STRUCTURAL AND DYNAMIC INVESTIGATIONS OF TYPE I COLLAGEN AND INTEGRIN I DOMAINS AND IMPLICATIONS FOR THEIR INTERACTIONS

by

JIE ZHU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemistry and Chemical Biology

Written under the direction of

Prof. Jean Baum & Prof. David A. Case

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2018

ABSTRACT OF THE DISSERTATION

Structural and Dynamic Investigations of Type I Collagen and Integrin I Domains and Implications for their Interactions

by JIE ZHU

Dissertation Director:

Professor Jean Baum

Integrin–collagen interactions play a critical role in numerous cellular functions. In this dissertation, the structures, dynamics, and interactions of the type I collagen and integrin I domains are investigated. The objective is to gain insight into the mechanism of the collagen–integrin interactions.

Collagen fibril interactions with cells and macromolecules in the extracellular matrix drive numerous cellular functions. Binding motifs for dozens of collagen-binding proteins have been determined on fully exposed collagen triple helical monomers. However, when the monomers are assembled into the functional collagen fibril, binding motifs become inaccessible, and yet critical cellular processes continue to occur. Here we use an integrative approach by combining molecular dynamics (MD) simulations with atomic force microscopy (AFM) experiments and show that fluctuations of

the collagen monomers within the complex fibril play a critical role in collagen interactions.

To better understand the mechanisms underlying collagen-induced conformational switches of integrin I domains, we employ NMR hydrogen-deuterium exchange (HDX) experiments to explore the impact of slower timescale dynamic events. NMR HDX results suggest a relationship between regions exhibiting a reduced local stability in the unbound I domain and those that undergo significant conformational changes upon binding. This study supports a model in which intrinsically destabilized regions predispose conformational rearrangement in the integrin I domain.

The morphology and mechanical properties of type I collagen fibrils vary greatly in different tissues. Integrins have been proposed to regulate the type I collagen fibrillogenesis *in vivo*. Here we report on the type I collagen fibrillogenesis affected by integrin I domains and mutants. The conducted experiments showed that integrins and variants slowed down the kinetics of type I collagen fibril formation and reduced the sizes of the immature fibrils. The gain-of-function mutants inhibited the fusions of fibrils. Enhanced viscosities of collagen gels were observed in the presence of integrin I domains, implying stronger interactions between collagen fibrils. We propose that *in vivo*, integrins of different activation states might regulate collagen fibrillogenesis.

Acknowledgments

I would like to express my sincere appreciation to my advisor Professor Jean Baum for her continuous support of my Ph.D. study and research. Her passion and enthusiasm for research were contagious and motivational to me, which encouraged me greatly during the tough days of my Ph.D. time. She gave me the freedom to start new projects and branch out into challenging research areas and continued to contribute valuable feedback, advice, and encouragement. Her critical thinking, problem-solving skills, and brilliant public speaking set excellent examples for me to learn from in my lifetime.

I would also like to thank my co-advisor Professor David Case. He is a super nice and supportive person. He raised many valuable points to help me proceed with my research. He even helped me set up some computation experiments in person. I also took one course that he taught, and his teaching style and enthusiasm made a strong impression on me.

Besides my advisors, I would like to thank my committee member Professor Sagar Khare who guided me through all of these years. I'm grateful to him for his encouragement and creative and comprehensive advice on my projects.

My sincere thanks also goes to Dr. Ana Monica Nunes. She guided me to gain knowledge about my project and trained me on the techniques in our lab when

I was a beginner. She was always kind and ready to help whenever I got confused and stuck by unexpected experimental results.

I would also like to acknowledge Dr. Cody Hoop, who is a super friendly person. It was very nice working with her when we wrote the paper together. She is always ready to answer my questions about the research and about the American culture. She also proofread this thesis and improved the English writing a lot.

I gratefully acknowledge the group members with whom I have been working: Dr. Jonathan Williams, Dr. Baifan Wang, Xue Yang, Tamr Atieh, Jonathan Roth, Jonathan Stoeber, Dr. Maria Janowska, Dr. Gina Moriarty, Jacqueline Jezioro, Robert Young, Michael Olson, and Jacky Wang. We had stimulating discussions about science, and they always provided tremendously valuable advice. I enjoyed the friendly and supportive atmosphere of this group, and will never forget the memorable moments of group dinners, mini golf, laser tag, the haunted houses, and birthday celebrations we had together.

I would also like to acknowledge NMR senior scientists Dr. Seho Kim, for his thorough NMR training, and Dr. Nagarajan Murali, for the wonderful ideas and suggestions he provided in the group meetings.

I must express my gratitude to Dr. Ming Huang from Prof. Darrin York's group for his expertise in computational chemistry and his kind help with installing and compiling software on my computer.

Nobody has been more important to me in the pursuit of my Ph.D. than my family. I would like to thank my parents and grandparents, who provide me with unfailing support and encouragement.

Finally, a special thanks goes to my husband Dr. Haoyuan Chen. He helped me through the awkward time as a beginner in computational chemistry with his patience and knowledge. When I encountered frustrations in research, he was always the first one who comforted me and helped me figure out what to do. My Ph.D. degree accomplishment would never be possible without him.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

α1 I	Alpha1 I domain
αI domain	α-inserted domain
AFM	Atomic force microscopy
BLI	Biolayer interferometry
BMRB	Biological magnetic resonance bank
CD	Circular dichroism
CMPs	Collagen mimetic peptides
DDR2	Discoidin domain receptor 2
ECM	Extracellular matrix
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
FID	Free induction decay
GAGs	Glycosaminoglycans
H-bonds	Hydrogen bonds
HDX	Hydrogen–deuterium exchange
HSQC	Heteronuclear single-quantum correlation
I-domain	Inserted domain
ITC	Isothermal titration calorimetry
MD	Molecular dynamics
MIDAS	Metal ion-dependent adhesion site

MMPs	Matrix metalloproteinases
NMR	Nuclear magnetic resonance
OD	Optical density
PDB	Protein data bank
PF	Protection factor
PPII	Polyproline type II
RMSD	Root mean square deviations
RMSF	Root mean square fluctuations
RT	Room temperature
SASA	Solvent-accessible surface area
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SPARC	Secreted protein acidic and rich in cycteine
SRU	Smallest repeating unit
TALOS	Torsion Angle Likeliness Obtained from Shift and Sequence Similarity
THP	Triple-helical peptides
UV-Vis	Ultraviolet-visible
WT	Wild type

Chapter 1 Introduction to this Thesis

1.1 Collagen

Collagen is the most abundant protein in vertebrates, making up 25 % to 35 % of protein content.[1] There are 28 subtypes in the collagen family, among which seven (type I, II, III, V, XI, XXIV, and XXVII) are able to form higher order fibrous super-structures, and are called fibrillar collagens. The most significant member of the collagen family is fibrillar type I collagen, making up 90% of the total collagen content, which is primarily found in connective tissues. [2-5]

1.1.1 Structural hierarchy of type I collagen

The fibrillar type I collagen structure has a complex hierarchy. The simplest level of collagen is a single polypeptide chain that consists of a repeating GXY sequence, where X is often proline and Y is often hydroxyproline. Three such polypeptide chains twist and intertwine to form a triple helix flanked by non-helical N- and C- telopeptides (3.5% of the sequence) to make up a full collagen molecule. The triple helical collagen is referred to as the collagen monomer and is the basic unit to build the collagen fibril. The three collagen polypeptide chains staggered by one residue are named the “leading”, “middle” and “trailing” chains according to their relative position from the N-terminus to C-terminus of the protein. There are three possible registries of the chains within heterotrimeric type I collagen (α1α1α2, α1α2α1, and α2α1α1), but the actual registry is not known. Collagen monomers pack in a staggered parallel manner with 67 nm distance from its direct neighbor along the fibril longitudinal

axis. This distance, termed the “D period”, divides a collagen monomer into 5 D segments (D1 to D5). The D1 to D4 are one D period long and D5 is only 0.46 D long. This short D5 gives rise to two successive regions in collagen fibrils: the “overlap region” where all five D segments are present and the “gap region” where there is a gap created by the missing D5. Studies over the past three decades have established the smallest repeating unit (SRU) of the collagen fibril, which contains D1 to D5 from consecutive collagen monomers packed in the quasi-hexagonal Hodge-Petruska scheme. [6] Rather than a linear packing, collagen fibrils have a circular cross-section made up by ~40 Å concentric layers. [7]

1.1.2 The type I collagen biosynthesis process

Two type I collagen polypeptide chains (α1(I) and α2(I)) are synthesized inside fibroblast cells by translating the genes (COL1A1 and COL1A2). Prolines and lysines are hydroxylated by peptidyl prolyl hydroxylase and peptidyl lysyl hydroxylase, respectively. Glycosylation of some lysine residues occurs after the hydroxylation. Two α1(I) and one α2(I) chains supercoil to form a triple helical structure that is flanked by two non-collagenous domains termed the “propeptides”. These procollagens are then secreted and released by fibroblast cells. Procollagen peptidases cleave the N- and C- propeptides and produce the triple helical domain flanked by random coiled N- and C-telopeptides, termed as “tropocollagen”. Multiple tropocollagen molecules

assemble near the cell surface and form rod-like collagen fibrils, which fuse and bundle into higher order collagen fibers. Cross-linking occurs on hydroxylysines and lysines to stabilize and mature the superstructure of the collagen fibril.

1.1.3 Collagen binding sequences for collagen receptors

Besides the characteristic glycine-proline-hydroxyproline (GPO) repeats, the collagen sequence contains other amino acid types to provide various binding sites for cell-membrane proteins and matrix molecules. Microscopy techniques have been used to image collagen-protein complexes and have found approximate binding domains within collagen sequences based on relative spatial binding locations of the binding partners within the collagen molecules. [8-10] Farndale et al. developed toolkits of collagen triple-helical peptides to determine specific binding sequences for several collagen-binding proteins. The toolkits contain a series of triple-helical peptides whose sequences span the entire ~1000 residues of the full collagen sequence. Each peptide is comprised of a 27-residue collagen fragment, overlaying with its direct neighbor by 9 residues. In this way, the collagen sequence is broken down into short fragments whose interactions with collagen binding proteins have been tested to find specific binding motifs. [11] Shorter collagen fragments are synthesized to find minimal binding motifs. By using the toolkits, many protein binding sequences have been precisely identified.

