



Publicly Accessible Penn Dissertations


Summer 8-14-2009

Structural Characterization of the Integrin Aiiib β 3 Transmembrane and Cytosolic Domains

Douglas G. Metcalf

University of Pennsylvania, doug.metcalf@gmail.com

Follow this and additional works at: <http://repository.upenn.edu/edissertations>

 Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Biochemistry, Biophysics, and Structural Biology Commons](#), and the [Oncology Commons](#)

Recommended Citation

Metcalf, Douglas G., "Structural Characterization of the Integrin Aiiib β 3 Transmembrane and Cytosolic Domains" (2009). *Publicly Accessible Penn Dissertations*. 23.

<http://repository.upenn.edu/edissertations/23>

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/edissertations/23>

For more information, please contact libraryrepository@pobox.upenn.edu.

Structural Characterization of the Integrin α IIB β 3 Transmembrane and Cytosolic Domains

Abstract

Integrins are the principal cell surface receptors that link the cytoskeleton to the extracellular matrix. They exist in active conformations that can bind extracellular ligands and resting conformations that cannot. The platelet integrin α IIB β 3 is a prototypical regulated integrin that is resting on a circulating platelet and becomes activated to adhere the platelet to the vascular endothelium or subendothelial matrix.

The integrin is composed of α and β subunits and each subunit contains a single transmembrane helix that form an α/β heterodimer in the resting state. Additionally, each subunit contains a cytosolic domain that binds signaling proteins that affect the resting-active equilibrium. Activation signals are transduced across the membrane by separating the transmembrane heterodimer.

The structure of the resting integrin α IIB β 3's transmembrane and cytosolic domains was characterized by molecular modeling and NMR spectroscopy. First, software was developed to model transmembrane helix dimers using experimental mutagenesis results as a modeling restraint. Next, the α IIB/ β 3 transmembrane heterodimer was modeled and the model was compared to published experimental data and other published models. The model correlated well with experimental findings and converged on the same structure as other top performing models, suggesting this conformation approximates the native interface. The model's interface includes α IIB residue Met987 and β 3 residue Leu712. These residues were mutated to cysteine to crosslink peptides corresponding to the α IIB and β 3 cytosolic tails, and the disulfide-linked construct was probed by NMR spectroscopy.

NMR revealed that the α IIB and β 3 cytosolic tails have a dynamic interface. The α IIB subunit is natively unstructured and the β 3 subunit consists of a hydrophobic helix followed by two amphiphilic helices. The amphiphilic portions of β 3 include domains that interact with cytosolic proteins, but the membrane embedding of its hydrophobic faces sequesters some of the interacting residues. This result suggests that the integrin's resting-active equilibrium is coupled to an equilibrium between membrane embedded and solvent exposed conformations of the β 3 cytosolic tail, providing new insight into integrin activation.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Biochemistry & Molecular Biophysics

First Advisor

William F. DeGrado

Keywords

integrin, a11bb3, transmembrane, molecular modeling, NMR

Subject Categories

Amino Acids, Peptides, and Proteins | Biochemistry, Biophysics, and Structural Biology | Oncology

STRUCTURAL CHARACTERIZATION OF THE INTEGRIN α IIb β 3
TRANSMEMBRANE AND CYTOSOLIC DOMAINS

Douglas G. Metcalf

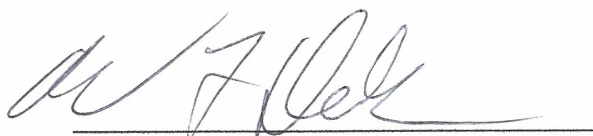
A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2009

A handwritten signature in black ink, appearing to read 'D. G. Metcalf', written over a horizontal line.

Supervisor of Dissertation

A handwritten signature in black ink, appearing to read 'J. M. Ferguson', written over a horizontal line. A red 'X' is written to the left of the signature.

Graduate Group Chairperson

dedicated to:

My Mom and Dad, Kathy and Doug Metcalf

ABSTRACT

STRUCTURAL CHARACTERIZATION OF THE INTEGRIN α IIb β 3 TRANSMEMBRANE AND CYTOSOLIC DOMAINS

Douglas G. Metcalf

William F. DeGrado

Integrin's are the principal cell surface receptors that link the cytoskeleton to the extracellular matrix. They exist in active conformations that can bind extracellular ligands and resting conformations that cannot. The platelet integrin α IIb β 3 is a prototypical regulated integrin that is resting on a circulating platelet and becomes activated to adhere the platelet to the vascular endothelium or subendothelial matrix.

The integrin is composed of α and β subunits and each subunit contains a single transmembrane helix that form an α/β heterodimer in the resting state. Additionally, each subunit contains a cytosolic domain that binds signaling proteins that affect the resting-active equilibrium. Activation signals are transduced across the membrane by separating the transmembrane heterodimer.

The structure of the resting integrin α IIb β 3's transmembrane and cytosolic domains was characterized by molecular modeling and NMR spectroscopy. First, software was developed to model transmembrane helix dimers using

experimental mutagenesis results as a modeling restraint. Next, the α IIb/ β 3 transmembrane heterodimer was modeled and the model was compared to published experimental data and other published models. The model correlated well with experimental findings and converged on the same structure as other top performing models, suggesting this conformation approximates the native interface. The model's interface includes α IIb residue Met987 and β 3 residue Leu712. These residues were mutated to cysteine to crosslink peptides corresponding to the α IIb and β 3 cytosolic tails, and the disulfide-linked construct was probed by NMR spectroscopy.

NMR revealed that the α IIb and β 3 cytosolic tails have a dynamic interface. The α IIb subunit is natively unstructured and the β 3 subunit consists of a hydrophobic helix followed by two amphiphilic helices. The amphiphilic portions of β 3 include domains that interact with cytosolic proteins, but the membrane embedding of its hydrophobic faces sequesters some of the interacting residues. This result suggests that the integrin's resting-active equilibrium is coupled to an equilibrium between membrane embedded and solvent exposed conformations of the β 3 cytosolic tail, providing new insight into integrin activation.

TABLE OF CONTENTS

CHAPTER 1

Background on Integrins and Methods to Define Their Three-Dimensional Structure

Introduction.	1
-----------------------	---

CHAPTER 2

Development of Software to Model the Integrin α IIb/ β 3 Transmembrane Heterodimer

Introduction.	8
Materials and Methods.	11
Results.	19
Discussion and Conclusions.	26

CHAPTER 3

Multiple Approaches Converge on the Structure of the Integrin α IIb/ β 3 Transmembrane Heterodimer

Introduction.	37
Materials and Methods.	39
Results.	46
Discussion and Conclusions.	57

CHAPTER 4

Structural Characterization of a Disulfide-Linked α IIb/ β 3 Cytosolic Domain

Introduction.	73
Materials and Methods.	78
Results.	84
Discussion and Conclusions.	99

CHAPTER 5

Discussion, Future Direction, and Concluding Remarks

Discussion.	115
---------------------	-----

LIST OF TABLES

I	Glycophorin A and BNIP3 mutations used as inputs for the Monte Carlo modeling algorithm.	14
II	Sequences of glycophorin A homologs, orthologs, and single nucleotide polymorphisms.	15
III	Qualitative comparison of six α IIb/ β 3 models with experimental findings and the structural similarity of these models reported as $C\alpha$ RMSD.	52
IV	Structural similarity between top performing models and two models that are currently in press, reported as $C\alpha$ RMSD.	63
V	NMR structure statistics.	95

LIST OF FIGURES

1	Energy landscapes for the glycoporphin A transmembrane dimer.	13
2	Flow chart for the Monte Carlo modeling algorithm.	16
3	Graphical representations of the penalties used to score models.	22
4	Models of glycoporphin A fit to its NMR structure.	24
5	The model of glycoporphin A generated with sequence homology data is nearly identical to a structure in the glycoporphin A NMR ensemble.	24
6	Model of BNIP3 and its sequence.	25
7	Cartoon of the equilibrium between resting and active integrins.	38
8	Sequences of the α IIb and β 3 transmembrane domains.	40
9	Monte Carlo model of the α IIb/ β 3 heterodimer depicting qualitative agreement with experimental mutagenesis results.	48
10	Slices through the Monte Carlo α IIb/ β 3 model depicting qualitative agreement with experimental cysteine crosslinking results.	49
11	Graphical representation of the quantitative correlations between different α IIb/ β 3 TM models and experimental mutagenesis results.	54
12	Graphical representation of the quantitative correlations between different α IIb/ β 3 TM models and experimental cysteine crosslinking results.	56
13	Structural comparison of different models for the resting α IIb/ β 3 transmembrane heterodimer.	61
14	Sequence and topology of the disulfide-linked α IIb/ β 3 NMR construct. .	76

LIST OF FIGURES (continued)

15 SDS-PAGE analysis of the disulfide-linked α IIb/ β 3 NMR sample. 82

16 Circular dichroism analysis of the disulfide-linked α IIb/ β 3 construct and its component monomers. 86

17 Analysis of β 3 C α chemical shifts in the disulfide-linked heterodimer and a β 3 monomer. 88

18 ^{13}C HSQC spectra for the disulfide-linked heterodimer and its component monomers. 89

19 H- ^{15}N NOE intensities in the disulfide-linked heterodimer. 91

20 ^{15}N HSQC spectra for the disulfide-linked heterodimer before and after hydrogen-deuterium exchange. 93

21 NMR structure of the β 3 subunit in the disulfide-linked heterodimer depicting interfaces that interact with α IIb, talin, kindlin-3, and Src kinase. . . . 96

22 Cartoon of the calculated membrane embedding for portions of the β 3 subunit. 100

23 Cartoons that illustrate the β 3/talin interaction. 105

LIST OF EQUATIONS

1	Scoring function used to penalize the calculated energy of a modeled conformation that is inconsistent with experimental mutagenesis results.	21
2	Function used to calculate penalties for modeled conformations that are inconsistent with one or more disruptive mutations.	21
3	Function used to calculate penalties for modeled conformations that are inconsistent with one or more neutral mutations.	22
4	Equation used to correlate the calculated C β -C β distances in a model with experimentally determined cysteine crosslinking yields.	44
5	Equation used to calculate a residue's fractional change in solvent accessible surface area f_{ASA} upon dimerization.	45

CHAPTER 1

Background on Integrins and Methods to Define Their Three-Dimensional Structure

INTRODUCTION

Integrins are cell surface receptors that enable cell-cell and cell-matrix interactions by engaging extracellular molecules. Integrin-ligand binding events mediate cell adhesion and migration and initiate intracellular signaling pathways that regulate key processes such as proliferation, differentiation, and apoptosis.¹ Thus it is not surprising that integrins play pivotal roles in health and disease. For example, they are targeted by snake venoms, including rattlesnake and cottonmouth venoms that can cause hemorrhage and occasionally death.² Additionally, viruses including adenovirus, rotavirus, hantavirus, and HIV recognize integrins to penetrate and hijack cells.³ In the research and clinical setting, integrins serve as biomarkers to distinguish different cell types and they're used for diagnostics to characterize pathological states.⁴ Also, various pharmaceuticals target integrins in approved treatments for multiple sclerosis, Crohn's disease, and certain coronary events (Tysabri™ generated \$589 million

in revenue for 2008),^{5; 6} and integrin antagonists are being developed as therapeutics for other conditions such as cardiovascular disease and cancer, the number 1 and 2 leading causes of death in America.^{6; 7; 8} Thus integrin structure and function is an intriguing and important field of study in biology and medicine.

Integrins are transmembrane (TM) glycoproteins composed of α and β subunits. In mammals, there are 18 different α subunits and 8 different β subunits that form 24 known α/β pairs through a heterodimeric interaction in their extracellular domains.¹ Each integrin subunit is a type 1 membrane protein that also contains a single TM helix and a cytosolic carboxy-terminus. The plasma membrane contains α/β heterodimers that exist in an equilibrium between resting conformations that have low affinity for extracellular ligands and active conformations that have high affinity.⁹ Signaling cascades that shift the integrin equilibrium toward the active state, termed “inside-out” signals, have received significant attention over the past few decades, including the recent identification of the cytoskeletal proteins talin and kindlin as essential mediators of integrin activation.^{10; 11} These molecules are postulated to bind the integrin’s cytosolic domains and disrupt a heterodimeric α/β interaction in the TM region. Separation of the resting integrin’s α/β TM heterodimer functions to transduce an activation signal across the membrane, ultimately causing a conformational change that exposes the integrin’s extracellular ligand binding sites.¹² Once an integrin is activated, it can initiate “outside-in” signals and recruit additional proteins to form large structural and signaling complexes such as focal adhesions that tightly bind

the underlying actin cytoskeleton.¹³ The goal of the research presented here is to test hypotheses that suggest mechanisms for integrin activation by defining the three-dimensional structure of the resting integrin's TM and cytosolic domains.

Three-dimensional models of protein structure are invaluable research tools because they provide paradigms to predict and validate experimental results, and their importance is underscored by the \$765 million Protein Structure Initiative.¹⁴ For integrins, three-dimensional structures of the extracellular domain in “bent” and “extended” conformations led to several testable hypotheses that suggest different mechanisms for integrin activation.^{15; 16; 17} Additionally, models of the integrin's TM and cytosolic domains have been developed, however until recently, they provided marginal insight into integrin function because they did not make useful predictions or spur further analysis.

This thesis describes and authenticates a model of the resting TM heterodimer for the $\alpha\text{IIb}\beta\text{3}$ integrin that aided in the engineering of a disulfide-linked $\alpha\text{IIb}/\beta\text{3}$ cytosolic domain. The model was calculated using a recently-developed Monte Carlo algorithm that includes a selective advantage for conformations that are consistent with experimental mutagenesis results, and this software is benchmarked and validated in chapter 2. The Monte Carlo model was subsequently confirmed by measuring several of its attributes and correlating them with published experimental results using other published models as controls, and chapter 3 presents this analysis. Correlations with cysteine crosslinking experiments were of particular interest because they

identified α IIb/ β 3 cysteine mutant pairs at the model's heterodimer interface that could covalently crosslink the α IIb and β 3 subunits.¹⁸ Based on this finding, cysteines were positioned to enforce the model's TM heterodimer interface in a construct consisting of the α IIb and β 3 cytosolic domains tethered by a disulfide bond. This construct was expressed, purified, and then probed by NMR spectroscopy, enabling the calculation of a solution structure for the β 3 subunit described in chapter 4. The subsequent structural analysis makes several predictions that weren't apparent *a priori*. Notably much of the β 3 cytosolic domain is pre-organized into conformations that are similar to structures depicted in β 3/talin interfaces,^{19;20} thereby minimizing the entropic cost of binding, however these regions are calculated to partition into the plasma membrane and an interaction with talin would trap β 3 in an alternate, exposed conformation, providing a mechanism for talin-induced conformational change. Further implications of the NMR structure and future directions are discussed in chapters 4 and 5.

References

1. Hynes, R. O. (2002). Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673-87.
2. Markland, F. S. (1998). Snake venoms and the hemostatic system. *Toxicon* **36**, 1749-800.

3. Nemerow, G. R. & Stewart, P. L. (2001). Antibody neutralization epitopes and integrin binding sites on nonenveloped viruses. *Virology* **288**, 189-91.
4. Moschos, S. J., Drogowski, L. M., Reppert, S. L. & Kirkwood, J. M. (2007). Integrins and cancer. *Oncology (Williston Park)* **21**, 13-20.
5. (2009). Biogen Idec Reports Full Year and Fourth Quarter 2008 Results. *The Wall Street Journal*.
6. Mandava, P., Thiagarajan, P. & Kent, T. A. (2008). Glycoprotein IIb/IIIa antagonists in acute ischaemic stroke: current status and future directions. *Drugs* **68**, 1019-28.
7. Twombly, R. (2005). Cancer surpasses heart disease as leading cause of death for all but the very elderly. *J Natl Cancer Inst* **97**, 330-1.
8. Tucker, G. C. (2006). Integrins: molecular targets in cancer therapy. *Curr Oncol Rep* **8**, 96-103.
9. Li, W., Metcalf, D. G., Gorelik, R., Li, R., Mitra, N., Nanda, V., Law, P. B., Lear, J. D., Degrado, W. F. & Bennett, J. S. (2005). A push-pull mechanism for regulating integrin function. *Proc Natl Acad Sci U S A* **102**, 1424-9.
10. Tadokoro, S., Shattil, S. J., Eto, K., Tai, V., Liddington, R. C., de Pereda, J. M., Ginsberg, M. H. & Calderwood, D. A. (2003). Talin binding to integrin beta tails: a final common step in integrin activation. *Science* **302**, 103-6.

11. Moser, M., Nieswandt, B., Ussar, S., Pozgajova, M. & Fassler, R. (2008). Kindlin-3 is essential for integrin activation and platelet aggregation. *Nat Med* **14**, 325-30.
12. Kim, M., Carman, C. V. & Springer, T. A. (2003). Bidirectional transmembrane signaling by cytoplasmic domain separation in integrins. *Science* **301**, 1720-5.
13. Romer, L. H., Birukov, K. G. & Garcia, J. G. (2006). Focal adhesions: paradigm for a signaling nexus. *Circ Res* **98**, 606-16.
14. Norvell, J. C. & Berg, J. M. (2007). Update on the protein structure initiative. *Structure* **15**, 1519-22.
15. Xiao, T., Takagi, J., Collier, B. S., Wang, J. H. & Springer, T. A. (2004). Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* **432**, 59-67.
16. Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A., Goodman, S. L. & Arnaout, M. A. (2001). Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* **294**, 339-45.
17. Xiong, J. P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S. L. & Arnaout, M. A. (2002). Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* **296**, 151-5.

18. Luo, B. H., Springer, T. A. & Takagi, J. (2004). A specific interface between integrin transmembrane helices and affinity for ligand. *PLoS Biol* **2**, e153.
19. Wegener, K. L., Partridge, A. W., Han, J., Pickford, A. R., Liddington, R. C., Ginsberg, M. H. & Campbell, I. D. (2007). Structural basis of integrin activation by talin. *Cell* **128**, 171-82.
20. Garcia-Alvarez, B., de Pereda, J. M., Calderwood, D. A., Ulmer, T. S., Critchley, D., Campbell, I. D., Ginsberg, M. H. & Liddington, R. C. (2003). Structural determinants of integrin recognition by talin. *Mol Cell* **11**, 49-58.

CHAPTER 2

Development of Software to Model the Integrin

α IIb/ β 3 Transmembrane Heterodimer

INTRODUCTION

The prediction of membrane protein structure is a particularly important endeavor given the relative difficulty of determining structures experimentally. Despite impressive progress in the development of force fields and energy scoring functions,^{1; 2; 3; 4} current modeling protocols cannot reliably identify the native conformation of most membrane proteins without additional information from experimental analysis. Successful protocols exploit information derived from sequence analysis,^{5; 6; 7; 8; 9; 10} spectroscopy,^{5; 8; 9; 11; 12; 13; 14} cross-linking,^{8; 11; 12} and/or mutagenesis,^{8; 15; 16} or the known structures of homologous¹⁷ or non-homologous proteins.¹

Brunger and coworkers developed a conformational searching algorithm to predict membrane helix oligomers, and this protocol has been applied to glycophorin A,^{15; 18} phospholamban,¹⁹ the M2 and CM2 proton channels in the influenza A and C viruses,²⁰ and the *vpu* protein from HIV-1.²⁰ In this protocol, conformational space is searched exhaustively and then low energy structures

are screened to identify conformations that are most consistent with experimental findings, frequently mutagenesis results. Alternatively, orthologous or homologous proteins can be modeled in parallel and the native conformation can be identified from the intersection of low energy structures generated for each ortholog/homolog.^{6; 7} Briggs et al. demonstrated that this approach will identify the native conformation of glycophorin A.⁷

The above methods use mutagenesis and phylogenetic information to identify correctly modeled conformations after an exhaustive computational search. Our methods use mutagenesis or phylogenetic information to simplify the computational search. Prior modeling efforts provide precedent for this approach. Most membrane protein modeling protocols, including the protocols described above, simplify their computational search by only considering conformations with α -helical secondary structure. Non-helical conformations can be excluded with structural restraints that maintain appropriate distances between the hydrogen bonding partners in an α -helix. The structural restraints are enforced by a term in the energy function which is analogous to the energy function that enforces distances predicted by NMR spectroscopy. Similar structural restraints can enforce tertiary or quaternary structure predicted by spectroscopy, crosslinking, mutagenesis, and/or sequence analysis.^{5; 8; 11; 12; 13; 14}

In contrast with structural restraints which enforce distances and angles, we use thermodynamic restraints which enforce relative differences in energy for an ensemble of mutations.

During each docking step of a Monte Carlo simulated annealing cycle, we compute the difference in dimerization energy ΔE between the wild type and an ensemble of point mutations for the step's conformation. The ΔE value should be near zero for mutations that do not affect protein stability and function. Likewise, we expect an unfavorable ΔE for destabilizing mutations. Conformations with computed ΔE values that are inconsistent with experimental findings are penalized by increasing their computed energies. Thus our modeling protocol creates a selective advantage for models that are consistent with experimental mutagenesis results.

We optimized our structure prediction protocol to study self associating transmembrane (TM) helices. Self associating TM helices are widespread and play vital functional and structural roles such as in the T cell receptor,²¹ the M2 proton channel,²² and phospholamban.²³ For reviews see Engelman et al.²⁴ and Senes et al.²⁵ We calibrated our structure prediction protocol with glycophorin A (GPA), a small, well-characterized protein that dimerizes along two TM helices. We applied our structure prediction protocol to the TM region of BNIP3 because several labs were working to determine its NMR structure which could validate our modeling protocol. Both GPA and BNIP3 contain GXXXG motifs.^{26; 27} The GXXXG motif is the most overrepresented sequence motif found in TM helices²⁸ and presents a strong dimerization signal.²⁹ The motif consists of two glycines separated by three amino acids. The glycines enable close contact which permits $C\alpha-H\cdots O$ hydrogen bonding between helix backbones.^{30; 31} $C\alpha-H\cdots O$

hydrogen bonds, which are found in many proteins, may be important for stabilizing membrane proteins.

Previously, our lab derived a conceptual basis for structure prediction guided by mutagenesis data using lattice models.³² Mutagenesis data can compensate for the limitations of a force field while permitting a significant increase in modeling speed. Here we extend the approach to the docking of TM helices, and we use the method to predict a molecular model for the TM region of the BNIP3 apoptosis factor which was subsequently validated by an experimental NMR structure.³³

MATERIALS AND METHODS

Potential Function

We define dimerization energy $E_{dimerization}$ as the potential energy of two helices in a docked conformation minus the energy of the two helices separated by 100 Å. Potential energies were calculated *in vacuo* with the AMBER united-atom force field for van der Waals interactions.³⁴ We softened the potential function to mitigate artifacts from rigid body docking. Favorable van der Waals interactions were calculated with a 12-6 Lennard-Jones potential in which the van der Waals radii were scaled to 95%. Unfavorable van der Waals interactions were dampened using a linear repulsive term with a 10 kcal per mole maximum repulsion.³⁵

Calibration

We calibrated our scoring function with glycoporphin A's (GPA) published structure and mutational data.^{26; 36} 10,000 GPA helix-helix dimer pairs with computed dimerization energies less than -10 kcal per mole were generated during ten Monte Carlo simulated annealing (MCSA) cycles similar to the MCSA protocol described below. A plot of these structures is shown in figure 1a. Our scoring function was parameterized to selectively penalize low energy dimers with an RMSD greater than 1.5 \AA with the GPA structure. Specifically, we averaged the scores of the ten best scoring models with an RMSD less than 1.5 \AA and the ten best scoring models with an RMSD greater than or equal to 1.5 \AA and maximized the difference between these two averages. Other more sophisticated optimization methods resulted in similar parameterization. The scoring function is described in greater detail in the results section.

MCSA Implementation

The GPA TM helices consisted of residues Ile73 through Ile95 and the BNIP3 helices consisted of residues Val164 through Gly184. A single proline occurs at position 167 near the N-terminus of the BNIP3 helix and should not affect the backbone geometries of downstream residues. It was modeled as alanine to eliminate its significance relative to the method we were testing. ϕ , φ , and ω angles were fixed at -65° , -40° , and 180° , respectively, for all amino acids.

