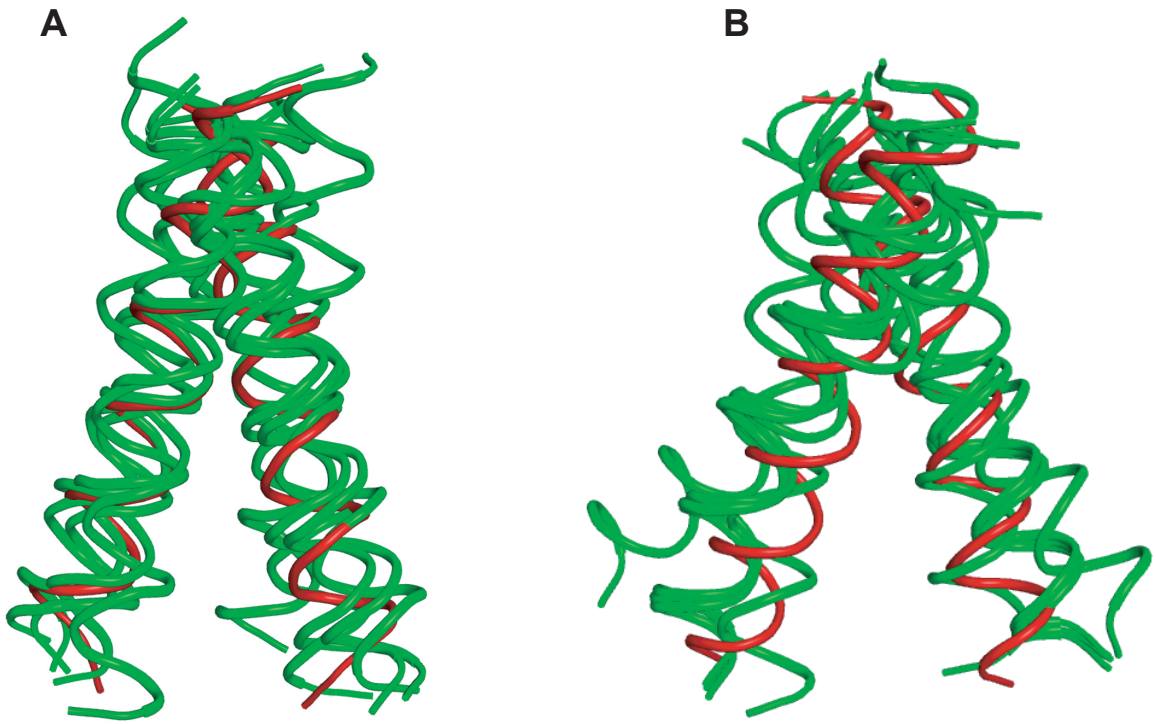


Figure S8

F) Supplemental Text and Figures

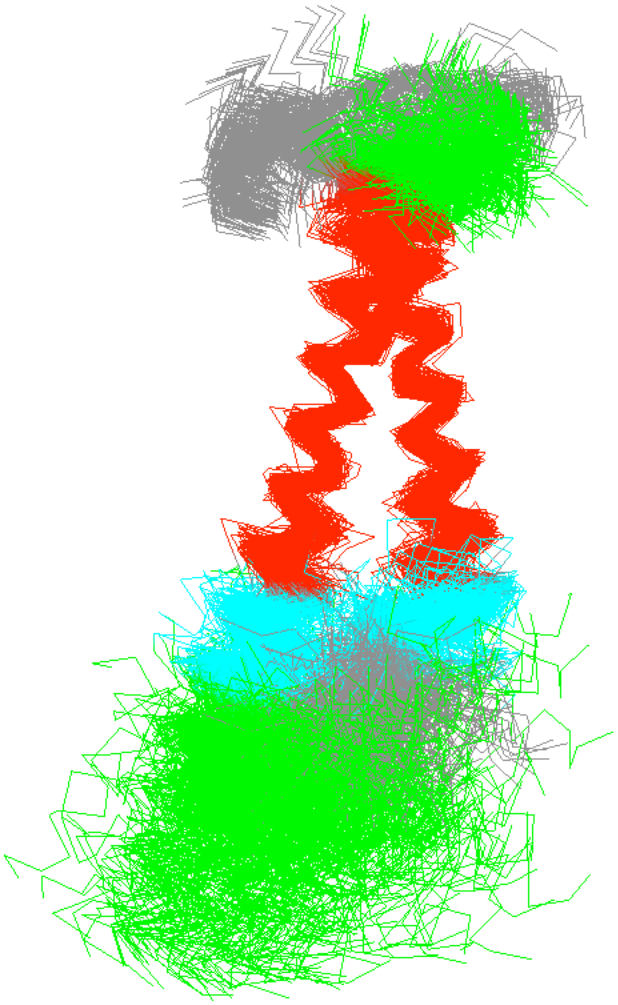


Figure S9

```
                    **   *   **   **   **          ****
αIIb   966   IWWVLVGVLGGLLLLTILVLAMWKVGFFKRNRP...
αV     964   VWVIILAVLAGLLLLAVLVFVMYRMGFFKRVRP...
α1     1114  LWVILLSAFAGLLLLMLLILALWKIGFFKRPLK...
α2     1103  TGVIIGSIIAGILLLLALVAILWKLGFFTRKYE...
α3a    960   LWLVLVAVGAGLLLLGLIILLLWKCGFFKRART...
α3b    960   LWLVLVAVGAGLLLLGLIILLLWKCDFFKRTRY...
α4     945   IVIISSSLLLGLIVLLLISYVMWKAGFFKRQYK...
α5     958   LWIIILAILFGLLLLGLLIYILYKLGFFKRSLP...
α6     992   WWIILVAILAGILMLALLVFILWKCGFFKRNKK...
α7     1005  WWVILLAVLAGLLVLALLVLLLWKMGFFKRAKH...
α8     974   LWVIILAILLGLLVLAILTLALWKCGFFDRARP...
α9     951   GWIIAISLLVGILIFLLLAVLLWKMGFFRRRYK...
α10    1101  LWILIGSVLGGLLLLALLVFCLWKLGFFAHKKI...
