STRUCTURAL CHARACTERIZATION OF THE INTEGRIN $\alpha_{IIb}\beta_3$

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

Supervisor of Dissertation

Graduate Group Chairperson
dedicated to:

My Mom and Dad, Kathy and Doug Metcalf
ABSTRACT

STRUCTURAL CHARACTERIZATION OF THE INTEGRIN \( \alpha{llb}\beta{3} \)
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 \( \alpha{llb}\beta{3} \) 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 \( \alpha \) and \( \beta \) subunits and each subunit contains a single transmembrane helix that form an \( \alpha/\beta \) 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 \( \alpha{llb}\beta{3} \)'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 $\alpha$IIb/$\beta$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 $\alpha$IIb residue Met987 and $\beta$3 residue Leu712. These residues were mutated to cysteine to crosslink peptides corresponding to the $\alpha$IIb and $\beta$3 cytosolic tails, and the disulfide-linked construct was probed by NMR spectroscopy. 

NMR revealed that the $\alpha$IIb and $\beta$3 cytosolic tails have a dynamic interface. The $\alpha$IIb subunit is natively unstructured and the $\beta$3 subunit consists of a hydrophobic helix followed by two amphiphilic helices. The amphiphilic portions of $\beta$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 $\beta$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 \(\alpha_{IIb}/\beta_3\) 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 \(\alpha_{IIb}/\beta_3\) Transmembrane Heterodimer
Introduction. ................................................................. 37
Materials and Methods. ................................................. 39
Results. ........................................................................ 46
Discussion and Conclusions. ............................................ 57
CHAPTER 4

Structural Characterization of a Disulfide-Linked αIIb/β3 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 αIIb/β3 models with experimental findings
and the structural similarity of these models reported as Cα RMSD .... 52
IV Structural similarity between top performing models and two models that
are currently in press, reported as Cα RMSD ......................... 63
V NMR structure statistics .................................................. 95
LIST OF FIGURES

1  Energy landscapes for the glycophorin 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 glycophorin A fit to its NMR structure......................... 24
5  The model of glycophorin A generated with sequence homology data is
nenearly identical to a structure in the glycophorin 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 αIIb and β3 transmembrane domains.................. 40
9  Monte Carlo model of the αIIb/β3 heterodimer depicting qualitative
    agreement with experimental mutagenesis results......................... 48
10 Slices through the Monte Carlo αIIb/β3 model depicting qualitative
    agreement with experimental cysteine crosslinking results............. 49
11 Graphical representation of the quantitative correlations between different
    αIIb/β3 TM models and experimental mutagenesis results............... 54
12 Graphical representation of the quantitative correlations between different
    αIIb/β3 TM models and experimental cysteine crosslinking results........ 56
13 Structural comparison of different models for the resting αIIb/β3
    transmembrane heterodimer.................................................. 61
14 Sequence and topology of the disulfide-linked αIIb/β3 NMR construct.. 76
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>SDS-PAGE analysis of the disulfide-linked αIIb/β3 NMR sample.</td>
<td>82</td>
</tr>
<tr>
<td>16</td>
<td>Circular dichroism analysis of the disulfide-linked αIIb/β3 construct and its component monomers.</td>
<td>86</td>
</tr>
<tr>
<td>17</td>
<td>Analysis of β3 Cα chemical shifts in the disulfide-linked heterodimer and a β3 monomer.</td>
<td>88</td>
</tr>
<tr>
<td>18</td>
<td>$^{13}$C HSQC spectra for the disulfide-linked heterodimer and its component monomers.</td>
<td>89</td>
</tr>
<tr>
<td>19</td>
<td>H-$^{15}$N NOE intensities in the disulfide-linked heterodimer.</td>
<td>91</td>
</tr>
<tr>
<td>20</td>
<td>$^{15}$N HSQC spectra for the disulfide-linked heterodimer before and after hydrogen-deuterium exchange.</td>
<td>93</td>
</tr>
<tr>
<td>21</td>
<td>NMR structure of the β3 subunit in the disulfide-linked heterodimer depicting interfaces that interact with αIIb, talin, kindlin-3, and Src kinase.</td>
<td>96</td>
</tr>
<tr>
<td>22</td>
<td>Cartoon of the calculated membrane embedding for portions of the β3 subunit.</td>
<td>100</td>
</tr>
<tr>
<td>23</td>
<td>Cartoons that illustrate the β3/talin interaction.</td>
<td>105</td>
</tr>
</tbody>
</table>
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 Cβ-Cβ 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 $f_{\text{ASA}}$ 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\(^\text{TM}\) generated $589 million
in revenue for 2008), 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. 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. 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. 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. 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. 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 $\alpha$IIb$\beta$3 integrin that aided in the engineering of a disulfide-linked $\alpha$IIb/$\beta$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 αIIb/β3 cysteine mutant pairs at the model's heterodimer interface that could covalently crosslink the αIIb and β3 subunits. Based on this finding, cysteines were positioned to enforce the model's TM heterodimer interface in a construct consisting of the αIIb and β3 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 β3 subunit described in chapter 4. The subsequent structural analysis makes several predictions that weren’t apparent a priori. Notably much of the β3 cytosolic domain is pre-organized into conformations that are similar to structures depicted in β3/talin interfaces, 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 β3 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


CHAPTER 2

Development of Software to Model the Integrin

$\alpha_{IIb}/\beta_3$ 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,\(^5\)\(^6\)\(^8\)\(^9\)\(^11\)\(^12\)\(^13\)\(^14\) cross-linking,\(^8\)\(^11\)\(^12\) and/or mutagenesis,\(^8\)\(^15\)\(^16\) or the known structures of homologous\(^17\) 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.\textsuperscript{6; 7} Briggs et al. demonstrated that this approach will identify the native conformation of glycophorin A.\textsuperscript{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 $\alpha$-helical secondary structure. Non-helical conformations can be excluded with structural restraints that maintain appropriate distances between the hydrogen bonding partners in an $\alpha$-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.\textsuperscript{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 $\Delta E$ between the wild type and an ensemble of point mutations for the step's conformation. The $\Delta E$ value should be near zero for mutations that do not affect protein stability and function. Likewise, we expect an unfavorable $\Delta E$ for destabilizing mutations. Conformations with computed $\Delta E$ 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,\textsuperscript{21} the M2 proton channel,\textsuperscript{22} and phospholamban.\textsuperscript{23} For reviews see Engelman et al.\textsuperscript{24} and Senes et al.\textsuperscript{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.\textsuperscript{26, 27} The GXXXG motif is the most overrepresented sequence motif found in TM helices\textsuperscript{28} and presents a strong dimerization signal.\textsuperscript{29} The motif consists of two glycines separated by three amino acids. The glycines enable close contact which permits $\text{C}_\alpha$-$\text{H} \cdots \cdot \text{O}$ hydrogen bonding between helix backbones.\textsuperscript{30, 31} $\text{C}_\alpha$-$\text{H} \cdots \cdot \text{O}$
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.\textsuperscript{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.\textsuperscript{33}