1.1.4 Atomic structures of Collagen

1.1.4.1 Collagenous protein structure

High-resolution structures of collagen-model peptides obtained by X-ray crystallography have been published within the last two decades, [12] allowing a close look at the atomic details of collagenous protein structures. Every third residue in the collagen sequence is occupied by a glycine with its sidechain pointing towards the inside of the triple helix and X and Y residues (frequently proline and hydroxyproline) pointing outside. The small glycine sidechain (R = H) ensures the tight winding of the collagen triple helix. The X-ray crystal structure of the collagen triple helix also reveals that it is held together by hydrogen bonds. The backbone amide proton of a Gly residue (residue index n) on one polypeptide chain (e.g. the “leading” chain) is hydrogen bonded to an X residue (residue index n-2) on its neighboring chain (e.g. the “middle” chain). Additionally, the Gly carbonyl (residue index n) on one polypeptide chain (e.g. the “leading” chain) forms a water bridging hydrogen bond with an X amide proton (residue index n+1) on its neighboring chain (e.g. the “middle” chain), shown in Figure 1.1. [13]

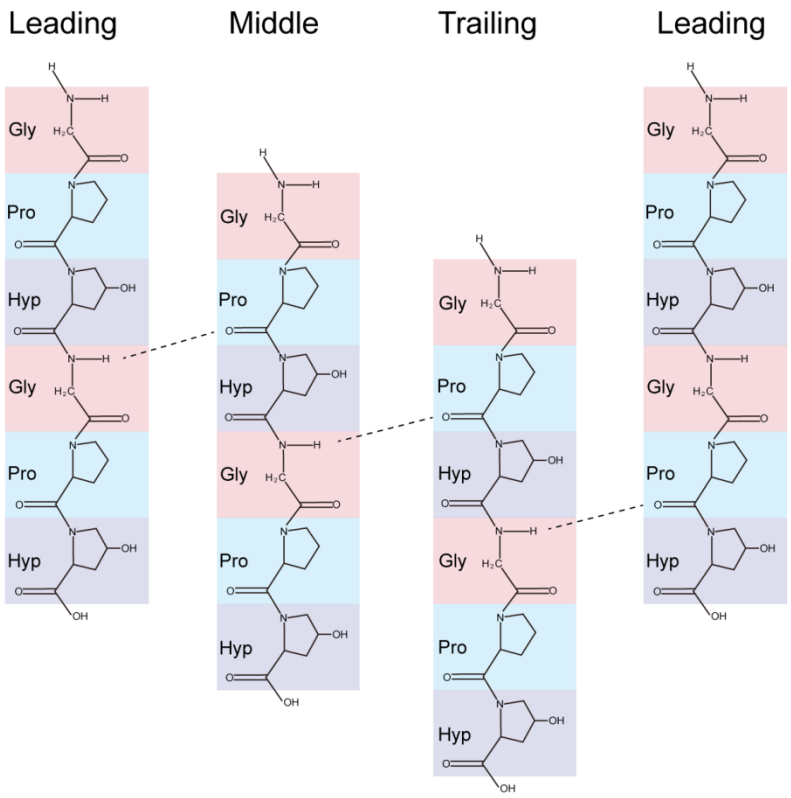

Figure 1.1 Hydrogen bond patterns within a collagen triple helix of sequence GPOGPO. Inter-chain backbone N-H hydrogen bonds are represented by dashed lines. Gly, Pro and Hyp are with pink, cyan, and purple background for clarity.

Each polypeptide chain of the triple helix adopts a 7/1 symmetry and therefore the unit twist (θ) is 51.4° (360°/7) and the unit height (h) is 8.57 Å (60 Å/7). A left-handed 7/2 helical model was found to be the main symmetry for collagenous proteins rich in imino acid residues (GPO repeats), which has an axial repeat of 20 Å. [14] The radial projection of this helical symmetry is shown in Figure 1.2.

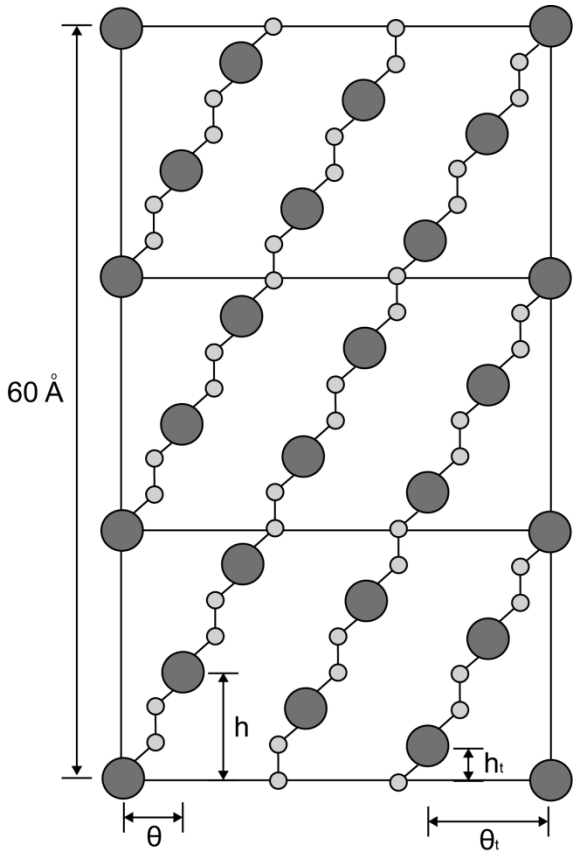

Figure 1.2 Radial projection of 7/2 triple helical model of collagen. Each polypeptide chain adopts a 7/1 symmetry. Dark gray circles represent glycines and the light gray circles represent the X and Y residues in a Gly-X-Y triplet. Unit heights of the polypeptide and the triple helix are shown by h and h_t, respectively. Unit twists of the polypeptide and the triple helix are shown by θ and $θ_t$, respectively.

1.1.4.2 Type I collagen fibril unit cell structure

Orgel et al. published an X-ray diffraction structure of the three-dimensional arrangement of collagen molecules in naturally occurring type I fibrils from rat tail tendon (PDB ID: 3HR2). [15] Although the resolution is not at the atomic level, it provides the electron density envelope of the overall shape and packing of collagen triple helices. Fitting to the collagen sequence and refinement of the

structure gives a model with all of the alpha carbon coordinates. This model provides a 678 Å long, 27 Å wide and 40 Å deep unit cell equivalent to the smallest repeating unit (SRU) containing all five D segments from successive collagen monomers. All-atom models of collagen fibrils can be built by fitting the collagen monomer model into the alpha carbon coordinates derived by X-ray diffraction.

1.1.4.3 Two possible collagen fibril surface models

Hulmes et al. proposed a concentric ring model for packing of collagen fibrils. [7] The radial orientation is approximately 4 nm spacing, consistent with the a axis of the unit cell described in Chapter 1.1.4.2. The packings of the unit cells in a concentric way render two possible surfaces (Surface A and Surface B) shown in Figure 1.3.

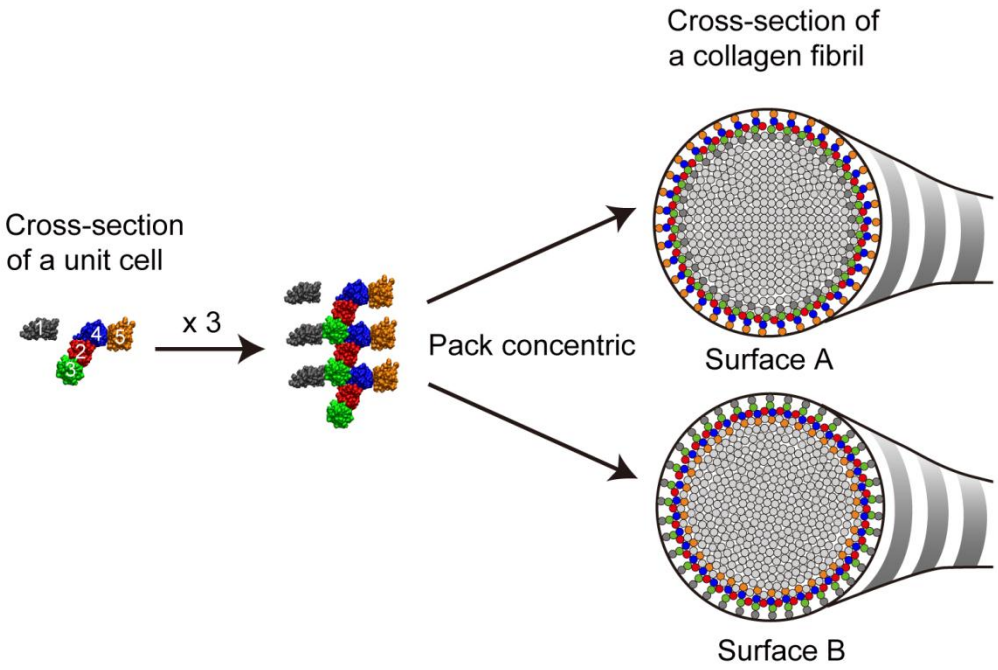

Figure 1.3 Two possible type I collagen fibril surfaces generated by concentric packing of collagen fibril unit cells. The collagen triple helix is divided by a D period (67 nm) into five segments: D1(grey), D2(red), D3(green), D4(blue), and D5(orange).

1.1.5 Molecular dynamics (MD) simulations of collagen

Molecular dynamics simulations study structural and physicochemical properties of proteins. It has been applied on collagens at the atomistic level to predict the dynamics, energetics and interactions within the local environment.

The model of the three-dimensional arrangement of collagen triple helices within the fibrillar environment (PDB ID: 3HR2) allows studies of collagen fibrils by MD simulations. [15] Streeter et al. built an all-atom model of a collagen fibril based on the atomic structure and periodic nature of the fibril. MD simulations were performed on the collagen model with periodic boundaries to

replicate the supermolecular arrangement [16]. The simulations were validated by comparison to experimental observations and theories of collagen's fibril structure. Streeter et al. also studied the inter-protein interactions within a collagen fibril [17]. Simulations of collagen fibrils and monomers were compared to study the interactions upon fibrillogenesis. The process of collagen fibrillogenesis was then compared with the protein folding process, which shares many similarities in concept with collagen fibrillogenesis. Varma et al. performed MD simulations on a collagen model with periodic boundaries to evaluate the effect of several factors on D-band shrinkage [18]. They found that the large shrinkage observed in earlier simulations was not eliminated by optimizations of temperature/pressure coupling algorithms, salt concentration or hydration level, or by cross-linking monomers. They concluded that the shrinkage was a force-field artifact. A new force field, Amber99sb-ildn, was found to produce a small D-band shrinkage of < 3%. Nair et al. performed full-atomistic calculations of a mineralized collagen protein matrix at different mineral densities [19]. They found that the tensile modulus of mineralized collagen increased with the mineral densities far beyond that of pure collagen fibrils. They concluded that the mineral crystals within collagen bear up stress while collagen is responsible for deformation response.