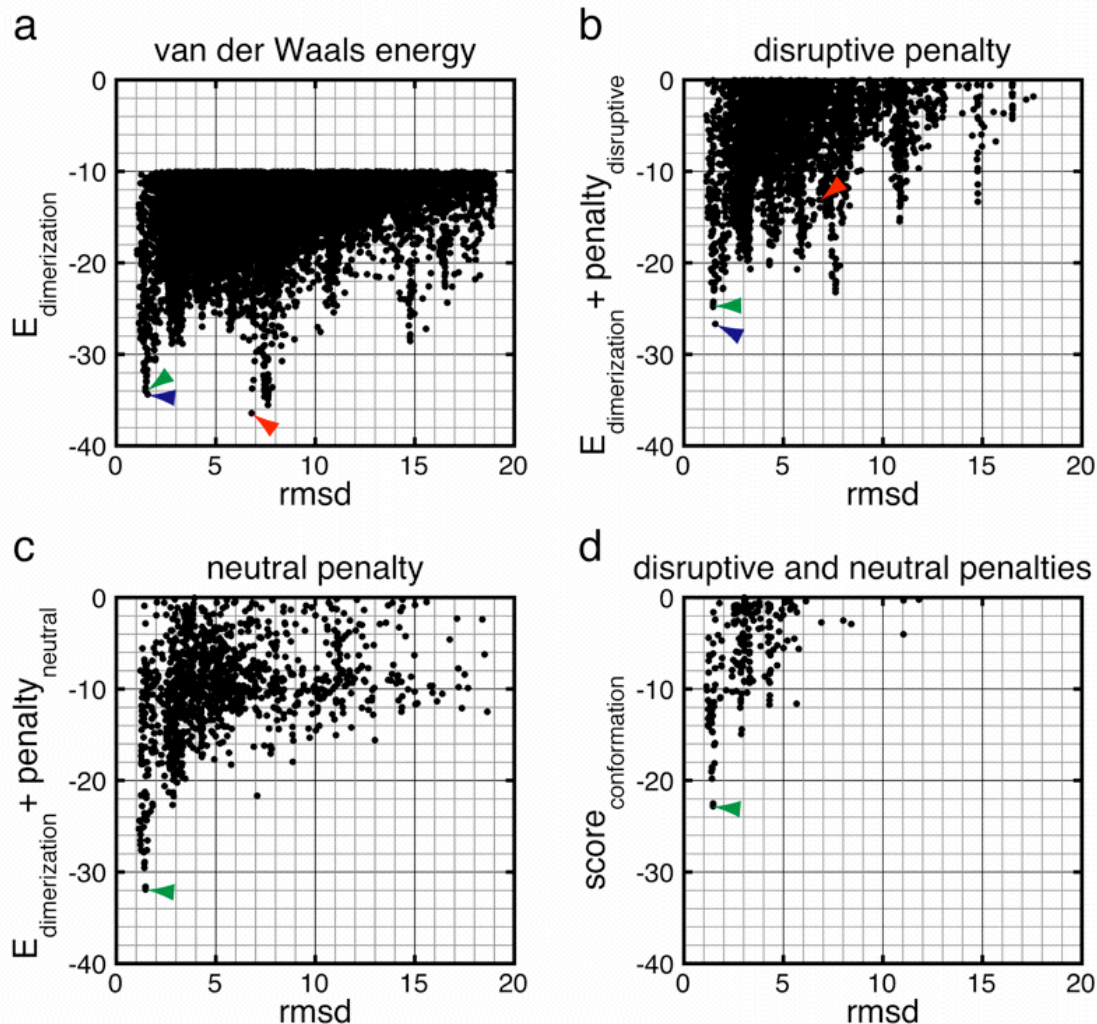


Fig 1. Each graph depicts the same set of conformations plotted with a different score on the y-axis. The arrowheads track three conformations through the graphs and do not appear when a score is greater than 0 kcal. Native-like conformations of glycoprotein A (blue and green arrowheads) cannot be distinguished from other low energy conformations (red arrowhead) using computed van der Waals energies (a). Penalties can be added to the van der Waals energies of models that are inconsistent with known disruptive mutations (b), neutral mutations (c), or both (d) to help distinguish native-like models from non-native conformations. The penalties eliminate the non-native energy wells at the expense of removing some conformations from the native energy well.

TABLE I. List of Mutations Used as Inputs to Model Glycophorin A and BNIP3

	neutral mutations	disruptive mutations
GPA, mutational data ^{26,a}	V80L, A82L, G86A, G86L, L89A, L90A, S92L, G94A, G94L	L75A, G79A, G79L, G83A, G83L, V84L
GPA, sequence data	T74A, I77V, F78L, M81I, V84I, V84T, L90F, L90S, S92Y, G94C, G94S	none
BNIP3, mutational data ²⁷	G178I, H173W, A176C	G180I, G180A, H173A

^aThe GPA mutational data was also used for calibration.

Our MCSA protocol utilizes previously characterized point mutations for protein structure prediction. The mutants used in this analysis are listed in table I. GPA mutants were taken from Lemmon et al.²⁶ and BNIP3 mutants were taken from Suljito et al.²⁷ The GPA mutations were selected to preferentially probe the effect of alanine and leucine scanning mutagenesis on glycine and leucine amino acids because these mutations are commonly made and often insightful. Less mutational data is available for BNIP3 and we used all point mutations that had unambiguous effects. Additionally, naturally occurring sequence variation can serve as a source of mutational information.³⁷ Single nucleotide polymorphisms (SNPs), orthologs, and homologs of GPA were identified with BLAST.³⁸ Each mismatched amino acid was scored as a neutral mutation such that each sequence could contribute multiple neutral mutations to the analysis.

A MCSA cycle began after finding a randomly generated conformation with a favorable dimerization energy (i.e. a dimerization energy less than zero). A conformation is defined by the six parameters that relate two rigid helices in space (figure 2). During each docking step of a MCSA cycle, there was an equal

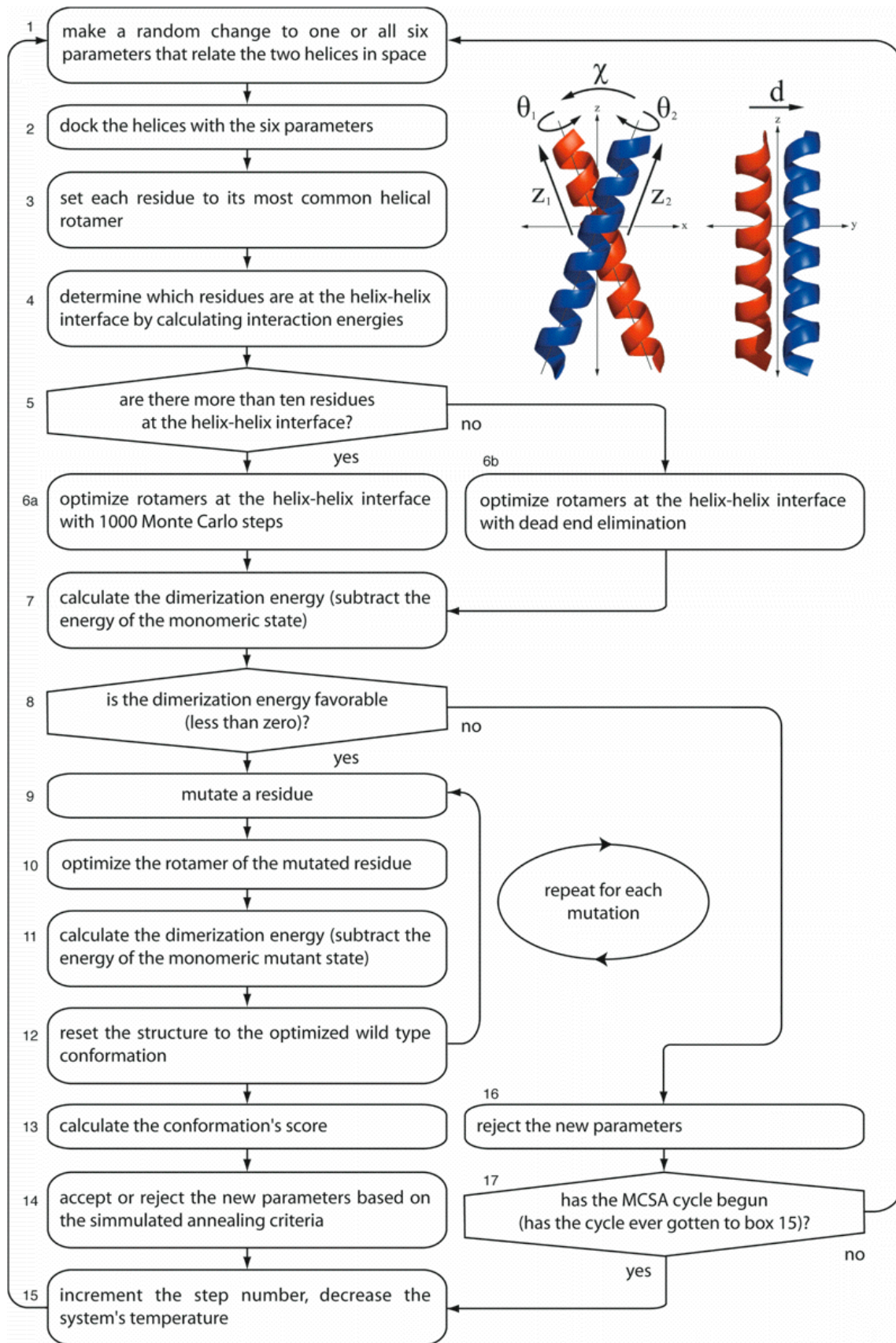
probability of changing any one parameter or all six parameters to random values. In nine initial MCSA cycles, all possible values were allowed; the phase of each helix θ_1 and θ_2 varied 360° , the crossing angle χ varied from -90° to 90° (parallel structures only), translations along a helical axis z_1 and z_2 varied from -15 to 15 \AA , and the diameter d varied from 5 to 9 \AA .

TABLE II.
List of Glycophorin A Single Nucleotide Polymorphisms (SNPs), Orthologs, and Homologs (i.e. GPB)
from which Mutational Information can be Inferred

GPA	⁷³ ITLIIIFGVMAGVIGTILLISYGI ⁹⁵
Human, SNP	ITLIIIFGV <u>I</u> AGVIGTILLISYGI
Human, SNP	ITLIIIFGVMAG <u>T</u> IGTILLISYGI
Japanese Macaque	I <u>A</u> LIIIFGVMAGVIGTIL <u>F</u> ISYGI
Orangutan	ITLII <u>V</u> FGVMAGVIGTILLISY <u>C</u> I
Gorilla	ITLIIIFGVMAG <u>I</u> IGTILLISY <u>S</u> I
Gibbon	ITLIIIFGVMAGVIGTIL <u>S</u> ISY <u>C</u> I
Chimpanzee, GPA	ITLIIIFGVMAGVIGTILLI <u>Y</u> <u>S</u> I
Chimpanzee, GPB	I <u>L</u> LII <u>L</u> GV <u>M</u> AG <u>I</u> IGTILLISY <u>C</u> I

Negative identities are underlined. Each negative identity was scored as a neutral mutation in our MCSA protocol.

Fig 2 (next page). The flowchart for a global Monte Carlo simulated annealing cycle. Idealized α -helices are docked with six orthogonal parameters: θ_1 and θ_2 are rotations of a helix about its helical axis, χ is the crossing angle, z_1 and z_2 are translations of a helix along its helical axis, and d is the distance between each helical axis. Global MCSA cycles explore all parameter space. Subsequent MCSA cycles restrict parameter space to the energy well defined by nine global cycles and more thoroughly explore rotamer space (see text). Each MCSA cycle consists of 50,000 docking steps which start at box 1 and end at box 15. The simulated annealing temperature undergoes exponential decay from 10,000 to 10 K over the 50,000 docking steps in each cycle.



After docking, rotamers were set to their most common helical rotamer. The helix-helix interface was defined by identifying residues with interaction energies greater than 10^{-3} kcal per mole in magnitude. The rotamers of these residues were optimized by Monte Carlo or dead end elimination (DEE)³⁹ using the Goldstein criterion⁴⁰ depending on how many residues were at the helix-helix interface. The rotamers of the monomeric states were optimized by DEE for every residue. Rotamers were selected from a library containing the most common helical rotamers which consists of one to three members for each amino acid.⁴¹

After rotamer optimization, the conformation's dimerization energy was calculated. If the dimerization energy was favorable, we calculated the dimerization energy of select point mutations. The dimerization energies were used to calculate a score for the conformation (see equations 1-3 in the results section). The score was used to accept or reject a conformation based on our simulated annealing criteria. If a conformation was accepted, its six parameters were passed to the next docking step. If a conformation was rejected, the parameters of the last accepted conformation were passed to the next step. Regardless of whether a conformation was accepted or rejected, its parameters and score were recorded to restrict conformational space in subsequent MCSA cycles (see below). Each MCSA cycle consisted of 50,000 docking steps with an exponential temperature decay from 10,000 to 10 K.

MCSA, Restricted Parameter Space

Nine global, independent MCSA cycles are sufficient to define an energy well that is consistent with an ensemble of mutations. Afterward, parameters are restricted to ± 2 standard deviations from their mean values for structures with a score within 10 kcal of the best structure. This typically restricts parameter space by three orders of magnitude and allows for a fine-grained search through conformational space. MCSA cycles were repeated as described above with additional optimization of χ angles: rotamers were optimized with DEE for every amino acid at the helix-helix interface and the rotamer's χ angles were further optimized with a grid/Monte Carlo search. Similarly, rotamers of the monomeric states were optimized by DEE and χ angles were further optimized with a grid/Monte Carlo search. Our figures and analyses are based on the single best scoring models for GPA/BNIP3 that were identified during the fine-grained search. These models are representative of the best scoring clusters.

Implementation of Penalties used to Score Mutations

The penalties used to score mutations (eq. 2 and 3 in the Results section) were implemented to handicap the energy of each conformation that was considered by the simulated annealing criterion. The penalties serve to artificially increase the dimerization energy of conformations that are in poor thermodynamic agreement with experimentally characterized point mutations. Thus a low energy structure that does not agree with experimental findings is less likely to pass the

simulated annealing criterion than a low energy structure that is in good agreement with experimental findings. This provides a selective advantage for parameters that are in good agreement with experimental results, resulting in final models that agree well with experimental data.

BNIP3 Hydrogen Bonding

Mutational analysis of the BNIP3 TM region predicts an intermolecular hydrogen bond involving His173.²⁷ Our model places His173 and Ser172 in proximity but not in contact. The χ angles of His173 and Ser172 can be changed to isoenergetic states that support a hydrogen bond between the His173 N δ or N ϵ and the Ser172 O γ . No other interhelical contact can support a hydrogen bond with His173 or Ser172.

The C-terminus of BNIP3 was not modeled because it is not embedded in the membrane. We extended the C-terminal helix of the final BNIP3 model and set each amino acid to its lowest energy rotamer. This positions both Thr188 residues in favorable van der Waals contact. The geometry and distance between the Thr188 hydroxyls predicts a hydrogen bonding interaction prior to any optimization step.

RESULTS

An ideal structure prediction protocol would identify a native protein fold based solely on energy calculations. A minimal Monte Carlo simulated annealing

protocol that computes only van der Waals energies cannot distinguish the native conformation of glycoporphin A (GPA) from other low energy conformations (figure 1a). However the protocol is sufficient to identify low energy conformations that are consistent with the published NMR structure. This result is similar to more sophisticated modeling protocols which cannot distinguish the native GPA fold from other low energy conformations without additional information.^{15; 18}

Mutagenesis data can help distinguish the native GPA structure from other low energy conformations. For example, residues that are important for folding, as assessed from experimental mutagenesis studies, should lie at the helix-helix interface of a correct model and show strong interaction energies within the structure. Treutlien et al. reported a GPA model which has residue interaction energies that best correlate with mutagenesis results.^{15; 18}

To automate the process of structure prediction using mutagenesis data, we compute the difference in dimerization energy ΔE between the wild type and select point mutants during each docking step. The ΔE value can be compared to experimental results to allow for the penalization of models that are inconsistent with mutagenesis data. In the current work we considered two phenotypes, disruptive and neutral mutations, which have been previously characterized experimentally. By definition, disruptive mutations have unfavorable dimerization energies relative to the wild type while neutral mutations are isoenergetic with the wild type.

We built a scoring function to penalize conformations that are inconsistent with known disruptive and neutral mutations. We calculate a score for each modeled conformation by summing its van der Waals energy and two penalties (equation 1).

$$score_{conformation} = E_{dimerization} + penalty_{disruptive} + penalty_{neutral} \quad 1$$

$E_{dimerization}$ is the energy of the dimer minus the energy of the monomeric state (see methods). $penalty_{disruptive}$ is a restraint that creates a selective advantage for conformations that are consistent with known disruptive mutations. $penalty_{neutral}$ is a restraint that creates a selective advantage for conformations that are consistent with known neutral mutations.

Each disruptive mutant should have a higher computed dimerization energy than the wild type, otherwise we penalize a conformation's score. After examining a number of functions, we found equation 2 to be most effective at penalizing low energy conformations that are inconsistent with one or more disruptive mutations. This function is plotted in figure 3a.

$$penalty_{disruptive} = \frac{\alpha_{disruptive}}{n} \sum_{i=1}^n \ln \left(\frac{1}{1 + e^{-\beta \Delta E_i}} \right) \quad 2$$

ΔE_i is the computed difference in energy between disruptive mutant i and the wild type protein for a given conformation. n is the total number of disruptive mutants considered. The α coefficient scales the magnitude of the penalty and the β coefficient shifts the curve from left to right. The optimized values used for these coefficients are -60.1 and 0.521 , respectively.

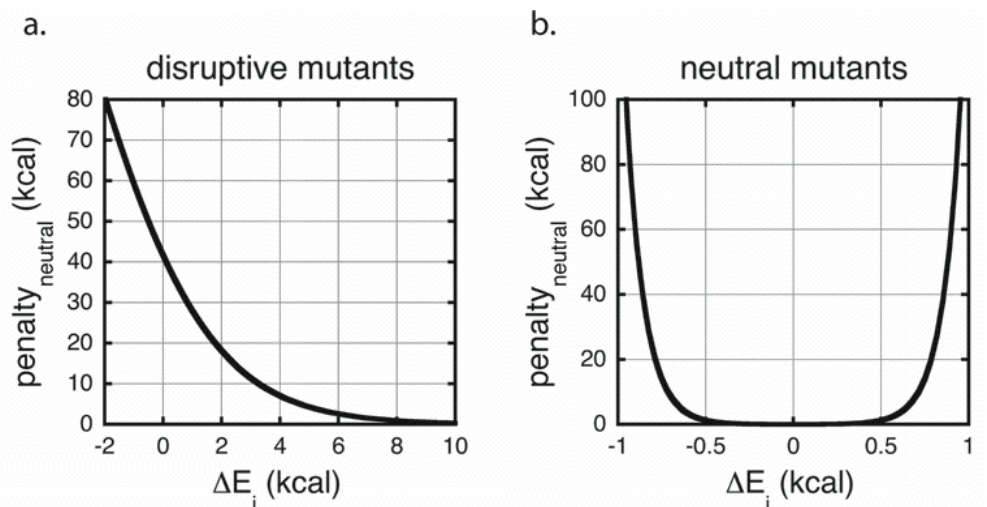


Fig 3. These functions penalize models that are inconsistent with a given mutation. A disruptive mutation should have a positive ΔE (a) and a neutral mutation should have a near zero ΔE (b). If a model is inconsistent with an experimental mutagenesis result, its computed energy is increased by adding a normalized penalty that corresponds to the computed ΔE for that mutant. This creates a selective advantage for models that are consistent with mutational data.

Each neutral mutant should have a computed dimerization energy that is comparable to the wild type. When considering GPA, we found that the optimal scoring function penalizes models that have one or more neutral mutations with computed stabilizing *or* destabilizing effects (equation 3).

$$penalty_{neutral} = \frac{\alpha_{neutral}}{n} \sum_{i=1}^n e^{\beta|\Delta E_i|} \quad 3$$

Our optimized α and β values are 1.02E-2 and 9.63, respectively.

We parameterized our scoring function to best penalize non-native low energy conformations (see methods). When considering the mutations in table I,

the scoring function creates a new energy landscape for GPA and the new global energy minimum corresponds to the native conformation (figure 1d). Each conformation is penalized by one or both penalties. It is necessary to penalize some native-like conformations in order to effectively penalize all non-native conformations. When the scoring function is applied to our MCSA protocol, some native-like conformations are eliminated from the analysis so that we can eliminate every low energy non-native conformation.

We repeated the MCSA protocol for GPA with the two penalties added to the van der Waals energy function. The best scoring model had a $C\alpha$ root mean squared deviation (RMSD) with GPA of 1.30 Å (figure 4a). This is an excellent result considering the lowest RMSD structure accessible to our search algorithm has an RMSD of 0.65 Å, and the TM regions of different NMR structures for GPA vary by up to 1.80 Å. (There are twenty NMR structures for GPA in its pdb file 1AFO; we used model 1 for calibration and RMSD calculations unless otherwise noted.) Finally, our model recovers the $C\alpha$ hydrogen bonding interactions described by Senes et al. even though no hydrogen bonding or electrostatic term was used to generate this model.³¹

Briggs et al. demonstrated that phylogenetic analysis is sufficient to identify the native conformation of GPA.⁷ We repeated the MCSA protocol with the naturally occurring sequence variants from table II. Encouragingly, the best scoring model had an RMSD with GPA of 1.37 Å (figure 4b). This model also recovers the $C\alpha$ hydrogen bonding interactions described by Senes et al.³¹

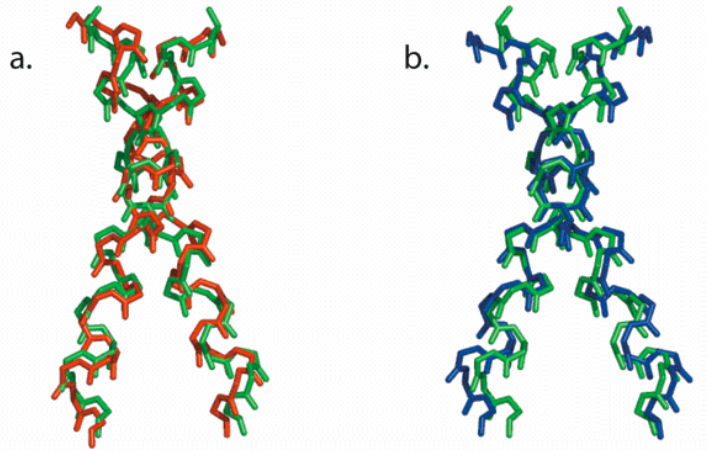


Fig 4. The glycoporphin A model generated with thermodynamics data (red) fit to the wild type backbone (green), RMSD 1.30 Å (a). The glycoporphin A model generated with naturally occurring sequence variation data (blue) fit to the wild type backbone (green), RMSD 1.37 Å (b).

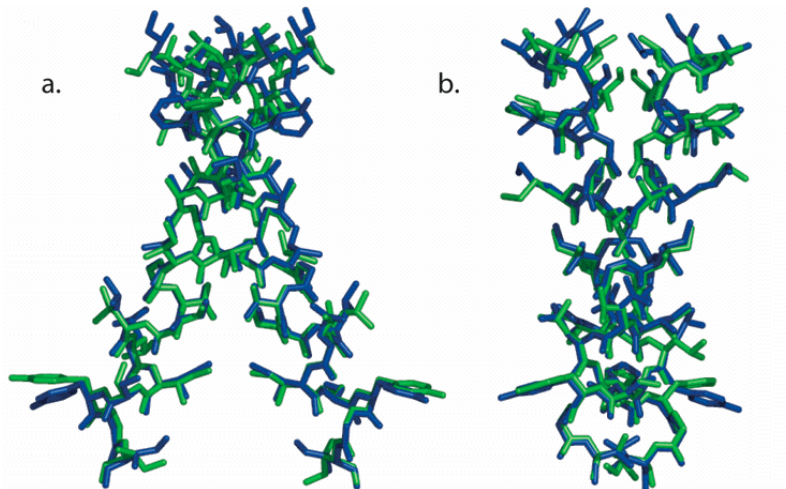


Fig 5. The glycoporphin A model generated with naturally occurring sequence variation data (blue) fit to structure 19 from the glycoporphin A pdb (green), RMSD 0.54 Å. The glycoporphin A pdb file 1AFO contains twenty different NMR structures and structure 19 shows the strongest $C\alpha-H\cdots O$ hydrogen bonding interactions. (Structure 1 was used for calibration and all other RMSD calculations.) The two figures are related by a rotation about the vertical axis.

In fact, this structure has a $C\alpha$ RMSD of 0.54 Å with the GPA NMR structure that shows the strongest $C\alpha$ -H•••O hydrogen bonding interactions (figure 5).