**MATERIALS AND METHODS**

**Potential Function**

We define dimerization energy $E_{\text{dimerization}}$ 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 \textit{in vacuo} with the AMBER united-atom force field for van der Waals interactions.\textsuperscript{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.\textsuperscript{35}
Calibration

We calibrated our scoring function with glycophorin A’s (GPA) published structure and mutational data.\textsuperscript{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 Å with the GPA structure. Specifically, we averaged the scores of the ten best scoring models with an RMSD less than 1.5 Å and the ten best scoring models with an RMSD greater than or equal to 1.5 Å 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. $\phi$, $\psi$, and $\omega$ angles were fixed at $-65^\circ$, $-40^\circ$, and $180^\circ$, 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 glycophorin 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.
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.\textsuperscript{26} and BNIP3 mutants were taken from Suljito et al.\textsuperscript{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.\textsuperscript{37} Single nucleotide polymorphisms (SNPs), orthologs, and homologs of GPA were identified with BLAST.\textsuperscript{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

<table>
<thead>
<tr>
<th></th>
<th>neutral mutations</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNIP3, mutational data\textsuperscript{27}</td>
<td>G178I, H173W, A176C</td>
<td>G180I, G180A, H173A</td>
</tr>
</tbody>
</table>

\textsuperscript{a} The GPA mutational data was also used for calibration.
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 $\theta_1$ and $\theta_2$ varied 360°, the crossing angle $\chi$ varied from -90° to 90° (parallel structures only), translations along a helical axis $z_1$ and $z_2$ varied from –15 to 15 Å, and the diameter $d$ varied from 5 to 9 Å.

### Table II

<table>
<thead>
<tr>
<th>GPA</th>
<th>73ITYIFGVMAGVGTILLISYG1^95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human, SNP</td>
<td>ITLIIIFGVIAGVGTILLISYGI</td>
</tr>
<tr>
<td>Human, SNP</td>
<td>ITLIIIFGVMAGTIGTILLISYGI</td>
</tr>
<tr>
<td>Japanese Macaque</td>
<td>ITLIIIFGVMAGVGTILLISYGI</td>
</tr>
<tr>
<td>Orangutan</td>
<td>ITLIIIFGVMAGVGTILLISYCI</td>
</tr>
<tr>
<td>Gorilla</td>
<td>ITLIIIFGVMAGIIGTILLISYI</td>
</tr>
<tr>
<td>Gibbon</td>
<td>ITLIIIFGVMAGVGTILSIYCI</td>
</tr>
<tr>
<td>Chimpanzee, GPA</td>
<td>ITLIIIFGVMAGVGTILLISYI</td>
</tr>
<tr>
<td>Chimpanzee, GPB</td>
<td>ILLLILGVMAGIIGTILLISYCI</td>
</tr>
</tbody>
</table>

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 $\alpha$-helices are docked with six orthogonal parameters: $\theta_1$ and $\theta_2$ are rotations of a helix about its helical axis, $\chi$ 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.
1. make a random change to one or all six parameters that relate the two helices in space

2. dock the helices with the six parameters

3. set each residue to its most common helical rotamer

4. determine which residues are at the helix-helix interface by calculating interaction energies

5. are there more than ten residues at the helix-helix interface?
   - yes
   - no

6a. optimize rotamers at the helix-helix interface with 1000 Monte Carlo steps

7. calculate the dimerization energy (subtract the energy of the monomeric state)

8. is the dimerization energy favorable (less than zero)?
   - yes
   - no

9. mutate a residue

10. optimize the rotamer of the mutated residue

11. calculate the dimerization energy (subtract the energy of the monomeric mutant state)

12. reset the structure to the optimized wild type conformation

13. calculate the conformation's score

14. accept or reject the new parameters based on the simulated annealing criteria

15. increment the step number, decrease the system's temperature

16. repeat for each mutation

17. has the MCSA cycle begun (has the cycle ever gotten to box 15)?
   - yes
   - no
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) using the Goldstein criterion 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.

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 $\pm 2$ 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 $\chi$ angles: rotamers were optimized with DEE for every amino acid at the helix-helix interface and the rotamer’s $\chi$ angles were further optimized with a grid/Monte Carlo search. Similarly, rotamers of the monomeric states were optimized by DEE and $\chi$ 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. Our model places His173 and Ser172 in proximity but not in contact. The $\chi$ angles of His173 and Ser172 can be changed to isoenergetic states that support a hydrogen bond between the His173 N$\delta$ or N$\varepsilon$ and the Ser172 O$\gamma$. 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 $\Delta E$ between the wild type and select point mutants during each docking step. The $\Delta E$ 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).

\[
\text{score}_{\text{conformation}} = E_{\text{dimerization}} + \text{penalty}_{\text{disruptive}} + \text{penalty}_{\text{neutral}}
\]

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

\[
\text{penalty}_{\text{disruptive}} = \frac{\alpha_{\text{disruptive}}}{n} \sum_{i=1}^{n} \ln \left( \frac{1}{1 + e^{-\beta \Delta E_i}} \right)
\]

\(\Delta E_i\) 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 \(\alpha\) coefficient scales the magnitude of the penalty and the \(\beta\) 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 $\Delta E$ (a) and a neutral mutation should have a near zero $\Delta E$ (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 $\Delta E$ 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).

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

Our optimized $\alpha$ and $\beta$ values are $1.02\times10^{-2}$ 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 Cα 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 Cα 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 Cα hydrogen bonding interactions described by Senes et al.31
Fig 4. The glycophorin A model generated with thermodynamics data (red) fit to the wild type backbone (green), RMSD 1.30 Å (a). The glycophorin 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 glycophorin A model generated with naturally occurring sequence variation data (blue) fit to structure 19 from the glycophorin A pdb (green), RMSD 0.54 Å. The glycophorin A pdb file 1AFO contains twenty different NMR structures and structure 19 shows the strongest Cα-H⋯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 Cα RMSD of 0.54 Å with the GPA NMR structure that shows the strongest Cα-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 Cα hydrogen bonds. These include symmetrical hydrogen bonds between the Ile177A Cα and Ala176B carbonyl, the Gly180A Cα and Ile181A Cα, and lle177B carbonyl, and the lle181A Cα 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. If a single conformation is considered, stabilizing mutations are rare and small in magnitude. Also, several mutations stabilize the integrin $\alpha_{IIb}$ TM homodimer but each is predicted to stabilize a non-native conformation. 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 glycophorin 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. Additionally, it identified both GPA-like and non-GPA-like structures for the integrin $\alpha_{IIb}/\beta_3$ TM heterodimer. Furthermore, it can distinguish between different subtypes of GXXXG motifs. Each of the above proteins contains one or more glycine-containing motifs which rely on geometric complementarity and potentially $C\alpha-H\cdots\cdots O$ 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. 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.