1.1.6 Significance of the collagen surface and binding site accessibilities

Although biophysical research has gained plenty of knowledge on

collagen-protein interactions by exploiting collagen model peptides (CMPs), as described in Chapter 1.1.3, protein interactions with collagen fibrils is of great significance due to the fact that the majority of type I collagen in human bodies is in the fibrillar form. The accessibilities of binding motifs within the collagen fibril to the interacting proteins are crucial for collagen-protein interactions. Due to the huge size and complexity of collagen, collagen fibril–protein interactions are very challenging to study by biophysical methods and thus remain unclear.

There are several papers that review the mapping of collagen receptor binding sites on the three-dimensional collagen fibril model built in silico. Two alternative collagen surfaces have been proposed and analyzed for protein accessibilities as described in Chapter 1.1.4.3. Interestingly, neither of the two surfaces provides access for all of the collagen binding proteins. It is proposed that collagen deformation or proteolysis is needed for some proteins to access their binding motifs. [20-22]

1.2 The extracellular matrix (ECM)

Cells are not the only component comprising tissues. A considerable part of the tissues is the extracellular matrix (ECM), consisting of proteins, sugars and other components. The ECM provides support for cells, guides important cellular processes, such as cell migration in wound healing, and transduces environmental signals to cells, and is therefore of biological significance.

1.2.1 Compositions of the ECM

The extracellular matrix is made up of water, minerals and two classes of macromolecules: fibrous proteins and glycosaminoglycans (GAGs). Fibrous proteins include collagen, fibronectin, and elastin, among which collagen is the most abundant component of the extracellular matrix scaffold in various tissues, including bones, tendons, skin, cartilage and cornea. [3,23] GAGs are long unbranched polysaccharide chains covalently bound to proteins to form proteoglycans. Collagen strengthens and also organizes the extracellular matrix whereas proteoglycans keep the ECM hydrated by binding and organizing water and serves as lubricants due to the high viscosity.

1.2.2 ECM function and protein bindings

ECM serves many functions including providing support, separating tissues and controlling cell activities. Type I collagen, the major component of ECM, plays both roles in ensuring tissue integrity through providing tensile strength and regulating cellular events, such as migration and differentiation, by interacting with its binding proteins. These roles ensure collagen's biological significance in human health. [24-26] Disruption of collagen–protein interactions is associated with diseases including cancer, thrombosis, and arthritis. [27,28]

1.2.3 ECMs in different tissues differ in morphologies

Collagen is the major component of the extracellular matrix, and the morphology of the ECM depends significantly on the nature of the collagen fibril and fiber organization. The diameters, structural organization, and mechanical properties of collagen fibers vary greatly in different tissues, resulting in tissue morphologies ranging from highly ordered, packed rigid bones, which support and protect various of organs, to loosely packed compliant skin, which exhibits key mechanical behaviors. [29-35]

1.2.4 Regulators of collagen fibrils in ECMs

In vivo, self-assembly is not sufficient to explain the diversity in type I collagen structures and properties, although *in vitro*, the fiber thickness, length, stiffness and elasticity can be manipulated by varying temperature, pH, ionic strength and collagen concentration.[36,37] In addition, the balance of type I collagen fibril formation and degradation is critical for some biological processes. The disruption of this equilibrium might result in diseases including fibrosis and tumor cell metastasis.

It has been studied for over a century what are the *in vivo* regulators of collagen fibrils formation. Quite a few regulators have been suggested, including proteoglycans in the ECM such as decorin, fibromodulin, lumican, osteoglycin, keratocan and biglycan [38-42]; enzymes required for post-translational modification of collagen such as 4-hydroxylase, lysyl

hydroxylases and oxidases [43,44]; proteins involved in the translations of collagen in the secretory pathway such as HSP47 [45]; matrix metalloproteinases (MMPs) responsible for the degradation of extracellular matrix proteins [46] and cell membrane proteins, which are involved in the cellular controls for collagen assembly, such as integrins. [47-49]

1.3 Integrin

Integrins are transmembrane proteins connecting the cell cytoskeleton and extracellular matrix (ECM). Integrins are dynamic communicators to convey the messages of cells to their extracellular environments and provide feedback from the ECM to guide intracellular events. The general function of integrins is in mediating cell-cell and cell-matrix interactions.

1.3.1 The integrin family

Integrins are heterodimer proteins comprised of an α subunit and a β subunit non-covalently linked together. Each subunit contains a large extracellular domain (~80-150kDa), a transmembrane domain, and a short intracellular cytoplasmic tail domain. There are 18 types of α subunits and 8 types of β subunits, pairing to form 24 different integrin types in mammals. [50] Integrins interact with ECM components such as collagen [50], laminin [51], and fibronectin [52] as listed in Table 1.1. Most integrins are extensively expressed in many tissues, while some are only expressed in certain tissues to carry out their

functions. In the scope of this thesis, attention is focused on collagen-binding integrins.

Table 1.1 Classification of integrin heterodimers and their binding ligands.

Integrin type		Binding ligand	Integrin type		Binding ligand
β1	α1	Collagen	β2	αD	Leukocyte
	α2	Collagen		αL	Leukocyte
	α10	Collagen		αM	Leukocyte
	α11	Collagen		αX	Leukocyte
	α4	N/A	β3	αIIb	RGD
	α9	N/A		αV	RGD
	αV	RGD	β4	α6	Laminin
	α8	RGD	β5	αV	RGD
	α5	RGD	β6	αV	RGD
	α3	Laminin	β7	αE	Leukocyte
	α6	Laminin		α4	N/A
	α7	Laminin	β8	αV	RGD

1.3.2 Integrin signaling and activations

Integrin signaling is bi-directional. Binding of ligands to the extracellular domain of integrin sends signals to the interior of the cells. The cells are then

further stimulated by other pathways and activate integrins to modulate the binding affinities to ligands in the ECM. [50]

There are two conformations of integrin in equilibrium: a resting state (closed, bent, low-affinity state) and an activated state (extended, high-affinity state). [53,54] The activation of integrin from resting state to activated state can be achieved by either “inside-out” or “outside-in” signaling. In “inside-out” signaling, intracellular activators (kindlin, talin, etc.) interact and destabilize the salt bridge between intracellular domains of the α and β subunits, converting integrin to an extended conformation with increased binding affinities to ligands. In “outside-in” signaling, a ligand binds to the resting state of integrin and activates integrin to the high-affinity state. This triggers integrin clustering with a series of intracellular signaling cascades. [55]

Collagen-binding integrins bind to collagen via the sequence GXX’GEX’’, determined by using CMPs and collagen toolkits described in Chapter 1.1.3. When integrins are in the resting states, they only bind to high-affinity binding motifs such as GFOGER and GLOGER. Once activated, integrins have less selectivity on binding sequence and bind to most GXX’GEX’’ motifs. The binding affinities are likely to be modulated by X in GXX’GEX’’. The binding affinity order at X position is F ≥ L ≥ M > A. [56]

1.3.3 The integrin I domain structure

Each of the four collagen binding integrins (α1β1, α2β1, α10β1, and α11β1) contains an inserted domain (I-domain) in their α subunits, which inserts into the β subunit. Integrins bind to collagen exclusively through this I-domain. The I-domain of integrin contains ~200 amino acids and adopts a Rossmann fold structure. One characteristic feature of integrin I domains is a metal ion dependent adhesion site (MIDAS), where a divalent metal cation, such as Mg^{2+} and Ca^{2+}, is bound. The amino acids directly bound to the metal ions are conserved in all integrin α subunits. For example, in the α1 I domain, D150, S152, S154, D253 and T220 (via a water molecule) coordinate with the metal cation in the MIDAS. A negatively charged glutamate residue on collagen binds to the metal ion by replacing a coordinated water molecule.

1.3.3.1 α1 and α2 Integrin I domain X-ray crystallography structures

Among all four collagen-binding integrins, α1β1 and α2β1 are the best characterized, and this thesis focuses on these two integrins. The structure of the integrin α2 I domain was resolved by X-ray crystallography (PDB ID 1aox) [57]; it adopts a classic Rossman-fold [58]. There are five parallel β-strands and one short antiparallel β-strand in the core surrounded by seven alpha helices, as shown in Fig. 1.4A. The sequential secondary structures from N- to C-termini are: βA(144-151), Loop1, α1(159-172), βB(180-187), βC(191-195), α3(203-212), Loop2, α4(223-233), α4'(237-239), βD(246-253), Loop3,

α5(260-272), βE(275-282), αC(284-288), α6(294-301), βF(312-315), and α7(318-333). The sequence indexes skipped are loops connecting the helices or strands. The MIDAS is located in the upper loop region with residues from Loop1, Loop2 and Loop3 coordinating the Mg^{2+}. The metal ion is directly coordinated to S153, S155, D254 and three water molecules in an octahedral arrangement (Fig. 1.4B). The alpha helix containing residues 284-288, called the C-helix, was found extending from the top loop region of the integrin I domain and protruding from the MIDAS. This C-helix is not found in many other integrin α subunit I domains. A characteristic salt bridge formed between R288 and E318 stabilizes the relative positions of the C-helix and helix 7 (Fig. 1.4C). [57]

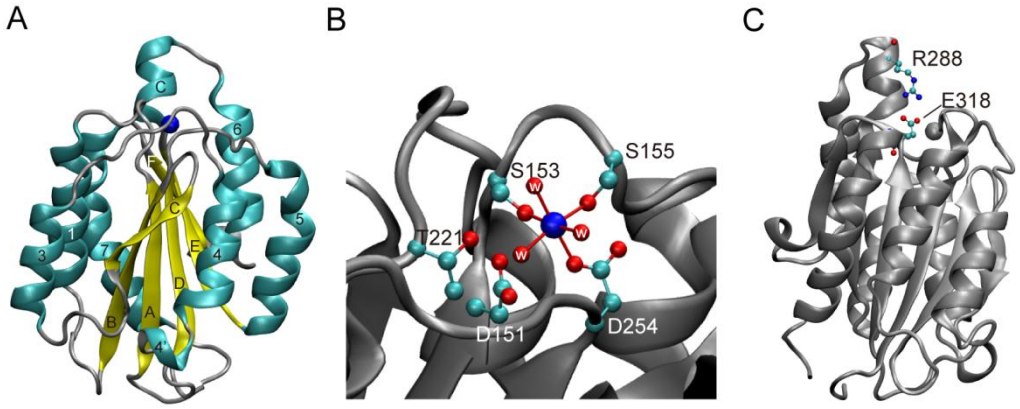

Figure 1.4 Integrin α2 subunit I domain structures. A. The Rossmann fold is shown in ribbon presentation with beta sheets in yellow, alpha helices in cyan and metal cation in blue. The sheets, helices and loops are labeled by numbers or letters. B. The MIDAS region of the upper surface of α2I. The metal cation is represented by a blue sphere. The amino acids from α2I that directly and indirectly coordinate with the metal are shown with ball-and-stick

representation. Oxygen atoms from water that are coordinated with the metal are shown as red spheres and labeled “w”. C. R288 and E318 that form a salt bridge between the C-helix and helix7 are shown in ball-and-stick representation.