To further explore our method, we modeled the TM region of the BNIP3 apoptosis factor which also contains a GXXXG motif. The resultant model predicts a right-handed crossing angle of -42° and a similar conformation as GPA. Our model accommodates an interhelix hydrogen bond between His173A and Ser172B. The model is inconsistent with other interhelical hydrogen bonding partners for either His173 or Ser172. Additionally, the model is consistent with an interhelix hydrogen bond between Thr188A and Thr188B (figure 6). Finally, it predicts six interhelical $C\alpha$ hydrogen bonds. These include symmetrical hydrogen bonds between the Ile177A $C\alpha$ and Ala176B carbonyl, the Gly180A $C\alpha$ and Ile177B carbonyl, and the Ile181A $C\alpha$ and Gly180B carbonyl.

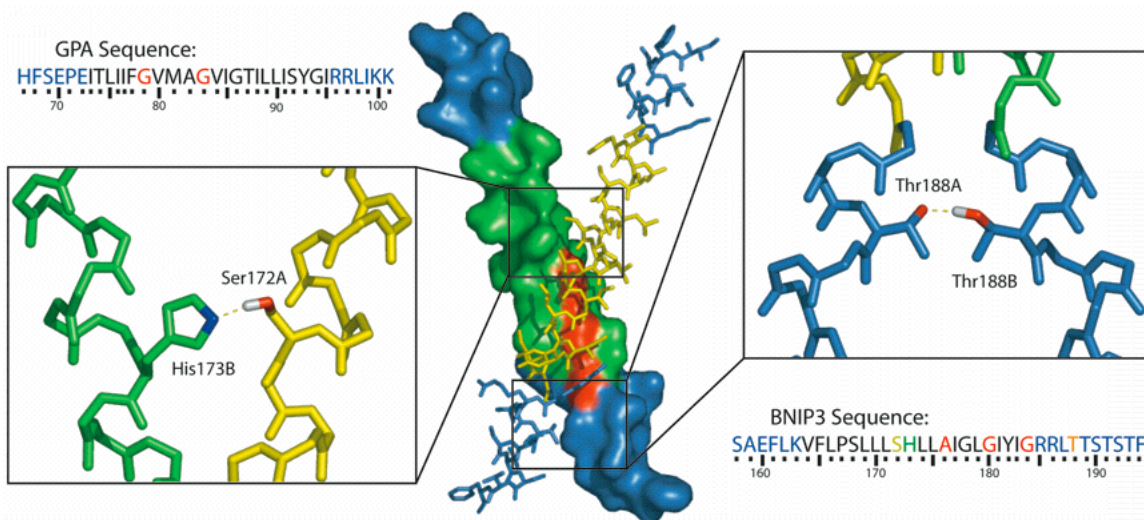


Fig 6. Molecular model of the BNIP3 homodimer. The TM helices dimerize along consecutive AXXXG and GXXXG motifs (red). The TM regions of the protein are colored green and yellow, and aqueous regions are colored blue.

DISCUSSION AND CONCLUSIONS

We developed a novel protein structure prediction strategy to model interactions between self-associating TM helices based on experimental mutagenesis results. The semi-quantitative determination of helix dimerization in response to mutagenesis was sufficient for this analysis. In the case of GPA, naturally occurring sequence variation can substitute for mutational information to drive structure prediction. However we suggest that experimental mutagenesis information be favored when available because sequence drift can be accompanied by structural drift.

Scoring Function Analysis

Our scoring functions provide a mathematical framework for the incorporation of experimental mutagenesis data into modeling protocols. The scoring function is the linear combination of a van der Waals term, a penalty that evaluates disruptive mutations, and a penalty that evaluates neutral mutations. The neutral penalty may be less intuitive than the disruptive penalty. If a neutral mutant has a computed destabilizing effect, we penalize the model because we expect disruptive mutants to have computed destabilizing effects. If a neutral mutant has a computed stabilizing effect, the interpretation is less straightforward because stabilizing mutations are often accompanied by structural changes. For example, many mutations stabilize the overall tetramerization of the M2 proton channel, however they stabilize different accessible conformations of the

channel.⁴² If a single conformation is considered, stabilizing mutations are rare and small in magnitude. Also, several mutations stabilize the integrin α IIb TM homodimer but each is predicted to stabilize a non-native conformation.⁴³ Therefore, if we calculate that a neutral mutation is stabilizing, the conformation may be accessible, however it is likely not native. Thus strongly stabilizing mutations are penalized in our method, and they indeed help discriminate the native fold from alternately folded conformations.

We calibrated our penalization functions with the well-characterized TM helix dimer from glycoporphin A. It is possible that the parameters we developed are most useful at predicting GPA-like structures. However, our protocol identified a conformation that is different from the GPA conformation for two TM helices in tetraspanin CD9.⁴⁴ Additionally, it identified both GPA-like and non-GPA-like structures for the integrin α IIb/ β 3 TM heterodimer.⁴⁵ Furthermore, it can distinguish between different subtypes of GXXXG motifs.⁴³ Each of the above proteins contains one or more glycine-containing motifs which rely on geometric complementarity and potentially C α -H...O hydrogen bonds for stabilization, and our methods might prove especially useful at predicting these types of interactions. Clearly, it would be desirable to include additional energetic terms to allow the scoring of hydrogen bonding interactions, electrostatics, rotamer strain, and other features that are important for association. We chose, however, to use a simple energy function to analyze its performance in combination with mutagenesis data. We demonstrated that a simple energy function is sufficient to

predict the association of TM helix dimers when used in combination with mutagenesis data. A more sophisticated energy function may be necessary for the accurate prediction of more complex interactions.

Qualitative thermodynamics information was sufficient to generate reasonable models for TM helix dimers, however quantitative information may increase structure prediction accuracy for larger proteins. For example, each disruptive mutant could be weighted proportionally to its experimentally determined degree of destabilization. A weighting coefficient could then be applied to the disruptive penalty to scale its magnitude for different mutations. However the penalties can be parameterized any number of ways to allow the addition of quantitative thermodynamics information.

Molecular Model of BNIP3

BNIP3 is a “BH3-only” member of the Bcl-2 family of apoptosis factors. When localized to the mitochondrial membrane, BNIP3 inhibits anti-apoptotic proteins which otherwise block pore formation.⁴⁶ Through this mechanism BNIP3 permits the release of mitochondrial contents causing cell death. The function of BNIP3 homodimerization remains unknown, however its tremendous stability suggests that its function requires a dimeric conformation.²⁷

The BNIP3 TM dimer is more stable than the prototypical TM dimer glycophorin A. Sulistijo et. al predicted that concurrent AXXXG and GXXXG motifs stabilize the BNIP3 homodimer in addition to electrostatic interactions

involving His173, and site directed mutagenesis corroborated their hypothesis.²⁷ Our model confirmed that the AXXXG and GXXXG motifs are at the homodimer interface and predicted an interchain hydrogen bond between His173 and Ser172. Following the publication of our model, the BNIP3 NMR structure was released and the two conformations have a $C\alpha$ RMSD of 1.06 Å, validating the model, and more importantly, the modeling method.³³

Conclusion

We developed a novel molecular modeling protocol that selects modeled protein conformations based on experimental mutagenesis results. In contrast to modeling protocols that enforce distance or angular restraints, we examine the relative stabilities for an ensemble of point mutations for each modeled conformation and create a selective advantage for conformations that are consistent with experimental findings. This approach is sufficient to identify the native conformations of the glycoporphin A and BNIP3 TM dimers without a sophisticated force field or an exhaustive search through conformational space. Importantly, the methodology was independently validated by the BNIP3 NMR structure.

References

1. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. (2006). Multipass membrane protein structure prediction using Rosetta. *Proteins* **62**, 1010-25.
2. Chen, J., Im, W. & Brooks, C. L., 3rd. (2005). Application of torsion angle molecular dynamics for efficient sampling of protein conformations. *J Comput Chem* **26**, 1565-78.
3. Lazaridis, T. (2003). Effective energy function for proteins in lipid membranes. *Proteins* **52**, 176-92.
4. Chen, Z. & Xu, Y. (2006). Energetics and stability of transmembrane helix packing: a replica-exchange simulation with a knowledge-based membrane potential. *Proteins* **62**, 539-52.
5. Fleishman, S. J., Harrington, S., Friesner, R. A., Honig, B. & Ben-Tal, N. (2004). An automatic method for predicting transmembrane protein structures using cryo-EM and evolutionary data. *Biophys J* **87**, 3448-59.
6. Gottschalk, K. E., Adams, P. D., Brunger, A. T. & Kessler, H. (2002). Transmembrane signal transduction of the alpha(IIb)beta(3) integrin. *Protein Sci* **11**, 1800-12.
7. Briggs, J. A., Torres, J. & Arkin, I. T. (2001). A new method to model membrane protein structure based on silent amino acid substitutions. *Proteins* **44**, 370-5.

8. Herzyk, P. & Hubbard, R. E. (1995). Automated method for modeling seven-helix transmembrane receptors from experimental data. *Biophys J* **69**, 2419-42.
9. Adair, B. D. & Yeager, M. (2002). Three-dimensional model of the human platelet integrin alpha IIb beta 3 based on electron cryomicroscopy and x-ray crystallography. *Proc Natl Acad Sci U S A* **99**, 14059-64.
10. Baldwin, J. M., Schertler, G. F. & Unger, V. M. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J Mol Biol* **272**, 144-64.
11. Gottschalk, K. E. (2005). A coiled-coil structure of the alpha IIb beta 3 integrin transmembrane and cytoplasmic domains in its resting state. *Structure* **13**, 703-12.
12. Sorgen, P. L., Hu, Y., Guan, L., Kaback, H. R. & Girvin, M. E. (2002). An approach to membrane protein structure without crystals. *Proc Natl Acad Sci U S A* **99**, 14037-40.
13. Sansom, M. S., Sankararamakrishnan, R. & Kerr, I. D. (1995). Modelling membrane proteins using structural restraints. *Nat Struct Biol* **2**, 624-31.
14. Kukol, A., Adams, P. D., Rice, L. M., Brunger, A. T. & Arkin, T. I. (1999). Experimentally based orientational refinement of membrane protein models: A structure for the Influenza A M2 H⁺ channel. *J Mol Biol* **286**, 951-62.

15. Treutlein, H. R., Lemmon, M. A., Engelman, D. M. & Brunger, A. T. (1992). The glycoporphin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry* **31**, 12726-32.
16. Simmerman, H. K., Kobayashi, Y. M., Autry, J. M. & Jones, L. R. (1996). A leucine zipper stabilizes the pentameric membrane domain of phospholamban and forms a coiled-coil pore structure. *J Biol Chem* **271**, 5941-6.
17. Forrest, L. R., Tang, C. L. & Honig, B. (2006). On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys J* **91**, 508-17.
18. Adams, P. D., Engelman, D. M. & Brünger, A. T. (1996). Improved prediction for the structure of the dimeric transmembrane domain of glycoporphin obtained through global searching. *Proteins* **26**, 257-261.
19. Adams, P. D., Arkin, I. T., Engelman, D. M. & Brunger, A. T. (1995). Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nature Structural Biology* **2**, 154-162.
20. Torres, J., Kukol, A. & Arkin, I. T. (2001). Mapping the energy surface of transmembrane helix-helix interactions. *Biophys J* **81**, 2681-92.

21. Call, M. E., Pyrdol, J., Wiedmann, M. & Wucherpfennig, K. W. (2002). The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell* **111**, 967-79.
22. Sugrue, R. J. & Hay, A. J. (1991). Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel. *Virology* **180**, 617-24.
23. Arkin, I. T., Adams, P. D., MacKenzie, K. R., Lemmon, M. A., Brünger, A. T. & Engelman, D. M. (1994). Structural organization of the pentameric transmembrane alpha-helices of phospholamban, a cardiac ion channel. *EMBO Journal* **13**, 4757-4764.
24. Engelman, D. M., Chen, Y., Chin, C. N., Curran, A. R., Dixon, A. M., Dupuy, A. D., Lee, A. S., Lehnert, U., Matthews, E. E., Reshetnyak, Y. K., Senes, A. & Popot, J. L. (2003). Membrane protein folding: beyond the two stage model. *FEBS Lett* **555**, 122-5.
25. Senes, A., Engel, D. E. & DeGrado, W. F. (2004). Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr Opin Struct Biol* **14**, 465-79.
26. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J. & Engelman, D. M. (1992). Sequence Specificity in the Dimerization of Transmembrane alpha Helices. *Biochemistry* **31**, 12719-12725.

27. Sulistijo, E. S., Jaszewski, T. M. & MacKenzie, K. R. (2003). Sequence-specific dimerization of the transmembrane domain of the "BH3-only" protein BNIP3 in membranes and detergent. *J Biol Chem* **278**, 51950-6.
28. Senes, A., Gerstein, M. & Engelman, D. M. (2000). Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* **296**, 921-36.
29. Russ, W. P. & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **296**, 911-9.
30. Javadpour, M. M., Eilers, M., Groesbeek, M. & Smith, S. O. (1999). Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. *Biophys J* **77**, 1609-18.
31. Senes, A., Ubarretxena-Belandia, I. & Engelman, D. M. (2001). The Calpha ---H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci U S A* **98**, 9056-61.
32. Nanda, V. & DeGrado, W. F. (2005). Automated use of mutagenesis data in structure prediction. *Proteins* **59**, 454-66.
33. Bocharov, E. V., Pustovalova, Y. E., Pavlov, K. V., Volynsky, P. E., Goncharuk, M. V., Ermolyuk, Y. S., Karpunin, D. V., Schulga, A. A., Kirpichnikov, M. P., Efremov, R. G., Maslennikov, I. V. & Arseniev, A. S. (2007). Unique dimeric structure of BNip3 transmembrane domain

- suggests membrane permeabilization as a cell death trigger. *J Biol Chem* **282**, 16256-66.
34. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr. & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765-784.
 35. Kuhlman, B. & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **97**, 10383-8.
 36. MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997). A transmembrane helix dimer: structure and implications. *Science* **276**, 131-133.
 37. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306-1310.
 38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-10.
 39. Desmet, J., De Maeyer, M., Hazes, B. & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539-542.
 40. Goldstein, R. F. (1994). Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys J* **66**, 1335-40.

41. Bower, M. J., Cohen, F. E. & Dunbrack, R. L., Jr. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J Mol Biol* **267**, 1268-82.
42. Stouffer, A. L., Nanda, V., Lear, J. D. & DeGrado, W. F. (2005). Sequence determinants of a transmembrane proton channel: an inverse relationship between stability and function. *J Mol Biol* **347**, 169-79.
43. Li, R., Gorelik, R., Nanda, V., Law, P. B., Lear, J. D., DeGrado, W. F. & Bennett, J. S. (2004). Dimerization of the transmembrane domain of Integrin α IIb subunit in cell membranes. *J Biol Chem* **279**, 26666-73.
44. Kovalenko, O. V., Metcalf, D. G., DeGrado, W. F. & Hemler, M. E. (2005). Structural organization and interactions of transmembrane domains in tetraspanin proteins. *BMC Struct Biol* **5**, 11.
45. Li, W., Metcalf, D. G., Gorelik, R., Li, R., Mitra, N., Nanda, V., Law, P. B., Lear, J. D., DeGrado, W. F. & Bennett, J. S. (2005). A push-pull mechanism for regulating integrin function. *Proc Natl Acad Sci U S A* **102**, 1424-9.
46. Kuwana, T., Mackey, M. R., Perkins, G., Ellisman, M. H., Latterich, M., Schneider, R., Green, D. R. & Newmeyer, D. D. (2002). Bid, Bax, and lipids cooperate to form supramolecular openings in the outer mitochondrial membrane. *Cell* **111**, 331-42.

CHAPTER 3

Multiple Approaches Converge on the Structure of the Integrin α IIb/ β 3 Transmembrane Heterodimer

INTRODUCTION

Integrins, the principal cell surface receptors responsible for linking the cytoskeleton to the extracellular matrix, are transmembrane (TM) heterodimers composed of non-covalently associated α and β subunits. Integrin molecules exist in an equilibrium between resting conformations that do not bind extracellular ligands and active conformations that both engage ligands and nucleate large intracellular complexes.^{1; 2} Agonist-induced intracellular signals shift integrins from resting to active conformations by exposing extracellular ligand-binding sites. To do so, signals must be transmitted across the membrane via the integrin's TM domain: an integrin is constrained in a resting conformation by the heteromeric association of its α and β subunits' TM domains. Moreover, disruption of this association is sufficient to induce integrin activation (figure 7).^{3; 4} Thus, the α/β TM heterodimer is a critical structure in regulating integrin function.

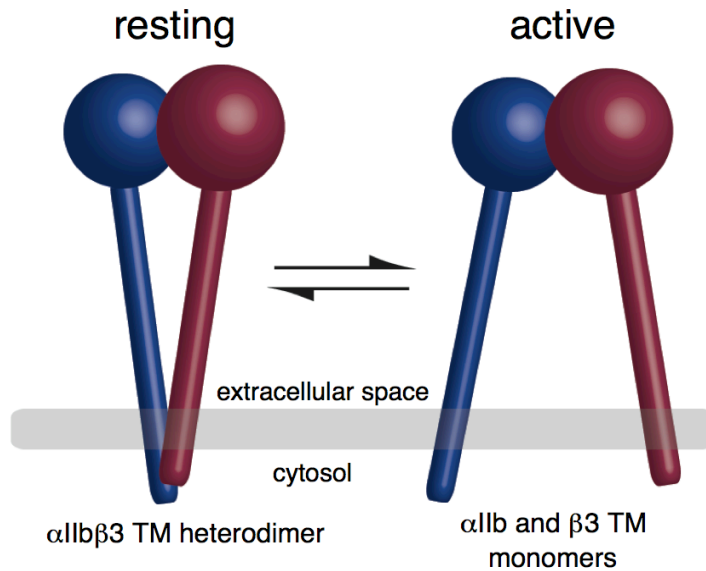


Fig 7. Integrins exist in equilibrium between resting and active conformations. In the resting conformation, the integrin's TM helices form an α/β heterodimer and the cytosolic domains are held in proximity. In the active conformation, the TM and cytosolic domains separate.

One of the most widely studied examples of regulated integrin function is the platelet integrin $\alpha\text{IIb}\beta\text{3}$. In its active conformation, $\alpha\text{IIb}\beta\text{3}$ binds fibrinogen, von Willebrand factor, or fibronectin and mediates platelet aggregation when these $\alpha\text{IIb}\beta\text{3}$ -bound ligands crosslink adjacent platelets.⁵ To prevent the deleterious formation of intravascular platelet aggregates, $\alpha\text{IIb}\beta\text{3}$ is maintained in a resting conformation on circulating platelets. Following vascular injury, $\alpha\text{IIb}\beta\text{3}$ is rapidly activated, enabling it to mediate the formation of a hemostatic platelet plug. The formation and disruption of the $\alpha\text{IIb}/\beta\text{3}$ TM domain heterodimer are key events in shifting $\alpha\text{IIb}\beta\text{3}$ between resting and active conformations. Thus,

there has been considerable effort to produce three-dimensional structural models of the TM domain heterodimer.^{6; 7; 8; 9; 10} However, each published model is substantially different and none have accounted well for the consequences of introducing mutations into the α IIb and β 3 TM domains. Because of the absence of a satisfactory model for the α IIb/ β 3 TM heterodimer, we explored two new and fundamentally different strategies to predict its structure.

In the first strategy, we utilized the Monte Carlo algorithm described in Chapter 2.¹¹ In the second strategy, we used a threading approach in which the sequences of the α IIb and β 3 TM domains were threaded onto a set of TM dimers parsed from high resolution structures in the protein data bank. Threaded structures were then scored according to their calculated energy and their agreement with experimental cysteine crosslinking results. Although the Monte Carlo and threading strategies relied on different sets of empirical data, they converged on a similar structure that likely approximates the native conformation of the α IIb/ β 3 TM heterodimer.

MATERIALS AND METHODS

Mutations used in the Monte Carlo Modeling Algorithm

We first modeled the α IIb/ β 3 TM domain heterodimer using the Monte Carlo-based structure prediction strategy described in Chapter 2.¹¹ The α IIb mutants Val969Asn, Leu970Asn, Leu974Asn, Gly975Asn, Leu983Ala and β 3 mutants Ser699Asn, Val700Asn, Gly702Asn, Ile704Asn, Leu705Asn were scored as

neutral mutations, and the α IIb mutants Gly972Asn, Gly972Ala, Gly972Leu, Gly976Leu and β 3 mutants Met701Asn and Gly708Asn were scored as disruptive mutations.^{2;3}

α IIb I₉₆₆WWVL VGVLG GLLLL TILVL AMW₉₈₈

β 3 I₆₉₃LV VLLSV MGAIL LIGLA ALLIW₇₁₅

Fig 8. Sequences of the α IIb and β 3 TM domains. Amino acids are highlighted if one or more of its mutants activate the integrin.

Comparing the Monte Carlo Interface with Other Published Structures

In order to sample every accessible dimer interface, our Monte Carlo method considers interfaces that are similar to those in published structures and theoretical interfaces that may not occur in nature. To assess whether the Monte Carlo model might reflect a natural interface, $C\alpha$ RMSDs were calculated between the model and conformations found in high resolution crystal structures from the orientations of proteins in membranes (OPM) database.¹² Of the parallel helix dimers parsed from OPM structures, 28% (113 of 400) had $C\alpha$ RMSDs less than 1.5 Å with the Monte Carlo model over at least ten residues from both the α IIb and β 3 helices which demonstrates that the Monte Carlo interface frequently occurs in nature.

Threading Known Structures with the Integrin's Sequence

Threading is the modeling of an unknown structure based on the experimentally determined structures of other proteins.^{13; 14} While it is usually applied to problems in which the protein of unknown structure has a sequence that is highly similar to a protein of known structure, we thought it could be useful for the prediction of membrane helix pairs due to the limited number of packing motifs found between membrane helices.¹⁵ The α IIb amino acids Ile966-Trp988 and β 3 amino acids Ile693-Trp715 were threaded through 214 parallel TM helix dimers parsed from pdb's 1c3w, 1e12, 1ehk, 1eul, 1fx8, 1h2s, 1iwg, 1j4n, 1jb0, 1k4c, 1kb9, 1kf6, 1kpl, 1kqf, 1l7v, 1l9h, 1m3x, 1m56, 1msl, 1nek, 1ocr, 1okc, 1pp9, 1pv6, 1pw4, 1q16, 1q90, 1qla, 1rc2, 1rh5, 1u7g, 1xfh, and 1yew, and the Monte Carlo model was threaded as an internal control. Sequences were threaded in all possible frames such that at least fifteen α IIb amino acids and fifteen β 3 amino acids overlapped at the same depth in the membrane. If the integrin sequence was longer than the template helix, only the portion of sequence for which a three-dimensional template was available was evaluated. When the template was longer than the integrin sequence, the additional amino acids were mutated to alanine to eliminate favorable contacts from the parent structure while maintaining a penalty for steric clashes. This procedure generated >50,000 models.

Each model was optimized prior to energy calculation: the side chain rotamers of each structure were selected with the SCWRL3 algorithm,¹⁶ then

each model was energy minimized in NAMD using the CHARMM force field.^{17; 18} NAMD minimization consisted of 2000 conjugant-gradient steps with a R=10 dielectric constant. Dimerization energies were calculated using the potential function described in Chapter 2, where a dimerization energy is defined as the energy of the optimized model minus the energy of the model's helices separated by 100 Å and re-optimized. The 500 lowest energy models were filtered based on whether they were consistent with experimental cysteine crosslinking results.