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. 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 Cα RMSD of 1.06 Å, validating the model, and more importantly, the modeling method.

**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 glycophorin 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


CHAPTER 3

Multiple Approaches Converge on the Structure of the Integrin αIIb/β3 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 αIIbβ3. In its active conformation, αIIbβ3 binds fibrinogen, von Willebrand factor, or fibronectin and mediates platelet aggregation when these αIIbβ3-bound ligands crosslink adjacent platelets. To prevent the deleterious formation of intravascular platelet aggregates, αIIbβ3 is maintained in a resting conformation on circulating platelets. Following vascular injury, αIIbβ3 is rapidly activated, enabling it to mediate the formation of a hemostatic platelet plug. The formation and disruption of the αIIb/β3 TM domain heterodimer are key events in shifting αIIbβ3 between resting and active conformations. Thus,
there has been considerable effort to produce three-dimensional structural models of the TM domain heterodimer. However, each published model is substantially different and none have accounted well for the consequences of introducing mutations into the αIIb and β3 TM domains. Because of the absence of a satisfactory model for the αIIb/β3 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. In the second strategy, we used a threading approach in which the sequences of the αIIb and β3 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 αIIb/β3 TM heterodimer.

MATERIALS AND METHODS

Mutations used in the Monte Carlo Modeling Algorithm

We first modeled the αIIb/β3 TM domain heterodimer using the Monte Carlo-based structure prediction strategy described in Chapter 2. The αIIb mutants Val969Asn, Leu970Asn, Leu974Asn, Gly975Asn, Leu983Ala and β3 mutants Ser699Asn, Val700Asn, Gly702Asn, Ile704Asn, Leu705Asn were scored as
neutral mutations, and the αIIb mutants Gly972Asn, Gly972Ala, Gly972Leu, Gly976Leu and β3 mutants Met701Asn and Gly708Asn were scored as disruptive mutations.\textsuperscript{2,3}

\begin{align*}
\alpha_{IIb} & \quad \text{I}_{966} \text{WWVL VGVGL GLLL L TILVL AMW}_{988} \\
\beta_{3} & \quad \text{I}_{693} \text{LV VLLSV MGAIL LIGLA ALLIW}_{715}
\end{align*}

Fig 8. Sequences of the αIIb and β3 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, Cα RMSDs were calculated between the model and conformations found in high resolution crystal structures from the orientations of proteins in membranes (OPM) database.\textsuperscript{12} Of the parallel helix dimers parsed from OPM structures, 28% (113 of 400) had Cα RMSDs less than 1.5 Å with the Monte Carlo model over at least ten residues from both the αIIb and β3 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.\textsuperscript{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.\textsuperscript{15} The $\alpha$IIb amino acids Ile966-Trp988 and $\beta_3$ amino acids Ile693-Trp715 were threaded through 214 parallel TM helix dimers parsed from pdb files 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 $\alpha$IIb amino acids and fifteen $\beta_3$ 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,\textsuperscript{16} then
each model was energy minimized in NAMD using the CHARMM force field.\textsuperscript{17; 18} NAMD minimization consisted of 2000 conjugate-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 αIIb-β3 amino acid pairs Gly972-Leu697, Gly972-Val700, Val969-Val696, Val971-Leu697, and Trp968-Val696 when the pair is mutated to cysteine.\textsuperscript{19} Accordingly, the distance between the Cβ atoms of each pair was calculated to determine whether a model was consistent with these results. Gly972 was mutated to alanine to add its Cβ atom, and any Cβ-Cβ 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 αIIb-β3 residue pairs. It consists of a template from the 1iwg pdb for the crystal structure of bacterial multidrug efflux transporter AcrB\textsuperscript{20} in which the αIIb TM amino acids Trp967-Trp988 were threaded onto 1iwg chain A residues 392-413, and the β3 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.\textsuperscript{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).  

**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 αIIb-β3 cysteine mutant pairs to determine the extent to which they could crosslink the αIIb and β3 subunits. 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 Cβ atoms of each pair was calculated for a given model. Glycine was mutated to alanine to add its Cβ atom. A plot of the Cβ 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 \( Y_{ij} \) of the disulfide between the i and j residues in a given double mutant to the distance between their Cβ atoms \( d_{ij} \) in a given model.
\[ Y_{i,j} = Y_{\text{max}} \frac{1}{1 + \left( \frac{d_{i,j} - 4.0}{d_{0.5} - 4.0} \right)^n} \]

in which \( Y_{\text{max}} \) 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 \( \alpha_{IIb}/\beta3 \) 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

\[
f_{ASA,i} = 1 - \left( \frac{\text{model}_{ASA,i}}{\text{monomer}_{ASA,i}} \right)
\]

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 αIIb Arg995 and β3 Asp723 stabilizes the integrin’s resting state. The αIIb and β3 helices in the Monte Carlo and threaded models were extended to Arg995 and Asp723 using ideal backbone geometries ($\phi = -65^\circ; \varphi = -40^\circ; \omega = 180^\circ$) 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 αIIb/β3 Model

In the Monte Carlo-based algorithm, two straight helices consisting of αIIb amino acids Ile966-Trp988 and β3 amino acids Ile693-Trp715 were docked by randomly altering the six orthogonal parameters that orient two cylinders in space. 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. This strategy enabled
us to accurately predict the structures of the TM homodimers for glycophorin A and BNIP3 (see Chapter 2).\textsuperscript{11, 29}

When applied to \(\alpha\)IIb and \(\beta\)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,\textsuperscript{15} and its conformation is similar to >100 different TM dimer interfaces reported in the protein data bank (see Methods). The heterodimer interface for \(\alpha\)IIb consisted of residues Trp968, Val969, Gly972, Gly976, Leu980, Leu983, and Met987 and the \(\beta\)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 \(\alpha\)IIb\(\beta\)3 TM region, even though cysteine crosslinking data was not considered in the modeling procedure (figures 10 and 12).\textsuperscript{19} The structure is also consistent with mutational analyses of the \(\alpha\)IIb and \(\beta\)3 TM domains with the exception of mutations involving \(\alpha\)IIb residue Thr981 that activate \(\alpha\)IIb\(\beta\)3 expressed in tissue culture cells, but reside on the opposite side of the \(\alpha\)IIb helix from other activating mutations.\textsuperscript{2, 3, 10, 22}