The structure of the α1 Integrin I domain resolved seven years later (PDB ID 1pt6) [59] shows that the α1 I domain shares all of the structural features with the α2 I domain (Fig. 1.5). All of the critical residues involved in structural contributions are conserved.

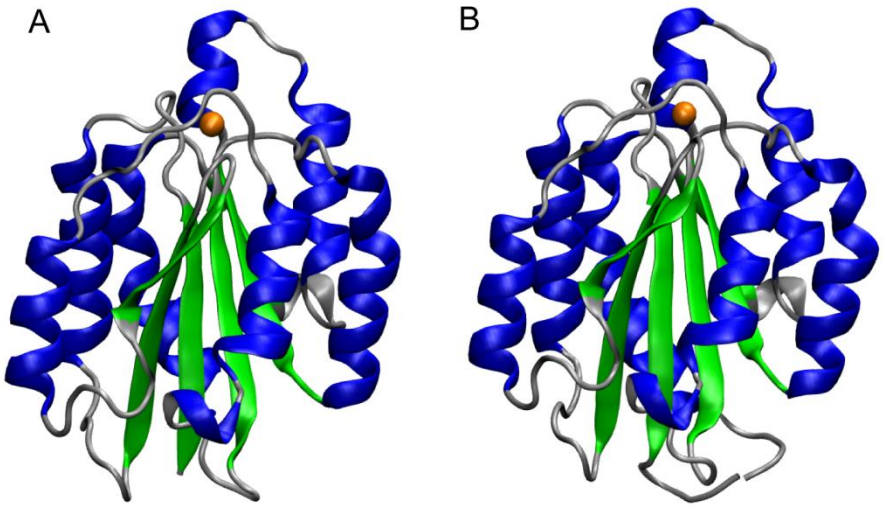

Figure 1.5 Comparison of α1 and α2 integrin I domains. α1 (A) and α2 (B) are shown in ribbon presentation with beta sheets in green, alpha helices in blue and metal cation in orange.

1.3.3.2 Integrin I domain–collagen complex structures

The structure of the integrin α2 I domain–collagen complex was resolved by exploiting collagen model peptides containing the critical integrin binding site GFOGER (PDB ID 1dzi). [60] In comparison to the free unbound integrin I domain, five major differences are found: 1) change in the coordination and

position of the Mg^{2+} ion at the MIDAS, which creates a complementary surface for the bound collagen; 2) the C-helix unwinds; 3) helix 6 extends four residues to the N-terminus; 4) helix 7 moves 10 Å downward; and 5) the characteristic salt bridge between R288 and E318 breaks. The binding of collagen at the upper MIDAS leads to conformational changes locally and propagates to the opposite side of the I domain (helix 7 downward movement). It is proposed that this conformational propagation might be the basis for integrin signal transduction. [60] By overlaying the free α2I ("closed" conformation) and the collagen-bound α2I ("open" conformation), Y285 on the C-helix is found to be in a position to partially block collagen binding in the "closed" form but shifts in the "open" form due to the C-helix unwinding. [60] In this way, the C-helix has been proposed to play a role as a "gatekeeper" to selectively allow collagen to access and bind to integrin. [61] A comparison of the unbound and bound forms of the α2 I domain is shown in Fig. 1.6.

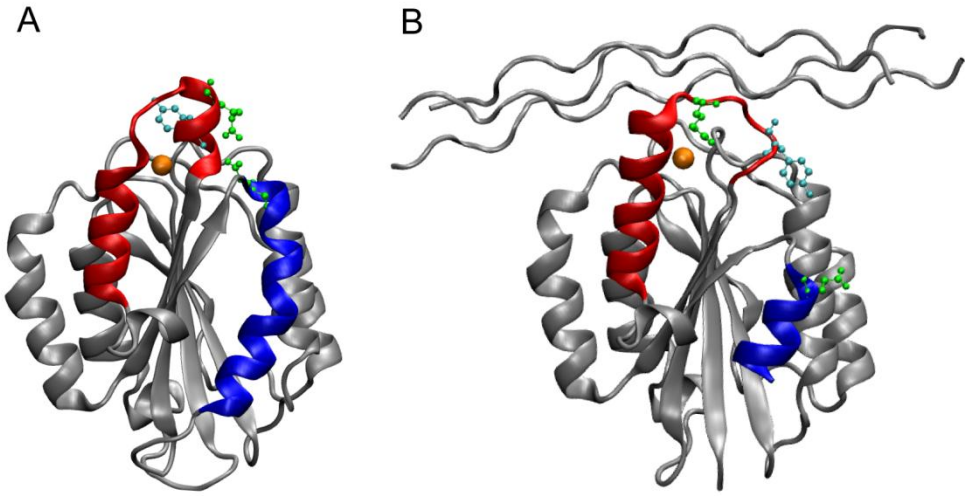

Figure 1.6 In comparison with the unbound from (A), the α2I bound form (B) adopts a more open conformation. The PDB IDs are 1aox [57] for the unbound form and 1dzi [60] for

the bound form. The three main conformational changes are highlighted. The C-helix and helix 6 are colored in red. Helix 7 is colored in blue. The metal ions are shown as orange spheres. The helix 7 shown in the bound form is not complete due to missing electron density, but in fact it is the same length as that in the unbound form. The two residues forming the salt bridge (R288 and E318) are colored green. The residue Y285, which blocks the access of collagen in the unbound form, is shown in cyan.

Collagen interacts with integrin in the MIDAS region of the integrin I domain, which forms a complementary contour of the collagen triple helix. The interactions on the binding interface include metal coordination, hydrogen bonding, salt bridges, hydrophobic interactions and others. A glutamate on the “middle chain” of collagen directly coordinates with the metal ion. Arginine on the “middle chain” forms a salt bridge with D219 of the integrin I domain. Hydrogen bonds form between N154 and a “middle” chain hydroxyproline, H258 and a “middle” chain arginine, and Y157 and a “leading” chain hydroxyproline. The “middle” chain phenylalanine sits in the dimple made by Q215 and N154 from integrin. In the “leading” chain, a phenylalanine makes hydrophobic contacts with Y157 and L286 of integrin, and an arginine lies in an acidic pocket close to E256.

In comparison with the unbound form, the bound form of the integrin α1 I domain (PDB ID 2m32) shows structural changes in the metal cation, C-helix,

helix 6 and helix 7, similar to those of the α2 I domain. [62]

1.3.3.3 Integrin α I domain gain-of-function mutants

The comparison of the free and collagen-bound forms of the α2 I domain of integrin reveals that upon activation, helix 7 travels downwards by ~10 Å. The linker segment of this helix contains a glutamic acid that is proposed to interact with the β subunit. [63] In addition, the C-helix blocks the access of collagen in the resting state and unwinds to allow access of collagen in the activated state. It is proposed that these two regions act as a bidirectional conduit for integrin signaling. [64] A salt bridge between R288 in the C-helix and E318 in helix 7 exists in the resting state and breaks in the activated state, and this salt bridge is therefore believed to stabilize the integrin α2 I domain.

Efforts have been made to mimic the active and inactive states of integrin αI domains by mutating specific amino acids. [56,64-68] A more active integrin α2 I domain is achieved by mutant E318W, which binds to collagen in a 2:1 stoichiometry. This mutant binds collagen with a higher binding affinity but less selectivity. [64] A similar, more active mutant for the α1 I domain, E317A, has also been reported. [68] Both gain-of-function mutants are found with an unwound C-helix, but inconsistent positions of helix 7. The structures and binding selectivities of unbound integrin α1 I domain, E317A and bound integrin α1 I domain are compared in Fig. 1.7.

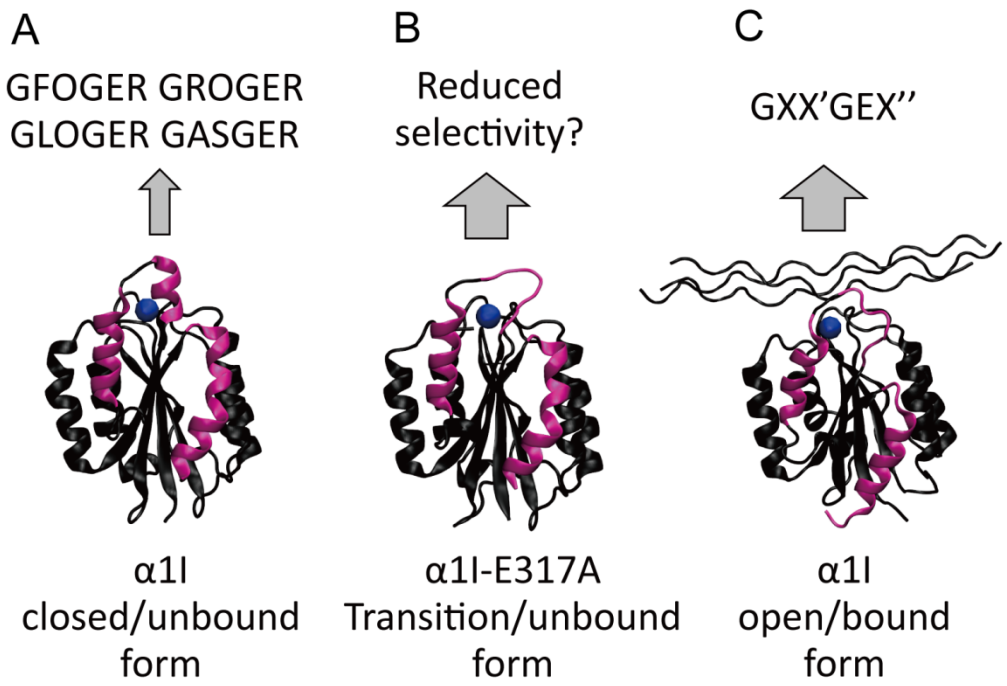

Figure 1.7 Unbound, bound and gain-of-function mutant E317A structures of α1 I domains. The metal ions are shown as blue spheres. The regions undergoing conformational changes are highlighted in magenta.