α11    1121  IWIIVGSTLGGLLLLALLVLALWKLGFFRSARR...
αD     1085  IPIIMGSSVGALLLLALITATLYKLGFFKRHYK...
αE     1107  LPIIIKGSVGGLLVLIVILVILFKCGFFKRKYQ...
αL     1065  LYLYVLSGIGGLLLLLLIFIVLYKVGFFKRNLK...
αM     1090  LPLIVGSSVGGLLLLALITAALYKLGFFKRQYK...
αX     1087  TPLIVGSSIGGLLLLALITAVLYKVGFFKRQYK...

          *   *   **   **   *   *          *   **   ***
β3     693   ILVVLLSVMGAILLIGLAALLIWKLLITIHDRKEFAKFEEERARA...
β1     709   IIPIVAGVVAGIVLIGLALLLIWKLLMIIHDRREFAKFEKEKMNA...
β2     679   IAAIVGGTVAGIVLIGILLLVIWKALIHLSDLREYRRFEKEKLKS...
β5     697   AMTILLAVVGSILLVGLALLAIWKLLVTIHDRREFAKFQSERSRA...
β6     691   IPMIMLGVSLAILLIGVVLLCIWKLLVSFHDRKEVAKFEAERSKA...
β7     705   TQAIVLGCVGGIVAVGLGLVLAYRLSVEIYDRREYSRFEKEQQQL...
β4     684   FWWLIPLLLLLLPLLALLLLLCWKYCACCKACLALLPCCNRGHMV...
β8     640   YLRIFFIIFIVTFLIGLLKVLIIRQVILQWNSNKIKSSSDYRVSA...
```