In addition to a TM heterodimer, \(\alpha\)IIb\(\beta\)3 function is thought to be constrained by a “clasp” involving membrane-proximal portions of the \(\alpha\)IIb and \(\beta\)3 cytoplasmic domains, a notable feature of which is a salt-bridge between Arg995 in \(\alpha\)IIb and Asp723 in \(\beta\)3.\textsuperscript{27} Several previous NMR models of the \(\alpha\)IIb and \(\beta\)3 cytoplasmic domains predict that Arg995 and Asp723 reside in helices, implying
that the αIIb and β3 TM helices might extend into the cytosol, at least through Arg995 and Asp723. When our Monte Carlo-derived model is propagated into the cytosol with straight helices, the distance between the Arg995 and Asp723 Cβ 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 αIIb/β3 TM heterodimer. On the left, the αIIb helix is depicted as a surface representation (red) and the β3 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 β3 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 αIIb/β3 Model**

To verify the Monte Carlo structure, we used threading as a different approach to derive a model for the αIIb/β3 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 αIIb and β3 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 SCWRL3.0\textsuperscript{16} followed by 2000 conjugant-gradient steps in NAMD,\textsuperscript{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 Cβ atoms of five αIIb-β3 residue pairs having a high propensity to form a disulfide bond when the pair is mutated to cysteine.\textsuperscript{19} (The remaining 115 experimentally evaluated cysteine mutant pairs were saved for structure validation, see below). The structure with the most consistent average Cβ distance came from the 1iwg pdb of the crystal structure for the bacterial multidrug efflux transporter AcrB\textsuperscript{20} consisting of αIIb Trp967-Trp988 threaded on chain A residues 392-414 and β3 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 αIIb residues Gly972-Gly975 and β3 residues Ser699-Gly702, and this conformation is characteristic of a canonical GXXXG interaction,
which is a dimerization motif found in TM helices. While the β3 helix is relatively straight, the αIIb helix is kinked by 35° between residues Gly975 and Gly976 which extends the αIIb-β3 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 β3 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 Cα 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 αIIb/β3 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). The other two models were generated from molecular dynamics simulations of integrin homologs that converged on two conformations,
TABLE III. Comparison of $\alpha$IIb/$\beta$3 TM models agreement with empirical data

<table>
<thead>
<tr>
<th>model</th>
<th>M</th>
<th>T</th>
<th>A</th>
<th>B</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>disulfides</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>+</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>mutagenesis</td>
<td>++</td>
<td>+++</td>
<td>++</td>
<td>-</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>salt bridge</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Model similarity, $C_\alpha$ RMSD in Å

<table>
<thead>
<tr>
<th>model</th>
<th>M RMSD</th>
<th>T RMSD</th>
<th>A RMSD</th>
<th>B RMSD</th>
<th>1 RMSD</th>
<th>2 RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monte Carlo</td>
<td>1.3</td>
<td>*</td>
<td>*</td>
<td>2.4</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Threading</td>
<td>*</td>
<td>1.3</td>
<td>*</td>
<td>*</td>
<td>2.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Model A</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Model B</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>2.4</td>
<td>2.4</td>
<td>*</td>
<td>*</td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.1</td>
<td>1.6</td>
<td>*</td>
<td>*</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>reference</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* denotes a $C_\alpha$ 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 $C_\alpha$ 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 $\alpha$IIb/$\beta$3 (literature models 1 and 2). Additional models of the $\alpha$IIb/$\beta$3 TM heterodimer have been reported but were
not considered here because they did not include atomic coordinates or contain a number of D-amino acids.

First, we considered how well each model correlates with the consequences of mutating either the \( \alpha_{IIb} \) or the \( \beta_3 \) TM domain, focusing on mutations that induce constitutive \( \alpha_{IIb}\beta_3 \) 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 \( \alpha_{IIb} \), blue for \( \beta_3 \)). 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 αIlb ( ) and β3 ( )

Model A

Model B

Model 1

Model 2

Monte Carlo Model

Threaded Model

αIlb residue number

β3 residue number

\[ R = 0.32 \]

\[ R = 0.06 \]

\[ R = 0.19 \]

\[ R = 0.39 \]

\[ R = 0.46 \]

\[ R = 0.57 \]
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 αIIb/β3 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 αIIb and β3 TM domains. Luo et al. expressed full length αIIbβ3 in 293T cells with single cysteine replacements in both the αIIb and β3 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. Thus, the distance between the Cβ 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 αIIb and β3 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. Cβ 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 Cα–Cβ bond vectors and the local dynamics of the structure. 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.

The Cβ distances in the Monte Carlo model correlate well with the cysteine crosslinking data (R = 0.82); furthermore, the equation 4 parameters of $d_{0.5} = 7.8$ Å and $n = 4$ are consistent with values reported in the literature. A similar good correlation was observed for the threaded model (R = 0.77; $d_{0.5} = 7.7$ Å; $n = 3.1$), and for literature model 2 (R = 0.73; $d_{0.5} = 8.5$ Å; $n = 3.4$). A less good correlation was observed for literature model 1 (R = 0.56; $d_{0.5} = 6.9$ Å; $n = 1.6$), 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, crosslinking, FRET experiments, electron microscopy, crystallographic and NMR analyses of integrin fragments, and molecular modeling. Notably, the extracellular portions of the integrins αvβ3 and αIβ3 have been crystallized in conformations that are believed to represent their resting and active states.
NMR has been used to obtain structures of peptides corresponding to the individual αIib and β3 TM domains,\textsuperscript{53; 54} individual cytosolic domains,\textsuperscript{30; 34} and complexes between the αIib and β3 cytosolic domains.\textsuperscript{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 αIib/β3 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.\textsuperscript{7; 56; 57} αIibβ3 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 αIib/β3 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.\textsuperscript{8} These simulations converged on a single structure that is consistent with a right-handed coiled coil. Recently, Gottschalk modeled the αIib/β3 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 \(\alpha\text{llb}\beta3\), Adair and Yeager proposed that the TM domains of resting \(\alpha\text{llb}\beta3\) 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 \(\alpha_{\text{IIb}}/\beta_3\) 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 function\(^{11}\) 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 \(\alpha_{\text{IIb}}\) 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 \(\alpha_{\text{IIb}}/\beta_3\) TM heterodimer is buttressed by one or more empirical findings; however, each model has substantially different structure.
Fig 13.  (a) The αIIb helices of each model were aligned and a single αIIb helix is displayed as a gray ribbon. Models have similar αIIb interfaces if their β3 helices overlap (cylinders).  (b) The β3 helices of each model were aligned and a single β3 helix is displayed as a gray ribbon. Models have similar β3 interfaces if their αIIb 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 αIIb/β3 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 αIIb and β3 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 Cα 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 αIIb/β3 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

<table>
<thead>
<tr>
<th></th>
<th>Monte Carlo</th>
<th>Threading</th>
<th>Model 2</th>
<th>NMR</th>
<th>Rosetta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monte Carlo</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Threading</td>
<td>1.3</td>
<td>1.6</td>
<td>1.3</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>1.1</td>
<td>1.6</td>
<td>1.5</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>NMR</td>
<td>1.2</td>
<td>1.3</td>
<td>1.5</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Rosetta</td>
<td>1.9</td>
<td>1.4</td>
<td>2.3</td>
<td>1.7</td>
<td></td>
</tr>
</tbody>
</table>

Reference values correspond to Cα RMSDs in Å

Conclusion

We generated two models of the αIIb/β3 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 αIIb/β3 TM heterodimer. These models have now been confirmed by its recently published NMR structure.