Disulfide bonds crosslink α IIb- β 3 amino acid pairs Gly972-Leu697, Gly972-Val700, Val969-Val696, Val971-Leu697, and Trp968-Val696 when the pair is mutated to cysteine.¹⁹ Accordingly, the distance between the C β atoms of each pair was calculated to determine whether a model was consistent with these results. Gly972 was mutated to alanine to add its C β atom, and any C β -C β distance closer than 4 Å was set to 4 Å because this distance approaches the maximum yield for cysteine crosslinking. Of the 500 low energy models, a "best" model was selected that has the shortest average distance for the five α IIb- β 3 residue pairs. It consists of a template from the 1iwg pdb for the crystal structure of bacterial multidrug efflux transporter AcrB²⁰ in which the α IIb TM amino acids Trp967-Trp988 were threaded onto 1iwg chain A residues 392-413, and the β 3 TM amino acids Ile693-Ala711 were threaded onto 1iwg chain A residues 466-484. The helices in this model were analyzed by HELANAL to characterize deviations from ideal structure and calculate interhelix crossing angles.²¹ Other cysteine mutant pairs have been analyzed in addition to the five robust

crosslinking pairs considered here, and these datapoints were saved for structure validation (below).¹⁹

Correlation with Cysteine Crosslinking Experiments

The Monte Carlo model, threaded model, and other published models were analyzed to determine whether they were consistent with the cysteine crosslinking experiments of Luo et al. that examined 120 α IIb- β 3 cysteine mutant pairs to determine the extent to which they could crosslink the α IIb and β 3 subunits.¹⁹ Cysteine mutant pairs that are efficient at crosslinking the integrin should be close in space relative to pairs that are less efficient at crosslinking, and the distances between residues in a model should correlate with published cysteine crosslinking efficiency. We correlated different models with cysteine crosslinking results as follows: for each cysteine mutant pair, the disulfide bond formation efficiency was calculated by sampling its published color density in Adobe Photoshop CS. Next, the distance between the C β atoms of each pair was calculated for a given model. Glycine was mutated to alanine to add its C β atom. A plot of the C β distance versus cysteine crosslinking efficiency was analyzed according to equation 4 using a non-linear least squares fitting routine implemented in KaleidaGraph. This formula relates the percent yield ($Y_{i,j}$) of the disulfide between the i and j residues in a given double mutant to the distance between their C β atoms $d_{i,j}$ in a given model.

$$Y_{i,j} = Y_{\max} * \frac{1}{1 + \left(\frac{d_{i,j} - 4.0}{d_{0.5} - 4.0} \right)^n} \quad 4$$

in which Y_{\max} is the maximal yield observed for the protein of interest (generally slightly less than 1.0 due to competing side reactions), $(d_{i,j} - 4.0)$ reflects the distance between the $C\beta$ atoms with their van der Waals radii subtracted, and $(d_{0.5} - 4.0)$ reflects the distance at which crosslinking is approximately 50%. The value of n reflects the fact that the crosslinking generally has a high order dependence on the distance. Finally a correlation coefficient was calculated for each fit to measure the accuracy of a given model.

Correlation with Mutagenesis Experiments

The Monte Carlo model, threaded model, and other published models were analyzed to determine whether they are consistent with published mutagenesis results. TM mutations that activate the integrin cause the $\alpha IIb/\beta 3$ TM heterodimer to separate, and these positions are likely to reside at the heterodimer interface.^{2; 3; 10; 22} To determine whether a residue is at a model's interface, we calculated its fractional change in solvent accessibility in the model and in the model's separated helices. First, the solvent accessibility of each amino acid was calculated using DSSP.²³ Then the solvent accessibility was

recalculated for the separated helices. The fractional change in solvent accessibility, $f_{asa,i}$ was calculated with equation 5

$$f_{ASA,i} = 1 - \left(\frac{\text{model}_{ASA,i}}{\text{monomer}_{ASA,i}} \right) \quad 5$$

in which $\text{model}_{ASA,i}$ is the solvent accessible surface area of the i^{th} residue in a model, and $\text{monomer}_{ASA,i}$ is the solvent accessibility of the same residue when the model's helices are isolated.²⁴ The f_{ASA} measurement was then correlated with experimental mutagenesis results using linear regression. For this analysis, a residue was assigned a value of 1 if at least one of its mutants activates the integrin. These positions should reside at the heterodimer interface and have fractional changes in solvent accessibility that approach 1. Other positions that have been probed by mutagenesis were valued 0. These positions should cluster away from the heterodimer interface and have fractional changes in solvent accessibility that approach zero. Mutations to hydrophilic amino acids were disregarded because they can affect oligomerization and orientation in a membrane, and mutations to threonine were disregarded because threonine can perturb the secondary structure of a helix.^{25; 26} In summary, a model's calculated f_{ASA} values were correlated with a binary index of positions that can activate the integrin when mutated.

Assessment of the Putative Arg995-Asp723 Salt Bridge

Reciprocal mutagenesis suggests that an interaction between α IIb Arg995 and β 3 Asp723 stabilizes the integrin's resting state.²⁷ The α IIb and β 3 helices in the Monte Carlo and threaded models were extended to Arg995 and Asp723 using ideal backbone geometries ($\phi = -65^\circ$; $\psi = -40^\circ$; $\omega = 180^\circ$) and the feasibility of a salt bridge was assessed by the manual manipulation of the Arg995 and Asp723 χ angles. Arg995 and Asp723 were proximal in both the Monte Carlo and threaded models, but only the threaded model allowed for the formation of a Arg995-Asp723 salt bridge.

RESULTS

Monte Carlo α IIb/ β 3 Model

In the Monte Carlo-based algorithm, two straight helices consisting of α IIb amino acids Ile966-Trp988 and β 3 amino acids Ile693-Trp715 were docked by randomly altering the six orthogonal parameters that orient two cylinders in space.¹¹ The algorithm's scoring function was designed to favor conformations that were consistent with published mutagenesis experiments by including a selective advantage for disruptive mutations having higher energies than the wild type and for neutral mutations that are iso-energetic. Inclusion of mutagenesis information compensates for approximations made during energy calculations and the limited conformational space accessible to the search algorithm.²⁸ This strategy enabled

us to accurately predict the structures of the TM homodimers for glycophorin A and BNIP3 (see Chapter 2).^{11; 29}

When applied to α IIb and β 3, the Monte Carlo-based algorithm converged on a structure with an angle of 18° between the two helical axes and a right-handed orientation (figure 9). This type of interaction occurs frequently in membrane proteins,¹⁵ and its conformation is similar to >100 different TM dimer interfaces reported in the protein data bank (see Methods). The heterodimer interface for α IIb consisted of residues Trp968, Val969, Gly972, Gly976, Leu980, Leu983, and Met987 and the β 3 interface consisted of Ile693, Val696, Leu697, Val700, Met701, Ile704, Gly708, Leu712, and Trp715. This structure is consistent with a published cysteine crosslinking analysis that examined 120 possible pairwise interactions in the α IIb β 3 TM region, even though cysteine crosslinking data was not considered in the modeling procedure (figures 10 and 12).¹⁹ The structure is also consistent with mutational analyses of the α IIb and β 3 TM domains with the exception of mutations involving α IIb residue Thr981 that activate α IIb β 3 expressed in tissue culture cells, but reside on the opposite side of the α IIb helix from other activating mutations.^{2; 3; 10; 22}

In addition to a TM heterodimer, α IIb β 3 function is thought to be constrained by a “clasp” involving membrane-proximal portions of the α IIb and β 3 cytoplasmic domains, a notable feature of which is a salt-bridge between Arg995 in α IIb and Asp723 in β 3.²⁷ Several previous NMR models of the α IIb and β 3 cytoplasmic domains predict that Arg995 and Asp723 reside in helices, implying

that the α IIb and β 3 TM helices might extend into the cytosol, at least through Arg995 and Asp723.^{30; 31; 32; 33; 34} When our Monte Carlo-derived model is propagated into the cytosol with straight helices, the distance between the Arg995 and Asp723 C β atoms is 12 Å, too far to form a salt bridge, however perturbations from uniform helical structure might allow for an Arg995-Asp723 interaction.

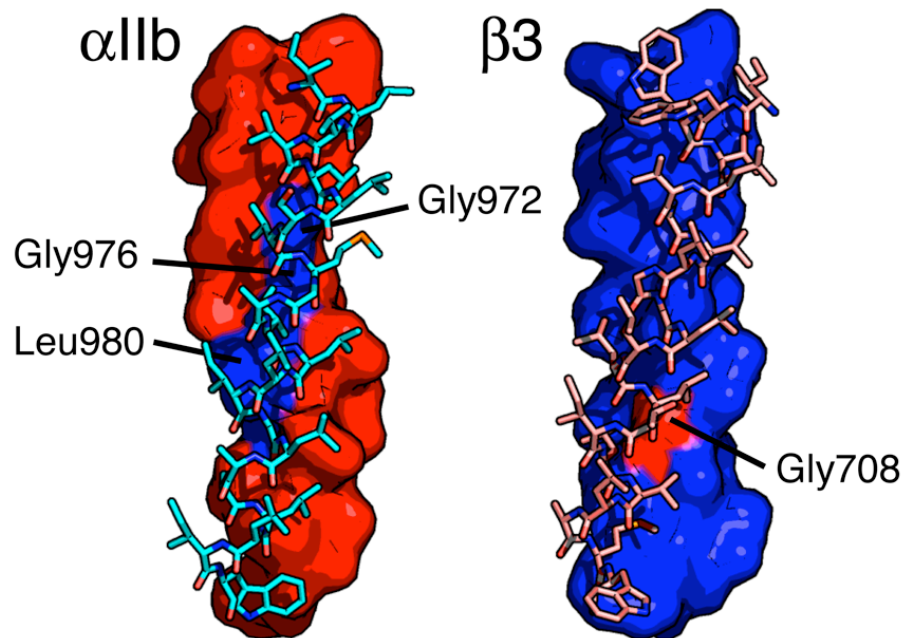


Fig 9. The Monte Carlo model of the α IIb/ β 3 TM heterodimer. On the left, the α IIb helix is depicted as a surface representation (red) and the β 3 helix is shown as a stick representation (cyan). Mutagenesis indicates that Gly972, Gly976, and Leu980 (blue) reside at the heterodimer interface. On the right, the β 3 helix is depicted as a surface representation (blue). Mutagenesis indicates that Gly708 (red) is at the heterodimer interface.

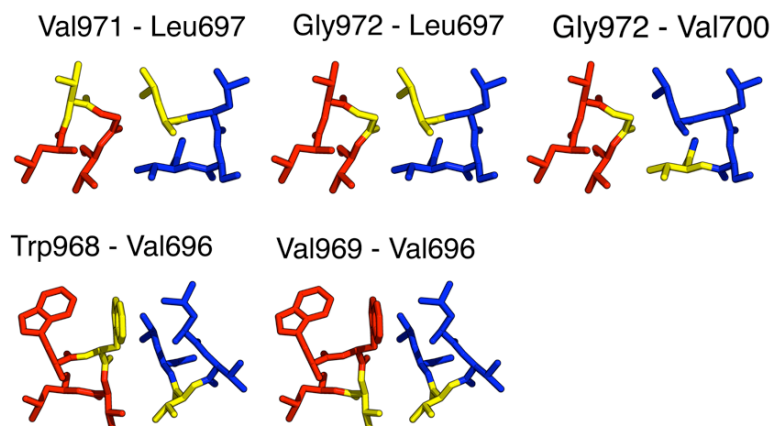


Fig 10. Slices through the Monte Carlo model with amino acids highlighted (yellow) that have a strong propensity to form a disulfide bond when the pair is mutated to cysteine. Leu697 lies between its crosslinking partners Val971 and Gly972. Gly972 lies between its crosslinking partners Leu697 and Val700. Finally, Val696 lies between its crosslinking partners Trp968 and Val969.

Threaded α IIb/ β 3 Model

To verify the Monte Carlo structure, we used threading as a different approach to derive a model for the α IIb/ β 3 TM heterodimer. In contrast with the Monte Carlo-based methods, threading makes use of experimentally determined structures, sampling real protein conformations rather than theoretical geometries. Thus, threaded models can account for kinks, bends, coiling, and other deviations from ideal helical structure with physically accessible conformations.

We threaded the α IIb and β 3 TM sequences through 214 parallel TM helix dimers found in high-resolution crystal structures. The sequences were threaded in multiple different frames to generate >50,000 structures. Each conformation was optimized by SCWRL3.0¹⁶ followed by 2000 conjugant-gradient steps in NAMD,¹⁷ and their dimerization energies were calculated. The top 1% lowest energy structures were analyzed to determine whether they were consistent with cysteine crosslinking results, which complements the use of mutagenesis results in the Monte Carlo strategy. Specifically, the distance was measured between the C β atoms of five α IIb- β 3 residue pairs having a high propensity to form a disulfide bond when the pair is mutated to cysteine.¹⁹ (The remaining 115 experimentally evaluated cysteine mutant pairs were saved for structure validation, see below). The structure with the most consistent average C β distance came from the 1iwg pdb of the crystal structure for the bacterial multidrug efflux transporter AcrB²⁰ consisting of α IIb Trp967-Trp988 threaded on chain A residues 392-414 and β 3 Ile693-Ala711 threaded on chain A residues 466-484.

As was the case for the Monte Carlo model, the threaded model has a right-handed crossing. However, due to the non-linearity of natural helical axes, the interhelix crossing angles in the threaded model range from 48° in the heart of the GXXXG interface to 3.5° near its C-terminus. The steepest crossing angle (48°) occurs between α IIb residues Gly972-Gly975 and β 3 residues Ser699-Gly702, and this conformation is characteristic of a canonical GXXXG interaction,

which is a dimerization motif found in TM helices.^{15; 35; 36; 37} While the $\beta 3$ helix is relatively straight, the αIIb helix is kinked by 35° between residues Gly975 and Gly976 which extends the αIIb - $\beta 3$ interface beyond the GXXXG motif and permits interactions near the membrane-cytosol boundary. Additionally, when the TM helices were propagated into the cytosol, the structure allowed for an interaction characteristic of the putative Arg995-Asp723 salt bridge. The αIIb interface consisted of residues Trp968, Val969, Gly972, Gly976, Leu980, Leu983, and Met987 and the $\beta 3$ interface consisted of Ile693, Val696, Leu697, Val700, Met701, Ile704, and Gly708, essentially identical to the Monte Carlo model and consistent with both mutational analyses and the additional cysteine crosslinking pairs that were not used to score the model (figures 11 and 12). Finally, the $C\alpha$ RMSD between the Monte Carlo and threaded models is 1.3 Å indicating similar structure.

Correlation with Experimental Results

The Monte Carlo and threaded models and other published models were correlated with experimental mutagenesis and cysteine crosslinking results. Four three-dimensional models have been previously reported for the $\alpha IIb/\beta 3$ TM heterodimer at atomic level resolution. Two of the models were generated by Monte Carlo methods that did not take into account experimental data (literature models A and B).¹⁰ The other two models were generated from molecular dynamics simulations of integrin homologs that converged on two conformations,

TABLE III. Comparison of α IIb/ β 3 TM models

agreement with empirical data						
model	M	T	A	B	1	2
disulfides	+++	+++	-	+	++	+++
mutagenesis	++	+++	++	-	+	++
salt bridge	-	+	-	-	-	+

Model similarity, C α RMSD in Å						
model	M	T	A	B	1	2
Monte Carlo		1.3	*	*	2.4	1.1
Threading	1.3		*	*	2.4	1.6
Model A	*	*		*	*	*
Model B	*	*	*		*	*
Model 1	2.4	2.4	*	*		2.3
Model 2	1.1	1.6	*	*	2.3	
reference			10	10	9	9

* denotes a C α RMSD greater than 3 Å

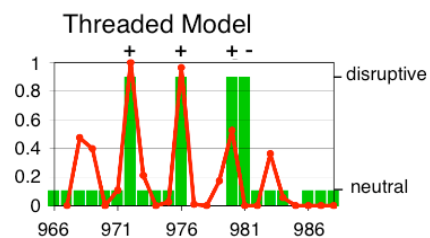
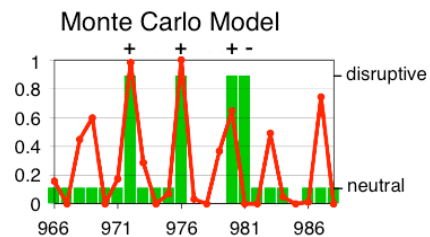
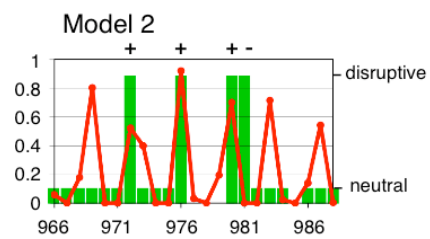
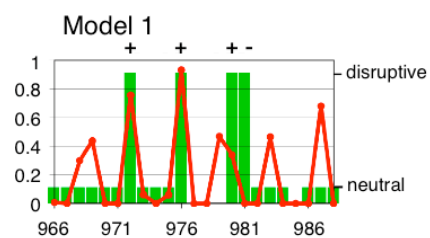
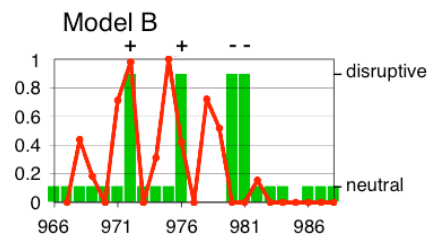
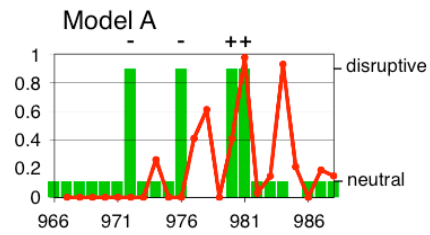
TABLE III.. A qualitative analysis of the Monte Carlo model, the threaded model, and four published models. The Monte Carlo and threaded models are most consistent with experimental results. “Disulfides” indicates whether a model is consistent with published cysteine crosslinking results, “mutagenesis” indicates whether a model is consistent with published mutational analyses, and “salt bridge” indicates whether a model is consistent with the putative Arg995-Asp723 interaction. Quantitative analyses are shown in figures 11 and 12. Additionally, the structural similarity of each model is reported as C α atom root mean squared deviations in angstroms (RMSD). A RMSD of 0 would indicate that two models are identical and a RMSD less than 2 indicates that two models are similar. The Monte Carlo model, threaded model, and literature model 2 are structurally similar.

with representative structures reported for α IIb/ β 3 (literature models 1 and 2).⁹ Additional models of the α IIb/ β 3 TM heterodimer have been reported but were

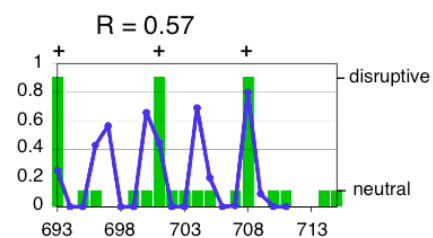
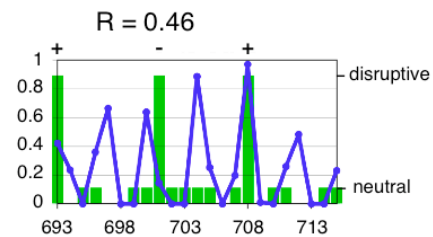
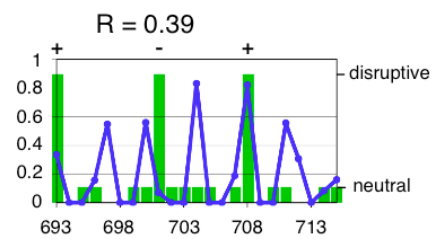
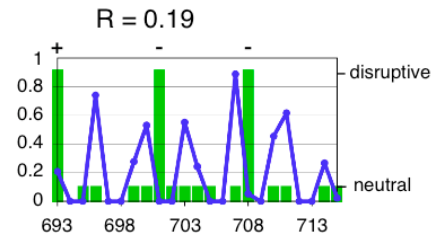
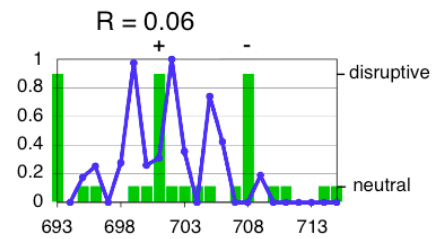
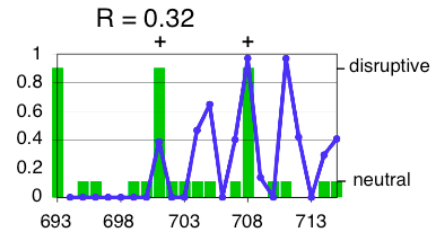
not considered here because they did not include atomic coordinates³⁸ or contain a number of D-amino acids.^{6; 7; 8}

First, we considered how well each model correlates with the consequences of mutating either the α IIb or the β 3 TM domain, focusing on mutations that induce constitutive α IIb β 3 activation and thus are likely present in the heterodimer interface. The fractional change in solvent accessibility was correlated with experimental mutagenesis results as shown in figure 11. For this analysis, residues were assigned a value of 1 if at least one of its mutants activates the integrin (large green bar; missing bars indicate points for which data is not available). These positions should reside at the heterodimer interface and have fractional changes in solvent accessibility that approach 1 (peaks in red for α IIb, blue for β 3). Mutations with no significant effect on activation were valued 0 (small green bar). These positions should cluster away from the heterodimer interface and have fractional changes in solvent accessibility that approach zero (red/blue minima). Disruptive mutations that occur at a model's heterodimer interface are marked with a "+" and indicate positive correlation. We also computed a correlation coefficient, R, for each model (figure 11), although we note that a perfectly correlating model would not have an R=1 because the mutagenesis results were treated in a binary manner. An example of a poorly correlating structure is literature model B, which displays poor overall correlation between $f_{asa,i}$ and experimental mutagenesis results (R = 0.06). The Monte Carlo and threaded models have the best correlation coefficients (R = 0.46 and 0.57).

Fractional Change in Solvent Accessible Surface for α IIb () and β 3 ()



α IIb residue number



β 3 residue number

Experimentally Characterized Disruptive and Neutral Mutations ()

Fig 11. Point mutations can activate the integrin (large green bars) or have no effect (small green bars; missing bars indicate positions for which mutagenesis information is not available). Activating mutations are likely to reside at the α IIb/ β 3 heterodimer interface. The interface of each model was defined using a calculation based on each amino acid's solvent accessible surface (red and blue lines, see equation 5). A model is consistent with experimental mutagenesis results if each activating mutation (large green bar) occurs at the model's interface (large change in solvent accessible surface). Experimental mutagenesis results were correlated with the fractional change in solvent accessible surface using linear regression, and each correlation coefficient R is reported.

Overall, the Monte Carlo and threaded models correlated with experimental mutagenesis results while other models correlated to a lesser extent, or not at all.

Next, the models were correlated with the results of introducing disulfide crosslinks between the α IIb and β 3 TM domains. Luo et al. expressed full length α IIb β 3 in 293T cells with single cysteine replacements in both the α IIb and β 3 TM helices and measured the efficiency of disulfide bond formation, based on the premise that positions forming disulfide crosslinks should be closer in space than positions that do not crosslink the integrin.¹⁹ Thus, the distance between the C β atoms of two cysteine residues in a model were correlated with the experimentally determined crosslinking efficiency for the pair (figure 12). For a quantitative comparison, it would be ideal to obtain the rates of the crosslinking reaction under carefully controlled conditions. Also, in comparing the

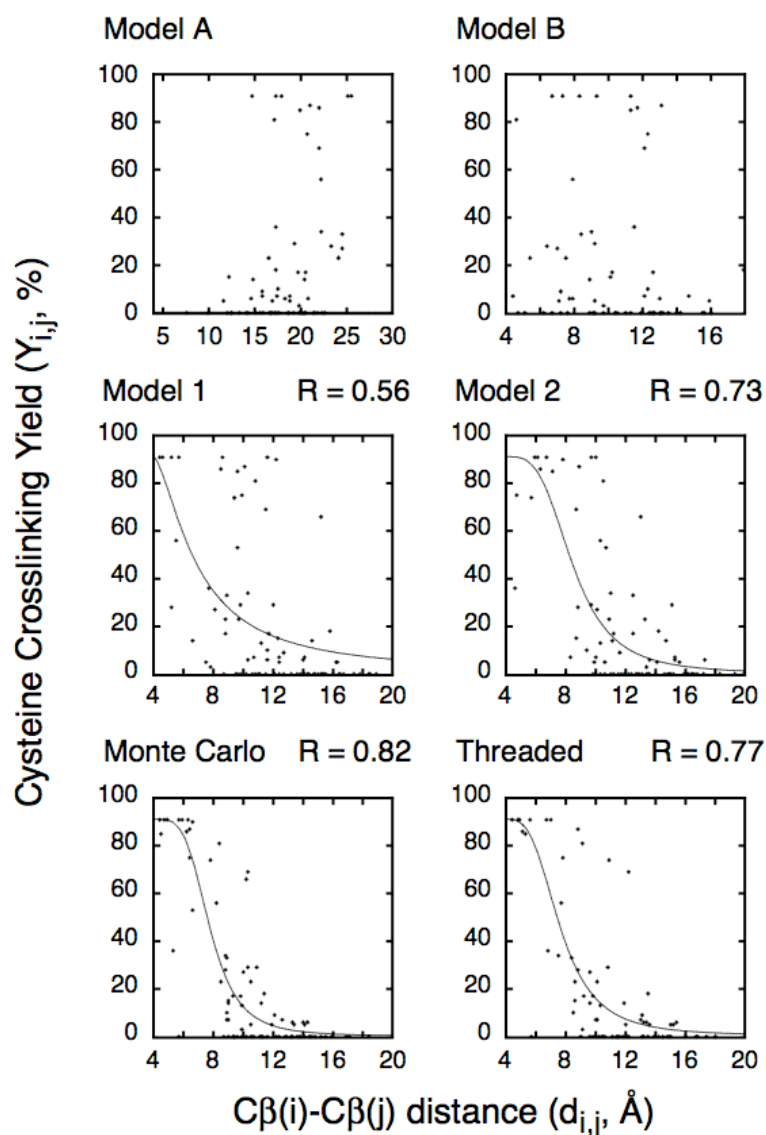


Fig 12. When single cysteine point mutations are introduced into both the α IIb and β 3 TM helices, a disulfide bond can crosslink the integrin subunits. Cysteine crosslinking yield correlates with the distance between two cysteines, and these distances can be measured in a given model. $C\beta$ distances for cysteine mutant pairs was plotted against the experimentally observed cysteine crosslinking yield and fit to equation 4. The correlation coefficient of each fit is reported as R.

experimental data to computational models, it would be ideal to consider not only interatomic distances, but also the angular relationship between $C\alpha$ – $C\beta$ bond vectors and the local dynamics of the structure.³⁹ However, even in the absence of this information, a modest correlation can be observed between the extent of disulfide formation and the distance between the interacting residues.⁴⁰

The $C\beta$ distances in the Monte Carlo model correlate well with the cysteine crosslinking data ($R = 0.82$); furthermore, the equation 4 parameters of $d_{0.5} = 7.8 \text{ \AA}$ and $n = 4$ are consistent with values reported in the literature.³⁹ A similar good correlation was observed for the threaded model ($R = 0.77$; $d_{0.5} = 7.7 \text{ \AA}$; $n = 3.1$), and for literature model 2 ($R = 0.73$; $d_{0.5} = 8.5 \text{ \AA}$; $n = 3.4$). A less good correlation was observed for literature model 1 ($R = 0.56$; $d_{0.5} = 6.9 \text{ \AA}$; $n = 1.6$), and a poor correlation was observed for literature models A and B.