1.3.4 Integrin collagen interactions *in vivo*

Besides the structural studies of recombinant collagen-binding integrin I domains, integrin-collagen interactions have been widely studied in cells and *in vivo*. Integrin-mediated cell adhesion to collagens are found to be involved in many important biological processes, such as platelet adhesion on the ECM during would healing [69], induction of collagenase gene expression [46], force generation to contract collagen gels [70], and mediation of collagen matrix reorganization [71,72]. Null mutations in mice showed embryonic lethality when an α2 integrin subunit was knocked out. The α1-deficient mice showed failures

in cell attachment to collagen and loss of collagen expression feedback regulation. [73,74]

1.4 Scope of this dissertation

As the most abundant protein in the human body and the main component of the ECM, type I collagen fibrils not only provide tensile strength to uphold tissue integrity, but also carry out biological functions through interactions with its many binding partners. Despite the importance of collagen-protein interactions, a number of receptor recognition sites on collagen are not accessible from collagen fibril surface. Outstanding questions remain: Is the collagen fibril surface rigid or dynamic? What are the structural fluctuations on the collagen fibril surface? How do collagen-binding proteins, such as integrins, get access to their binding sites on collagen when their binding sites are buried inside the fibril and not obviously accessible from the collagen surface? These questions are addressed in Chapter 3.

The interactions between collagen and α1β1 integrin play key roles in numerous biological processes. Upon activation, integrins enhance the binding affinities to collagen and are able to bind low-affinity binding sites on collagen. Questions still remain: What is the mechanism of the integrin I domain activation process induced by collagen? Are slow timescale dynamics (minutes to months) of integrin critical for allosteric events and collagen

recognition. Does the local stability of integrin I domains affect the conformational changes upon binding? These questions are addressed in Chapter 4.

Collagen fibrils in different tissue ECMs show great diversity in their diameters, structural organization, and mechanical properties. Integrins have been proposed to be one of the regulators for collagen fibril formation. Questions still remain: How does integrin regulate collagen fibril assembly kinetics, morphologies, and mechanical properties? This question is addressed in Chapter 5.

In summary, my Ph.D. research has been focused on structure, dynamics and functions of the collagen triple helices and fibrils, collagen-binding integrins and collagen–integrin interactions. The global objective has been to uncover the mechanism of the collagen–integrin interaction from the perspective of both interacting partners.

Chapter 2 Research Methodology

2.1 Molecular dynamics (MD) simulation

Molecular dynamics (MD) is a computer simulation method to mimic the movements of atoms in real life. The basic idea of MD simulation is dividing the time into discrete time steps (often a few femtoseconds) and in each time step, the motion of a single atom is calculated by solving Newton's equation. At the starting of this time step, the force acting on a single atom is calculated based on the force field and the positions of this atom and the nearby atoms. The acceleration of this atom is calculated from the force and the velocity at this time step. The velocity and the time step decides the direction and distance this atom travels so that the position of this atom is updated at the end of this time step. The MD simulation contains a number of these cycles (position – force – acceleration – speed – position) and accumulates the discrete femtosecond time steps to a longer time (often on the nanosecond time scale).

In real time, atoms are in constant motion and will not stay in the energy minimum states for all time. Given a long enough time, MD simulations can sample all states of each atom. The probability of finding an atom in a certain state is a function of potential energy and therefore the potential energy surface can be mapped. In reality, no MD simulations are long enough to sample all of the energy minima and might be stuck in the local energy minima. Thus, a good starting structure is crucial for MD simulations.

MD simulation has wide applications and is frequently used to study structures, dynamics and energies of proteins. The dynamics and the ensembles of a protein obtained by MD simulations complement the static single structure from X-ray crystallography. MD simulations are also used to determine where small drug molecules bind to proteins and how they affect the proteins' structure and functions. This method is used to simulate the type I collagen fibril model with fibril surfaces and study the dynamics in Chapter 3.2.2.

2.2 Solvent-accessible surface area (SASA)

Solvent-accessible surface area refers to the surface area of biomolecules that can be accessed by the solvent. The outline of the solvent-accessible surface area is drawn by the continuous position of the center of a solvent molecule as it rolls over the van der Waals surface of the protein. The solvent molecule is represented by a probe sphere and its radius is relevant to the solvent type (1.4 Å for water). The solvent accessible surface of each atom is calculated by the sphere surface of its van der Waals radius extended by the probe radius excluding the hidden surface by the neighboring atoms. [75,76] This method is used with a larger probe size (8.0 Å) to characterize the receptor accessibilities from collagen fibril surface in Chapter 3.2.5.

2.3 Atomic force microscopy (AFM)

Atomic force microscopy is a scanning probe microscopy (SPM) technique,

designed to measure local properties such as heights, conductivity and force maps. A very sharp tip (end radius < 10 nm) mounted on a cantilever is used to “feel” the sample with a feedback loop. While the cantilever is scanning, a laser pulse is applied to the cantilever and the reflected laser point is detected by a position-sensitive photo-detector to track the deflections of the cantilever. Feedback is given accordingly to control the height of the tip to maintain it at a constant distance from the sample surface. Recording the height of each scanned pixel makes the height profile of the scanned sample.

The deflection of the cantilever is caused by the interactions between the sample and the tip. The tip-sample distance is within the atomic force range and the atomic forces (repulsive or attractive) are calculated by Hooke’s law:

$$F = k \cdot x$$

where *F* is the force, *k* is the cantilever force constant and *x* is the deflection of the cantilever. The atomic force is precisely converted to the distance between the sample surface and tip by a Lennard-Jones potential map, which gives the relationship between force and distance between two atoms. The feedback system adjusts the height of the tip according to the deviations in tip-sample distance from the set value. According to the set distance between tip and sample, AFM is roughly classified into contact mode (repulsive force region, < 10 Å) and non-contact mode (attractive force region, > 10 Å). This method is used to obtain height profiles of collagen fibrils in Chapter 3.2.6 and 5.2.4.

2.4 Site-directed mutagenesis

Site-directed mutagenesis is a method to introduce mutations in double-stranded DNA plasmids. This method is applied widely in molecular biology to study the changes of protein activity when specific residues are modified and to insert or remove functional tags. The basic procedure of site-directed mutagenesis requires a pair of forward and inverse primers that are complementary to the template DNA around the mutation site and contain the desired mutations (single point, multiple points, insertion or deletion). The primer is then mixed with wild type DNA template and extended by DNA polymerase to the rest of the gene. The amplification products are straight DNA stands and then treated with a kinase to phosphorylate the 5’ ends and a ligase to connect the two ends of DNA stands to make them circles. DpnI enzyme is used to digest the methylated wild type DNA template (amplification products containing the mutation sites are not methylated). The treated DNA products are then ready to transform into host cells and clone. The purified DNAs from host cells are checked by DNA sequencing for the success of mutation. This method is used to introduce single residue mutation on DNA of recombinant integrin αI domains in Chapter 4.2.1 and 5.2.1.

2.5 Recombinant protein expressions

Recombinant protein expression refers to the way a specific protein is

produced in microbial systems. Previously, kilograms of tissues from animals or plants were required to produce small amounts of given proteins. Recombinant protein expression allows researchers to obtain the desired protein in a large quantity in a simple and quick way and therefore has wide applications in life science, biotechnology, and medicine. Common protein expression systems include bacteria, yeast, baculovirus/insect, mammalian cells and filamentous fungi. Among the expression systems, E.coli is the most widely used expression hosts because it is efficient and inexpensive, and the system has the machinery for folding and some post-translational modifications. Recombinant protein expression by E. coli includes the following steps: 1. Clone the gene of interest in an expression vector. 2. Transform the plasmids into the competent E. coli cells. 3. Incubate and induce the cells so the recombinant proteins are over-expressed for future purifications. [77] This method is used to express the recombinant integrin αI domains in E. coli. system in Chapter 4.2.1 and 5.2.1.

2.6 Protein purifications

Pure proteins are vital for functional and structural studies. The overexpressed recombinant proteins by E. coli systems remain in the cells. Therefore, the objective of protein purification is to get the protein of interest out of cells and obtain the largest amount of protein with fewest contaminants. The steps in purification include cell lysis, separating soluble proteins from cell debris, and

separating the protein of interest from the endogenous proteins and DNA of the cells, which is the most challenging step.

Column chromatography is a common technique for separating soluble proteins from other soluble components. Various column chromatographies separate proteins based on different characteristics, such as sizes (size exclusion chromatography), charges (ion exchange chromatography), and affinity chromatography (antibody or protein-tag). One strategy is to add the sequence of a functional tag (e.g. His tag, GST tag) to either terminus of the protein sequence when encoding the recombinant DNAs, which can be specifically recognized by some affinity chromatography. His-trap chromatography is a commonly used affinity chromatography, in which a column filled with Ni-NTA agarose, coordinates poly-histidine tags in proteins. When cell lysate passes through the column, only his-tagged proteins retain on the surface of the agarose. After washing off the loosely bound impurities, a buffer with a high concentration of imidazole or free histidine is used to elute the His-tagged protein through competition for Ni^{2+} coordination and releases the bound protein. This method is used to purify recombinant integrin αI domains from the E.coli. expressing system in Chapter 4.2.1 and 5.2.1.

2.7 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is an electrophoresis technique to analytically separate proteins components in the mixture solution by their molecular weights. SDS is mixed with proteins and acts as a surfactant, covering the intrinsic charges of proteins and unfolding proteins. Thus, all proteins have similar shapes and charge to mass ratios. When a constant electric field is applied, proteins travel to the anode in the polyacrylamide gel medium at different speeds depending on their molecular weights and separate into multiple bands on the gel. A marker comprised of proteins of known molecular weights is loaded on the polyacrylamide gel for accurate molecular weight determinations. Staining of the polyacrylamide gel is needed to turn the protein into observable bands. Besides the molecular weight information, the thicknesses of the bands allow the comparison of protein concentrations qualitatively. This method is used to inspect the molecular weights and purities of integrin αI domains after purifications in Chapter 4.2.1 and 5.2.1.

