©2014

Jacqueline Sikora

ALL RIGHTS RESERVED

TOWARDS DETERMINING THE MECHANISM OF RECOGNITION AND BINDING TO COLLAGEN BY ALPHA 1 BETA 1 AND ALPHA 2 BETA 1 INTEGRINS

by

JACQUELINE SIKORA

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Chemistry and Chemical Biology

written under the direction of

Professor Jean Baum

and approved by

New Brunswick, New Jersey

October, 2014

ABSTRACT OF THE THESIS

Towards determining the mechanism of recognition and binding to collagen by alpha 1 beta 1 and alpha 2 beta 1 integrins

By JACQUELINE SIKORA

Thesis Director: Professor Jean Baum

Collagen is the most abundant protein in the human body. There are several collagen-binding integrins, which exist as transmembrane proteins and serve to transmit cellular signals via inside-out and outside-in pathways, via a cascade of conformational changes. Collagen is capable of initiating outside-in signaling by binding to the extracellular I-domain of integrin, in a cation-dependent manner. While much is known about collagen subtype and sequence specificity of varying integrins, the mechanism of binding has yet to be determined. Our lab studies the interaction between collagen and the I-domains of integrins $\alpha_1\beta_1$ (α_1I), and $\alpha_2\beta_1$ $(\alpha_2 I)$, via molecular biology and nuclear magnetic resonance (NMR) techniques. We are also investigating recombinant collagen-like peptides for the purpose of binding assays between collagen and integrin. The goal is to determine a structuralfunctional relationship occurring in the interaction, and to use this in order to determine the mechanism of binding. The interaction of collagen and integrin has roles in cancer cell proliferation and thrombosis, making the determination of this mechanism vital to the generation of future treatments for these diseases.

ii

Abstract ii
Table of contents iii
List of Figures iv
Chapter 1 Introduction 1
Chapter 2 Aims and Significance
2. 1 Characterize the dynamics of integrins $\alpha 1I$ and $\alpha 2I$ in solution by NMR 3
2.2 Characterize the dynamics of the collagen-integrin $\alpha 1I$ or $\alpha 2I$ complex 4
2.3 Characterize the role of the I-domain in regulating integrins affinities
2.4 Develop recombinant collagen peptides for NMR studies5
Chapter 3 Results and Progress
3.1 Overexpression and purification of labeled $\alpha 2I$
3.2 NMR chemical shift assignments of $\alpha 2I$
3.3 ^{15}N relaxation experiments probing the dynamics of $\alpha 2I$
3.4 Structure-function relationship13
3.5 Recombinant collagen peptides14
Chapter 4 Proposed Research 19
4.1 Characterize the dynamics of integrins $\alpha 1I$ and $\alpha 2I$ in solution by NMR 19
4.2 Characterize the dynamics of the collagen-integrin $\alpha 1I$ or $\alpha 2I$ complex 20
4.3 Characterize the role of the I-domain in regulating integrins affinities
4.4 Develop recombinant collagen peptides for NMR studies
Chapter 5 Conclusion 27
References

TABLE OF CONTENTS

LIST OF FIGURES

Figure 1. Structures of collagen-bound and unbound $\alpha 2I$	3
Figure 2. Chemical shift changes in $\alpha 2I$ as a result of metal binding	9
Figure 3. R2 values for α 2I at various field strengths	. 10
Figure 4. Chemical exchange plots for $\alpha 1I$ and $\alpha 2I$ with and without metal	. 12
Figure 5. Amino acid sequence and CD spectra for V-GFPGER	. 15
Figure 6. MALDI spectra of thrombin and trypsin cleaved V-GFPGER	. 16
Figure 7. CD spectra and melting curve of synthetic GFPGER	. 17

Chapter 1 Introduction:

Integrins are a family of transmembrane receptor proteins that act to connect cytoskeletal molecules such as actin and talin with various extracellular matrix proteins including collagen, laminin, and fibronectin among others.¹⁻⁶ This association of intracellular and extracellular molecules occurs through a bidirectional signaling cascade, during which the integrin undergoes various structural rearrangements rendering the molecule active.^{1, 3-4} Some integrins, including those that bind collagen ($\alpha_1\beta_1,\alpha_2\beta_1,\alpha_{10}\beta_1,\alpha_{11}\beta_1$), have an inserted domain (I-domain) in the α -subunit of their extracellular region that mediates ligand binding.^{1-2, 4, 6-15}

Integrins $\alpha_1\beta_1$ and $\alpha_2\beta_1$ are the most extensively studied of the collagenbinding integrins. The binding of integrin $\alpha_1\beta_1$ or $\alpha_2\beta_1$ to collagen via its I-domain (α_1 I and α_2 I respectively), induces several intracellular signals that play roles in diverse cellular processes, including cellular adhesion, spreading, and migration.^{1, 3-4, 14-17} Both molecules are potentially interesting drug targets because they play important roles in crucial cellular processes, but their underexpression does not cause severe impact on the body.^{16, 18-21} While the crystal structures of α_2 I alone²², and bound to a collagen model peptide (CMP) have been determined^{13, 23}, the mechanism by which integrin recognizes and specifically binds collagen is not yet known. Knowledge of this recognition mechanism is critical for the development of reaction inhibitors. Here, we propose studies that will aid in the elucidation of this mechanism.

The isolated I-domains of the collagen-binding integrins retain all of the specificities and affinities of the parent integrin for collagen, and undergo the same structural rearrangements upon ligand binding.²²⁻²³ They can therefore be used as a model of the full integrin for structural and functional studies.^{6, 9, 12, 24} There is increasing evidence that protein dynamics and conformational flexibility play a critical role in function, recognition and interactions.²⁵⁻²⁹ Therefore, we propose to use molecular biology techniques and solution nuclear magnetic resonance (NMR) experimentation to study the structural and dynamic features of integrin α_1 I and α_2 I. Recently, various models of protein-protein recognition have been proposed. These include the induced-fit model, conformational selection model, and models that combine conformational selection recognition followed by induced-fit binding.²⁶ Previous studies have also revealed that changes in the strength of the collagen-integrin interaction can lead to dysfunction and disease²³, making it essential to uncover knowledge of the integrin recognition mechanism and affinities for various collagen sequences. The goal of this proposal is to uncover the mechanism by which integrins $\alpha_1\beta_1$ and $\alpha_2\beta_1$ bind collagen, using NMR as our primary tool for structural and dynamic studies.