Figure S10

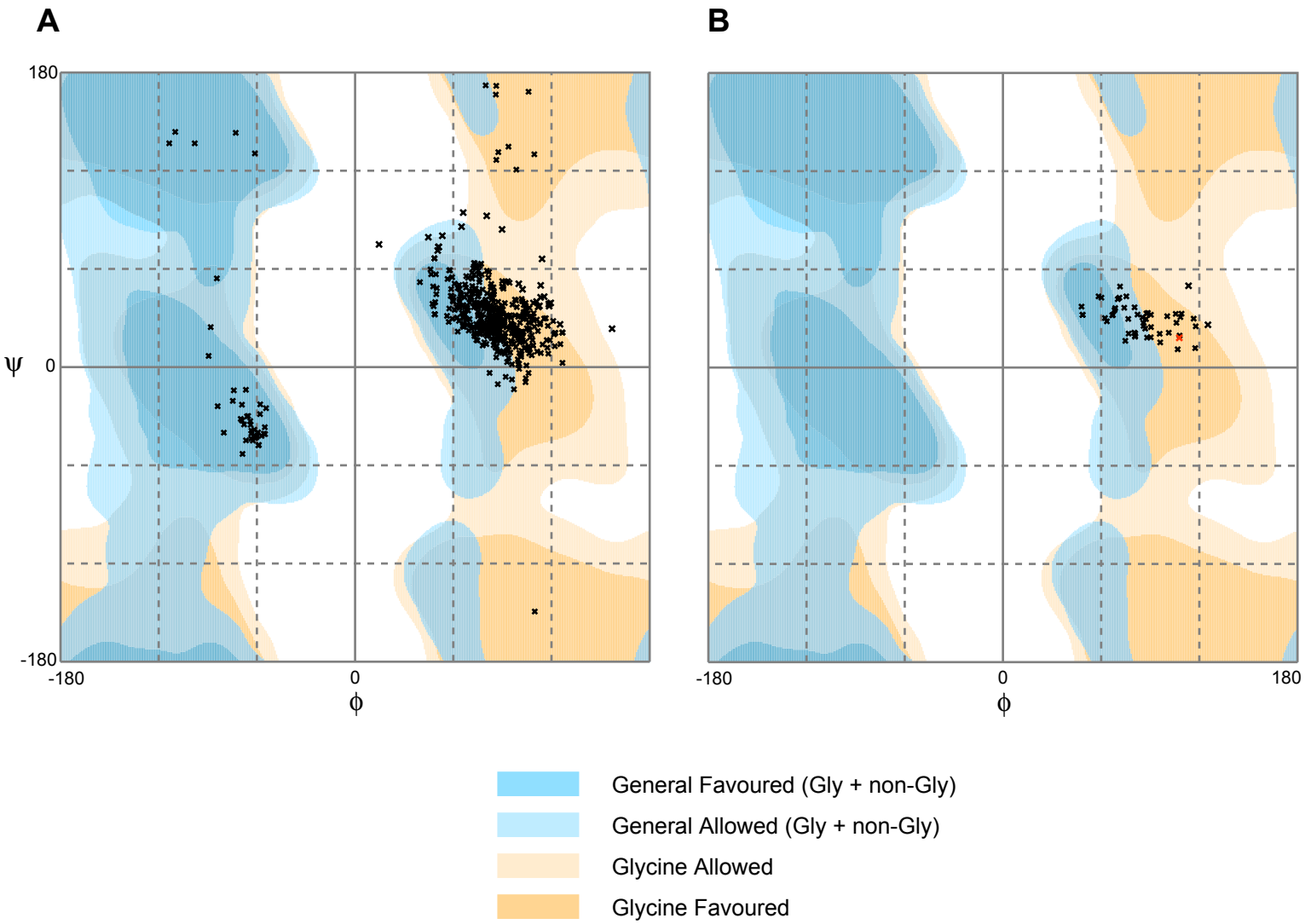

A
B
180
0
-180
ψ
-180
0
ϕ
-180
0
180
ϕ
General Favoured (Gly + non-Gly)
General Allowed (Gly + non-Gly)
Glycine Allowed
Glycine Favoured