DISCUSSION AND CONCLUSIONS

Although it has not yet been possible to determine the complete integrin structure at high resolution, partial structural information has been derived from mutagenesis,^{2; 3; 10; 22; 27} crosslinking,^{19; 41; 42; 43; 44} FRET experiments,^{4; 45} electron microscopy,^{44; 46; 47; 48} crystallographic and NMR analyses of integrin fragments,^{30; 31; 32; 33; 34; 49; 50; 51; 52; 53; 54} and molecular modeling.^{6; 7; 8; 9; 10} Notably, the extracellular portions of the integrins $\alpha v\beta 3$ and $\alpha IIb\beta 3$ have been crystallized in conformations that are believed to represent their resting and active states.^{49; 50; 52} Additionally,

NMR has been used to obtain structures of peptides corresponding to the individual α IIb and β 3 TM domains,^{53; 54} individual cytosolic domains,^{30; 34} and complexes between the α IIb and β 3 cytosolic domains.^{31; 33; 55} However the experimental determination of structures for a TM heterodimer has proven to be challenging. Here, we describe two fundamentally different modeling approaches that converged on the same structure for the α IIb/ β 3 TM heterodimer. This conformation differs from previously published models and favorably compares with experimental data.

Review of Published Integrin TM Heterodimer Models

Gottschalk et al. performed the first structural analysis of the integrin TM heterodimer using a grid/molecular dynamics protocol pioneered by Axel Brunger.^{7; 56; 57} α IIb β 3 was modeled in parallel with homologous integrins in order to identify an evolutionarily conserved structure. Twelve different conformations were identified and a right-handed structure with a small crossing angle was judged to be in best qualitative agreement with the then-available experimental data. Subsequently, Gottschalk and Kessler modeled a portion of the α IIb/ β 3 TM heterodimer via a molecular dynamics simulation that utilized interchain distance restraints derived from NMR and an additional restraint imposed for the putative Arg995-Asp723 salt bridge.⁸ These simulations converged on a single structure that is consistent with a right-handed coiled coil. Recently, Gottschalk modeled the α IIb/ β 3 TM heterodimer with a simulated annealing protocol that utilized

distance restraints extrapolated from a cysteine crosslinking analysis and an additional restraint for the Arg995-Asp723 salt bridge.⁶ This method converged on a single conformation consistent with a right-handed coiled coil. However, Gottschalk's models contain a number of D-amino acids, possibly because of unfavorable contacts in the starting coordinates, so the resultant models contain a number of inverted stereo-centers.

On the basis of reconstructed electron cryomicroscopy images for low affinity α IIb β 3, Adair and Yeager proposed that the TM domains of resting α IIb β 3 form a coiled coil and modeled it as either a left-handed or a right-handed heterodimer by placing the Arg995-Asp723 salt bridge at the interface.³⁸ They noted that the right-handed coiled coil positioned more conserved amino acids at the heterodimer interface; however these models were not considered in this analysis.

Substantially different structures were proposed by Partridge et al.¹⁰ Four hundred conformations were generated by Monte Carlo and representative structures were selected from two heavily populated clusters that passed geometric filters (literature models A and B). One of the conformations predicted the effect of subsequent point mutations (model A).

Finally, Lin and coworkers performed a grid search of conformational space followed by molecular dynamics for each grid point using the sequences of each human integrin homolog in order to identify an evolutionarily conserved structure.⁹ This method is similar to the original work of Gottschalk et al, however

proper chirality was maintained. Two conformations were identified (literature models 1 and 2) and model 2 was predicted to reflect the resting α IIb/ β 3 TM heterodimer.

Previously, we published a model of the integrin TM heterodimer using a Monte Carlo strategy that included a selective advantage for conformations that were consistent with experimental mutagenesis results, using an earlier generation of the Monte Carlo software described in Chapter 2.² In our original publication, we were able to identify the same interface reported here, but were unable to distinguish models with “shallow” interhelix crossing angles (-18°) from those with “glycophorin-like” crossing angles (-40°). We have since re-parameterized the scoring function¹¹ and the revised protocol consistently identifies structures with a crossing angle around -18°. Finally, this chapter describes an additional threading method used to generate a model that is consistent with experimental cysteine crosslinking results. The resulting model is essentially identical to the Monte Carlo model, except it introduces a slight kink in the α IIb subunit which allows for a larger crossing angle near the GXXXG interface, similar to a canonical, glycophorin-like interaction.

Analysis of Different Models

An accurate model successfully predicts experimental results, and each published model of the α IIb/ β 3 TM heterodimer is buttressed by one or more empirical findings; however, each model has substantially different structure.

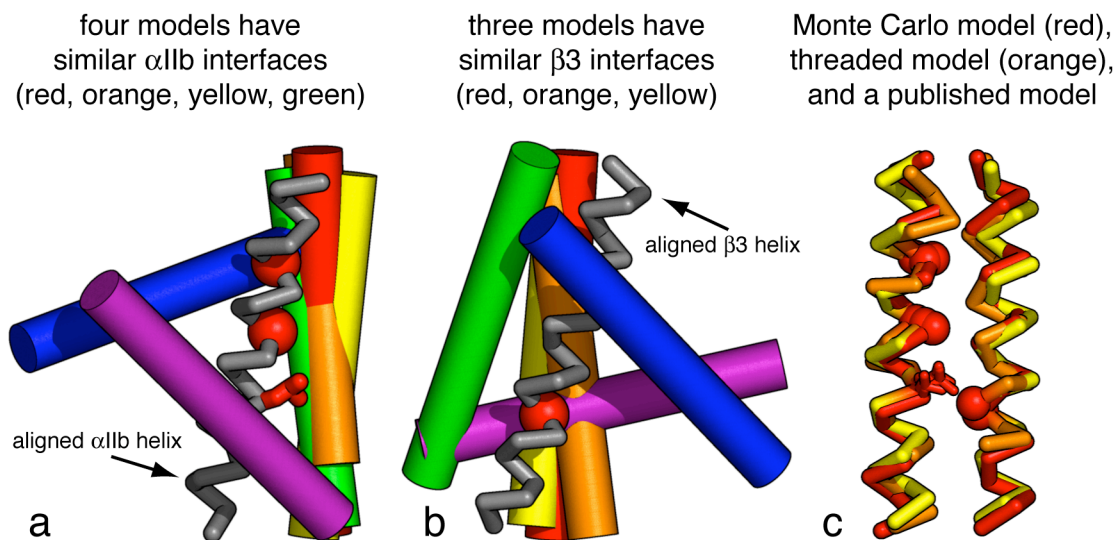


Fig 13. (a) The α IIb helices of each model were aligned and a single α IIb helix is displayed as a gray ribbon. Models have similar α IIb interfaces if their β 3 helices overlap (cylinders). (b) The β 3 helices of each model were aligned and a single β 3 helix is displayed as a gray ribbon. Models have similar β 3 interfaces if their α IIb helices overlap (cylinders). (c) Alignment of the Monte Carlo model, the threaded model, and a previously published model (literature model 2). These models have similar structure.

Residues Gly972, Gly976, Leu980, and Gly708 are highlighted in red. Mutagenesis indicates that these residues are at the α IIb/ β 3 heterodimer interface. The models are color coded as follows: Monte Carlo model, red; threaded model, orange; model A, purple; model B, blue; model 1, green; model 2, yellow.

To quantitatively assess the accuracy with which a model predicts experimental results, we performed objective measurements on each model and correlated these measurements with published experimental findings. First, fractional changes in solvent accessibility were correlated with published experimental

mutagenesis results; the Monte Carlo and threaded models reported here had the highest correlation coefficients. Additionally, the distance between different α IIb and β 3 residues was correlated with published cysteine crosslinking results; again the Monte Carlo and threaded models had the highest correlation coefficients. Of the other models, literature model 2 had the strongest correlation with experimental results and this model was structurally similar to the Monte Carlo and threaded models with $C\alpha$ RMSDs of 1.1 and 1.6 Å, respectively. It is noteworthy that model 2 was generated by a molecular dynamics method fundamentally different from the Monte Carlo and threading methods, and selected models based on integrin homology rather than functional or crosslinking data.⁹

Comparison with Models in Press

There are currently two additional models of the α IIb/ β 3 TM heterodimer in press including a NMR structure⁵⁵ and a model generated by the Rosetta algorithm using restraints derived from experimental cysteine crosslinking results. Of the models considered here, the Monte Carlo and threaded models are most consistent with the NMR structure with RMSDs of 1.2 and 1.3 Å, respectively. The Rosetta model has divergent structure that is more similar to the NMR structure of glycophorin A than the models reported here.

Table IV. Structural comparison of top performing models and models in press

	Monte Carlo	Threading	Model 2	NMR	Rosetta
Monte Carlo		1.3	1.1	1.2	1.9
Threading	1.3		1.6	1.3	1.4
Model 2	1.1	1.6		1.5	2.3
NMR	1.2	1.3	1.5		1.7
Rosetta	1.9	1.4	2.3	1.7	
Reference			9	55	58

values correspond to $C\alpha$ RMSDs in Å

Conclusion

We generated two models of the α IIb/ β 3 TM domain heterodimer using fundamentally different methods: a Monte Carlo algorithm that selected conformations based on their agreement with published mutagenesis results and a threading method that selected conformations based on their agreement with cysteine crosslinking results. The two methods converged on a similar structure, and when compared to previously published models, the Monte Carlo and threaded models were most consistent with reported experimental findings, suggesting they are most likely to reflect the native structure of the α IIb/ β 3 TM heterodimer. These models have now been confirmed by its recently published NMR structure.

References

1. Hynes, R. O. (2002). Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673-87.

2. Li, W., Metcalf, D. G., Gorelik, R., Li, R., Mitra, N., Nanda, V., Law, P. B., Lear, J. D., DeGrado, W. F. & Bennett, J. S. (2005). A push-pull mechanism for regulating integrin function. *Proc Natl Acad Sci U S A* **102**, 1424-9.
3. Li, R., Mitra, N., Gratkowski, H., Vilaire, G., Litvinov, R., Nagasami, C., Weisel, J. W., Lear, J. D., DeGrado, W. F. & Bennett, J. S. (2003). Activation of integrin α IIb β 3 by modulation of transmembrane helix associations. *Science* **300**, 795-8.
4. Kim, M., Carman, C. V. & Springer, T. A. (2003). Bidirectional transmembrane signaling by cytoplasmic domain separation in integrins. *Science* **301**, 1720-5.
5. Bennett, J. S. (2005). Structure and function of the platelet integrin α IIb β 3. *J Clin Invest* **115**, 3363-9.
6. Gottschalk, K. E. (2005). A coiled-coil structure of the α IIb β 3 integrin transmembrane and cytoplasmic domains in its resting state. *Structure* **13**, 703-12.
7. Gottschalk, K. E., Adams, P. D., Brunger, A. T. & Kessler, H. (2002). Transmembrane signal transduction of the α (IIb) β (3) integrin. *Protein Sci* **11**, 1800-12.
8. Gottschalk, K. E. & Kessler, H. (2004). Evidence for hetero-association of transmembrane helices of integrins. *FEBS Lett* **557**, 253-8.

9. Lin, X., Tan, S. M., Law, S. K. & Torres, J. (2006). Unambiguous prediction of human integrin transmembrane heterodimer interactions using only homologous sequences. *Proteins* **65**, 274-9.
10. Partridge, A. W., Liu, S., Kim, S., Bowie, J. U. & Ginsberg, M. H. (2005). Transmembrane domain helix packing stabilizes integrin alphaIIb beta3 in the low affinity state. *J Biol Chem* **280**, 7294-300.
11. Metcalf, D. G., Law, P. B. & DeGrado, W. F. (2007). Mutagenesis data in the automated prediction of transmembrane helix dimers. *Proteins* **67**, 375-84.
12. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. (2006). OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623-5.
13. Bowie, J. U., Luthy, R. & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164-70.
14. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). A new approach to protein fold recognition. *Nature* **358**, 86-9.
15. Walters, R. F. & DeGrado, W. F. (2006). Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* **103**, 13658-63.
16. Canutescu, A. A., Shelenkov, A. A. & Dunbrack, R. L., Jr. (2003). A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* **12**, 2001-14.

17. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-802.
18. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.* **4**, 187-217.
19. Luo, B. H., Springer, T. A. & Takagi, J. (2004). A specific interface between integrin transmembrane helices and affinity for ligand. *PLoS Biol* **2**, e153.
20. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587-93.
21. Bansal, M., Kumar, S. & Velavan, R. (2000). HELANAL: a program to characterize helix geometry in proteins. *J Biomol Struct Dyn* **17**, 811-9.
22. Luo, B. H., Carman, C. V., Takagi, J. & Springer, T. A. (2005). Disrupting integrin transmembrane domain heterodimerization increases ligand binding affinity, not valency or clustering. *Proc Natl Acad Sci U S A* **102**, 3679-84.
23. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-637.

24. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55**, 379-400.
25. Deupi, X., Olivella, M., Govaerts, C., Ballesteros, J. A., Campillo, M. & Pardo, L. (2004). Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices. *Biophys J* **86**, 105-15.
26. Moore, D. T., Berger, B. W. & DeGrado, W. F. (2008). Protein-protein interactions in the membrane: sequence, structural, and biological motifs. *Structure* **16**, 991-1001.
27. Hughes, P. E., Diaz-Gonzalez, F., Leong, L., Wu, C., McDonald, J. A., Shattil, S. J. & Ginsberg, M. H. (1996). Breaking the integrin hinge. A defined structural constraint regulates integrin signaling. *J Biol Chem* **271**, 6571-4.
28. Nanda, V. & DeGrado, W. F. (2005). Automated use of mutagenesis data in structure prediction. *Proteins* **59**, 454-66.
29. Bocharov, E. V., Pustovalova, Y. E., Pavlov, K. V., Volynsky, P. E., Goncharuk, M. V., Ermolyuk, Y. S., Karpunin, D. V., Schulga, A. A., Kirpichnikov, M. P., Efremov, R. G., Maslennikov, I. V. & Arseniev, A. S. (2007). Unique dimeric structure of BNip3 transmembrane domain suggests membrane permeabilization as a cell death trigger. *J Biol Chem* **282**, 16256-66.

30. Vinogradova, O., Vaynberg, J., Kong, X., Haas, T. A., Plow, E. F. & Qin, J. (2004). Membrane-mediated structural transitions at the cytoplasmic face during integrin activation. *Proc Natl Acad Sci U S A* **101**, 4094-9.
31. Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T., Plow, E. & Qin, J. (2002). A structural mechanism of integrin alpha(IIb)beta(3) "inside-out" activation as regulated by its cytoplasmic face. *Cell* **110**, 587-97.
32. Li, R., Babu, C. R., Valentine, K., Lear, J. D., Wand, A. J., Bennett, J. S. & DeGrado, W. F. (2002). Characterization of the monomeric form of the transmembrane and cytoplasmic domains of the integrin beta 3 subunit by NMR spectroscopy. *Biochemistry* **41**, 15618-24.
33. Weljie, A. M., Hwang, P. M. & Vogel, H. J. (2002). Solution structures of the cytoplasmic tail complex from platelet integrin alpha IIb- and beta 3-subunits. *Proc Natl Acad Sci U S A* **99**, 5878-83.
34. Vinogradova, O., Haas, T., Plow, E. F. & Qin, J. (2000). A structural basis for integrin activation by the cytoplasmic tail of the alpha IIb-subunit. *Proc Natl Acad Sci U S A* **97**, 1450-5.
35. MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997). A transmembrane helix dimer: structure and implications. *Science* **276**, 131-3.
36. Senes, A., Gerstein, M. & Engelman, D. M. (2000). Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs

- frequently and in association with beta-branched residues at neighboring positions. *J Mol Biol* **296**, 921-36.
37. Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brunger, A. T. & Engelman, D. M. (1994). A dimerization motif for transmembrane alpha helices. *Nature, Structural biology* **1**, 157-163.
 38. Adair, B. D. & Yeager, M. (2002). Three-dimensional model of the human platelet integrin alpha IIb beta 3 based on electron cryomicroscopy and x-ray crystallography. *Proc Natl Acad Sci U S A* **99**, 14059-64.
 39. Careaga, C. L. & Falke, J. J. (1992). Structure and dynamics of Escherichia coli chemosensory receptors. Engineered sulfhydryl studies. *Biophys J* **62**, 209-16; discussion 217-9.
 40. Bass, R. B., Butler, S. L., Chervitz, S. A., Gloor, S. L. & 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.
 41. Shimaoka, M., Lu, C., Palframan, R. T., von Andrian, U. H., McCormack, A., Takagi, J. & Springer, T. A. (2001). Reversibly locking a protein fold in an active conformation with a disulfide bond: integrin alphaL I domains with high affinity and antagonist activity in vivo. *Proc Natl Acad Sci U S A* **98**, 6009-14.

42. Luo, B. H., Takagi, J. & Springer, T. A. (2004). Locking the beta3 integrin I-like domain into high and low affinity conformations with disulfides. *J Biol Chem* **279**, 10215-21.
43. Lu, C., Shimaoka, M., Zang, Q., Takagi, J. & Springer, T. A. (2001). Locking in alternate conformations of the integrin alphaLbeta2 I domain with disulfide bonds reveals functional relationships among integrin domains. *Proc Natl Acad Sci U S A* **98**, 2393-8.
44. Takagi, J., Petre, B. M., Walz, T. & Springer, T. A. (2002). Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* **110**, 599-11.
45. Chigaev, A., Buranda, T., Dwyer, D. C., Prossnitz, E. R. & Sklar, L. A. (2003). FRET detection of cellular alpha4-integrin conformational activation. *Biophys J* **85**, 3951-62.
46. Takagi, J., Strokovich, K., Springer, T. A. & Walz, T. (2003). Structure of integrin alpha5beta1 in complex with fibronectin. *Embo J* **22**, 4607-15.
47. Weisel, J. W., Nagaswami, C., Vilaire, G. & Bennett, J. S. (1992). Examination of the platelet membrane glycoprotein IIb-IIIa complex and its interaction with fibrinogen and other ligands by electron microscopy. *J Biol Chem* **267**, 16637-43.
48. Litvinov, R. I., Nagaswami, C., Vilaire, G., Shuman, H., Bennett, J. S. & Weisel, J. W. (2004). Functional and structural correlations of individual alphaIIb beta3 molecules. *Blood* **104**, 3979-85.

49. Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L., Joachimiak, A., Goodman, S. L. & Arnaout, M. A. (2001). Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* **294**, 339-45.
50. Xiong, J. P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S. L. & Arnaout, M. A. (2002). Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* **296**, 151-5.
51. Beglova, N., Blacklow, S. C., Takagi, J. & Springer, T. A. (2002). Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation. *Nat Struct Biol* **9**, 282-7.
52. Xiao, T., Takagi, J., Collier, B. S., Wang, J. H. & Springer, T. A. (2004). Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* **432**, 59-67.
53. Lau, T. L., Dua, V. & Ulmer, T. S. (2008). Structure of the integrin alphaIIb transmembrane segment. *J Biol Chem* **283**, 16162-8.
54. Lau, T. L., Partridge, A. W., Ginsberg, M. H. & Ulmer, T. S. (2008). Structure of the integrin beta3 transmembrane segment in phospholipid bicelles and detergent micelles. *Biochemistry* **47**, 4008-16.
55. Lau, T. L., Kim, C., Ginsberg, M. H. & Ulmer, T. S. (2009). The structure of the integrin alphaIIbbeta3 transmembrane complex explains integrin transmembrane signalling. *Embo J*.

56. Treutlein, H. R., Lemmon, M. A., Engelman, D. M. & Brunger, A. T. (1992). The glycoporphin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry* **31**, 12726-32.
57. Adams, P. D., Engelman, D. M. & Brunger, A. T. (1996). Improved prediction for the structure of the dimeric transmembrane domain of glycoporphin A obtained through global searching. *Proteins* **26**, 257-61.
58. Zhu, J., Luo, B. H., Barth, P., Schonbrun, J., Baker, D. & Springer, T. A. (2009). The Structure of a Receptor with Two Associating Transmembrane Domains on the Cell Surface: Integrin α IIb β 3. *Mol Cell* **34**, 234-249.

CHAPTER 4

Structural Characterization of a Disulfide-Linked α IIb/ β 3

Cytosolic Domain

INTRODUCTION

One of the widely studied examples of regulated integrin function is the platelet fibrinogen receptor α IIb β 3.¹ On a circulating platelet, α IIb β 3 is maintained in a resting conformation to prevent deleterious interactions with other circulating cells or the vascular endothelium that could cause thrombosis. Upon vascular injury, platelets respond by activating α IIb β 3 to engage soluble ligands that crosslink adjacent platelets and form a hemostatic platelet plug. The activation of resting α IIb β 3 proceeds through intracellular “inside-out” signaling cascades that enable the cytoskeletal proteins talin and kindlin-3 to bind the β 3 cytosolic domain, favoring conformations of β 3 that displace α IIb from heteromeric α IIb/ β 3 interactions in the cytosolic and TM regions.^{2;3} Separation of the TM heterodimer transduces the intracellular activation signal across the membrane to expose the integrin’s extracellular ligand-binding sites.^{4;5} Active integrins can then nucleate large complexes that initiate “outside-in” signaling cascades including the activation of Src kinase which constitutively binds the β 3 cytosolic domain.⁶

The $\beta 3$ cytosolic domain interacts with α IIb, talin, kindlin-3 and Src kinase, and the structural basis of these interactions has been previously probed by NMR and crystallography.^{7; 8; 9; 10; 11; 12} Each three-dimensional structure is markedly different and it is unclear the extent to which differences reflect discrete conformational states or different average structures from a dynamic equilibrium. In either case, portions of both the α IIb and $\beta 3$ cytosolic domains are dynamic,^{13; 14; 15; 16} a functional necessity for interactions with multiple different partners, and small changes in experimental conditions might shift the dynamic equilibrium between different conformations.