Chapter 2 Aims and Significance:

2.1 Characterize the dynamics of integrin α₁I and α₂I in solution by NMR: Is there evidence for a conformational selection mechanism?

Collagen binding to $\alpha 1I$ and $\alpha 2I$ is divalent cation dependent, and occurs at the metal ion-dependent adhesion site (MIDAS) of the I-domain^{7, 22} (Figure 1). We propose that the mechanism of collagen binding involves a conformational selection process, and that the metal may play a role in I-domain activation. A small population of the "excited" conformation may exist in equilibrium with the "resting" closed form. We will probe for these conformationally dynamic residues by: 1) producing labeled, recombinant wild-type $\alpha_1 I$ and $\alpha_2 I$; 2) characterizing the motions of the proteins in solution, using NMR, in the presence and absence of metal; 3) deriving the chemical exchange rates for the dynamic residues.

Figure 1. Structures of $\alpha_2 I$ in the (A) un-bound closed and (B) bound open forms. There are several major conformational rearrangements: 1) the unwinding of helix-C; 2) a gain of a turn in helix-6; 3) the downward shift of helix-7; and 4) a change in metal coordination, to include the direct coordination of the side chain of a Glu residue from collagen.²²⁻²³



2.2 Characterize the structure and dynamics of the collagen-integrin $\alpha_1 I$ or $\alpha_2 I$ complex, in order to determine the biologically relevant sites.

We can determine the residues involved in the protein-protein interaction by: 1) exploring the binding of $\alpha_1 I$ and $\alpha_2 I$ to collagen peptides by monitoring NMR peak perturbations; 2) determining the impact of collagen binding on $\alpha_1 I$ and $\alpha_2 I$ structure and dynamics using NMR; 3) testing the biological relevance of the residues identified by NMR as dynamic, or critical to the binding interaction, using ELISA (enzyme-linked immunosorbent assay).

2.3 Characterize the role of the I-domain in regulating integrins affinities for various collagen sequences.

Further insight into the recognition of specific collagen sequences by $\alpha_1 I$ and $\alpha_2 I$ will be provided by: 1) elucidating the structure of recombinantly produced $\alpha_1 I$ and $\alpha_2 I$ variants (containing point mutations of residues of interest as determined by NMR studies of the wild-type protein); 2) studying the dynamics of the $\alpha_1 I$ and $\alpha_2 I$ variants to determine if the mutants have different conformational heterogeneities from wild-type; 3) test these mutants against our collaborator Richard Farndale's Collagen Toolkits in order to determine a) differences in affinity and selectivity and b) the biological relevance of the sites, allowing for the determination of the relationship between structure and function.

2.4 Develop recombinant collagen peptides for NMR studies.

In order to fully investigate the mechanism of the integrin-collagen binding interaction by NMR, we will need to form the integrin-collagen complex. Collagen, however, is an extremely large protein (~300-400 kDa). This large size makes the protein difficult to study by solution NMR, because the effect of line broadening increases with increasing size, making individual residues difficult to find and distinguish from one another. Collagen model peptides (CMPs) of a much smaller molecule weight can be used to model specific regions of collagen. They have been shown to be good models of collagen in structural and biological studies.³⁰⁻³² The cost of ordering the labeled peptides required for NMR studies, however, can be extremely high, especially if attempting to complex the peptides with a protein that only remains stable for a short time. Therefore, the ability to develop recombinant collagen peptides would be extremely useful to us.

Chapter 3 Results and Progress:

3.1 Overexpression and purification of labeled $\alpha_2 I$.

The recombinant I-domain of integrin $\alpha 2$ was expressed in the E. coli bacterial strain BL21DE3 in both ¹⁵N-labeled and ¹⁵N-¹³C-labeled M9 minimal media. Two different expression plasmids containing the $\alpha 2I$ construct were expressed separately, in an attempt to improve yield and protein stability. Both constructs included a His-tag at the N-terminus, to aid in eventual protein purification. The first attempt at expression was made using the Single Protein Production (SPP) system³³. The SPP system offered a benefit of reducing cost by cutting the culture volume, and therefore the amount of isotopes needed for the expression of a uniformly labeled protein, by ten-fold. While α 2I was successfully expressed using this construct, the overall yield was very low. Producing an NMR sample from this construct would have required many liters of growth, thereby negating the previously mentioned cost benefit. The construct was modified by truncating the final eight residues from the C-terminus, in hopes that it would improve expression levels. These were "extra" residues beyond what was contained in the protein secondary structure as determined by the crystal structure. Expression levels, however, remained low, and an NMR sample could not be made from these conditions.

Bacterial expression was next attempted from a construct in the pET expression vector. This construct maintained the full length of the original SPP construct, however it possessed a longer His-tag at the N-terminus, followed by a shorter linker between the tag and protein sequence. With this construct, expression yields were much higher for both ¹⁵N-labeled and ¹⁵N-¹³C-labeled cultures, as shown by SDS-gel. Protein expressed from the pET construct was used in all future studies.

Prior to purification, two different methods of extracting the protein from the bacterial cells were employed, in order to determine which would provide the highest protein yield. In the first method, the protein was physically removed from the bacterial cells by homogenization. This method was quick, but involved a harsh, physical breaking of the cell membranes in order to release the contents of the cell. Since this method completely tore apart the bacteria, it also released proteases capable of chopping up and degrading the protein of interest. A second method, involving the chemical breakdown of the cell membrane was employed. In this chemical method, the bacterial cells are overwhelmed with sucrose (in a TES solution supplemented with lysozyme, benzonase nuclease and a protease inhibitor tablet), so much so that the contents of the periplasm (including the protein of interest) are extracted, but the cells aren't degraded any farther, preventing the release of harmful proteases.