References


CHAPTER 4

Structural Characterization of a Disulfide-Linked αIIb/β3 Cytosolic Domain

INTRODUCTION

One of the widely studied examples of regulated integrin function is the platelet fibrinogen receptor αIIbβ3.\(^1\) On a circulating platelet, αIIbβ3 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 αIIbβ3 to engage soluble ligands that crosslink adjacent platelets and form a hemostatic platelet plug. The activation of resting αIIbβ3 proceeds through intracellular "inside-out" signaling cascades that enable the cytoskeletal proteins talin and kindlin-3 to bind the β3 cytosolic domain, favoring conformations of β3 that displace αIIb from heteromeric αIIb/β3 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 β3 cytosolic domain.\(^6\)
The β3 cytosolic domain interacts with αIIb, talin, kindlin-3 and Src kinase, and the structural basis of these interactions has been previously probed by NMR and crystallography. 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 αIIb and β3 cytosolic domains are dynamic, 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 β3 structure contains a helix that extends to approximately Asp723, and analyses that include the β3 TM domain depict this region as an extension of the TM helix. This portion of β3 is important for stabilizing the integrin’s resting state, including a putative salt bridge between β3 Asp723 and αIIb Arg995. Downstream of Asp723, published three-dimensional models have divergent structure, including regions of β3 that interact with talin, kindlin-3, and Src kinase. These regions are dynamic, which might account for structural differences between various models. Similarly, the αIIb cytosolic domain is dynamic, which might account for differences between its published conformations.

The structural analysis of the αIIb/β3 cytosolic heterodimer is challenging because of its dynamics and because peptides corresponding to the αIIb and β3
cytosolic domains do not form a stable complex. Never-the-less, NMR structures of the αllb/β3 cytosolic interaction have been solved for heteromeric interactions observed in equilibrium with other states. Notably, Vinogradova et al developed a NMR model for the monomeric αllb and β3 cytosolic domains docked using 13 distance restraints derived predominantly from transferred NOEs. More recently, Lau et al developed a NMR model of the αllb/β3 TM and membrane proximal regions that agrees with many empirical findings, including mutagenesis of residues in the TM and cytosolic regions. 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.

Because published NMR models of the αllb/β3 cytosolic interaction have not converged on similar structures in part because the αllb/β3 interaction is not stable, we built a stable αllb/β3 heterodimer by positioning a disulfide bond in the αllb/β3 interface (figure 14). This strategy was previously employed to study the resting integrin’s αllb and β3 cytosolic domains, however no interaction was observed. Since these first generation constructs, we published a model of the αllb/β3 TM heterodimer 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 αllb and β3 subunits in the full length
Fig 14. Design of the disulfide-linked αIIb/β3 cytosolic heterodimer. (top) The sequences of the αIIb and β3 subunits; each subunit includes a SPE helix cap followed by the cysteine C used for crosslinking. After the cysteine, the αIIb and β3 cytosolic domains consist of residues W715-E1008 and L713-T762, respectively. Additionally, the β3 subunit contained an upstream histidine tag that was used during purification steps. The cysteines were positioned based on the αIIb/β3 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 αIIb Arg995-β3 Asp723 salt bridge, which is postulated to stabilize the integrin’s resting conformation, and β3 cytosolic domains that bind talin, kindlin-3, and Src kinase.
integrin. 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 αIIb and β3 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 αIIb and β3 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 β3 interacts with αIIb along the face of a helix containing β3 residues Lys716, Ile719, and Asp723, similar to the published structure of the αIIb/β3 TM heterodimer. However there were no NOE interactions between αIIb and β3 downstream of the disulfide bond, the αIIb subunit had no unique conformation, and the αIIb interface could not be defined. Taken together, these results suggest a physical yet likely disordered interaction between a β3 helix and a natively unstructured region of αIIb.

The β3 solution structure was solved with the disulfide-linked αIIb 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 β3 TM and cytosolic regions. Portions of the β3 interfaces that bind talin had conformations that are similar to those observed in
structures of β3/talin complexes, suggesting that the β3 cytosolic domain is pre-organized to bind talin. In contrast, the β3 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 β3 cytosolic domain RGT, and this region was found to be unstructured and dynamic, possibly primed for interactions with Src.

Finally, different portions of the β3 structure were analyzed with a membrane insertion potential to define their calculated membrane embedding. 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 β3 structure defines conformations of the resting integrin that are inaccessible to either talin or kindlin-3.

MATERIALS AND METHODS

Plasmid Constructs
The αIIb cytosolic domain was cloned into a pET-16b vector (Novagen) as a C-terminal fusion with the designed protein α3D, a small, highly stable protein that
expresses well in bacteria. A linking region was engineered to include proline to disrupt secondary structure, two glycines as a flexible spacer, and a thrombin cleavage site. The αIIb construct immediately followed including a SPE N-terminal helix cap and the cysteine used to crosslink the αIIb and β3 subunits. The construct’s sequence is: α<sub>3</sub>D-P-GG-LVPR-SPE-C-WKVGFKRNRPPL EEDDEEGE. The italicized residues were used for expression, purification, and crosslinking while the non-italicized sequence corresponds to αIIb residues Trp988-Glu1008.

The β3 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 αIIb and β3 subunits. The sequence of the β3 peptide is: MGSSHHHHHHSSGLVPRGSHE-SPE-C-LIWKLITIHDRKEFAKFEERARAKWDTANNPLYKEATSTFTNITYRGT. The italicized residues were used for expression, purification, and crosslinking while the non-italicized sequence corresponds to β3 residues Leu713-Thr761.

All constructs were confirmed by DNA sequencing.

Expression and Purification

The α<sub>3</sub>D-αIIb 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 µg/ml Ampicillin, then induced with 1 mM IPTG for four hours at 37°C. The media contained <sup>13</sup>C glucose and/or <sup>15</sup>N 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 α₃D-αIIb 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 -80ºC. This protocol generates >10mg of the αIIb cytosolic domain per liter culture. The product’s molecular weight was confirmed by mass spectroscopy.