Figure S11

F) Supplemental Text and Figures

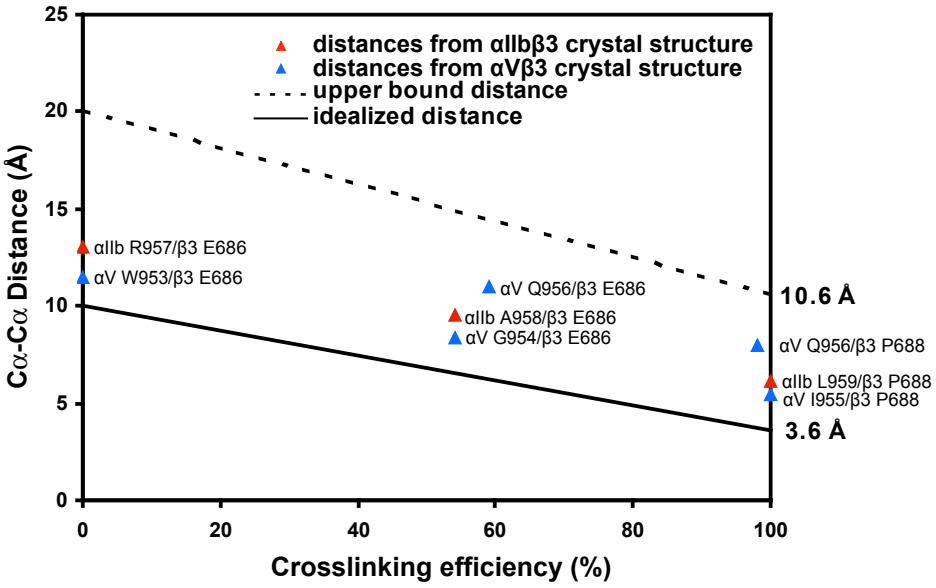


Figure S12

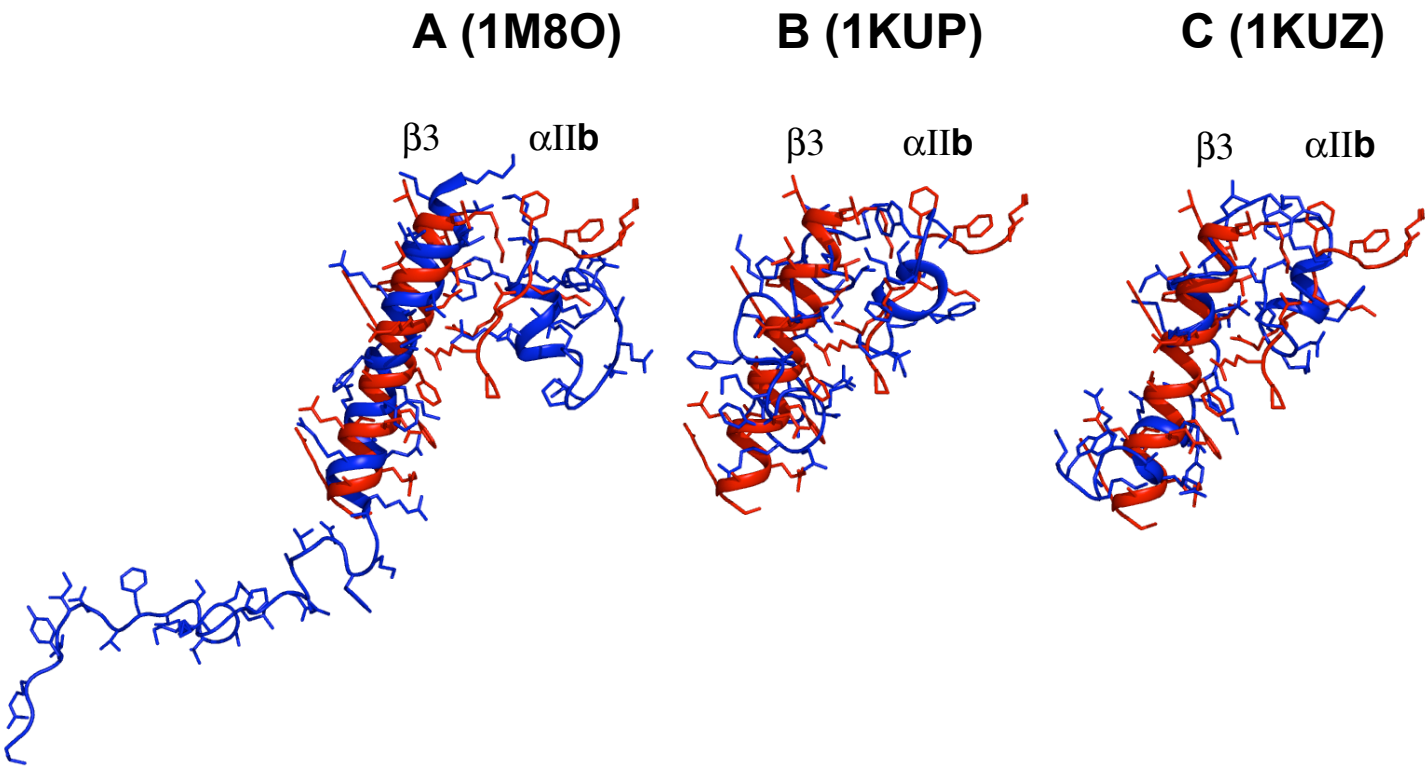


Figure S13

Table S1. Comparison of omitted distance restraints from active integrin cysteine mutants and the average Cα-Cα distance from the cluster 1 structure ensemble.

αIIb	β3	Crosslinking (%)	Corrected crosslinking (%)	Idealized distance (Å)	upper bound distance (Å)	Cα-Cα distance (51 models) (Å)
G976	M701	62.0	49.0	6.03	13.43	4.29±0.51
G976	L698	34.0	31.0	7.82	15.85	8.67±0.59
G976	G702	53.7	30.7	6.56	15.89	7.26±0.53
G976	V700	41.3	28.3	7.36	16.21	7.10±0.58
G976	L697	36.0	22.0	7.70	17.05	8.81±0.64
G976	I707	33.5	22.5	7.86	17.25	10.26±0.50
L978	G708	34.7	20.2	7.78	17.29	11.40±0.50
L979	G708	37.0	28.0	7.63	16.25	9.75±0.61
L980	G708	58.0	39.5	6.29	14.71	6.67±0.56
L983	G708	41.5	37.5	7.34	14.98	9.18±0.86
V984	G708	41.5	37.5	7.34	14.98	9.25±0.76
K994	I719	27.5	27.5	8.24	16.32	9.88±1.28
K994	D723	21.5	20.0	8.62	17.32	11.25±1.10
R995	D723	33.3	28.3	7.87	16.21	8.09±1.14

F) Supplemental Text and Figures

Table S2. Integrin juxtamembrane structure characteristics among the 10% lowest energy (500 models) and cluster 1 ensemble

	β3 K716 ε-amino hydrogen bonding to αIIb F992 or K994 carbonyl oxygen[a]	αIIb R995 to β3 D723 side chain hydrogen bonding[a, b]	Turn at αIIb G991[c]	αIIb F992 or F993 membrane embedding[d]	β3 K716 and αIIb F992 or F993 side chain contact at 6 (5, 4) Å[e]
All 500 models	31%	16%	91%	81%	72% (60%, 38%)
Cluster 1 [f] (52 models)	65%	30%	100%	92%	90% (73%, 53%)

[a] Distance for hydrogen bonding cutoff is 3.5 Å.
[b] At least one hydrogen bond between the three guanido N atoms of R995 and two side chain oxygens of D723.
[c] Turn at G991 was defined by dihedral angle -60° <ψ<180° and 0°<φ<180°.
[d] Membrane embedding was defined as a distance of the C_ζ atom of F992 or F993 from the membrane center of less than 17 Å.
[e] At least one contact between K716 side chain carbons and F992 or F993 side chain carbons less than the indicated distance.
[f] The largest cluster with 2.0 Å cutoff.