Despite their differences, the N-terminal portion of each $\beta 3$ structure contains a helix that extends to approximately Asp723, and analyses that include the $\beta 3$ TM domain depict this region as an extension of the TM helix.^{7; 8; 9; 10; 12; 14; 15; 16} This portion of $\beta 3$ is important for stabilizing the integrin's resting state, including a putative salt bridge between $\beta 3$ Asp723 and α IIb Arg995.¹⁷ Downstream of Asp723, published three-dimensional models have divergent structure, including regions of $\beta 3$ that interact with talin, kindlin-3, and Src kinase.^{7; 8; 9; 10; 11; 15} These regions are dynamic, which might account for structural differences between various models.^{14; 15; 16} Similarly, the α IIb cytosolic domain is dynamic,^{13; 16} which might account for differences between its published conformations.^{7; 8; 9; 12; 13; 18}

The structural analysis of the α IIb/ $\beta 3$ cytosolic heterodimer is challenging because of its dynamics and because peptides corresponding to the α IIb and $\beta 3$

cytosolic domains do not form a stable complex. Never-the-less, NMR structures of the α IIb/ β 3 cytosolic interaction have been solved for heteromeric interactions observed in equilibrium with other states.^{8; 9; 12} Notably, Vinogradova et al developed a NMR model for the monomeric α IIb and β 3 cytosolic domains docked using 13 distance restraints derived predominantly from transferred NOEs.⁸ More recently, Lau et al developed a NMR model of the α IIb/ β 3 TM and membrane proximal regions that agrees with many empirical findings, including mutagenesis of residues in the TM and cytosolic regions.¹² This structure does not form a stable complex, rather it interchanges between heteromeric and monomeric states, similar to the equilibrium observed for the full length integrin. However, the overlapping regions in the Vinogradova and Lau structures are substantially different and vary from other reported conformations.^{8; 9; 12}

Because published NMR models of the α IIb/ β 3 cytosolic interaction have not converged on similar structures in part because the α IIb/ β 3 interaction is not stable, we built a stable α IIb/ β 3 heterodimer by positioning a disulfide bond in the α IIb/ β 3 interface (figure 14). This strategy was previously employed to study the resting integrin's α IIb and β 3 cytosolic domains, however no interaction was observed.¹⁶ Since these first generation constructs, we published a model of the α IIb/ β 3 TM heterodimer¹⁹ that has been validated by experiments in several subsequent reports. Notably the model was confirmed by cysteine scanning mutagenesis near the TM heterodimer's N-terminus in which cysteine mutations at the model's interface could crosslink the α IIb and β 3 subunits in the full length

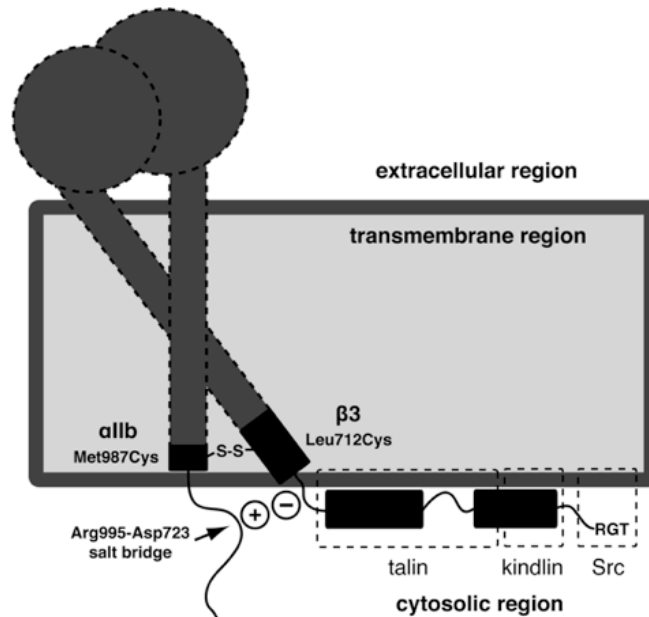
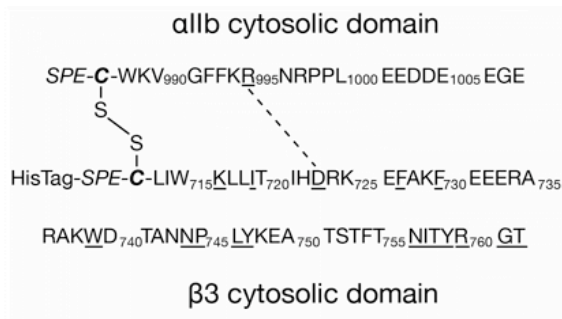


Fig 14. Design of the disulfide-linked α Ib/ β 3 cytosolic heterodimer. (top) The sequences of the α Ib and β 3 subunits; each subunit includes a *SPE* helix cap followed by the cysteine *C* used for crosslinking. After the cysteine, the α Ib and β 3 cytosolic domains consist of residues W715-E1008 and L713-T762, respectively. Additionally, the β 3 subunit contained an upstream histidine tag that was used during purification steps. The cysteines were positioned based on the α Ib/ β 3 interface depicted in the Monte Carlo model from Chapter 3. (bottom) A cartoon depicting the position of the engineered disulfide bond and its relation to other integrin domains including the α Ib Arg995- β 3 Asp723 salt bridge, which is postulated to stabilize the integrin's resting conformation, and β 3 cytosolic domains that bind talin, kindlin-3, and Src kinase.

integrin.²⁰ Building upon this finding, cysteines were positioned in the model's C-terminal interface, which is closer to the cytosolic domains, to crosslink a NMR construct consisting of the α IIb and β 3 cytosolic domains. The disulfide bond enforces 1:1 stoichiometry and eliminates many conformations that are inaccessible to the native integrin.

The construct was dissolved in dodecyl phosphocholine (DPC) micelles to mimic the membrane environment and analyzed at 37° with physiological protonation (pH = 6.0). The α IIb and β 3 subunits had different circular dichroism (CD) spectra before and after reducing the cysteine crosslink which demonstrates that they interact when constrained by a disulfide bond. The analysis of chemical shift suggests that β 3 interacts with α IIb along the face of a helix containing β 3 residues Lys716, Ile719, and Asp723, similar to the published structure of the α IIb/ β 3 TM heterodimer. However there were no NOE interactions between α IIb and β 3 downstream of the disulfide bond, the α IIb subunit had no unique conformation, and the α IIb interface could not be defined. Taken together, these results suggest a physical yet likely disordered interaction between a β 3 helix and a natively unstructured region of α IIb.

The β 3 solution structure was solved with the disulfide-linked α IIb subunit to enforce the conformation it has in the resting integrin. The structure consists of an N-terminal helix and two additional cytosolic helices, similar to the solution structure of the monomeric β 3 TM and cytosolic regions.¹⁵ Portions of the β 3 interfaces that bind talin had conformations that are similar to those observed in

structures of $\beta 3$ /talin complexes,^{10; 11} suggesting that the $\beta 3$ cytosolic domain is pre-organized to bind talin. In contrast, the $\beta 3$ NPXY motif that binds kindlin-3 resides in an α -helix, inconsistent with known interacting conformations for NPXY motifs. However this helix was highly dynamic and its NPXY motif contained a number of violated NOEs indicating the helical conformation exists in equilibrium with other states. Lastly, Src kinase constitutively binds the last three residues of the $\beta 3$ cytosolic domain RGT,⁶ and this region was found to be unstructured and dynamic, possibly primed for interactions with Src.

Finally, different portions of the $\beta 3$ structure were analyzed with a membrane insertion potential to define their calculated membrane embedding.²¹ The N-terminal helix embedded in a membrane spanning orientation, as expected for a TM helix, and the two cytosolic helices partitioned into the membrane/cytosol interface, suggesting amphiphilic character. Notably, the two cytosolic helices contain residues that interact with talin and kindlin-3, and membrane embedding can compete with these interactions. Thus the calculated membrane embedding of the $\beta 3$ structure defines conformations of the resting integrin that are inaccessible to either talin or kindlin-3.

MATERIALS AND METHODS

Plasmid Constructs

The α IIb cytosolic domain was cloned into a pET-16b vector (Novagen) as a C-terminal fusion with the designed protein α_3 D, a small, highly stable protein that

expresses well in bacteria.²² A linking region was engineered to include proline to disrupt secondary structure, two glycines as a flexible spacer, and a thrombin cleavage site. The α IIb construct immediately followed including a SPE N-terminal helix cap and the cysteine used to crosslink the α IIb and β 3 subunits. The construct's sequence is: *α_3 D-P-GG-LVPR-SPE-C-WKVGFFKRNRPPL EEDDEEGE*. The italicized residues were used for expression, purification, and crosslinking while the non-italicized sequence corresponds to α IIb residues Trp988-Glu1008.

The β 3 cytosolic domain was cloned into the pET-15b vector (Novagen) as a C-terminal fusion to a histidine tag. This construct also includes a SPE N-terminal helix cap followed by the cysteine used for crosslinking the α IIb and β 3 subunits. The sequence of the β 3 peptide is: *MGSSHHHHHHSSGLVPRGSHM-SPE-C-LIWKLLITIHDRKEFAKFEEERARAKWDTANNPLYKEATSTFTNITYRGT*. The italicized residues were used for expression, purification, and crosslinking while the non-italicized sequence corresponds to β 3 residues Leu713-Thr761. All constructs were confirmed by DNA sequencing.

Expression and Purification

The α_3 D- α IIb fusion protein was expressed in *E. coli* strain BL21. Cells were grown to an OD600 of 0.8 AU in M9 minimal media supplemented with 200 μ g/ml Ampicillin, then induced with 1 mM IPTG for four hours at 37°C. The media contained ¹³C glucose and/or ¹⁵N ammonium chloride to label the peptide for

NMR experiments. The cells were pelleted, resuspended in 50ml PBS per liter culture, and lysed with three freeze-thaw cycles. Lysozyme was added to 100 μg per ml and the lysate was sonicated to further break apart the bacterial membrane. The cell debris was pelleted and discarded. The supernatant was heated to 70° C for 30 minutes to denature soluble proteins, then cooled to 4° C which precipitates denatured proteins while leaving the $\alpha_3\text{D-}\alpha\text{IIb}$ fusion protein in solution. The mixture was centrifuged to remove the precipitate and the supernatant was dialyzed into PBS buffer at pH 7.5. Absorbance at 280nm was used to approximate the protein concentration, and the fusion protein was cleaved overnight with 10 units of thrombin per milligram protein. The cleaved peptide was reduced with excess TCEP and further purified by reverse phase HPLC. Finally, the eluted protein was lyophilized and stored at -80°C. This protocol generates >10mg of the αIIb cytosolic domain per liter culture. The product's molecular weight was confirmed by mass spectroscopy.

Likewise, the β_3 peptide was expressed in *E. coli* strain BL21. The cells were grown to an OD600 of 0.8 AU in M9 minimal media supplemented with 200 $\mu\text{g/ml}$ Ampicillin then induced with 1 mM IPTG for four hours at 37°C. The media contained ^{13}C glucose and/or ^{15}N ammonium chloride to label the peptide for NMR experiments. The β_3 peptide was purified on Ni-NTA resin using the QIAExpressionist protocol for denaturing conditions (Qiagen). Eluted β_3 peptide was reduced with excess TCEP and further purified by reverse phase HPLC. Purified β_3 was lyophilized and stored at -80°C. The histidine tag was not

removed from the $\beta 3$ cytosolic domain, and this protocol makes 10mg of the $\beta 3$ construct per liter culture. The product's molecular weight was confirmed by mass spectroscopy.

Coupling α IIb and $\beta 3$

The $\beta 3$ peptide was dissolved at 5mg/ml in 2M guanidine, buffered to pH 6.0 with 50mM MES. The reduced $\beta 3$ cysteine thiol was activated for coupling with 5-fold molar excess 2-2'dithio-bis-(5-nitropyridine) dissolved DMSO.²³ Activated $\beta 3$ was purified from excess 2-2'dithio-bis-(5-nitropyridine) on a PD-10 column using 5% acetic acid as the mobile phase. The purified $\beta 3$ -thio-nitropyridine was lyophilized. Activated $\beta 3$ was reacted with a slight molar excess of reduced α IIb in guanidine-MES buffer, pH 6.5 for several hours. Crosslinked α IIb $\beta 3$ was purified from the reaction mix using reverse phase HPLC and the eluted protein was lyophilized. The heterodimer's molecular weight was confirmed by mass spectroscopy.

Circular Dichroism (CD) Spectroscopy

CD spectra of 10 μ M disulfide-linked α IIb/ $\beta 3$ heterodimer in 5mM sodium phosphate buffer containing 10mM DPC, pH 6.5 was recorded in a 1mm path length cell before and after the addition of the reducing agent TCEP. TCEP was added from a 0.5mM stock solution, pH 6.5 at 5-fold molar excess. As a control, buffer without reducing agent was added to a similar sample.

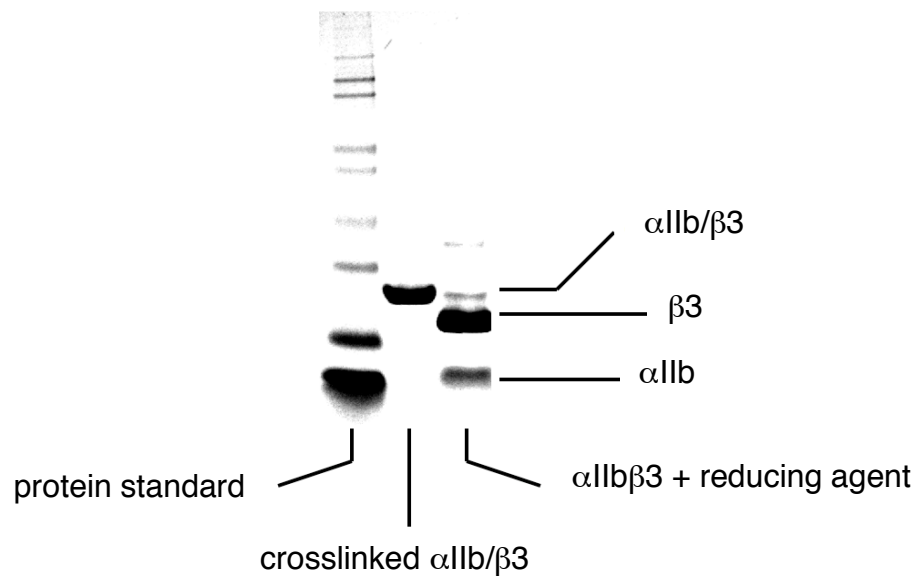


Fig 15. SDS-PAGE analysis of a NMR sample, crosslinked α IIb/ β 3, and the same sample with the addition of reducing agent. The NMR sample has a molecular weight that corresponds to α IIb/ β 3, and it is composed of subunits with molecular weights that correspond to the individual α IIb and β 3 peptides. Samples were further validated using a battery of more rigorous techniques including mass spectroscopy.

NMR Spectroscopy

The disulfide linked α IIb β 3 heterodimer, as isolated from HPLC, was dissolved at 1mM in 5mM sodium phosphate buffer containing 100mM perdeuterated dodecyl phosphocholine (DPC, Cambridge Isotope Laboratories), 10% deuterium oxide, and 0.02% sodium azide as a preservative. The pH was adjusted to 6.0 with sodium hydroxide. DSS was added to some samples at 1mM to reference proton

chemical shifts. Samples were analyzed at 37°C on a 750 MHz NMR spectrophotometer equipped with a Varian probe. A standard battery of experiments allowed us to identify the proton, carbon, and nitrogen resonances for most backbone, aliphatic, phenyl, phenol, indole, imidazole, amide, and guanidinium groups.²⁴

β3 Structure Calculation

NOE distance restraints were derived from a 3D ¹⁵N-edited NOESY, 3D aliphatic ¹³C-edited NOESY, 3D aromatic ¹³C-edited NOESY, and a 4D ¹³C-edited NOESY. Hydrogen bonding distance and geometry restraints²⁵ were implemented for the backbone amides of β3 residues Leu713-Ile721 which are protected from hydrogen-deuterium exchange and predicted to reside in an α-helix. Additionally, the φ and ψ dihedrals of β3 residues Ile714-His722, Phe727-Arg736, and Tyr747-Ile757 were restrained based on the statistical analysis of chemical shift using the TALOS algorithm²⁶ which predicts that these regions are helical. For other residues, chemical shift restraints were implemented for their C_α and C_β atoms,²⁷ and chemical shift restraints were implemented for every H_α atoms.²⁸ Finally, HN-H_α J-couplings were measured in the HNHA experiment and implemented as φ dihedral angle restraints for helical regions or J-coupling restraints for regions with less defined secondary structure which restricts the φ dihedral to 1-4 possible angles.²⁹

6000 structures were calculated with XPLOR-NIH using a protocol similar to the anneal.py example script that comes with the software package.^{30; 31} The top 60 structures were refined using the same protocol with increased weighting for proper bond geometries. The top 20 refined structures were aligned over residues Leu713-Ala735, which converged on a similar conformation, and this structural ensemble will be submitted to the protein databank upon publication.

Hydrogen-deuterium exchange (HDX)

HDX was performed by re-dissolving a lyophilized NMR sample in deuterium oxide. Exchange was monitored in the ¹⁵N HSQC experiment and the first timepoint was available at 11 minutes.

Calculated Membrane Insertion

β 3 residues 713-735 and 744-762 from the calculated structure were embedded in the membrane using a grid search that sampled membrane depth and two orthogonal rotations. Energies were calculated using the E(z) potential,³² and the conformations with the best E(z) energies are depicted in figure 22.

RESULTS

Design of the disulfide-linked α IIb and β 3 cytosolic domains

The NMR construct was engineered to enforce the interface observed in the Monte Carlo model from Chapter 3. Specifically, the TM residues Met987 in α IIb

and Leu712 in $\beta 3$ occur at the model's interface and these positions were substituted with cysteine to crosslink the NMR construct (figure 14).¹⁹ The model predicts an ideal disulfide bond between the two cysteines, and the disulfide bond contains three torsional angles that can accommodate conformational adjustment and dynamics. The subsequent NMR structure of the α IIb/ $\beta 3$ TM heterodimer confirmed this model and that Met987 and Leu712 occur at the α IIb/ $\beta 3$ TM interface.¹²

Additionally, N-terminal SPE helix caps were added upstream of each cysteine to enforce helical structure that might be propagated from the TM helices to the cytosolic domains. The N-terminal portion of the $\beta 3$ subunit was helical, however, the α IIb subunit had no stable secondary structure, consistent with the conformations observed in two recent NMR analyses of the α IIb and $\beta 3$ TM and membrane proximal regions in phospholipid bicelles.^{13; 14} Finally, the construct was characterized with a histidine tag attached to the N-terminus of $\beta 3$ which was used in purification steps.

Circular dichroism spectroscopy demonstrates that the α IIb and $\beta 3$ cytosolic domains interact

We first used circular dichroism (CD) spectroscopy to determine whether crosslinking the α IIb and $\beta 3$ cytosolic domains affects their structure. The CD spectrum of disulfide-linked α IIb/ $\beta 3$ was dominated by signals at 190, 208, and 222 nm which is characteristic of helical secondary structure (figure 16). When

the disulfide bond was reduced, these signals had smaller magnitudes which suggests the disulfide-linked heterodimer has more helical content than its component monomers. The different CD spectra demonstrate that α IIb and β 3 interact when constrained by the engineered disulfide bond.

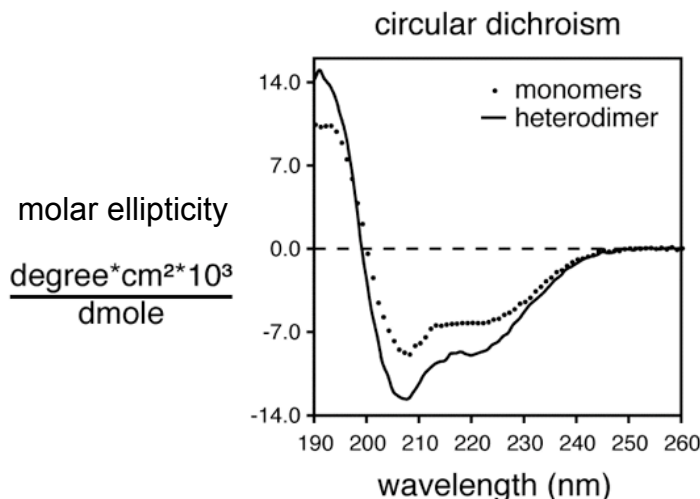


Fig 16. The circular dichroism (CD) spectrum of the disulfide-linked α IIb/ β 3 cytosolic heterodimer (solid line) was dominated by signals at 190, 208, and 222 nm, characteristic of helical secondary structure. When the α IIb and β 3 subunits were separated by reducing the disulfide bond, the monomers had a different CD spectrum (dotted line) which demonstrates that α IIb and β 3 interact when constrained by the disulfide bond, and that this interaction affects their structure.

C α chemical shift indicates that crosslinking the α IIb cytosolic domain to β 3 induces conformational changes in the β 3 subunit

To further probe the interaction between α IIb and β 3, we identified differences in C α chemical shift between the disulfide-linked α IIb/ β 3 heterodimer and a

previously characterized $\beta 3$ monomer. $C\alpha$ chemical shift is an NMR parameter that is sensitive to protein secondary structure and protein interactions.³³ Prior analysis of $C\alpha$ chemical shift found that this parameter demarcated three helical regions in a construct consisting of the monomeric $\beta 3$ TM and cytosolic domains.¹⁵ Similarly, this parameter demarcates three helical regions of $\beta 3$ in the disulfide-linked $\alpha 11b/\beta 3$ heterodimer (figure 17). However, there were significant differences in $C\alpha$ chemical shift between the $\beta 3$ monomer and the $\alpha 11b/\beta 3$ heterodimer encompassing twelve residues that define the N-terminal $\beta 3$ helix in the heterodimer construct. This region includes residues that are considerably downstream of the disulfide bond and demonstrates that the disulfide bond enforces other cytosolic $\alpha 11b/\beta 3$ interactions. Lastly, the $C\alpha$ chemical shifts of $\alpha 11b$ displayed no preference for helical secondary structure, consistent with a recent NMR structure of this region that includes its neighboring TM domain embedded in phospholipid bicelles.¹³

Aliphatic chemical shifts define a $\beta 3$ interface that interacts with $\alpha 11b$

The $\beta 3$ interface that interacts with $\alpha 11b$ was identified by comparing the ^{13}C HSQC spectra of the disulfide-linked $\alpha 11b/\beta 3$ heterodimer and the reduced monomers. Reducing the disulfide bond altered side chain chemical shifts in $\beta 3$ residues Lys716 and Ile719 which indicates that these residues interact with $\alpha 11b$ (figure 18). This interface resides on a face of the N-terminal $\beta 3$ helix that includes Asp723 (figure 21), consistent with predictions that Asp723 is at the

α IIb/ β 3 interface.¹⁷ Additionally, it is consistent with the β 3 interfaces reported in previous NMR structures of the α IIb/ β 3 cytosolic interaction including the structure of the α IIb/ β 3 TM heterodimer that found 7 NOEs between Ile719 and different α IIb residues.^{8; 9; 12}

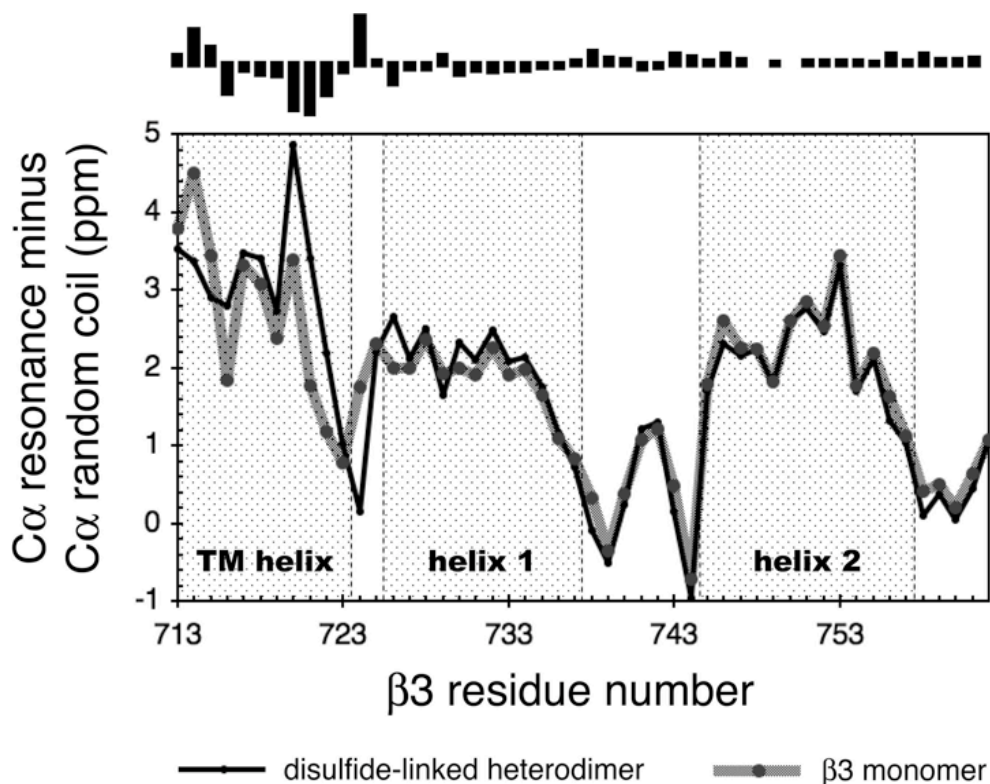


Fig 17. $C\alpha$ resonance is an NMR parameter that is sensitive to protein secondary structure and protein interactions. This parameter demarcated three helical regions of β 3 in the disulfide-linked heterodimer, similar to a previous structural analysis of the β 3 monomer; helices end when the $C\alpha$ resonance approach its random coil value ($C\alpha$ resonance – $C\alpha$ random coil = 0). The α IIb subunit affects $C\alpha$ resonances over the first twelve residues of β 3 in the disulfide-linked construct, suggesting that the disulfide bond enforces additional interactions in this region.