Likewise, the β₃ 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 µg/ml Ampicillin then induced with 1 mM IPTG for four hours at 37ºC. The media contained ¹³C glucose and/or ¹⁵N ammonium chloride to label the peptide for NMR experiments. The β₃ peptide was purified on Ni-NTA resin using the QIAExpressionist protocol for denaturing conditions (Qiagen). Eluted β₃ peptide was reduced with excess TCEP and further purified by reverse phase HPLC. Purified β₃ was lyophilized and stored at -80ºC. The histidine tag was not
removed from the β3 cytosolic domain, and this protocol makes 10mg of the β3 construct per liter culture. The product’s molecular weight was confirmed by mass spectroscopy.

**Coupling αIIb and β3**

The β3 peptide was dissolved at 5mg/ml in 2M guanidine, buffered to pH 6.0 with 50mM MES. The reduced β3 cysteine thiol was activated for coupling with 5-fold molar excess 2-2'dithio-bis-(5-nitropyridine) dissolved DMSO. Activated β3 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 β3-thio-nitropyridine was lyophilized. Activated β3 was reacted with a slight molar excess of reduced αIIb in guanidine-MES buffer, pH 6.5 for several hours. Crosslinked αIIbβ3 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 10µM disulfide-linked αIIb/β3 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 αIIb/β3, and the same sample with the addition of reducing agent. The NMR sample has a molecular weight that corresponds to αIIb/β3, and it is composed of subunits with molecular weights that correspond to the individual αIIb and β3 peptides. Samples were further validated using a battery of more rigorous techniques including mass spectroscopy.

NMR Spectroscopy

The disulfide linked αIIbβ3 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.  

β3 Structure Calculation

NOE distance restraints were derived from a 3D $^{15}$N-edited NOESY, 3D aliphatic $^{13}$C-edited NOESY, 3D aromatic $^{13}$C-edited NOESY, and a 4D $^{13}$C-edited NOESY. Hydrogen bonding distance and geometry restraints 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 $\phi$ and $\psi$ dihedrals of β3 residues Ile714-His722, Phe727-Arg736, and Tyr747-Ile757 were restrained based on the statistical analysis of chemical shift using the TALOS algorithm which predicts that these regions are helical. For other residues, chemical shift restraints were implemented for their C$\alpha$ and C$\beta$ atoms, and chemical shift restraints were implemented for every H$\alpha$ atoms. Finally, HN-H$\alpha$ J-couplings were measured in the HNHA experiment and implemented as $\phi$ dihedral angle restraints for helical regions or J-coupling restraints for regions with less defined secondary structure which restricts the $\phi$ dihedral to 1-4 possible angles.
6000 structures were calculated with XPLOR-NIH using a protocol similar to the anneal.py example script that comes with the software package.\textsuperscript{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 $^{15}$N HSQC experiment and the first timepoint was available at 11 minutes.

Calculated Membrane Insertion

$\beta$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,\textsuperscript{32} and the conformations with the best E(z) energies are depicted in figure 22.

RESULTS

Design of the disulfide-linked $\alpha$IIb and $\beta$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 $\alpha$IIb
and Leu712 in β3 occur at the model’s interface and these positions were substituted with cysteine to crosslink the NMR construct (figure 14). 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 αIib/β3 TM heterodimer confirmed this model and that Met987 and Leu712 occur at the αIib/β3 TM interface.

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 β3 subunit was helical, however, the αIib subunit had no stable secondary structure, consistent with the conformations observed in two recent NMR analyses of the αIib and β3 TM and membrane proximal regions in phospholipid bicelles. Finally, the construct was characterized with a histidine tag attached to the N-terminus of β3 which was used in purification steps.

**Circular dichroism spectroscopy demonstrates that the αIib and β3 cytosolic domains interact**

We first used circular dichroism (CD) spectroscopy to determine whether crosslinking the αIib and β3 cytosolic domains affects their structure. The CD spectrum of disulfide-linked αIib/β3 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 αIIb and β3 interact when constrained by the engineered disulfide bond.

![Circular dichroism](image)

Fig 16. The circular dichroism (CD) spectrum of the disulfide-linked αIIb/β3 cytosolic heterodimer (solid line) was dominated by signals at 190, 208, and 222 nm, characteristic of helical secondary structure. When the αIIb and β3 subunits were separated by reducing the disulfide bond, the monomers had a different CD spectrum (dotted line) which demonstrates that αIIb and β3 interact when constrained by the disulfide bond, and that this interaction affects their structure.

**Cα chemical shift indicates that crosslinking the αIIb cytosolic domain to β3 induces conformational changes in the β3 subunit**

To further probe the interaction between αIIb and β3, we identified differences in Cα chemical shift between the disulfide-linked αIIb/β3 heterodimer and a
previously characterized β3 monomer. Cα chemical shift is an NMR parameter that is sensitive to protein secondary structure and protein interactions. Prior analysis of Cα chemical shift found that this parameter demarcated three helical regions in a construct consisting of the monomeric β3 TM and cytosolic domains. Similarly, this parameter demarcates three helical regions of β3 in the disulfide-linked αIIb/β3 heterodimer (figure 17). However, there were significant differences in Cα chemical shift between the β3 monomer and the αIIb/β3 heterodimer encompassing twelve residues that define the N-terminal β3 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 αIIb/β3 interactions. Lastly, the Cα chemical shifts of αIIb 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.

**Aliphatic chemical shifts define a β3 interface that interacts with αIIb**

The β3 interface that interacts with αIIb was identified by comparing the 13C HSQC spectra of the disulfide-linked αIIb/β3 heterodimer and the reduced monomers. Reducing the disulfide bond altered side chain chemical shifts in β3 residues Lys716 and Ile719 which indicates that these residues interact with αIIb (figure 18). This interface resides on a face of the N-terminal β3 helix that includes Asp723 (figure 21), consistent with predictions that Asp723 is at the
\( \alpha_{IIb/\beta3} \) interface. Additionally, it is consistent with the \( \beta3 \) interfaces reported in previous NMR structures of the \( \alpha_{IIb/\beta3} \) cytosolic interaction including the structure of the \( \alpha_{IIb/\beta3} \) TM heterodimer that found 7 NOEs between Ile719 and different \( \alpha_{IIb} \) residues.

![Graph showing \( C_\alpha \) resonance minus \( C_\alpha \) random coil (ppm) vs \( \beta3 \) residue number]

**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 \( \beta3 \) in the disulfide-linked heterodimer, similar to a previous structural analysis of the \( \beta3 \) monomer; helices end when the \( C_\alpha \) resonance approach its random coil value (\( C_\alpha \) resonance – \( C_\alpha \) random coil = 0). The \( \alpha_{IIb} \) subunit affects \( C_\alpha \) resonances over the first twelve residues of \( \beta3 \) 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 $\alpha$IIb/$\beta$3 heterodimer, and the red and green overlay corresponds to the $\alpha$IIb 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 $\alpha$IIb. (Purple and green signal is folded in the carbon dimension by 20 ppm).