The products obtained from either physical or chemical method were loaded onto Ni-NTA resin, in order to purify α 2I. While purification using the chemical breakdown method was more time consuming, the final purified product was found by SDS-gel to be more pure, more concentrated, and possessed greater stability in the long-term. The eluate from the chemical method of purification was analyzed by SDS-gel, which showed a single band indicating a pure product. Protein identity was confirmed by MALDI-TOF mass spectrometry. Multiple buffer conditions were employed in order to determine how best to keep the protein stable in suitable conditions for NMR studies. Two modes of buffer exchange were also tested: exchange by dialysis and exchange by desalting columns. The desalting columns were found to be both quicker and more effective at completely exchanging the buffer. In the end, the conditions chosen for producing labeled sample were as follows: the cells from two liters of bacterial culture grown in labeled M9 minimal media were broken by chemical means (TES); protein was purified on a 1mL His-Trap column (Ni-NTA resin), and exchanged in the NMR buffer (50mM NaPi, 150mM NaCl, 10mM BME, pH 7.0, with 10% D₂O) via desalting column The buffer contained 5mM MgCl₂ (metal-bound samples) or 10mM EDTA (metal-free). Samples were concentrated to ~500µM for NMR experimentation.

3.2 NMR chemical shift assignments of $\alpha_2 I$.

The NMR fingerprints of metal-free and bound $\alpha 2I$, found from the transverse relaxation optimized heteronuclear single quantum correlation (¹H-¹⁵N-TROSY-HSQC) spectrum, consistently showed high-quality spectra having well-dispersed and well-defined peaks (Figure 2). The triple resonance experiments HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCACONH, and HNCACB, in addition to information obtained from previous assignments for metal-bound $\alpha 2I^{34}$, allowed for the assignment of 95% of the ¹HN, ¹³C α , ¹³C β , and ¹³CO chemical shifts for metal-bound $\alpha 2I$ and 89% for metal-free $\alpha 2I$.



Figure 2. Some regions of $\alpha 2I$ experience chemical shift changes as a result of metal binding. The overlay of 15N-TROSY-HSQC spectra (left) of metal-free (red) and metal-bound (black) $\alpha 2I$ reveals multiple residues are perturbed by the metal coordination. Spectra were taken at 25C, 700 MHz. A plot of the chemical shift perturbations (right) quantifies the total combined shift experienced by the backbone amide of each residue upon addition of Mg²⁺. The total chemical shift was calculated by adding the changes in chemical shifts of ¹H (Δ H) and ¹⁵N (Δ N), according to the equation Δ = ((Δ H)²⁺ (Δ N*0.154)²)^{1/2}.

3.3 ¹⁵N relaxation experiments probing the dynamics of $\alpha_2 I$.

We have hypothesized that α 2I possesses a dynamic, rather than static, structure. Relaxation rates of the ¹⁵N-labeled protein backbone are a function of backbone motions, on a µs to ms timescale²⁵, and can be used to determine the dynamic regions of α 2I. Results from preliminary ¹⁵N-TROSY-R2 experiments, used to measure the transverse relaxation rate, indicated to us that the metal-bound protein was still unstable. Comparing the results of multiple R2 experiments, completed on the same sample but at various field strengths over time, we can clearly see that the baseline of the relaxation rate is rising (Figure 3). The rise is greater than the variance between machines of different field strengths. This rise in the baseline is indicative of protein aggregation. A similar rise in the baseline was experienced for R2 data, recorded for metal-free α 2I (data not shown). The formation of aggregates was confirmed by native gel. More work towards stabilizing the protein is needed. In spite of the rising baseline, we can see clear trends in the spectra taken at all three field strengths. The results show multiple regions having elevated relaxation rates suggesting that α 2I does indeed have dynamic heterogeneity, and possibly experiences conformational exchange in these regions. By mapping the high R2 values on the crystal structure of α 2I, it can be seen that the alpha helices that lie on the exterior experience either slower backbone motions than the core, or are subject to conformational exchange on the micro to millisecond timescale. Further experiments were needed in order to determine the extent of conformational exchange experienced by α 2I in solution.



Figure 3. ¹⁵N-R2 experiments indicate that metal-bound α 2I possesses dynamic heterogeneity, and that the sample may be experiencing aggregation over time (left). Spectra were recorded first on the 700MHz NMR. There is an upward baseline shift for spectra taken in the following days on both the 600 and 800 MHz machines. Data from all machines show similar trends. Regions having elevated R2 values have been highlighted in red on the structural representation of α 2I (right).

In order to determine if the dynamic regions of $\alpha 2I$ were subject to conformational exchange on the microsecond to millisecond timescale, the chemical

exchange rate (R_{ex}) was derived for each residue²⁷. In order to derive the R_{ex}, the

difference between the relaxation rates determined from two experiments was required. The first experiment, the transverse relaxation R2, mostly suppresses the chemical exchange effects to relaxation rates. The second experiment, the Hahn-Echo R2, is designed to measure relaxation rates under conditions of full exchange. R_{ex} values derived from the differences between the results obtained from these two experiments should allow for some indication of the extent of chemical exchange experienced by individual residues of the protein.

The R_{ex} experiment was completed in our lab, by Dr. Ana Monica Nunes, on the I-domain of $\alpha 1\beta 1$ in both the presence and absence of the Mg²⁺ ion (Figure 4 top). Based on these experiments, it can be seen that metal-free $\alpha 1I$ experiences little to no conformational exchange on the millisecond to microsecond timescale. It is interesting to see, however, that when the metal ion is added there are large jumps in the value of R_{ex} in several regions of the protein. These variations indicate that the metal induces conformational heterogeneity in some residues of $\alpha 1I$, activating the domain dynamically, on this timescale.