2.8 Nuclear magnetic resonance (NMR)

Certain isotopes have a quantum mechanical property called nuclear spin that can interact with an external magnetic field. When placed into an external magnetic field, the nuclear spins align with and precesses around the axis of the external magnetic field. This precession is termed as Larmor precession. The frequency of Larmor precession is measured by applying a radio frequency pulse to convert the equilibrium nuclear magnetization of the sample

(which is collinear to the external field) to the plane transverse to the static magnetic field and recording the nuclear precession, which is measured as a current on the NMR probe. This NMR signal, a fluctuating curve with declining amplitude over time, is termed as the free induction decay (FID). Fourier transform is used to convert the FID which is a function of time to a function of frequency.

The electrons of the atom placed in the external magnetic field, is induced to produce local currents, resulting in an inverse magnetic field to shield the external magnetic field. This screening effect is known as chemical shift. Chemical shifts are determined by the nuclei and influenced by the micro-environment of each nucleus. Although like nuclei (such as H atoms) have the same gyromagnetic ratios, within a molecule their precession frequencies are different due to the effect of chemical shielding. Chemical shielding is the phenomenon where the effective magnetic field at the nucleus is modulated by the electron clouds of chemical bonds. Chemical shielding modulates the measured precession frequency of a nucleus.

2.8.1 Heteronuclear correlation spectroscopy

Two nuclei in one molecule connected by a few chemical bonds can be observed to be coupled in NMR spectra as splitting of NMR signals. This phenomenon is known as J-coupling, which is applied to identify the

correlations of nuclei. ^{1}H-^{15}N HSQC (heteronuclear single quantum correlation) is a one-bond coupling NMR method, which is the most widely used spectrum to study proteins by NMR. In the HSQC spectrum, each cross-peak originates from a pair of coupled ^{1}H-^{15}N nuclei. The chemical shifts of these two nuclei are shown on the x and y axis of HSQC spectrum. The H-N bond on the backbone of each amino acid in a protein is represented as a peak on the HSQC spectrum, which allows the separation of residues. ^{1}H-^{15}N HSQC spectrum is like a “protein fingerprint”. The chemical shifts contain structural information of well-ordered vs. mobile regions. This spectrum is usually the first heteronuclear experiment performed on proteins for further NMR experiments such as titrations and dynamics. This method is used to obtain HSQC spectra of integrin I domains alone and in complex with collagen model peptide in Chapter 4.2.3 and 4.2.7.

2.8.2 Sequential assignment of protein residues by heteronuclear coupling

Although HSQC allows the separations of spin system representing each residue along the sequence, it is not informative of the sequential assignment of each spin system. A 3-D triple resonance heteronuclear correlation spectrum correlating the one-bond couplings between ^{1}H and^{15}N and one or two bonds couplings between ^{15}N and ^{13}C are used for the assignments of NMR spin systems. The continuous order of backbone atoms H, N, C, O C^{α}

and C^{β} can be obtained by various NMR methods. The spectrum HNCACB provides chemical shifts information of N, H, C^{α} and C^{β} of the i^{th} residue and C^{α} and C^{β} of the $(i-1)^{th}$ residue. The spectrum HN(CO)CACB gives chemical shifts of N, H of the i^{th} residue and C^{α} and C^{β} of the $(i-1)^{th}$ residue. Combining information from each, neighboring residues are linked by the matching the C^{α} and C^{β} chemical shifts of the i and (i-1) residue. Despite that HNCACB and HN(CO)CACB provides information of both C^{α} and C^{β}, the peak intensities are quite weak and sometimes missing. HN(CO)CA and HNCA spectra are usually the spectra to start with due to the higher peak intensities. The chemical shifts of C(O) can also be used to link the neighboring residues by HNCO and HN(CA)CO spectra. This method is used to assign the spin systems of integrin I domain HSQC spectra to residues in Chapter 4.2.5.

2.8.3 Secondary structure prediction by Talos+

Talos+ (Torsion Angle Likeliness Obtained from Shift and Sequence Similarity +) program is a hybrid method to predict protein secondary structures from NMR chemical shifts. [78,79] This program is based on a database of 200 proteins, in which nearly complete backbone NMR chemical shifts assignments from Biological magnetic resonance bank (BMRB) and coordinates from X-ray crystallography or NMR structures from Protein data bank (PDB) are available. The concept of Talos+ is to search this database for a sequence of triplet residues of interest with similar chemical shifts to yield the

secondary structures of the 10 closest triplets. It also includes a neural network to enhance the conventional database search and reports a reliable estimate of the prediction. If the central amino acid in these 10 triplets exhibits similar Φ and Ψ angles, their averaged values are accepted as the predicted Φ and Ψ angles for that central amino acid. This method is used to predict the secondary structures of integrin α1 I domains and the gain-of-function mutant E317A from the NMR chemical shifts in Chapter 4.2.2.

2.8.4 Hydrogen-deuterium exchange (HDX) studied by NMR

Hydrogen-deuterium exchange is a reaction of a hydrogen atom exchanging with a deuterium atom. When a protein is diluted into a solution of D_2O, deuterium starts to exchange with the hydrogen atoms, and due to the large number of deuterium atoms in the solvent, it is hard for the deuterium to exchange back to a hydrogen atom. Hydrogen-deuterium exchange studies focus mainly on the main chain amide group (-CO-NH-) because it is placed at every residue in a protein except for prolines, and is involved in hydrogen-bonding networks, which are important factors for protein secondary structure. The hydrogen exchange rates for small structureless amide models under different conditions (pH, temperature, and effects of direct neighbor residue side-chains) have been carefully calibrated and stored in a database for calculating the intrinsic hydrogen exchange rate (k_{int}). [80-82]

Amide protons on the surface of the protein, which have contacts with external solvent, can be exchanged with deuterium readily. Residues protected by the secondary structure of the protein, whose amide protons do not have direct contact with solvents, can still be deuterium exchanged when the secondary structure is transiently broken or “open” following the equation below:

$$Closed \underset{k_{cl}}{\overset{k_{op}}{\longleftrightarrow}} open \xrightarrow{k_{int}} exchanged$$

where k_{op} and k_{cl} are the opening and closing constants, respectively. When the structure is stable ($k_{cl} >> k_{op}$), the apparent HDX rate k_{ex} is given by:

$$k_{ex} = \frac{k_{op}k_{int}}{k_{op} + k_{cl} + k_{int}} \approx \frac{k_{op}k_{int}}{k_{cl} + k_{int}}$$

In the biomolecular exchange EX2 regime, where reclosing is faster than exchange ($k_{cl} > k_{int}$), the apparent HDX rate k_{ex} is given by:

$$k_{ex} = \frac{k_{op}k_{int}}{k_{cl}} = K_{op}k_{int} = \frac{k_{int}}{P_f}$$

where K_{op} is the pre-equilibrium opening constant, and P_f is the protection factor. Local unfolding free energy ΔG_{op} is given by:

$$\Delta G_{op} = RTlnP_f$$

where R is the ideal gas constant whose value is 8.314 J mol^{-1} K^{-1}, and T is the temperature. ΔG_{op} is usually within the range of 6.6 – 8.2 kcal/mol. [83]

The observable hydrogen-deuterium exchange rate can be obtained by NMR. Protein solutions are lyophilized into powders and dissolved in D_2O, and this time is recorded as t_0. N-H HSQC spectra are performed of this NMR sample

at different times relative to t_0. As amide protons are continuously exchanged by deuterium and rarely exchange back, the N-H HSQC resonance peak intensities decay with time. The peak intensities of each amino acid vs. time are fitted to obtain the k_{ex}, which is used to calculate the protection factor P_f and local unfolding free energy ΔG_{op}. This method is used to obtain and map the per-residue protection factors of integrin α1 I domains from the NMR chemical shifts in Chapter 4.2.2.

2.9 Circular Dichroism (CD)

Circular dichroism (CD) is a spectroscopic technique measuring the difference in absorptions of left- and right-handed circularly polarized light of a molecule. Molecules with chiralities have circular dichroism signals, which are wavelength dependent. CD has extensive applications in biochemistry to study the secondary structure of proteins. CD spectra between 260 and 180 nm can be analyzed for secondary structure content of proteins, such as alpha helix, beta sheet, collagen triple helix. Standard CD spectra of each pure secondary structure are available [84] for structural comparison, determination and deconvolution of percentages of each secondary structural element. Monitoring the CD signal of a protein at a characteristic wavelength (usually maximum or minimum of the standard CD spectrum) while gradually increasing the temperature renders the unfolding profile, in which the melting temperature can be determined at the transition midpoint. The melting

temperature of a fully reversible folding/unfolding process is directly related to the conformational stability of a protein. This method is used to characterize the secondary structure and to measure the melting temperature of a collagen-like triple-helical peptide GLOGEN in Chapter 4.2.6.

2.10 Enzyme-linked immunosorbent assay (ELISA)

ELISA is a solid-phase bioassay that tests the identity of a specific protein through antibody recognitions. ELISA has been widely used because it is simple and quick to perform. In a direct ELISA assay, the protein of interest is directly immobilized on polystyrene microplate wells and allowed to be recognized by antibodies. The steps of ELISA include coating (immobilizations of antigen on polystyrene microplate wells), blocking (applications of irrelevant proteins to cover all the unsaturated spots of microplate wells), detection (incubation with tagged antigen-specific antibodies to bind to antigens) and signal measurements (detections of signals transformed from the bound antibodies). The signal observed is proportional to the amount of signal-generating enzyme tagged antibodies in the microplate wells.

A “sandwich” ELISA is used to measure the bindings of two proteins. Instead of directly immobilizing antigens, protein A is immobilized and allowed to bind the protein B (antigens). The amount of protein B are transformed to signals to reveal the binding capacities of protein B to immobilized protein A. A “sandwich”

ELISA has applications of comparing the binding affinities of variants to wild type proteins. A “sandwich” ELISA is also applied to quantitatively determine the dissociation constant (K_d) by fitting signals of a series of concentrations of protein B to immobilized protein A.