F) Supplemental Text and Figures

Table S3A. Crosslinking efficiency and calculated Cα-Cα distance

Integrin TM+JM+CT					
αIIb	β3	Crosslinking(%)	Corrected[a] crosslinking(%)	idealized[b] Cα-Cα distance(Å)	Upper boundary[c] Cα-Cα distance(Å)
W968	I693	100.0	100.0	3.60	6.60
G972	L697	77.0	77.0	5.07	9.68
G972	V700	76.0	63.0	5.14	11.56
W968	L697	62.3	62.3	6.01	11.65
V971	L697	61.5	61.5	6.06	11.76
V969	V700	62.5	59.5	6.00	12.03
G972	V696	56.7	56.7	6.37	12.40
G972	M701	67.7	54.7	5.67	12.67
G972	L698	66.3	53.3	5.76	12.86
L983	L709	54.5	53.0	6.51	12.90
V984	L712	55.0	52.0	6.48	13.03
P998	E731	80.0	49.0	4.88	13.43
V969	L697	48.5	44.0	6.90	14.10
V973	V700	69.3	44.3	5.56	14.06
W968	V696	48.3	48.3	6.91	13.53
G975	M701	51.3	44.0	6.72	14.10
N996	D723	65.0	43.5	5.84	14.17
R997	F730	60.0	41.0	6.16	14.51
P998	F730	75.0	41.0	5.20	14.51
R997	I719	49.0	40.0	6.86	14.64
L979	L705	44.3	38.3	7.16	14.87
R997	D723	70.0	37.0	5.52	15.04
L979	M701	51.0	34.0	6.74	15.44
R997	E726	63.0	34.0	5.97	15.44
V973	L697	47.3	33.3	6.97	15.54
L980	L705	38.0	32.0	7.57	15.71
R997	F727	61.0	32.0	6.10	15.71
P998	F727	82.0	32.0	4.75	15.71
N996	E726	52.0	31.0	6.67	15.85
N996	F727	50.0	31.0	6.80	15.85
G975	L697	43.0	29.0	7.25	16.11
L980	A711	51.7	28.7	6.69	16.15
V984	A711	29.5	28.0	8.11	16.25
G975	V700	51.3	26.3	6.72	16.48
V969	V696	32.0	27.5	7.95	16.32
V969	L698	30.5	25.0	8.05	16.65
V971	V700	25.0	25.0	8.40	16.65
L980	L712	34.5	25.0	7.79	16.65
V984	L713	34.5	25.0	7.79	16.65
R997	E731	43.0	24.0	7.25	16.78
L983	L713	25.5	23.5	8.37	16.85
L980	I704	28.5	22.5	8.18	16.99
L983	L712	25.5	22.5	8.37	16.99
W968	L698	30.3	22.0	8.06	17.05
N996	A728	41.0	22.0	7.38	17.05
N996	F730	34.0	22.0	7.82	17.05
G972	L694	20.3	20.3	8.70	17.28
N996	R724	41.0	20.0	7.32	17.32

GPA					
GPA TM in					
αIIb	β3	Crosslinking(%)	Corrected[a] crosslinking(%)	idealized[b] Cα-Cα distance(Å)	Upper boundary[c] Cα-Cα distance(Å)
G79	V80	93.4	88.9	4.02	8.09
G79	G79	91.8	87.3	4.13	8.31
G83	G83	86.3	83.0	4.47	8.88
G83	V84	74.5	67.2	5.23	11.00
V80	V80	73.3	66.7	5.31	11.07
V80	G83	65.3	61.7	5.82	11.74
T87	T87	62.3	59.7	6.01	12.00
I88	I91	60.5	58.5	6.13	12.16
I91	I91	60.0	53.0	6.16	12.90
I88	L90	65.0	52.0	5.84	13.03
G79	G83	52.3	47.8	6.65	13.59
L90	I91	63.0	45.5	5.97	13.90
L75	I76	54.3	44.8	6.53	14.00
L75	L75	63.0	44.5	5.97	14.04
V84	V84	41.7	38.0	7.33	14.91
I88	I88	48.0	38.0	6.93	14.91
L90	L90	61.0	38.0	6.10	14.91
I76	G79	35.0	35.0	7.76	15.31
T87	I88	37.3	31.7	7.61	15.76
I76	I76	33.0	29.3	7.89	16.08
G86	G86	28.3	28.0	8.19	16.25
A82	G83	29.0	26.5	8.14	16.45
G86	T87	23.8	23.4	8.48	16.86
I91	S92	30.0	23.0	8.08	16.92
S92	G94	31.5	22.0	7.98	17.05
G79	A82	26.3	21.8	8.31	17.07
I73	L75	35.0	21.0	7.76	17.19

[a] Corrected crosslinking efficiency was calculated as described in the supplement and Fig. S5; only ≥20% crosslinking was used for restraint generation.
[b] Idealized Cα-Cα distance was calculated with the uncorrected crosslinking efficiency.
[c] Upper boundary Cα-Cα distance was calculated with the corrected crosslinking efficiency.