Figure 4. Chemical exchange rates experiences by $\alpha 1I$ (top) and $\alpha 2I$ (bottom) in the presence (blue) and absence (red) of the Mg²⁺ ion. Spectra were recorded at 800 MHz.

Since $\alpha 1I$ and $\alpha 2I$ have high sequence similarity (66%), nearly identical structures (95%), and perform similar roles (although are localized in different cell types), we hoped to see similar R_{ex} results for $\alpha 2I$. R_{ex} experiments were completed for $\alpha 2I$ in the presence and absence of metal (Figure 4 bottom); however, the data maintains a relatively flat line in both cases. This might indicate that $\alpha 2I$ does not experience conformational exchange upon the addition of the metal, or perhaps the effect of chemical exchange is only hampered by the aggregation issues we are experiencing. R_{ex} experiments will be repeated when the protein is stabilized.

3.4 Structure-Function relationship.

Upon collagen binding to the I-domain of $\alpha 2$, the C-helix unwinds.^{13, 22, 23} This conformational change is suspected to be the case for $\alpha 1I$ as well, based on the crystal structure of the α 1I activated mutant, E317A, which shows the absence of a wound C-helix.³⁵ Furthermore, the amino acid sequence of the C-helix differs between α 1I and α 2I. As a result, we hypothesize that the C-helix may play a role in the pre-selection of collagen sequences or subtypes for binding. While the C-helices of both $\alpha 1$ and $\alpha 2$ have a tyrosine residue at position 285, this position exists at a different place on the helix in the different integrins. We believe that this tyrosine may be the element that acts as the pre-selector of the C-helix, and as a result this residue was the first to be modified. An investigation of the role of the C-helix of $\alpha 11$ has begun in our lab, in order to determine a structure-function relationship. The pET α 11 plasmid was modified to include single-point mutations at the Y285 position. Using PCR mutagenesis, our lab has developed two modified plasmids. These constructs include pET α 1 Y285G and pET α 1 Y285F. Mutation to glycine is known to weaken helices. Phenylalanine is a large steric site, similar to tyrosine but without the hydroxyl moiety. This mutation in $\alpha 2$ results in total loss of function including the inhibition of collagen binding.³⁶ These new mutant constructs, and others, will have to be recombinantly produced in our lab in order to determine the structure and allow for functional studies.

3.5 Recombinant collagen peptides.

The bacterial collagen construct was also designed in the SPP system, in the pCOLD vector. An N-terminal His tag was followed by the V-domain, a globular protein known to aide in trimerization³⁷⁻³⁸. The trimerization domain was proceeded by a thrombin cleavage site, and finally the collagen-like sequence (GPP)₄-GFPGER-(GPP)₅GY, which includes the high affinity binding sequence (GFPGER) for α 2I. The high proline content was necessary in order to make up for the stability that bacterial collagen loses as a result of it's inability to produce hydroxyproline. Expression was attempted, initially failed due to the high content of proline in the construct, and the known toxicity of proline to E. coli (personal communication with Dr. M. Inouye). The construct was inserted into a different vector, the pCOLD secretion vector, to eliminate this toxicity. The secretion vector provided the plasmid with an additional component – the ompA signaling peptide, prior to the Vdomain. Upon protein expression in the cell, the ompA signaling peptide promotes the secretion of the protein out of the cytoplasm and into the periplasm where proline toxicity is no longer a factor. Upon secretion, the ompA signaling peptide is cleaved, leaving us with the bacterial collagen-like protein composed of the his-tag, V-domain, thrombin cleavage site, and collagen-like sequence as described above. This secretion vector construct is known as V-GFPGER, and was used in all subsequent experimentation in our lab. Expression conditions were optimized in LB media. Bacterial cells were broken by homogenization and the protein was purified on Ni-NTA resin. The bacterial collagen sequence was cleaved from the Vdomain. Protein identity was confirmed by MALDI-TOF mass spectrometry. Circular

Dichromism (CD) data of samples before and after thrombin cleavage revealed that the protein was alpha helical in structure, and did not show the presence of a triple helix (Figure 5). It is probable that the spectrum is overwhelmed by the presence of the helical V-domain in solution. Further purification following thrombin cleavage is necessary, in order to completely remove the V-domain from the bacterial collagen peptide. Expression of the V-domain peptide construct was also found to be successful in ¹⁵N-labeled M9 minimal media, as determined by SDS gel.



Figure 5. Amino acid sequence of the V-domain collagen-like peptide construct (left), after ompA-promoted secretion, indicates the location of thrombin cleavage (*). The V domain is shown in bold and the His Tag is italicized. Circular dichromism spectra of V-GFPGER (red, right) is indicative of a peptide alpha helical in structure. The circular dichromism spectra of cleaved GFPGER (blue, right) maintains the same trend as the helical V-domain still exists in solution.

Subsequent purification steps on a Ni-NTA column allowed for the separation

of the cleaved peptide from the V-domain, which maintained the His tag after

thrombin cleavage. However, upon MALDI analysis, it was seen that the full length

V-GFPGER is a major component of the Ni²⁺-bound product, even after thrombin cleavage (Figure 6, left). Therefore, the thrombin cleavage was ineffective and the small amount of peptide that had been cleaved was inadequate for CD analysis, preventing the determination of its helicity. After personal communication with Dr. B. Brodsky, we were encouraged to try cleavage with trypsin. Trypsin is a less specific enzyme, but should be capable of cleaving at the thrombin cleavage site in our V-GFPGER construct. Cleavage with trypsin was effective in yielding the GFPGER peptide, and abolishing the presence of the full-length V-GFPGER. However, due to its unspecific enzymatic nature, trypsin also cleaved at many other sites resulting in a very unclean sample seen by MALDI (Figure 6, right). Unfortunately, the trypsin chopped up the his-tag bound V-domain, and with it any hope of further purification by Ni-NTA resin. Further examination of other cleavage or purification methods must be explored.