The separations of specific and non-specific interactions that occur on solid surfaces are important. Washing steps are required to get rid of non-specific bindings to the largest extent. Replacing the immobilized protein A by irrelevant proteins generates the negative control signals to illustrate the non-specific bindings. Optimizations of ELISA are crucial to maximize the specific bindings and minimize the non-specific bindings. This method is used to compare the adhesions of integrin α1 I domains and the gain-of-function mutant E317A to collagen in Chapter 4.2.8.

2.11 Biolayer interferometry (BLI)

BLI is a label-free optical technique to measure protein-protein interactions. It analyzes the interference pattern of reflected light from two surfaces of the biosensors, the internal reference layer and the external layer. A ligand is immobilized to a biosensor in the first step. When an analyte binds to the ligand, the changes in external layer thickness result in wavelength shifts. Wavelength shifts are monitored in real time to record the association and dissociation events, which can be fit to different models (e.g. 1:1, 1:2, 2:1) to

obtain reaction rate constants and dissociation constant of analyte from immobilized ligand. This method is used to measure the binding events of integrin I domains on immobilized type I collagen in Chapter 5.2.2.

2.12 Ultraviolet-visible (UV-vis) spectroscopy

Amino acids with aromatic rings (tyrosine, phenylalanine and tryptophan) absorb ultraviolet light at 280 nm. This allows convenient quantifications of protein concentrations by directly measuring absorbance without additional reagents or incubations. The extinction coefficients of proteins are empirically estimated by the number of aromatic residues in proteins. The relationship between protein concentration and the absorbance observes Beer-Lambert's Law:

$$A = -logT = -log(I_t/I_0) = \varepsilon \cdot c \cdot l$$

where *A* is the absorbance; *T* is the transmittance; I_t is the emergent beam intensity; I_0 is the incident beam intensity; ε is the extinction coefficient; *c* is the protein concentration; *l* is the path length

Beer-Lambert's Law applies only on a certain range of absorbance depending on the proteins (generally Abs 0.2 – 0.8). Proteins need to be diluted or concentrated to this range to obtain accurate concentrations. This method is used to measure the concentrations of integrin αI domains in Chapter 4 and 5.

UV-Vis also measures the optical density (OD). Rather than absorbance of the light, OD describes the light scattered by a suspension. OD is used to determine the turbidity of a suspension, such as monitoring the growth of aggregates formation in a solution. The wavelength 313 nm is within the protein/peptide non-absorbing region, and thus OD_{313nm} has been used to reflect the turbidity of a fibril-forming solution (such as collagen). Wavelengths in the 350-600 nm range are also used to measure ODs but 313 nm is the most common one. The increase in turbidity might result from the increasing amounts or sizes of collagen fibrils in the solution. [85] This method is used to obtain type I collagen fibril formation curves in Chapter 5.2.3.

2.13 Rheology

Rheology describes the deformation and flow behavior of materials. All materials are composed of molecules. When a liquid is put in motion, the molecules within the liquid are forced to slide along each other and develop a resistance to the flow. The viscosity of a fluid measures this resistance to the deformation and therefore reflects the internal friction between molecules in the fluid. Rheometers test the shear by parallel plate geometry: the fluid fills the gap between two parallel plates with one plate stationary and the other plate rotating in a controllable shear rate, and performing stress measurements. Viscosity is defined as shear stress divided by shear rate and is an absolute value independent of the measuring system. For the ideal

viscous fluid (Newtonian fluid), the viscosity is independent of the shear rate. There are many non-newtonian fluids having shear-thinning behavior (lower viscosity with higher shear rate) and shear-thickening behavior (higher viscosity with higher shear rate). Solutions containing large molecules are usually shear-thinning fluid, which is related to the internal structures of samples. This method is used to evaluate the influences of integrin αI domains on the viscosity of type I collagen gels in Chapter 5.2.5.

Chapter 3 Cryptic Binding Sites become accessible through Surface Reconstruction of the Type I Collagen Fibril

3.1 Introduction

The extracellular matrix (ECM) in connective tissues contains a mixture of biological components that regulate cell migration, growth, and differentiation through cellular interactions. Making up 90% of all collagen in the human body, type I collagen forms large fibrillar structures that not only provide tensile strength to uphold tissue integrity, but also maintain biological functions through interactions with its many binding partners, including cell surface receptors, enzymes, and other ECM components[1,3,26,86]. For example, collagen interactions with integrin cellular receptors are important for platelet aggregation, cell development, differentiation, and hemostasis[50,87,88]. Collagen fibril degradation and turnover is dependent upon cleavage by matrix metalloproteinases (MMPs). Defects in collagen interactions are associated with fatal diseases, such as heart disease, cancer, and arthritis[27,28].

Interactions with full-length collagen monomers and fibrils are extremely challenging to study due to their huge size and complexity. Broad interaction domains on collagen monomers and fibrils have been identified through visualization of protein binding by atomic force microscopy (AFM) and electron microscopy (EM)[8-10,89,90]. More specific recognition sequences for dozens of type I collagen binding partners have been determined through elegant use of synthetic collagen mimetic peptides (CMPs)[11,65,91] and recombinant bacterial expression systems that contain partial collagen sequences[92-94]. Through

adhesion to triple helical CMPs, a minimal binding sequence for collagen-binding integrins has been established, GXX'GEX", in which the Glu of the collagen motif coordinates a divalent metal cation with the metal ion-dependent adhesion site of the integrin inserted (αI) domain[66,95]. In the context of the linear triple helix, in which all possible binding sites are exposed (Figure 3.1a), αI domains show preferential binding to a subset of these motifs[56]; high and moderate affinity binding motifs for α1I and α2I are colored yellow in Figure 3.1. However, in the ECM, collagen monomers assemble into cylindrical D-banded fibrils via microfibrils[6,96] (Figure 3.1b-d). The bundling of monomers into the quasi-hexagonal arrangement[15,97] buries many of these sites, making them unavailable for interaction (Figure 3.1c). The approximate locations of the six highlighted integrin binding motifs are shown within the smallest repeating unit (SRU) of the fibril, which is one D-period length of the microfibril and contains a bundle of five unique segments from different collagen monomers (Figure 3.1c). Collectively, these "D-segments" contain the entire type I collagen sequence. As the microfibrils assemble in all dimensions, forming a long cylindrical fibril superstructure with a circular cross-section of concentric layers[7], only one face is left exposed for interaction (Figure 3.1d). There are two possible models of the fibril surface; "surface A," represented by D5 and D4 as shown in Figure 3.1[98], and "surface B," represented by D1[21] (see Figure 3.2). Previous studies support the view that the surface proposed by Perumal et al.[98] is a better fit of the corrugated profile of the type I collagen

fibril from rat tail tendon observed by scanning electron microscopy and AFM[99,100] and potential exposure of certain binding sites, such as those of decoron and matrix metalloproteinases (MMPs)[20,22,98,101]. Despite many of its binding motifs being obstructed, integrin α2β1 has been shown to indeed interact with mature type I collagen fibrils as visualized by immuno-EM imaging and mediation of cell spreading, yet through undetermined binding sites[47]. Remarkably, although critical binding sites are buried inside the fibril, these and numerous other cellular processes reliant on collagen–protein interactions are accomplished. It has been suggested that specific packing of triple helical monomers within the fibril makes important protein-binding sites cryptic to become available only in specific instances[20,22]. However, how partner binding sites become available in the supermolecular fibril assembly is not understood.

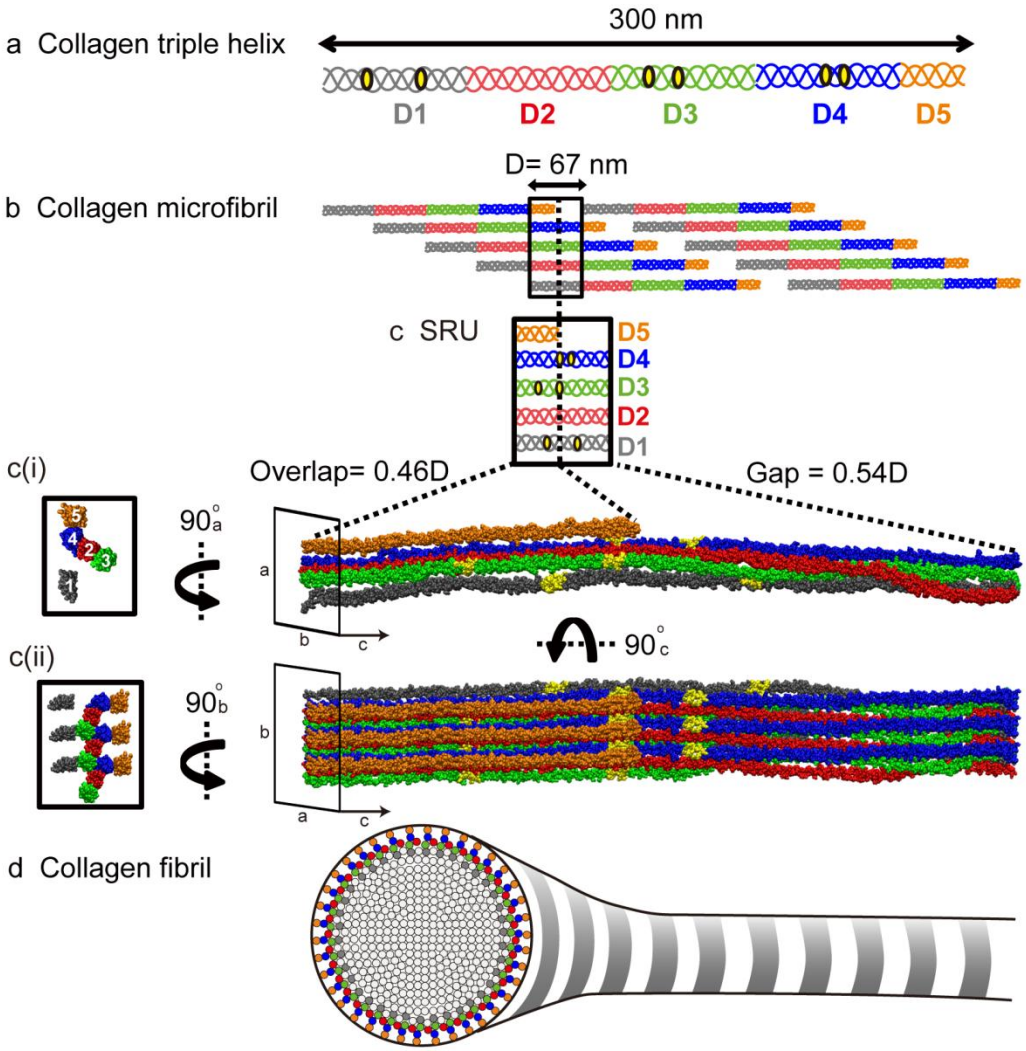

Figure 3.1 Type I Collagen structural hierarchy. (a) Collagen monomer: The type I collagen monomer is a heterotrimer triple helix consisting of two α1 and one α2 chains with approximate dimensions of 300 × 1.5 nm. The monomer is divided into five D-segments with D1-D4 having a length of 67 nm and D5 equal to 0.46D. (b) Microfibril: Five monomers pack in parallel and stagger by one D-period into microfibrils. Based on PDB: 3HR2[15]. (c) Smallest repeating unit (SRU): Isolating one D-period length of the microfibril gives the SRU, which contains the entire sequence of all five D-segments of the monomer in the configuration of the microfibril bundle. (ci) All-atom model of SRU rendered by VMD. (cii) Three replicates of the SRU along the b-axis are created to define