Table S3B. Crosslinking efficiency and calculated Cα-Cα distance of linker region

Integrin extracellular linker					
αIIb	β3	Crosslinking(%)	Corrected[a] crosslinking(%)	idealized[b] Cα-Cα distance(Å)	Upper boundary[c] Cα-Cα distance(Å)
L959	P688	100.0	100.0	3.60	10.60
E960	P688	98.0	98.0	3.73	10.79
R962	P691	56.0	22.0	6.42	17.93
A963	P691	34.0	30.5	7.82	17.13
I964	P691	57.5	37.5	6.32	16.48
I964	I693	89.5	84.5	4.27	12.06
I964	V696	42.0	42.0	7.31	16.05
P965	P691	96.0	87.5	3.86	11.78
P965	I693	97.0	94.0	3.79	11.16

[a] Corrected crosslinking efficiency was calculated as described in the supplement and Fig. S5; only ≥20% crosslinking in redox buffer was used for restraint generation.
[b] Idealized Cα-Cα distance was calculated with the uncorrected crosslinking efficiency.
[c] Upper boundary Cα-Cα distance was calculated with the corrected crosslinking efficiency.

F) Supplemental Text and Figures

Table S4. Format of Cα-Cα distance constraints for GPA transmembrane structure generation

res1	atom1	res2	atom2	upper	native
7	CA	31	CA	8.09	#G79 V80
7	CA	30	CA	8.31	#G79 G79
11	CA	34	CA	8.88	#G83 G83
11	CA	35	CA	11.00	#G83 V84
8	CA	31	CA	11.07	#V80 V80
8	CA	34	CA	11.74	#V80 G83
15	CA	38	CA	12.00	#T87 T87
16	CA	42	CA	12.16	#I88 I91
19	CA	42	CA	12.90	#I91 I91
16	CA	41	CA	13.03	#I88 L90
7	CA	34	CA	13.59	#G79 G83
18	CA	42	CA	13.90	#L90 I91
3	CA	27	CA	14.00	#L75 I76
3	CA	26	CA	14.04	#L75 L75
12	CA	35	CA	14.91	#V84 V84
16	CA	39	CA	14.91	#I88 I88
18	CA	41	CA	14.91	#L90 L90
4	CA	30	CA	15.31	#I76 G79
15	CA	39	CA	15.76	#T87 I88
4	CA	27	CA	16.08	#I76 I76
14	CA	37	CA	16.25	#G86 G86
10	CA	34	CA	16.45	#A82 G83
14	CA	38	CA	16.86	#G86 T87
19	CA	43	CA	16.92	#I91 S92
20	CA	45	CA	17.05	#S92 G94
7	CA	33	CA	17.07	#G79 A82
1	CA	26	CA	17.19	#I73 L75
30	CA	8	CA	8.09	#G79 V80
30	CA	7	CA	8.31	#G79 G79
34	CA	11	CA	8.88	#G83 G83
34	CA	12	CA	11.00	#G83 V84
31	CA	8	CA	11.07	#V80 V80
31	CA	11	CA	11.74	#V80 G83
38	CA	15	CA	12.00	#T87 T87
39	CA	19	CA	12.16	#I88 I91
42	CA	19	CA	12.90	#I91 I91
39	CA	18	CA	13.03	#I88 L90
30	CA	11	CA	13.59	#G79 G83
41	CA	19	CA	13.90	#L90 I91
26	CA	4	CA	14.00	#L75 I76
26	CA	3	CA	14.04	#L75 L75
35	CA	12	CA	14.91	#V84 V84
39	CA	16	CA	14.91	#I88 I88
41	CA	18	CA	14.91	#L90 L90
27	CA	7	CA	15.31	#I76 G79
38	CA	16	CA	15.76	#T87 I88
27	CA	4	CA	16.08	#I76 I76
37	CA	14	CA	16.25	#G86 G86
33	CA	11	CA	16.45	#A82 G83
37	CA	15	CA	16.86	#G86 T87
42	CA	20	CA	16.92	#I91 S92
43	CA	22	CA	17.05	#S92 G94
30	CA	10	CA	17.07	#G79 A82
24	CA	3	CA	17.19	#I73 L75