Figure 6. MALDI spectra of V-GFPGER after thrombin cleavage (left), and trypsin cleavage (right). Thrombin cleavage results in a clean spectrum in the peptide range (left inset), but inefficient cleavage leaves much uncleaved starting product at approximately 13,000 m/z. Trypsin eliminates all starting product but results in many small impurities in the peptide range (right inset).

Another potential route for peptide production is to express the peptide as a monomer, eliminating the trimerization domain altogether. If the peptide is compotent, meaning capable of forming the triple helix, we should to be able to express and purify the monomer, and then allow triple helix formation in a suitable solution. In order to ensure that the GFPGER was a competent peptide, capable of forming the triple helix, we obtained synthetic GFPGER, having an identical sequence to our post-cleavage peptide from the original construct. After allowing the peptide to incubate at 4°C in 0.1M acetic acid for 24 hours at a concentration of 0.22mg/ml, the product was analyzed by CD. The resulting spectra indicate a compound, triple helical in nature, with a melting temperature of around 22°C (Figure 7). These results indicate that GFPGER does not pose a folding problem and a competent peptide.



Figure 7. The CD spectra of the GFPGER synthetic peptide (sequence: GSPGPPGPPGPPGPPGFPGERGPPGPPGPPGPPGPPGP), with a maximum around 224 and a minimum around 198, is indicative of the triple helix structure (left). Melting of the synthetic GFPGER peptide indicates that the melting point of triple helix to monomer occurs around 22°C (right).

In another effort to utilitze the V-GFPGER construct, enzyme-linked immunosorbent assays (ELISA) were used to test the ability of the full-length construct to be used as a α 2I binding ligand. Casein, at a concentration of 0.5% was used as an effective blocking buffer after many failed attempted with BSA. After many trials, coating of collagen type I was achieved on Nunc 96-well plates (Thermo Scientific). All binding assays were done under conditions of both Mg²⁺ and EDTA. EDTA conditions were used as a negative control, as collagen binding to the Idomain is metal-dependent. Unfortunately, in all cases there was high background signal from binding under EDTA conditions. Therefore, it was not possible to test the binding efficacy of α 2I to V-GFPGER versus the V domain alone due to the high background. Further work towards a more effective assay with less background signal is needed.

Chapter 4 Proposed Research:

The way in which ligand and target recognize and bind one another is difficult to determine based on structural information alone. Preliminary results from NMR indicate that $\alpha 1I$ and $\alpha 2I$ are dynamic in solution. This leads us to hypothesize that collagen recognition and binding occurs through a conformational selection process, and there is a small amount of the "excited" state, or open binding conformation, present in equilibrium with the unbound conformation. In order to investigate mechanism of recognition and binding, further structural, dynamic, and functional studies are needed.

4.1 Characterize the dynamics of integrin $\alpha_1 I$ and $\alpha_2 I$ in solution by NMR: Is there evidence for a conformational selection mechanism?

Further work towards producing stable α 2I is needed, in order to determine if the incorporation of magnesium produces a pre-binding "excited" conformation of the I-domain in solution. Changes to the protein sequence have been considered in order to improve protein stability. A single point mutation will be made at the N-terminus of α 2I. Site-directed mutagenesis will be used to convert Cys-150 to Leu. This mutation has been used in previous NMR studies of the I-domain³⁹, and the group was able to obtain stable protein, even at higher concentrations than what has been used in our dynamic studies. Additionally, a cleavage site will be added between the N-terminal His Tag and the α 2I sequence, by insertion mutagenesis. Since the His tag possesses no secondary structure and is extremely flexible in solution, its motions may be disrupting the stability of the protein. Cleavage of the tag following

purification should aid in stabilizing α 2I. It will also reduce the pI, thereby lowering the allowable pH of the NMR buffer solution. This will also benefit us in relaxation studies, which should be completed at lower pHs for best results in minimizing hydrogen exchange.

Since α 2I and α 1I have high sequence and structural similarities (66% and 95% respectively), we expected to see greater similarities in the results of the relaxation and chemical exchange experiments. We believe that the lack of similarity may be due to the aggregation of α 2I. The formation of aggregates may be preventing the molecule from experiencing the same exchange phenomena as a stable monomer. The major difference in the conditions of the experiments was the pH of the buffer solutions. All experimentation for α 2I was done at pH 7.0, while α 1I was buffered at pH 6.7. It is possible that the slight variation in pH would increase or decrease the stability of the proteins. R2 and R_{ex} experiments will be repeated for α 2I at pH 6.7 in order to keep the conditions constant for both proteins, for a better comparison of results and to possibly prevent or delay protein aggregation, which may be causing the baseline shift.

4.2 Characterize the structure and dynamics of the collagen-integrin $\alpha_2 I$ complex, in order to determine the biologically relevant sites.

GFOGER is a high affinity binding site in collagen, capable of activating integrin upon binding $\alpha 2I.^{22-24}$ The binding occurs via the divalent cation in the MIDAS. The crystal structure of $\alpha 2I$ bound to a triple helical collagen peptide containing the GFOGER sequence²³ gives detailed information on the conformational changes undergone by α 2I as it binds collagen, including changes in the metal position and coordination sphere. However, the crystal structure can only reveal static information. The structure revealed the conformation of α 2I in it's activated conformation, however we know nothing more about the dynamic binding mechanism. Using NMR we will monitor the conformational and dynamic changes that α 2I undergoes upon binding to the GFOGER motif, and investigate the role of the metal in collagen binding.

The α 2I-GFOGER complex will be formed by adding equimolar amounts of the triple helical CMP, (GPO)₄GFOGER(GPO)₅, and α 2I in a buffer solution containing MgCl₂. The use of the CMP is necessary, as collagen is too large to be studied by NMR. The formation of the complex will be confirmed by size exclusion chromatography.

The solution structure of the α 2I-GFOGER complex will first be determined, to confirm the structure that was previously solved by X-ray crystallography. The crystal structure was solved using a α 2I construct truncated at the C-terminus.²³ This deletion removed some of helix-7 from the I-domain. We will need to confirm the conformation of the α 2I-complex, without the truncation, in solution.