**Analysis of NOEs at the $\alpha$IIb/$\beta$3 heterodimer interface**

The analysis of aliphatic chemical shifts demonstrates that $\beta$3 residues Lys716 and Ile719 interact with $\alpha$IIb, however these residues displayed no interchain
NOEs with αIIb. Furthermore, the only NOEs between the αIIb and β3 subunits occur near the disulfide crosslink. Upon first inspection, this result appears to conflict with published NMR structures that found NOEs between the αIIb and β3 subunits. These analyses relied on innovative strategies to detect interchain NOEs, including transferred NOE experiments and selective labeling, because the heterodimer could not be isolated from conformations that otherwise masked αIIb/β3 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 αIIb/β3 TM heterodimer. The lack of interchain NOEs at the β3 interface suggests that the αIIb/β3 cytosolic interaction is intrinsically disordered, at least in the disulfide-linked construct described here.

Also, the guanidine group of αIIb Arg995 is postulated to form a salt bridge with β3 Asp723, 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 β3 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 αIIb and β3 subunits.
Fig 19. The H\textsuperscript{15}N 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\textsuperscript{15}N NOE correlates with each residue’s order parameter and it was analyzed to qualitatively assess the dynamics of the disulfide-linked construct.\textsuperscript{36; 37} The αIIb
residues Val990-Asn996 and β3 residues Leu713-Glu726 have similar H-15N NOE intensities which supports our finding that these regions interact (figure 19). However, the αIIb intensities trended toward higher dynamics, consistent with a natively unstructured region of αIIb interacting with a structured, helical portion of β3. Residues downstream of this interacting region were increasingly dynamic in both the αIIb and β3 subunits.

**Stability of β3 secondary structure determined by hydrogen-deuterium exchange**

The increasing dynamics in β3 corresponds to increasingly divergent structure in published models of β3, and since the β3 subunit in disulfide-linked αIIb/β3 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 β3 helix was protected from exchange which demonstrates it has stable helical structure (figure 20). In contrast, β3’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 β3 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 αIIb/β3 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 HN-Hα 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 αIIb.

In contrast with αIIb, the β3 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-NMR 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 β3 structural ensemble has near
ideal Ramachandran statistics when the loop regions are not considered.

The β3 cytosolic domain is primed for interactions with talin, kindlin-3, and
Src kinase
The structure and dynamics of β3 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>
<thead>
<tr>
<th>NOE restraints</th>
<th>NOE classification</th>
<th>number of NOEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>residue 1, 2</td>
<td>i to i</td>
<td>self</td>
</tr>
<tr>
<td></td>
<td>i to i+1</td>
<td>neighbor</td>
</tr>
<tr>
<td></td>
<td>i to i+2</td>
<td>turn/loop</td>
</tr>
<tr>
<td></td>
<td>i to i+3/4</td>
<td>helix</td>
</tr>
<tr>
<td></td>
<td>i to i&gt;4</td>
<td>turn/loop</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NOEs per residue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>violations &gt; 0.5 Å</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dihedral angle restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>TALOS $\phi/\psi$</td>
</tr>
<tr>
<td>HN-H$\alpha$ $\phi$</td>
</tr>
<tr>
<td>violations &gt; 10°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chemical shift restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C\alpha/C\beta$</td>
</tr>
<tr>
<td>$H\alpha$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hydrogen bond O…H-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance restraints</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residues in most favored</th>
</tr>
</thead>
<tbody>
<tr>
<td>regions of Ramachandran space</td>
</tr>
</tbody>
</table>

*107 NOEs were omitted from the structure calculation and violated by the final structural ensemble.
Fig 21. The β3 NMR structural ensemble was divided into two regions consisting of residues Leu713-Trp739 and Asp740-Thr762 and the Cα atoms of each region were aligned. Domains that interact with αIIb, talin, kindlin-3, and Src kinase are circled and side chains that make up the interfaces are depicted as spheres. Also, the two β3 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 αIIb, and these positions reside on the same face of a β3 helix as Asp723, which is postulated to form a salt bridge with αIIb.

including Trp739 and a β3 NPXY motif consisting of residues Asn744, Pro745, Leu746, and Tyr747. Every Trp739 indole resonance was unambiguously assigned and probed in $^{13}$C and $^{15}$N NOESY experiments edited for indole $^{13}$C/$^{15}$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, 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 $\alpha$-helix. The N-terminal helix cap is similar to conformations observed in previous structures of the $\beta$3/talin interaction, 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, and this portion of $\beta$3 is helical in the structure presented here, suggesting 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, 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 β3 Arg760, Gly761, and Thr762 which are called an RGT motif, 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 β3 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). Talin contains a novel PTB domain that binds the unphosphorylated β3 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 β3 bound to talin is consistent with both a type I β-turn and a N-terminal helix cap.

In the NMR structure of the disulfide-linked αIIb/β3 heterodimer described here, NOEs define a conformation that is most consistent with a N-terminal helix cap that nucleates the second cytosolic helix in β3, suggesting the conformation of the β3 NPXY motif is pre-organized to bind talin thereby minimizing its entropic cost of binding (figure 23).
Calculated membrane insertion predicts the regions of β3 that interact with talin and kindlin-3 partition into the membrane.

The membrane embedding of the β3 subunit was calculated using a membrane insertion potential. 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. Intriguingly, the membrane embedding of these helices would sequester them from interactions with talin and kindlin-3. This finding suggests that the β3 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 αIIb/β3 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 αIIb/β3 cytosolic domain was engineered that has 1:1 stochiometry and enforces an interface observed in models of the αIIb/β3 TM heterodimer. The αIIb and β3 cytosolic domains interact in the construct and their interface is intrinsically disordered, consisting of a natively unstructured
region of αIIb interacting with a β3 helix. The structure on the β3 subunit was calculated, and we further characterized β3 interfaces that interact with αIIb, talin, kindlin-3, Src kinase, and the cell membrane.

Fig 22. Portions of the calculated β3 structure were oriented using a membrane insertion potential. The first helix in the β3 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. Notably, Vinogradova et al. observed transferred NOEs between peptides corresponding to the integrin’s cytosolic domains and calculated a structure based on these interactions. 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 αIIb and β3 cytosolic domains in a parallel orientation with 1:1 stoichiometry, however no heteromeric interaction was observed, suggesting that the αIIb and β3 cytosolic domains do not interact or that the upstream constraints failed to approximate native interactions. To account for native TM domain interactions, Li et al. analyzed peptides containing both the cytosolic and TM domains of αIIb and β3, however the peptides formed homo-oligomers instead of heterodimers. 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. However this analysis only considered a fraction of the αIIb and β3 cytosolic domains and did not report on their dynamics.