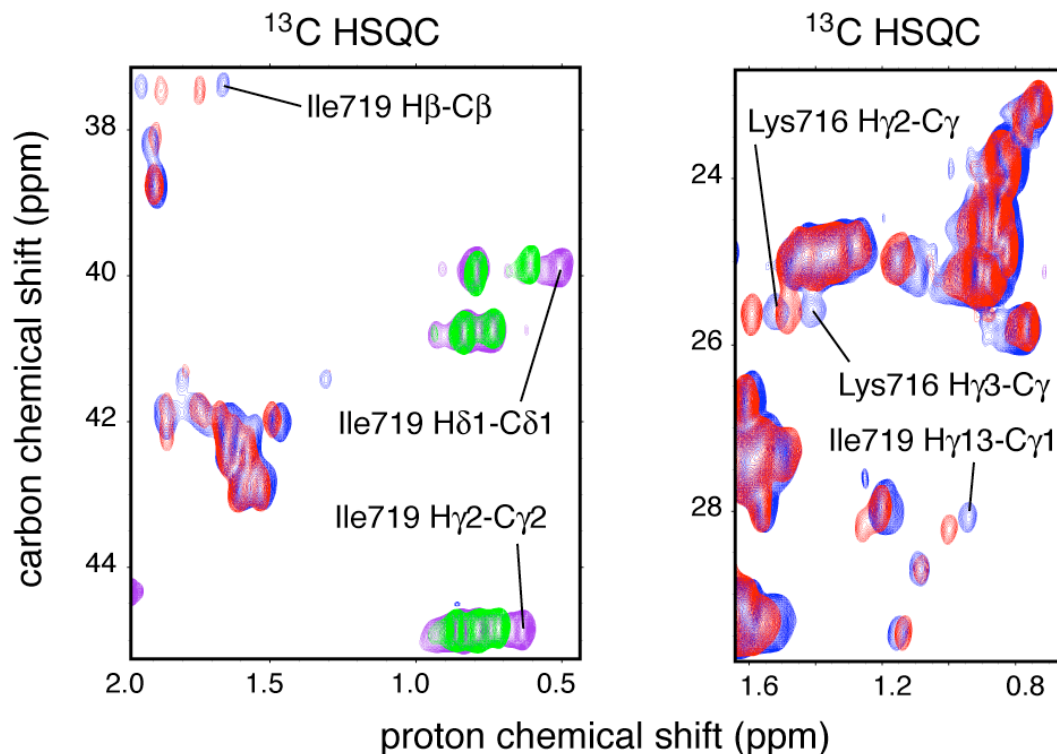


Fig 18. Side chain chemical shifts are sensitive to protein interactions, and the chemical shifts of aliphatic atoms can be visualized in the ^{13}C HSQC spectrum. The blue and purple signals correspond to aliphatic atoms in the disulfide-linked $\alpha\text{IIb}/\beta 3$ heterodimer, and the red and green overlay corresponds to the αIIb and $\beta 3$ monomers. The chemical shifts of $\beta 3$ residues Lys716 and Ile719 change when the disulfide bond is reduced which suggests these residues interact with αIIb . (Purple and green signal is folded in the carbon dimension by 20 ppm).

Analysis of NOEs at the $\alpha\text{IIb}/\beta 3$ heterodimer interface

The analysis of aliphatic chemical shifts demonstrates that $\beta 3$ residues Lys716 and Ile719 interact with αIIb , however these residues displayed no interchain

NOEs with α IIb. Furthermore, the only NOEs between the α IIb and β 3 subunits occur near the disulfide crosslink. Upon first inspection, this result appears to conflict with published NMR structures that found NOEs between the α IIb and β 3 subunits. These analyses relied on innovative strategies to detect interchain NOEs, including transferred NOE experiments⁸ and selective labeling,¹² because the heterodimer could not be isolated from conformations that otherwise masked α IIb/ β 3 interactions. However our experimental design effectively isolates the heterodimer by positioning a covalent tether at its interface. In this construct, a stable interface would be readily defined by standard NOE experiments, and we employed a 4D NOESY that can unambiguously identify NOEs including every interchain NOE observed in the published structure of the α IIb/ β 3 TM heterodimer.¹² The lack of interchain NOEs at the β 3 interface suggests that the α IIb/ β 3 cytosolic interaction is intrinsically disordered, at least in the disulfide-linked construct described here.

Also, the guanidine group of α IIb Arg995 is postulated to form a salt bridge with β 3 Asp723,¹⁷ and our experimental conditions allowed us to unambiguously assign guanidine resonances (figure 20) and detect NOEs between the guanidine groups and other adjacent protons. However the Arg995 guanidine protons did not have any NOEs with the β 3 sidechain. This result is consistent with an intrinsically disordered interface, and the interaction likely reflects a solvent exposed acid/base pair that does not have a unique conformation, but could still impart some specificity between the α IIb and β 3 subunits.^{34; 35}

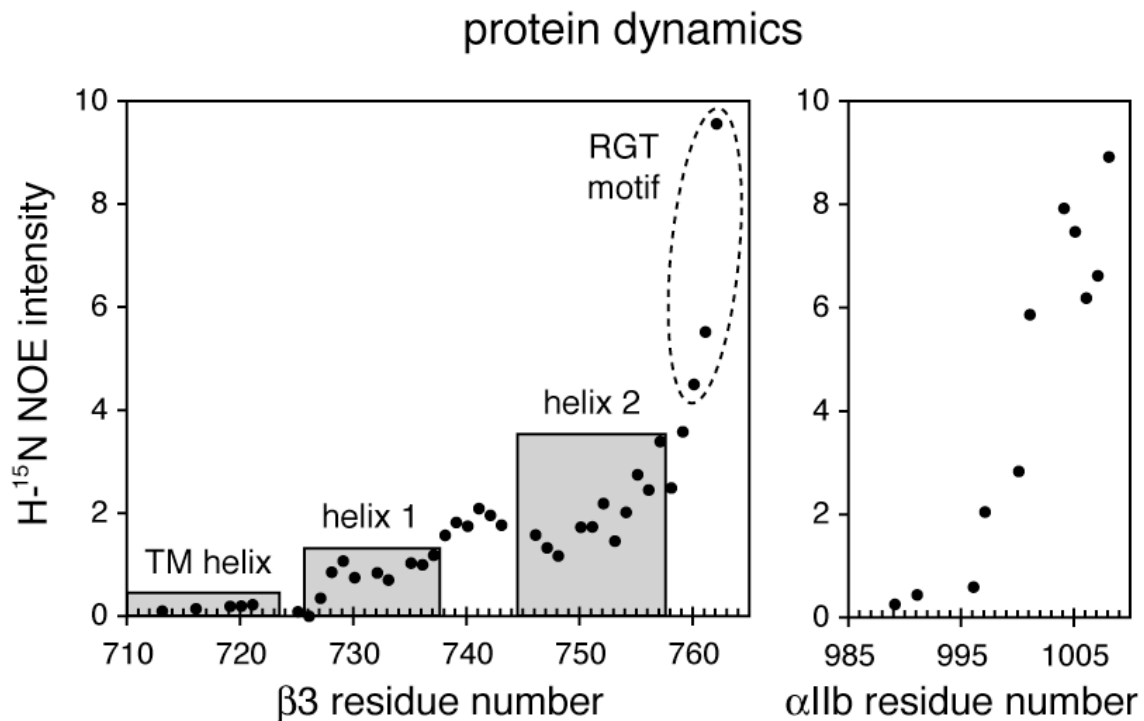


Fig 19. The H-¹⁵N NOE intensity correlates with backbone dynamics. β3 residues Leu713-Glu726 and αIIb residues Val990-Asn996 undergo similar dynamics which suggests that these regions interact. Downstream residues are increasingly dynamic, including two β3 helices. Additionally, the β3 RGT motif is dynamic which suggests it is accessible to cytosolic proteins.

Backbone dynamics are consistent with an αIIb/β3 interaction

Since there were no NOEs between the αIIb and β3 subunits, suggesting an intrinsically disordered interaction, we determined whether the construct's dynamics were consistent with an αIIb/β3 interaction. The strength of the H-¹⁵N NOE correlates with each residue's order parameter and it was analyzed to qualitatively assess the dynamics of the disulfide-linked construct.^{36; 37} The αIIb

residues Val990-Asn996 and $\beta 3$ residues Leu713-Glu726 have similar H-¹⁵N NOE intensities which supports our finding that these regions interact (figure 19). However, the α IIb intensities trended toward higher dynamics, consistent with a natively unstructured region of α IIb interacting with a structured, helical portion of $\beta 3$. Residues downstream of this interacting region were increasingly dynamic in both the α IIb and $\beta 3$ subunits.

Stability of $\beta 3$ secondary structure determined by hydrogen-deuterium exchange

The increasing dynamics in $\beta 3$ corresponds to increasingly divergent structure in published models of $\beta 3$, and since the $\beta 3$ subunit in disulfide-linked α IIb/ $\beta 3$ is predominantly helical, hydrogen-deuterium exchange (HDX) was used to determine whether dynamics can account for differences in helical content. We performed HDX by dissolving a lyophilized NMR sample in deuterium oxide and identifying amide protons that do not exchange with deuterium, indicating they are likely to form stable hydrogen bonds in a helix. The first $\beta 3$ helix was protected from exchange which demonstrates it has stable helical structure (figure 20). In contrast, $\beta 3$'s second two helices were not protected from exchange which demonstrates they have transient helical structure. These results suggest that dynamics might account for differences between $\beta 3$ structures reported in the literature which each depict an N-terminal helix, but have varying degrees of helical content downstream of Asp723.

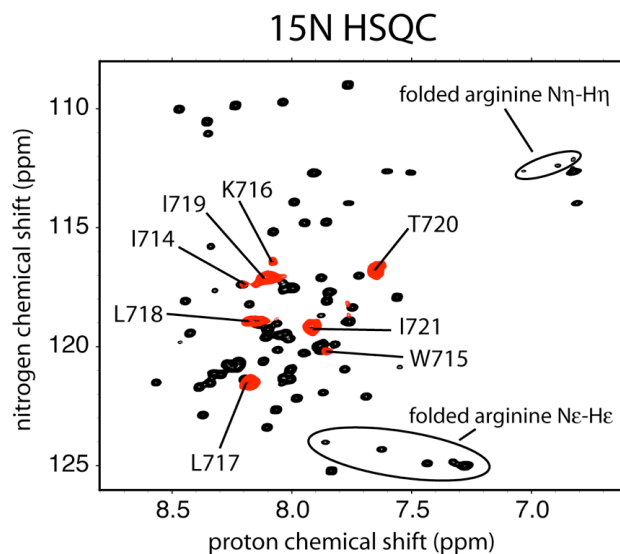


Fig 20. The ^{15}N HSQC detects backbone amines, and in some cases, the arginine guanidine group. When the $\alpha\text{IIb}/\beta\text{3}$ heterodimer is transferred from water to deuterium oxide, most of the protons exchange with deuterium but the red peaks corresponding to β3 residues Ile714-Ile721 are protected from exchange which suggests they form stable hydrogen bonds in a helix. In contrast, other helical regions exchange rapidly with deuterium, suggesting they are dynamic.

β3 structure calculation

NMR structure calculations were used to further define the three-dimensional conformations of αIIb and β3 based on NOEs, the statistical analysis of chemical shift, and $\text{HN-H}\alpha$ J-couplings. However calculations did not converge on a unique conformation for the αIIb subunit because it does not contain secondary structure and has incompatible NOEs, presumably due to a dynamic equilibrium between multiple conformations. Additionally, residual dipolar couplings (RDCs) were measured for use as structural restraints, however they did not aid structure

calculation in part because the structure is dynamic and the RDC alignment tensors vary throughout the construct. This result is consistent with a natively unstructured conformation for α IIb.

In contrast with α IIb, the β 3 subunit contained NOE patterns that defined three helical regions. Additionally, NOEs defined the orientation of the first two helices in the construct. However many unambiguous NOEs are violated in the calculated structure which define (1) alternate rotamer conformations, (2) alternate conformations for the loop linking the last two helices, and (3) alternate conformations for the NPXY motif that terminates the last helix. These violated NOEs were identified in initial structure calculations and eliminated from the calculations used for the structural ensemble presented here to reduce contradictory energy gradients that might otherwise falsely restrain the structure. The quality of the structural ensemble was assessed using PROCHECK-NMR³⁸ and 86.8% of the residues have conformations in the “most favored” region of Ramachandran space; a structure with ideal statistics has >90% but <100% of its residues in the most favored region, and the β 3 structural ensemble has near ideal Ramachandran statistics when the loop regions are not considered.

The β 3 cytosolic domain is primed for interactions with talin, kindlin-3, and Src kinase

The structure and dynamics of β 3 sequence motifs that interact with talin, kindlin-3, and Src kinase were further analyzed to determine whether they could assume

conformations that are compatible with binding (figure 21). First, motifs that bind the cytoskeletal protein talin were compared to crystal and NMR structures that depict $\beta 3$ /talin interactions.^{10; 11} Several $\beta 3$ residues are critical for talin binding

Table V. Summary of the $\beta 3$ structure statistics

NOE restraints residue 1, 2	NOE classification	number of NOEs
i to i	self	87
i to i+1	neighbor	355
i to i+2	turn/loop	57
i to i+3/4	helix	360
i to i>4	turn/loop	29
total		888
NOEs per residue		17.8
violations > 0.5 Å		1*
<hr/>		
Dihedral angle restraints		
TALOS ϕ/ψ		30/30
HN-H α ϕ		16
violations > 10°		0
<hr/>		
Chemical shift restraints		
C α /C β		19/19
H α		50
<hr/>		
HN-H α J coupling restraints		
		11
<hr/>		
Hydrogen bond O...H-N distance restraints		
		5
<hr/>		
Residues in most favored regions of Ramachandran space		
		86.8%

*107 NOEs were omitted from the structure calculation and violated by the final structural ensemble.

structural ensemble of the $\beta 3$ subunit in the disulfide-linked α IIB/ $\beta 3$ heterodimer

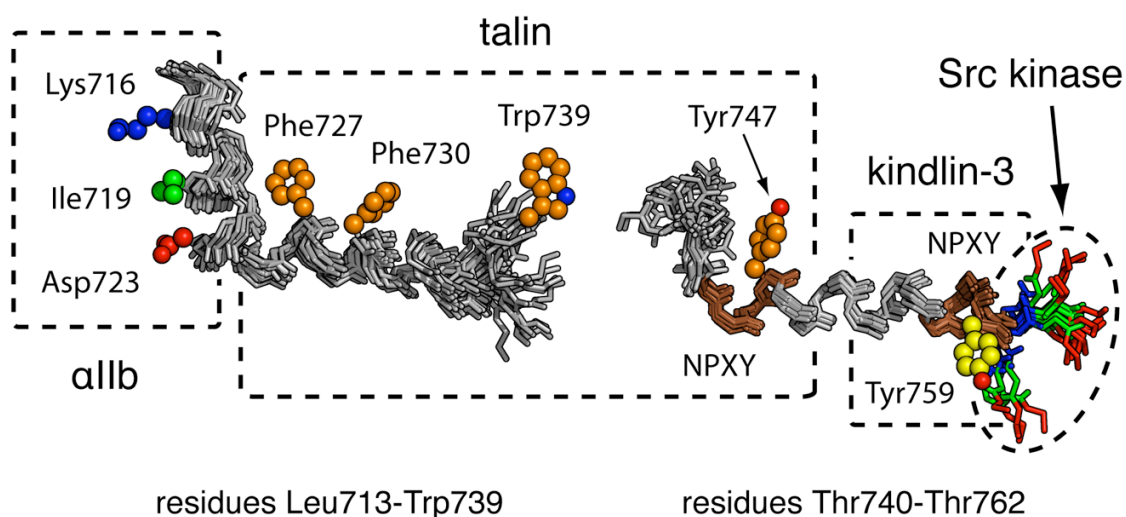


Fig 21. The $\beta 3$ NMR structural ensemble was divided into two regions consisting of residues Leu713-Trp739 and Asp740-Thr762 and the $C\alpha$ atoms of each region were aligned. Domains that interact with α IIB, talin, kindlin-3, and Src kinase are circled and side chains that make up the interfaces are depicted as spheres. Also, the two $\beta 3$ NPXY motifs are colored brown and the RGT motif that binds Src kinase is colored blue/green/red. Notably, the analysis of chemical shift suggests that Lys716 and Ile719 interact with α IIB, and these positions reside on the same face of a $\beta 3$ helix as Asp723, which is postulated to form a salt bridge with α IIB.

including Trp739 and a $\beta 3$ NPXY motif consisting of residues Asn744, Pro745, Leu746, and Tyr747.^{3; 11} Every Trp739 indole resonance was unambiguously assigned and probed in ^{13}C and ^{15}N NOESY experiments edited for indole $^{13}\text{C}/^{15}\text{N}$ frequencies. Several protons displayed NOEs at the solvent resonance, but there were few NOEs with other atoms in the construct. This lack of interaction is

apparent in the structural ensemble and quite rare for a tryptophan indole,³⁹ suggesting that Trp739 is both accessible and primed for interaction. Next, the structure of the $\beta 3$ NPXY motif that binds talin was defined by NOEs, the analysis of chemical shift, and J-coupling which predict that the NPXY is a N-terminal helix cap that nucleates an α -helix. The N-terminal helix cap is similar to conformations observed in previous structures of the $\beta 3$ /talin interaction,¹¹ suggesting that the $\beta 3$ NPXY motif pre-organized to bind talin, however the capping interaction has never been described for an NPXY motif and will be discussed in further detail below. Finally, a previous NMR analysis found that talin interacts with a helical portion of $\beta 3$ containing Phe727 and Phe730,¹⁰ and this portion of $\beta 3$ is helical in the structure presented here, suggestion that this interface is pre-organized to bind talin as well.

The $\beta 3$ interface that binds the cytoskeletal protein kindlin-3 has been defined as a NPXY motif consisting of $\beta 3$ residues Asn756, Ile757, Thr758, and Tyr759 and possibly the upstream residue Ser752,² although the kindlin-3 interface has not been as extensively examined as the talin interface and there are no experimental structures depicting the interaction. The kindlin-binding NPXY motif terminates the final $\beta 3$ helix in the structure presented here, which is not a known interacting conformation. However, this region of $\beta 3$ is highly dynamic and contains NOEs that are incompatible with a single conformation which suggests that the NPXY motif exists in equilibrium between conformations have different affinities for kindlin-3.

Lastly Src kinase binds the last three C-terminal residues of $\beta 3$ Arg760, Gly761, and Thr762 which are called an RGT motif,⁶ although the structural basis for this interaction has not been described in the literature. Our NMR analysis found that the RGT motif is highly dynamic (figure 19), does not have a unique conformation, and does not interact with other portions of the construct which suggests it is accessible and primed for interactions with Src.

NOEs define the $\beta 3$ NPXY Motif as a N-terminal Helix Cap

An NPXY motif consists of residues Asn-Pro-X-Tyr and this sequence binds phosphotyrosine binding (PTB) domains upon phosphorylation (X can be any of several different amino acids).⁴⁰ Talin contains a novel PTB domain that binds the unphosphorylated $\beta 3$ NPXY motif consisting of residues Asn744, Pro745, Leu746, and Tyr747. Previous structures of NPXY motifs found that they adopt a type I β -turn when bound to a PTB domain, however the crystal structure of $\beta 3$ bound to talin is consistent with both a type I β -turn and a N-terminal helix cap.¹¹ In the NMR structure of the disulfide-linked α IIb/ $\beta 3$ heterodimer described here, NOEs define a conformation that is most consistent with a N-terminal helix cap that nucleates the second cytosolic helix in $\beta 3$, suggesting the conformation of the $\beta 3$ NPXY motif is pre-organized to bind talin thereby minimizing its entropic cost of binding (figure 23).

Calculated membrane insertion predicts the regions of $\beta 3$ that interact with talin and kindlin-3 partition into the membrane.

The membrane embedding of the $\beta 3$ subunit was calculated using a membrane insertion potential.³² The N-terminal helix partitions into the bilayer in a membrane spanning orientation which suggests it is an extension of the TM helix (figure 22). Additionally, the two downstream helices partition into the membrane in amphiphilic conformations, derived in part from hydrophobic residues Phe727 and Phe730 in one helix and Tyr747 in the other which also make up portions of the interfaces that interact with talin.^{10; 11} Intriguingly, the membrane embedding of these helices would sequester them from interactions with talin and kindlin-3. This finding suggests that the $\beta 3$ cytosolic domain exists in equilibrium between a membrane bound conformations that cannot bind talin or kindlin-3 and accessible, solvent exposed conformations.

DISCUSSION AND CONCLUSIONS

Previous NMR analyses hypothesized that the native $\alpha IIb/\beta 3$ interface would stabilize a complex between peptides corresponding to their cytosolic and/or TM domains, however a stable 1:1 complex has never been observed. Based on these findings, a disulfide-linked $\alpha IIb/\beta 3$ cytosolic domain was engineered that has 1:1 stoichiometry and enforces an interface observed in models of the $\alpha IIb/\beta 3$ TM heterodimer. The αIIb and $\beta 3$ cytosolic domains interact in the construct and their interface is intrinsically disordered, consisting of a natively unstructured

region of α IIb interacting with a β 3 helix. The structure on the β 3 subunit was calculated, and we further characterized β 3 interfaces that interact with α IIb, talin, kindlin-3, Src kinase, and the cell membrane.

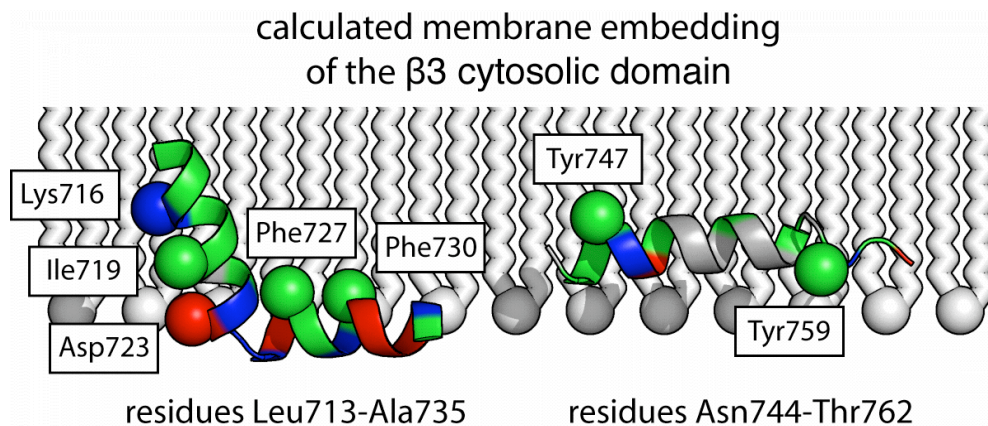


Fig 22. Portions of the calculated β 3 structure were oriented using a membrane insertion potential. The first helix in the β 3 construct embedded in a membrane-spanning orientation and two subsequent helices embedded in amphiphilic orientations. The membrane-embedding of Phe727, Phe730, and Tyr747 would prevent interactions with talin and the membrane embedding of Tyr759 would prevent interactions with kindlin-3.

Prior analysis of the integrin cytosolic domains

The resting integrin's cytosolic heterodimer has been probed by several NMR analyses.^{8; 9; 16; 41} Notably, Vinogradova et al. observed transferred NOEs between peptides corresponding to the integrin's cytosolic domains and calculated a structure based on these interactions.⁸ Several features of the resting integrin support this model, however transferred NOEs require a kinetically unstable interaction, so they are more likely to reflect non-specific

interactions than the standard NMR experiments used for protein structure determination.

Vinogradova's research suggests that the integrin's TM heterodimer is necessary to stabilize the cytosolic heterodimer, and two strategies attempted to account these interactions. First, Ulmer et al. engineered several constructs with N-terminal disulfide bonds and coiled coils that fixed the α IIb and β 3 cytosolic domains in a parallel orientation with 1:1 stoichiometry, however no heteromeric interaction was observed, suggesting that the α IIb and β 3 cytosolic domains do not interact or that the upstream constraints failed to approximate native interactions.¹⁶ To account for native TM domain interactions, Li et al. analyzed peptides containing both the cytosolic and TM domains of α IIb and β 3, however the peptides formed homo-oligomers instead of heterodimers.⁴¹ More recently, Lau and coworkers found experimental conditions that favored heteromeric TM associations and developed an NMR model for this interaction that agrees well with mutagenesis results.¹² However this analysis only considered a fraction of the α IIb and β 3 cytosolic domains and did not report on their dynamics.