NMR titration experiments will reveal further information about the location of the binding residues in both α 2I and the CMP, and the extent of non-binding residues also influenced by complex formation. These experiments will be completed by gradually adding small amounts of non-labeled CMP to ¹⁵N-labeled α 2I. ¹⁵N-TROSY-HSQC spectra will be taken at each ratio, and chemical shift perturbations will be measured. R2 experiments will also be completed at each ratio in order to determine if collagen binding increases or decreases the flexibility of α 2I, and in which regions these changes occur. The experiments will be completed in the opposite fashion as well, where non-labeled α 2I is added to ¹⁵N-labeled CMP. This will allow us to observe which are the binding residues in collagen, and how binding influences collagen dynamics.

Various single point mutations will be introduced into the α 2I sequence at the residues indicated by NMR studies to be dynamically involved in the mechanism of collagen binding. Non-labeled mutant proteins will be expressed in LB media and purified over Ni-NTA resin as described previously. In order to confirm their biological significance, ELISA will be performed to evaluate their binding affinities towards commercially available collagen. The observed changes in activity and affinity of these variants, relative to wild-type α 2I, will confirm the biological relevance of these sites in the regulation of the selective binding of α 2I to collagen.

4.3 Characterize the role of the I-domain in regulating integrins affinities for various collagen sequences.

Our collaborator, Richard Farndale, has developed collagen Toolkits, made up of overlapping collagen peptides, spanning the entire collagen sequence of human collagens type II and III.⁴⁰ Using the toolkit assays, protein binding can be mapped against specific collagen sequences. This allows for the efficient testing of a protein's selectivity and binding affinity towards a vast number of collagen peptide sequences. Preliminary results from Farndale's lab have shown that select I-domain mutants have varying affinities for the toolkit peptides, suggesting that mutants can be used to mimic various states of I-domain activation. We propose that the changes in binding affinities for the various mutants may arise from conformational or dynamic changes that effect the equilibrium populations of closed and activated Idomain in solution. Complementary NMR structure and functional studies will grant great new insights into the molecular mechanism of how the I-domain recognizes and specifically binds collagen sequences with varying affinities.

Recombinant variants of α 1I and α 2I, chosen based on dynamic information obtained from NMR and functional information from ELISA, will be produced in labeled M9 minimal media as described previously. Triple resonance experiments will be performed in order to determine the ${}^{13}C\alpha$, ${}^{13}C\beta$, and ${}^{13}CO$ chemical shifts of α 1I and α 2I mutants. These chemical shifts will be compared to the wild-type protein in order to determine changes in secondary structure. When necessary, full structure determination will be completed. Our lab has developed a list of potential α 1I mutants of interest, focusing on one key area: the C-helix. As mentioned above, the C-helix is suspected to unwind upon collagen binding, therefore leading us to believe that is plays some role in collagen pre-selection. The bulky tyrosine at position 285 may play a large role in this pre-selection. So far we have developed the constructs for α 1I Y285G and Y285F. For further analysis of the C-helix, we are proposing several other point mutations to examine the structure-function relationship. The first is Y285A, as alanine is known to strengthen helices. The next would be the substitution of the tyrosine for another large steric site, in the mutation Y285W. Our lab has the primers for these mutagens, but thus far PCR has been unsuccessful to produce the correct plasmid sequence. The final proposed mutant is the development of an $\alpha 1I/\alpha 2I$ chimera, in which the C-helix of $\alpha 1I$ is

replaced with the C-helix of α 2I. If the chimera displays binding to collagen characteristic to that of α 2I, we will learn the extent of the role of the C-helix in preselection.

Relaxation experiments will be performed to determine if the mutants express any changes in conformational heterogeneity from wild-type α 1I and α 2I. These experiments, which will reveal the dynamics of the protein backbone, will provide insight into the small populations of α 1I and α 2I that may exist in equilibrium, but that we would not see through structure determination alone. Perhaps it is these small populations of "excited" α 1I and α 2I that regulate the recognition of, and affinity towards, collagen.

These mutants will be tested against the Toolkits in order to determine their selectivities and affinities towards specific collagen peptide sequences. The observed changes in secondary structure will be correlated to the functional results obtained from testing against the Toolkits, in order to determine which structural features are important for regulating the affinity of α 11 and α 21 towards specific collagen sequences. Preliminary results from the Farndale lab have indicated that mutants have the tendency to bind a greater number of normally low affinity peptides. We will study, structurally and dynamically, by NMR the mutants complexed with peptides of various binding affinities in order to uncover if the mechanism of binding, or residues involved in complex formation, change for various peptides. The results will grant new insight into the recognition and binding mechanisms. When similar studies are completed for the I-domain of α 1, we will be

able to uncover how the two I-domains have such different binding specificities and affinities, despite their almost identical structure.

4.4 Develop recombinant collagen peptides for NMR studies.

Our lab was one of the first to study collagen by biophysical methods, through the use of CMPs.^{30, 32} Collagen is too large a protein to be studied by solution NMR methods, but through the use of CMPs, we can explore the molecular structure and dynamics of collagen and collagen-ligand complexes. The cost of producing multiple synthetic, labeled CMPs however can quickly become very expensive. For this reason, it would be to our benefit to develop methods for recombinantly producing collagen-like peptides in a bacterial system.

Collagen-like sequences have been identified in prokaryotic genomes, and have been recombinantly expressed, are triple helical in structure, and maintain a similar thermal stability to animal collagens.³⁷⁻³⁸ These bacterial collagen-like proteins contain a globular domain, located either at the N- or C-terminus, that is essential for protein trimerization. Our goal is to produce short collagen model peptides, recombinantly, with the aid of these globular trimerization domains.

The construct used for our preliminary experiments employs the N-terminal V-domain, fused with the collagen-like sequence (GPP)₄GFPGER(GPP)₅. Further trials with trypsin cleavage and a new mode of purification must be investigated and employed in order to obtain pure peptide. CD studies will again be employed, in order to determine the structure and thermal stability of the collagen-like domain. By removing the V-domain, we hope to see CD spectra indicative of a triple helix.