We combined prior strategies to stabilize the αIIb/β3 heterodimer by including a portion of the αIIb and β3 TM regions and introducing a disulfide-crosslink at the TM heterodimer interface. This construct is the first stable αIIb/β3 cytosolic heterodimer that has an observable αIIb/β3 interaction, and it allowed us to define the structure and dynamics of the αIIb and β3 cytosolic domains in a single conformation that approximates the integrin’s resting state.
The αIIb/β3 cytosolic interface is intrinsically disordered

CD and NMR analyses revealed that the engineered disulfide-crosslink in the αIIb and β3 TM region enforced an interaction between the αIIb and β3 cytosolic domains. However while the chemical shift differences of the β3 Lys716 and Ile719 side chains demonstrated that these residues interacted with αIIb, they displayed no NOEs with αIIb suggesting the interaction is dynamic on the millisecond timescale. Also, we found that the αIIb subunit had no stable helical structure, consistent with a previous NMR structure of this region in phospholipids bicelles. Never-the-less, αIIb interacted with β3 over the length of approximately twelve residues including portions of the β3 helix that are significantly downstream from the TM crosslink. The interaction suggests that the αIIb subunit does not have random structure, instead it has conformational bias toward structures that interact with β3. 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, phosphorylation, and proteolysis, and the cytosolic domains convert this dynamic information into a binary signal: resting or active.
The $\alpha$IIb Arg995-$\beta$3 Asp723 salt bridge is a solvent exposed interaction

Mutations affecting cytosolic residues $\alpha$IIb Arg995 and $\beta$3 Asp723 activate the integrin, but reciprocal mutations Arg995Asp and Asp723Arg restore the integrin’s resting state. 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 $\beta$3 interface that interacts with $\alpha$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. This type of interaction is especially useful for orienting a dynamic interface without providing a driving force that could independently stabilize the $\alpha$IIb/$\beta$3 complex.

The $\beta$3 NPXY motif is a N-terminal helix cap

Previous structures of NPXY motifs have been characterized as type I $\beta$-turns and have slightly different structure than the $\beta$3 NPXY motif which was found to be a N-terminal helix cap that nucleates an $\alpha$-helix. The capping conformation is similar to previous structures of talin bound to $\beta$3, PIPK$,\gamma$, and a $\beta$3/PIPK$,\gamma$ chimera.
that are consistent with both $\beta$-turns and N-terminal helix capping motifs, however the NPXY motifs in these structures do not nucleate helices.\textsuperscript{10; 11; 44} It is possible that talin binding favors unfolded conformations of the $\beta$3 helix, which might function to expose the second NPXY motif to bind kindlin-3.

**The membrane embedding of $\beta$3 prevents interactions with talin and kindlin-3**

The E(z) membrane depth potential predicts that the $\beta$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.\textsuperscript{7} Membrane insertion sequesters these residues from cytosolic proteins so they cannot bind talin as depicted in structures of the $\beta$3/talin complex, and the membrane embedding of the kindin-3 binding site would prevent $\beta$3 interactions with kindlin-3 as well. However the cytosolic helices exist in dynamic equilibrium with unfolded states that likely correspond with solvent exposed conformations. Amphiphilic domains that switch their degree of membrane exposure are commonly found proximal to TM helices in proteins including voltage gated channels,\textsuperscript{45; 46} phospholamban,\textsuperscript{47} and the M2 proton channel from influenza.\textsuperscript{48; 49} Recently, the cytosolic domain of the T-cell receptor’s CD3$\varepsilon$ subunit was found to bind membrane which sequesters aromatic residues from cytosolic proteins including a tyrosine in its NPXY motif.\textsuperscript{50} When the T-cell receptor is activated, the amphiphilic domain is
Fig 23. Analysis of talin binding to the β3 cytosolic domain. (a) The analysis of chemical shift suggests that β3 residues Lys716 and Ile719 interact with αIIb, and these residues lie along the same face of the β3 helix that includes Asp723, which is postulated to form a salt bridge with αIIb residue Arg995. Also, prior structural analysis of β3 and talin depict interfaces that contain β3 residues Phe727, Phe730, Trp739, and Tyr747. (b) Talin interacts with a β3 NPXY motif, and the calculated β3 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 β3 structure reported here because it is amphiphilic and embeds in the membrane. When β3 is modeled as a helix, similar to a previous NMR model that depicts a β3/talin interaction, the αIIb interface resides on the same face of the β3 helix as the talin interface. These findings are consistent with the hypothesis that talin shifts the integrin equilibrium toward active conformations and displaces αIIb from heteromeric interactions with β3.
released from the membrane and binds signaling proteins. The β3 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 β3 cytosolic domain requires a conformational change**

The membrane embedding of the β3 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 β3/talin complex which suggests the kink straightens when β3 binds to talin. The helical conformation positions Phe727 and Phe730 on the same face of the β3 helix as residues that interact with αIIb (figure 23c and d), consistent with the hypothesis that talin displaces αIIb from the β3 cytosolic domain.

**The β3 RGT motif is accessible to Src kinase**

Integrin activation and clustering induces the autophosphorylation of Src kinase which initiates intercellular signaling cascades. Src constitutively binds the β3 cytosolic domain and the β3 RGT motif is critical for this interaction. Although most of the β3 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 αIIb/β3 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


CHAPTER 5

Discussion, Future Direction, and Concluding Remarks

DISCUSSION

The previous chapters describe the development of software used to model the αIIb/β3 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. 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. 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 αIIb/β3 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 αIIb/β3 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 αIIb/β3 heterodimer remain unresolved: First, the β3 TM domain contains a conserved SXXXA dimerization motif that is not involved in the αIIb/β3 interaction (corresponding to a GXXXG dimerization motif in other β subunits), and its function has not been discovered. Second, mutations to αIIb Thr981 create a constitutively active state, but its role in stabilizing the resting integrin remain unknown.5; 6 Lastly, although the structure of the αIIb 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 αIIb and the β3 membrane proximal, cytosolic domain has not been defined at high-resolution, and the significance of a possible αIIb Argg995-β3 Asp723 interaction remains subject to debate.3; 4
The solution structure of the β3 subunit in the disulfide-linked construct sheds light on the membrane proximal αIIb/β3 interaction. Our analysis demonstrates that the αIIb and β3 cytosolic domains interact and suggests a specific interface for β3, consistent with the β3 interface defined by mutagenesis, cystiene 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 αIIb/β3 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 β3 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 β3 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 αIIb/β3 TM and cytosolic heterodimer. These models corroborate experimentally defined αIIb/β3 interfaces and provide a foundation for new, testable hypotheses that back avant-garde mechanisms of integrin activation.
References