We combined prior strategies to stabilize the α IIb/ β 3 heterodimer by including a portion of the α IIb and β 3 TM regions and introducing a disulfide-crosslink at the TM heterodimer interface. This construct is the first stable α IIb/ β 3 cytosolic heterodimer that has an observable α IIb/ β 3 interaction, and it allowed us to define the structure and dynamics of the α IIb and β 3 cytosolic domains in a single conformation that approximates the integrin's resting state.

The α IIb/ β 3 cytosolic interface is intrinsically disordered

CD and NMR analyses revealed that the engineered disulfide-crosslink in the α IIb and β 3 TM region enforced an interaction between the α IIb and β 3 cytosolic domains. However while the chemical shift differences of the β 3 Lys716 and Ile719 side chains demonstrated that these residues interacted with α IIb, they displayed no NOEs with α IIb suggesting the interaction is dynamic on the millisecond timescale. Also, we found that the α IIb subunit had no stable helical structure, consistent with a previous NMR structure of this region in phospholipids bicelles.¹³ Never-the-less, α IIb interacted with β 3 over the length of approximately twelve residues including portions of the β 3 helix that are significantly downstream from the TM crosslink. The interaction suggests that the α IIb subunit does not have random structure, instead it has conformational bias toward structures that interact with β 3. This type of intrinsically disordered interaction is important for proteins that couple multiple different signaling events to an overall equilibrium. For the integrin, an intrinsically disordered interface allows it to couple many different cytosolic events to the overall resting-active equilibrium. The integrin's equilibrium can be affected by talin and kindlin binding,^{2;3} phosphorylation,⁴² and proteolysis,⁴³ and the cytosolic domains convert this dynamic information into a binary signal: resting or active.

The α IIb Arg995- β 3 Asp723 salt bridge is a solvent exposed interaction

Mutations affecting cytosolic residues α IIb Arg995 and β 3 Asp723 activate the integrin, but reciprocal mutations Arg995Asp and Asp723Arg restore the integrin's resting state.¹⁷ This result is widely interpreted as evidence that a salt bridge between Arg995 and Asp723 stabilizes the resting conformation. The current analysis is consistent with the Arg995-Asp723 salt bridge and provides some evidence supporting the interaction. First, Asp723 is at the β 3 interface that interacts with α IIb, and second, the regions that contain Arg995 and Asp723 undergo similar dynamics. However, HDX suggests that Asp723 is solvent exposed, so any interaction between Asp723 and Arg995 would be more similar to an electrostatic interaction on the surface of a protein than a buried interaction that might have a single stable conformation. Solvent exposed electrostatic interactions are dynamic and can be very important for specificity without locking an acid/base pair into a single conformation.^{34; 35} This type of interaction is especially useful for orienting a dynamic interface without providing a driving force that could independently stabilize the α IIb/ β 3 complex.

The β 3 NPXY motif is a N-terminal helix cap

Previous structures of NPXY motifs have been characterized as type I β -turns⁴⁰ and have slightly different structure than the β 3 NPXY motif which was found to be a N-terminal helix cap that nucleates an α -helix. The capping conformation is similar to previous structures of talin bound to β 3, PIPK γ , and a β 3/PIPK γ chimera

that are consistent with both β -turns and N-terminal helix capping motifs, however the NPXY motifs in these structures do not nucleate helices.^{10; 11; 44} It is possible that talin binding favors unfolded conformations of the β 3 helix, which might function to expose the second NPXY motif to bind kindlin-3.

The membrane embedding of β 3 prevents interactions with talin and kindlin-3

The E(z) membrane depth potential predicts that the β 3 cytosolic domain is amphiphilic and residues Phe727, Phe730, and Tyr747 partition into the cell membrane (figure 22), consistent with previous NMR experiments that observed NOEs between DPC and Tyr747.⁷ Membrane insertion sequesters these residues from cytosolic proteins so they cannot bind talin as depicted in structures of the β 3/talin complex, and the membrane embedding of the kindlin-3 binding site would prevent β 3 interactions with kindlin-3 as well. However the cytosolic helices exist in dynamic equilibrium with unfolded states that likely correspond with solvent exposed conformations. Amphiphilic domains that switch their degree of membrane exposure are commonly found proximal to TM helices in proteins including voltage gated channels,^{45; 46} phospholamban,⁴⁷ and the M2 proton channel from influenza.^{48; 49} Recently, the cytosolic domain of the T-cell receptor's CD3 ϵ subunit was found to bind membrane which sequesters aromatic residues from cytosolic proteins including a tyrosine in its NPXY motif.⁵⁰ When the T-cell receptor is activated, the amphiphilic domain is

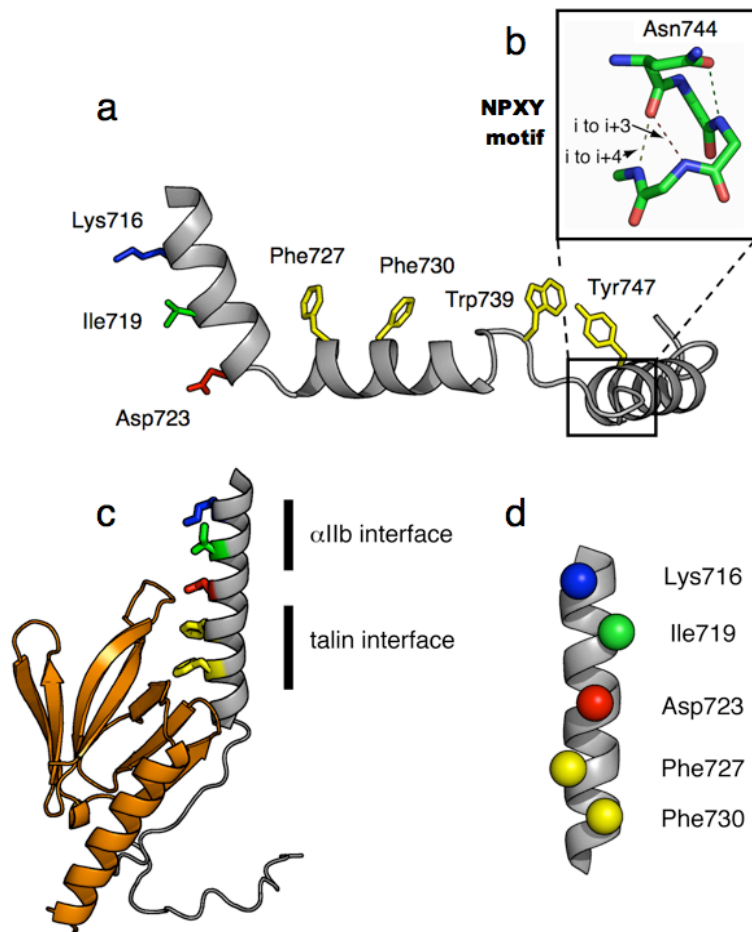


Fig 23. Analysis of talin binding to the $\beta 3$ cytosolic domain. (a) The analysis of chemical shift suggests that $\beta 3$ residues Lys716 and Ile719 interact with α IIb, and these residues lie along the same face of the $\beta 3$ helix that includes Asp723, which is postulated to form a salt bridge with α IIb residue Arg995. Also, prior

structural analysis of $\beta 3$ and talin depict interfaces that contain $\beta 3$ residues Phe727, Phe730, Trp739, and Tyr747. (b) Talin interacts with a $\beta 3$ NPXY motif, and the calculated $\beta 3$ structure suggests that the NPXY motif is a N-terminal helix cap that nucleates an α -helix. (c and d) Talin (orange) cannot bind the $\beta 3$ structure reported here because it is amphiphilic and embeds in the membrane. When $\beta 3$ is modeled as a helix, similar to a previous NMR model that depicts a $\beta 3$ /talin interaction, the α IIb interface resides on the same face of the $\beta 3$ helix as the talin interface. These findings are consistent with the hypothesis that talin shifts the integrin equilibrium toward active conformations and displaces α IIb from heteromeric interactions with $\beta 3$.

released from the membrane and binds signaling proteins. The $\beta 3$ cytosolic domain might have analogous interactions with membrane where the cytosolic domain favors a membrane associated conformation in the resting integrin and talin and/or kindlin-3 shifts the conformation toward a solvent exposed state.

Talin binding to the $\beta 3$ cytosolic domain requires a conformational change

The membrane embedding of the $\beta 3$ cytosolic domain is partially mediated by a kink near the membrane/cytosol interface. However, the kink adopts a helical conformation in an NMR structure of the $\beta 3$ /talin complex which suggests the kink straightens when $\beta 3$ binds to talin.¹⁰ The helical conformation positions Phe727 and Phe730 on the same face of the $\beta 3$ helix as residues that interact with αIIb (figure 23c and d), consistent with the hypothesis that talin displaces αIIb from the $\beta 3$ cytosolic domain.³

The $\beta 3$ RGT motif is accessible to Src kinase

Integrin activation and clustering induces the autophosphorylation of Src kinase which initiates intercellular signaling cascades.⁶ Src constitutively binds the $\beta 3$ cytosolic domain and the $\beta 3$ RGT motif is critical for this interaction. Although most of the $\beta 3$ cytosolic domain is membrane-embedded and inaccessible to cytosolic proteins, the RGT motif is unstructured and should be accessible to Src.

CONCLUSION

The NMR analysis of the disulfide-linked α IIb/ β 3 construct reveals that stable TM interactions are coupled to an intrinsically disordered cytosolic interaction. This structural organization allows the cytosolic domains to integrate many different cytosolic events into a binary signal that activates the integrin or maintains its resting state.

References

1. Bennett, J. S. (2005). Structure and function of the platelet integrin α IIb β 3. *J Clin Invest* **115**, 3363-9.
2. Moser, M., Nieswandt, B., Ussar, S., Pozgajova, M. & Fassler, R. (2008). Kindlin-3 is essential for integrin activation and platelet aggregation. *Nat Med* **14**, 325-30.
3. Tadokoro, S., Shattil, S. J., Eto, K., Tai, V., Liddington, R. C., de Pereda, J. M., Ginsberg, M. H. & Calderwood, D. A. (2003). Talin binding to integrin beta tails: a final common step in integrin activation. *Science* **302**, 103-6.
4. Li, R., Mitra, N., Gratkowski, H., Vilaire, G., Litvinov, R., Nagasami, C., Weisel, J. W., Lear, J. D., DeGrado, W. F. & Bennett, J. S. (2003). Activation of integrin α IIb β 3 by modulation of transmembrane helix associations. *Science* **300**, 795-8.

5. Kim, M., Carman, C. V. & Springer, T. A. (2003). Bidirectional transmembrane signaling by cytoplasmic domain separation in integrins. *Science* **301**, 1720-5.
6. Arias-Salgado, E. G., Lizano, S., Sarkar, S., Brugge, J. S., Ginsberg, M. H. & Shattil, S. J. (2003). Src kinase activation by direct interaction with the integrin beta cytoplasmic domain. *Proc Natl Acad Sci U S A* **100**, 13298-302.
7. Vinogradova, O., Vaynberg, J., Kong, X., Haas, T. A., Plow, E. F. & Qin, J. (2004). Membrane-mediated structural transitions at the cytoplasmic face during integrin activation. *Proc Natl Acad Sci U S A* **101**, 4094-9.
8. Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T., Plow, E. & Qin, J. (2002). A structural mechanism of integrin alpha(IIb)beta(3) "inside-out" activation as regulated by its cytoplasmic face. *Cell* **110**, 587-97.
9. Weljie, A. M., Hwang, P. M. & Vogel, H. J. (2002). Solution structures of the cytoplasmic tail complex from platelet integrin alpha IIb- and beta 3-subunits. *Proc Natl Acad Sci U S A* **99**, 5878-83.
10. Wegener, K. L., Partridge, A. W., Han, J., Pickford, A. R., Liddington, R. C., Ginsberg, M. H. & Campbell, I. D. (2007). Structural basis of integrin activation by talin. *Cell* **128**, 171-82.
11. Garcia-Alvarez, B., de Pereda, J. M., Calderwood, D. A., Ulmer, T. S., Critchley, D., Campbell, I. D., Ginsberg, M. H. & Liddington, R. C. (2003). Structural determinants of integrin recognition by talin. *Mol Cell* **11**, 49-58.

12. Lau, T. L., Kim, C., Ginsberg, M. H. & Ulmer, T. S. (2009). The structure of the integrin α IIb β 3 transmembrane complex explains integrin transmembrane signalling. *Embo J*.
13. Lau, T. L., Dua, V. & Ulmer, T. S. (2008). Structure of the integrin α IIb transmembrane segment. *J Biol Chem* **283**, 16162-8.
14. Lau, T. L., Partridge, A. W., Ginsberg, M. H. & Ulmer, T. S. (2008). Structure of the integrin β 3 transmembrane segment in phospholipid bicelles and detergent micelles. *Biochemistry* **47**, 4008-16.
15. Li, R., Babu, C. R., Valentine, K., Lear, J. D., Wand, A. J., Bennett, J. S. & DeGrado, W. F. (2002). Characterization of the monomeric form of the transmembrane and cytoplasmic domains of the integrin β 3 subunit by NMR spectroscopy. *Biochemistry* **41**, 15618-24.
16. Ulmer, T. S., Yaspan, B., Ginsberg, M. H. & Campbell, I. D. (2001). NMR analysis of structure and dynamics of the cytosolic tails of integrin α IIb β 3 in aqueous solution. *Biochemistry* **40**, 7498-508.
17. Hughes, P. E., Diaz-Gonzalez, F., Leong, L., Wu, C., McDonald, J. A., Shattil, S. J. & Ginsberg, M. H. (1996). Breaking the integrin hinge. A defined structural constraint regulates integrin signaling. *J Biol Chem* **271**, 6571-4.
18. Vinogradova, O., Haas, T., Plow, E. F. & Qin, J. (2000). A structural basis for integrin activation by the cytoplasmic tail of the α IIb-subunit. *Proc Natl Acad Sci U S A* **97**, 1450-5.

19. Li, W., Metcalf, D. G., Gorelik, R., Li, R., Mitra, N., Nanda, V., Law, P. B., Lear, J. D., DeGrado, W. F. & Bennett, J. S. (2005). A push-pull mechanism for regulating integrin function. *Proc Natl Acad Sci U S A* **102**, 1424-9.
20. Luo, B. H., Springer, T. A. & Takagi, J. (2004). A specific interface between integrin transmembrane helices and affinity for ligand. *PLoS Biol* **2**, e153.
21. Walters, R. F. & DeGrado, W. F. (2006). Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* **103**, 13658-63.
22. Walsh, S. T., Cheng, H., Bryson, J. W., Roder, H. & DeGrado, W. F. (1999). Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proc Natl Acad Sci U S A* **96**, 5486-91.
23. Rabanal, F., DeGrado, WF, Dutton, PL. (1996). Use of 2,2'-dithiobis(5-nitropyridine) for the heterodimerization of cysteine containing peptides. *Tetrahedron Letters* **37**, 1347-1350.
24. Sattler M, S. J., Griesinger C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93-158.
25. Lipsitz, R. S., Sharma, Y., Brooks, B. R. & Tjandra, N. (2002). Hydrogen bonding in high-resolution protein structures: a new method to assess NMR protein geometry. *J Am Chem Soc* **124**, 10621-6.

26. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302.
27. Kuszewski, J., Qin, J., Gronenborn, A. M. & Clore, G. M. (1995). The impact of direct refinement against ¹³C alpha and ¹³C beta chemical shifts on protein structure determination by NMR. *J Magn Reson B* **106**, 92-6.
28. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. (1995). The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J Magn Reson B* **107**, 293-7.
29. Vuister, G. W. & Bax, A. (1993). Quantitative J correlation: A new approach for measuring homonuclear three-bond J(HNHA) coupling constants in ¹⁵N-enriched proteins. *J Am Chem Soc* **115**, 7772-7777.
30. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**, 65-73.
31. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. (2006). Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **48**, 47-62.
32. Senes, A., Chadi, D. C., Law, P. B., Walters, R. F., Nanda, V. & Degrado, W. F. (2007). E(z), a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: derivation and

- applications to determining the orientation of transmembrane and interfacial helices. *J Mol Biol* **366**, 436-48.
33. Wishart, D. S. & Sykes, B. D. (1994). The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* **4**, 171-80.
 34. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1992). Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell* **68**, 699-708.
 35. Kohn, W. D., Kay, C. M. & Hodges, R. S. (1998). Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils. *J Mol Biol* **283**, 993-1012.
 36. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-9.
 37. Alexandrescu, A. T. & Shortle, D. (1994). Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J Mol Biol* **242**, 527-46.
 38. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477-86.

39. Samanta, U., Pal, D. & Chakrabarti, P. (2000). Environment of tryptophan side chains in proteins. *Proteins* **38**, 288-300.
40. Uhlik, M. T., Temple, B., Bencharit, S., Kimple, A. J., Siderovski, D. P. & Johnson, G. L. (2005). Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *J Mol Biol* **345**, 1-20.
41. Li, R., Babu, C. R., Lear, J. D., Wand, A. J., Bennett, J. S. & DeGrado, W. F. (2001). Oligomerization of the integrin α IIb β 3: roles of the transmembrane and cytoplasmic domains. *Proc Natl Acad Sci U S A* **98**, 12462-7.
42. Phillips, D. R., Prasad, K. S., Manganello, J., Bao, M. & Nannizzi-Alaimo, L. (2001). Integrin tyrosine phosphorylation in platelet signaling. *Curr Opin Cell Biol* **13**, 546-54.
43. Bialkowska, K., Kulkarni, S., Du, X., Goll, D. E., Saido, T. C. & Fox, J. E. (2000). Evidence that β 3 integrin-induced Rac activation involves the calpain-dependent formation of integrin clusters that are distinct from the focal complexes and focal adhesions that form as Rac and RhoA become active. *J Cell Biol* **151**, 685-96.
44. de Pereda, J. M., Wegener, K. L., Santelli, E., Bate, N., Ginsberg, M. H., Critchley, D. R., Campbell, I. D. & Liddington, R. C. (2005). Structural basis for phosphatidylinositol phosphate kinase type I γ binding to talin at focal adhesions. *J Biol Chem* **280**, 8381-6.

45. Elinder, F., Nilsson, J. & Arhem, P. (2007). On the opening of voltage-gated ion channels. *Physiol Behav* **92**, 1-7.
46. Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. (2007). Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* **450**, 376-82.
47. Traaseth, N. J., Ha, K. N., Verardi, R., Shi, L., Buffy, J. J., Masterson, L. R. & Veglia, G. (2008). Structural and dynamic basis of phospholamban and sarcolipin inhibition of Ca(2⁺)-ATPase. *Biochemistry* **47**, 3-13.
48. Stouffer, A. L., Acharya, R., Salom, D., Levine, A. S., Di Costanzo, L., Soto, C. S., Tereshko, V., Nanda, V., Stayrook, S. & DeGrado, W. F. (2008). Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* **451**, 596-9.
49. Schnell, J. R. & Chou, J. J. (2008). Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **451**, 591-5.
50. Xu, C., Gagnon, E., Call, M. E., Schnell, J. R., Schwieters, C. D., Carman, C. V., Chou, J. J. & Wucherpfennig, K. W. (2008). Regulation of T cell receptor activation by dynamic membrane binding of the CD3epsilon cytoplasmic tyrosine-based motif. *Cell* **135**, 702-13.

CHAPTER 5

Discussion, Future Direction, and Concluding Remarks

DISCUSSION

The previous chapters describe the development of software used to model the $\alpha\text{IIb}/\beta\text{3}$ TM heterodimer, the verification of the model, and the use of the model to engineer an NMR construct that approximates the conformation of the resting integrin's cytosolic domains. Considering first the construction, benchmarking, and validation of modeling algorithms, an ideal structure prediction method would require only a primary amino acid sequence to accurately calculate a protein's three-dimensional coordinates. However, as evidenced in the CASP competition, currently no method can reliably predict a protein's structure without additional information from empirical analyses.¹ Regardless, structure prediction is more of an art than a science when it is not supported by experimentation, and so prediction and experimentation go hand-in-hand. In chapter 2, we describe the implementation of restraints that enforce calculated differences in energy between a wild type model and the same model containing selected point mutations in order to provide a selective advantage for conformations that are consistent with experimental mutagenesis results.² While it is unlikely that the Monte Carlo algorithm will be employed in additional structure prediction efforts,

the energy function and calibration methods provide a framework to implement this new type of thermodynamic restraint in next generation protocols.

Next, chapter 3 provided a quantitative assessment for various models of the α IIb/ β 3 TM heterodimer, and ultimately substantiated the Monte Carlo model. This chapter provides strong corroboratory evidence in favor of the interface reported in the Monte Carlo model, and recently published NMR and Rosetta models likely mark the final global push to determine the α IIb/ β 3 TM heterodimer's structure with every major player agreeing to the consensus model, declaring victory, and elevating the interface from hypothesis to dogma.^{3; 4} Still, several features of the α IIb/ β 3 heterodimer remain unresolved: First, the β 3 TM domain contains a conserved SXXXA dimerization motif that is not involved in the α IIb/ β 3 interaction (corresponding to a GXXXG dimerization motif in other β subunits), and its function has not been discovered. Second, mutations to α IIb Thr981 create a constitutively active state, but its role in stabilizing the resting integrin remain unknown.^{5; 6} Lastly, although the structure of the α IIb GFFKR motif converged on similar conformations in both the NMR and Rosetta models,^{3; 4} and the orientation of these residues agrees well with experimental results--and makes good physiological sense--the interaction between this region of α IIb and the β 3 membrane proximal, cytosolic domain has not been defined at high-resolution, and the significance of a possible α IIb Argg995- β 3 Asp723 interaction remains subject to debate.^{3; 4}

The solution structure of the $\beta 3$ subunit in the disulfide-linked construct sheds light on the membrane proximal $\alpha \text{IIb}/\beta 3$ interaction. Our analysis demonstrates that the αIIb and $\beta 3$ cytosolic domains interact and suggests a specific interface for $\beta 3$, consistent with the $\beta 3$ interface defined by mutagenesis, cysteine crosslinking, and multiple NMR models. In contrast, the literature record has not reported a reproducible αIIb interface; instead our results suggest that αIIb is natively unstructured, and the $\alpha \text{IIb}/\beta 3$ interface is intrinsically disordered. This type of interaction may have evolved to couple multiple different intracellular events to the binary resting/active signal relayed by the TM domains. Lastly, we analyzed the structure the $\beta 3$ cytosolic domain and calculated that helical conformations of the talin and kindlin-3 binding sites partition into the membrane, a topology that is inaccessible to either protein. Thus the binding of talin or kindlin-3 would trap the $\beta 3$ cytosolic domain in an exposed conformation, providing a mechanism for conformational change that could shift the integrin equilibrium toward an active state.

In conclusion, this thesis describes three-dimensional structures for the $\alpha \text{IIb}/\beta 3$ TM and cytosolic heterodimer. These models corroborate experimentally defined $\alpha \text{IIb}/\beta 3$ interfaces and provide a foundation for new, testable hypotheses that back avant-garde mechanisms of integrin activation.

References

1. Cozzetto, D., Giorgetti, A., Raimondo, D. & Tramontano, A. (2008). The evaluation of protein structure prediction results. *Mol Biotechnol* **39**, 1-8.
2. Metcalf, D. G., Law, P. B. & DeGrado, W. F. (2007). Mutagenesis data in the automated prediction of transmembrane helix dimers. *Proteins* **67**, 375-84.
3. Zhu, J., Luo, B. H., Barth, P., Schonbrun, J., Baker, D. & Springer, T. A. (2009). The Structure of a Receptor with Two Associating Transmembrane Domains on the Cell Surface: Integrin α IIb β 3. *Mol Cell* **34**, 234-249.
4. Lau, T. L., Kim, C., Ginsberg, M. H. & Ulmer, T. S. (2009). The structure of the integrin α IIb β 3 transmembrane complex explains integrin transmembrane signalling. *Embo J*.
5. Partridge, A. W., Liu, S., Kim, S., Bowie, J. U. & Ginsberg, M. H. (2005). Transmembrane domain helix packing stabilizes integrin α IIb β 3 in the low affinity state. *J Biol Chem* **280**, 7294-300.
6. Luo, B. H., Carman, C. V., Takagi, J. & Springer, T. A. (2005). Disrupting integrin transmembrane domain heterodimerization increases ligand binding affinity, not valency or clustering. *Proc Natl Acad Sci U S A* **102**, 3679-84.