If the V-domain constructs will not produce triple helical peptides, we propose to try another known trimerization domain to aid in their expression. The N-terminus of the bacteriophage T4 fibritin contains a small trimeric globular domain, known as the foldon-domain.⁴¹⁻⁴² Fibritin lacking the foldon domain does not fold correctly into its triple helical coiled coil conformation. Therefore, it is thought that function of the foldon-domain is to aid in correct strand alignment and subsequent triple helix formation. It has been shown that the foldon-domain can also be fused, C-terminally, to a collagen-like peptide in order to increase the peptides thermal stability.⁴² We propose to develop a construct in which we fuse the foldon-domain to the C-terminus of our collagen-like peptide (GPP)₄GFPGER(GPP)₅. The peptide-foldon construct will be developed in the pET vector system that we are familiar with from I-domain expression, and expressed in LB media. It will also contain a His-Tag for purification over Ni-NTA resin, and thrombin cleavage sites for the removal of the tag and foldon-domain. Conformation will again be confirmed by CD spectroscopy.

We believe that we will be able to obtain stable, triple helical, collagen-like peptides from one or both of the above mentioned constructs. Collagen-like peptides from either construct can easily be labeled for NMR studies via expression in ¹⁵N- or ¹³C-¹⁵N-labeled M9 minimal media. When we have optimized the stability of the peptides, they can be used in place of synthetic CMPs for many of the structural, dynamic, and functional studies proposed here.

Chapter 5 Conclusion

The interaction of integrins $\alpha 1\beta 1$ and $\alpha 2\beta 2$ with collagen is of critical importance, due to its implications in human disease. In spite of this, little is known about the mechanism by which integrins selectively bind collagen. Our approach towards the issue, which involves the integration of NMR with molecular biology techniques, has the potential to reveal great new insight into the interaction, providing new targets for drug design and the treatment of disease.

References

- 1. Luo, B., Carman, C.V. and Springer, T.A. Structural Basis of Integrin Regulation and Signaling. Annu. Rev. Immunol. 2007. 25: 619-647.
- 2. Heino, J. The collagen receptor integrins have distinct ligand recognition and signaling functions. Matrix Biology. 2000. 19: 319-323.
- 3. Leitinger, B. and Hohenester, E. Mammalian collagen receptors. Matrix Biology. 2007. 26: 146-155.
- 4. Hynes, R. Integrins: Bidirectional, Allosteric Signaling Machines. Cell. 2002. 110: 673-687.
- 5. Anthis, N.J., Wegener, K.L., Ye, F., Kim, C., Goult, B.T., Lowe, E.D., Vakonakis, I., Bate, N., Critchley, D.R., Ginsberg, M.H. and Campbell, I.D. The structure of an integrin/talin complex reveals the basis of inside-out signal transduction. The EMBO Journal. 2009. 28: 3623-3632.
- 6. Tuckwell, D., Calderwood, D.A., Green, L.J. and Humphries, M.J. Integrin a2 Idomain is a binding site for collagens. Journal of Cell Science. 1995. 108: 1629-1637.
- 7. Calderwood, D.A., Tuckwell, D.S., Ebles, J., Kuhn, K. and Humphries, M.J. The Integrin a1 A-domain is a Ligand Binding Site for Collagen and Laminin. The Journal of Biological Chemisty. 1997. 272 (19): 12311-12317.
- 8. Tulla, M., Pentikainen O.T., Viitasalo, T., Kapyla, J., Impola, U., Nykvist, P., Nissinen, L., Johnson, M.S. and Heino, J. Selective Binding of Collagen Subtypes by Integrin a1I, a2I, and a10I Domains. The Journal of Biological Chemistry. 2001. 276 (51): 48206-48212.
- 9. Bahou, W.F., Potter, C.L. and Mirza, H. The VLA-2 (a2b1) I domain functions as a ligand-specific recognition sequence for endothelial cell attachment and spreading: molecular and functional characterization. Blood. 1994. 84: 3734-3741.
- 10. Wegener, H., Leineweber, S. and Seeger, K. The vWFA2 domain of type VII collagen is responsible for collagen binding. Biochemical and Biophysical Research Communications. 2012. 430 (2): 449-453.
- 11. Kapyla, J., Ivaska, J., Riikonen, R., Nykvist, P., Pentikainen, O. Johnson, M. and Heino, J. Integrin a2I Domain Recognizes Type I and Type IV Collagens by Different Mechanisms. The Journal of Biological Chemistry. 2000. 275 (5): 3348-3354.
- 12. Kamata, T. and Takada, Y. Direct Binding of Collagen to the I Domain of Integrin a2b1 (VLA-2, CD49b/CD29) in a Divalent Cation-independent Manner. The Journal of Biological Chemistry. 1994. 269 (42): 26006-26010.
- Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J. and Liddington, R.C. Structural Basis of Collagen Recognition by Integrin a2b1. Cell. 2000. 101: 47-56.
- Shimaoka, M., Takagi, J. and Springer, T.A. Conformational Regulation of Integrin Structure and Function. Annu. Rev. Biomol. Struct. 2002. 31: 485-516.
- 15. Barczyk, M., Carracedo, S. and Gullberg, D. Integrins. Cell Tissue Res. 2010. 339: 269-280.

- 16. Deb, M., Sengupta, D. and Patra, S.K. Integrin-epigenetics: a system with imperative impact on cancer. Cancer Metastasis Rev. 2012. 31: 221-234.
- 17. Heino, J. The collagen family members as cell adhesion proteins. BioEssays. 2007. 29: 1991-1010.
- Nieuwenhuis, K.H., Sakariassen, K.S., Houdijk, W.P.M., Nievelstein, P.F.E.M. and Sixma, JJ. Deficiency of Platelet Membrane Glycoprotein Ia Associated With a Decreased Platelet Adhesion to Subendothelium: A Defect in Platelet Spreading. Blood. 1986. 68: 692-695.
- 19. Aoudjit, F. and Vuori, K. Integrin Signaling in Cancer Cell Survival and Chemoresistance. Chemotherapy Research and Practice. 2012. 2012: 283181.
- Desgrosellier, J.S. and Cheresh, D.A. Integrins in cancer: biological implications and therapeutic opportunities. Nature Reviews: Cancer. 2010. 10: 9-23.
- 21. Leventhal, K.R., Yu, H., Kass, L., Lakins, J.N., Egeblad, M., Erler, J.T., Kong, S.F.T., Csiszar, K., Giaccia, A., Weninger, W., Yamauchi, M., Gasser, D.L. and Weaver, V.M. Matrix Crosslinking Forces Tumor Progression by Enhancing Integrin Signaling. Cell. 2009. 139 (5): 891-906.
- Emsley, J. King, S.L., Bergelson, J.M. and Liddington, R.C. Crystal Structure of the I Domain from Integrin a2b1. The Journal of Biological Chemistry. 1997. 272 (45): 28512-28517.
- 23. Emsley, J., Knight, C.G., Farndale, R.W. and Barnes, M.J. Structure of the Integrin a2b1-binding Collagen Peptide. J. Mol. Biol. 2004. 335: 1019-1028.
- 24. Knight, C.G., Morton, L.F., Onley, D.J., Peachey, A.R., Messent, A.J., Smethurst, P.A., Tuckwel, D.S., Farndale, R.W. and Barnes, M.J. Identification in Collagen Type I of an Integrin a2b1-binding Site Containing an Essential GER Sequence. The Journal of Biological Chemistry. 1998. 273 (50): 33287-33294.
- 25. Henzler-Wildman, K. and Kern, D. Dynamic personalities of proteins. Nature. 2007. 450:964-972.
- 26. Csermely, P., Palotai, R. and Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010. 25 (10): 539-546.
- 27. Boehr, D.D., Nussinov, R. and Wright, P.E. The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol. 2009. 5 (11): 789-796.
- 28. Tzeng, S. and Kalodimos, C.G. Protein dynamics and allostery: an NMR view. Current Opinion in Structural Biology. 2011. 21: 62-67.
- 29. Kern, D. and Zuiderweg, E.R.P. The role of dynamics in allosteric regulation. Current Opinion in Structural Biology. 2003. 13: 748-757.
- 30. Baum, J. and Brodsky, B. Folding of peptide models of collagen and misfolding in disease. Current Opinion in Structural Biology. 1999. 9:122-128.
- Brodsky, B., Thiagarajan, G., Madhan, B. and Kar, K. Triple-Helical Peptides: An Approach to Collagen Conformation, Stability, and Self-Association. Biopolymers. 2008. 89 (5): 345-354.
- 32. Baum, J. and Brodsky, B. Real-time NMR investigations of triple-helix folding and collagen folding diseases. Folding & Design. 1997. 2 (4): 53-60.

- 33. Suzuki, M., Mao, L. and Inouye, M. Single protein production (SPP) system in Escherichia coli. Nature Protocols. 2007. 2 (7): 1802-1810.
- 34. Elshort, B., Jacobs, D.M., Schwalbe, H. and Langer, T. 1H, 13C and 15N backbone resonance assignment of the integrin alpha 2 I-domain. Journal of Biomolecular NMR. 2003. 27 (2): 191-192.
- 35. Lahti, M., Bligt, E., Niskanen, H., Parkash, V., Brandt, A.M., Jokinen, J., Patrikainen, P., Kapyla, J., Heino, J., and Salminen, T.A. Structure of collagen receptor integrin $\alpha(1)$ I domain carrying the activating mutation E317A. Journal of Biological Chemistry. 2011. 286 (50): 43343-51.
- 36. Nissinen, L., Koivunen, J., Kapyla, J., Salmela, M., Nieminen, J., Jokinen, J., Sipila, K., Pihlavisto, M., Pentikainen, O.T., Marjamaki, A. and Heino, J. Novel α2β1 integrin inhibitors reveal that integrin binding to collagen under shear stress conditions does not require receptor preactivation. Journal of Biological Chemistry. 2012. 287 (53): 44694-702.
- 37. Yu, Z., Mirochnitchenko, O. Xu, C., Yoshizumi, A., Brodsky, B. and Inouye, M. Noncollagenous region of the streptococcal collagen-like protein is a trimerization domain that supports refolding of adjacent homologous and heterologous collagenous domains. Protein Science. 2010. 19: 775-785.
- 38. Yu, Z., Brodsky, B. and Inouye, M. Dissecting a Bacterial Collagen Domain from Steptococcus pyogenes: sequence and length-dependent variations in triple helix stability and folding. Journal of Biological Chemistry. 2011. 286 (21): 18960-18969.
- Lambert, L.J., Bobkov, A.A., Smith, J.W. and Marassi, F.M. Competitive Interactions of Collagen and a Jararhagin-derived Disintegrin Peptide with the Integrin a2-I Domain. Journal of Biological Chemistry. 2008. 283 (24): 16665-16672.
- 40. Farndale, R.W., Lisman, T., Bihan, D., Hamaia, S., Smerling, C.S., Pugh, N., Konitsiotis, A., Leitinger, B., Groot, P.G., Jarvis, G.E. and Raynal, N. Cellcollagen interactions: the use of peptide Toolkits to investigate collagenreceptor interactions. Biochemical Society Transactions. 2008. 36: 241-250.
- 41. Frank, S., Kammerer, R.A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bachinger, H.P. and Engel, J. Stabalization of Short Collagen-like Triple Helices by Protein Engineering. Journal of Molecular Biology. 2001. 308 (5): 1081-1089.
- 42. Niland, S., Westerhausen, C., Schneider, S.W., Eckes, B., Schneider, M.F. and Eble, J.A. Biofunctionalization of a generic collagenous triple helix with the a2b1 integrin binding site allows molecular force measurements. The International Journal of Biochemistry & Cell Biology. 2011. 43: 721-731.