a representative fibril surface, shown in longitudinal view (right) and cross-section view (left). The short D5 divides the D-period into two regions; the “overlap” region contains segments D1 to D5, while the “gap” region only contains D1 to D4. In a and c, integrin binding motifs are indicated in yellow. (d) Fibril: The alternating overlap and gap regions create the characteristic “bright and dark” D-banding pattern viewed by electron microscopy when the collagen fibril is stained with heavy metal. The concentric packing of collagen monomers within a single fibril for the overlap region is viewed in the cross-section. Colored circles represent the estimated positions of collagen monomers on the surface layer and are color-coded by D-segments.

Here we use an integrative approach that combines all-atom molecular dynamics (MD) simulations with atomic force microscopy (AFM) of type I collagen fibrils to characterize the interaction surface of the type I collagen fibril from rat tail tendon. Through MD, we probe the dynamics and surface reconstruction of the surface layer of the type I collagen fibril from a starting model derived from the X-ray fiber diffraction model of the type I collagen fibril unit cell from rat tail tendon[15]. Although the X-ray fiber diffraction model provides only the Cα positions and does not have the resolution to make conclusions about atomic-level details of triple helical conformations within the fibril, it importantly provides the arrangement of collagen monomers within the repeating unit of the fibril, which allows us to model relative positions of interaction sites near the fibril surface. High resolution AFM experiments give

nanoscale insight into the surface structure. Important contributions from X-ray diffraction and microscopy provide static snapshots of a supermolecular assembly, but dynamics, especially at the binding interface, which are potentially critical for cellular function, have not been investigated and are inaccessible by these methods. Previous computational studies of infinite periodic models of collagen fibrils provide structural and mechanical information of the fibril core[16-18,102], but do not distinguish the interaction surface of the fibril from the interior. We have performed an all-atom MD simulation of an early stage, fully solvated type I collagen fibril model with an explicit interaction surface that allows sampling of rare conformational events on the surface. In our model, we used a matrix of SRUs to represent the fibril surface, since the surface is identical around the cylindrical fibril superstructure[7](Figure 3.1d). Through the MD simulation, we observe that the fibril surface is not merely a rigid rod, but exhibits large fluctuations and displacements of particular segments within the D-period. The reconstructed fibril results in the inward contraction of the gap region and the outward expansion of the overlap region creating an overall denser packing of monomers in the surface layer and exposing certain previously hidden interaction sites. The conformational fluctuations change the accessibility of certain binding regions over time and suggest that the dynamics on the surface are critical for collagen fibril interactions and dependent cellular processes.

3.2 Materials and Methods

3.2.1 Constructing and solvating the all-atom collagen fibril periodic model.

We created the all-atom starting structure by combining high resolution models of crystallized collagen-like peptides and the low resolution X-ray fiber diffraction model of the type I collagen fibril from rat tail tendons, adapting the methodologies from de Leeuw et al. [16] We used THeBuScr [103] and Scwrl [104] programs to predict the all-atom model of the triple helical domain of type I collagen based on the sequence translated from genes COL1A1 (P02452) and COL1A2 (P08123) in the UniProt Knowledgebase (www.uniprot.org). This creates a perfectly straight triple helix without supermolecular structure. The Cα atoms of the N- and C-telopeptides were added to the straight triple helix model based on the X-ray fiber diffraction structure (PDB ID: 3HR2) [15] and all other atoms were added by the program LEaP in AMBER package. [105] The all-atom model was then fitted to the supermolecular structure of PDB entry 3HR2 by a best-fit rotation and translation of the Cα atomic coordinates. In vacuum minimizations were carried out to remove bad contacts. We used the AddToBox utility in the AMBER package, which is designed for crystal simulations, to add 11980 explicit water molecules to our system based on de Leeuw's trial and error result. [16] We performed 15,000 steps of minimizations and 100 ps of heating from 0K to 310K, followed by production with gradually decreasing restraints from 10 to 0.1 kcal/mol × $Å^2$ applied to all protein atoms.

3.2.2 Constructing the collagen fibril surface model.

The X-ray fiber diffraction structure of the three-dimensional arrangement of collagen molecules in naturally occurring type I fibrils from rat tail tendon (PDB ID: 3HR2) [15] determined that the smallest repeating unit (SRU) of the collagen fibril contains all five D-segments from successive collagen monomers. The SRU unit cell is 678 Å long, 27 Å wide and 40 Å deep. We built an all-atom model of a single SRU by truncating the last frame of the all-atom periodic collagen model MD simulation every one D-period and packing them into a single unit cell. The five D-segments are indicated by colors in Figure 3.1. The 3a3b collagen fibril model was built to be one unit cell long, three unit cells wide and three unit cells deep, containing two possible surfaces: surface A and surface B (Figure 3.2). The other two boundaries (a-c plane) are not fibril surfaces since full-length fibrils keep expanding in the b dimension. We built only three units along the b axis to represent the periodically extending surface to minimize the computing expense. The 3a3b model was neutralized by Cl^- ions and solvated as a solute in a 12 Å buffer of explicit TIP3P water molecules by the program tLEaP in the AMBER package. [105]

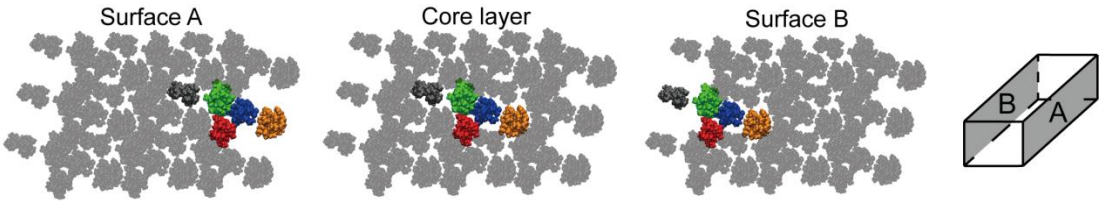

Figure 3.2 The three SRUs analyzed in the MD simulations to represent surface A, B

and core layer are highlighted in the cross-sectional view of collagen fibril model. The box shows the orientation of the fibril surfaces, shaded in gray, relative to the model.

3.2.3 Molecular dynamics (MD) simulation.

All MD simulations were performed in the AMBER2017 package. [105] The protein was treated with the ff14SB force field [106] and the solvent water molecules were described using the TIP3P model. [107] A minimization was first carried out to remove bad contacts in the initial structure. The system was then gradually heated up to 300 K for 100 ps in the NVT ensemble using the Berendsen thermostat. [108] Equilibration in NPT ensemble was then performed with gradually decreasing weak restraints from 5.0 to 0.5 kcal/mol x $Å^2$ applied on all protein atoms until the density of the system reached 1.0 g/mL. Finally, the production run was carried out for 250 ns in the NVT ensemble with 2.0 kcal/mol x $Å^2$ restraints on three residues of the N- and C-termini of each D segment except for the N- and C- telopeptides. These restraints were used to maintain the interactions between D-segments in adjacent D-periods in full-length collagen fibrils. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms during the simulations. [109] Hydrogen masses were repartitioned onto bonded heavy atoms using the algorithm by Hopkins et al. [110], which allowed a long integration time step of 4 fs to be used to accelerate the simulations.

3.2.4 Analysis of simulation results.

The trajectories were visualized using the VMD software, [111] in which the mapping of collagen receptor binding motifs were performed by changing colors of their indexes. Three microfibrils in the 3a3b model were selected to represent either surface A, surface B or the core layer (Figures 3.3C and 3.2) because they are surrounded by all the neighbors present in full length collagen fibrils. The cpptraj utility in the AMBER package [112] was used to perform RMSD, RMSF and hydrogen bonding analyses with default settings.

3.2.5 Solvent Accessible Surface Area (SASA).

The Molecular Surface (MS) program [75,76] was adopted to perform solvent accessible surface area (SASA) calculations, which give the surface area of individual residues that can be approached by a probe of an indicated radius. The default probe size is 1.4 Å which is the radius of a water molecule. We used a larger 8.0 Å probe size to characterize the accessibilities of collagen binding partners. The SASA values of residues on the same position of the three chains within a triple helix were averaged and plotted.

3.2.6 Atomic force microscopy (AFM) of type I collagen fibrils.

Type I collagen from rat tail (Discovery Labware Inc., Bedford, MA) was diluted in 10mM phosphate buffered saline (PBS) pH 7.4 to a final concentration of 2.0 mg/mL and incubated at 37 °C for 2 hours for fibril self-assembly. A sample

volume of 20 µL was deposited on a 1 cm × 1 cm square of freshly cleaved mica (Ted Pella Inc., Redding, CA) and incubated at room temperature for 15 min. Then the surface of the sample was washed with 1 mL of deionized water and left to dry at room temperature for 1 h before being imaged. The samples were imaged by NX-10 instrument (Park Systems, Suwon, South Korea) in non-contact mode with PPP-NCHR tips (force constant 42 N/m; 330 kHz frequency; Nanosensors, Neuchatel, Switzerland). Image processing was conducted using XEI (Park Systems, Suwon, South Korea).

3.3 Results

3.3.1 The type I collagen interaction surface undergoes conformational fluctuations on the nanosecond timescale.

We monitored the motions of the collagen monomers at the fibril surface during MD simulations to characterize the surface reconstruction. MD simulations were performed on a model that contains three copies of the smallest repeating unit (SRU) along the a-axis and three copies along the b-axis, the “3a3b” model. The three layers along the a-axis are surface layers A and B that each have one face exposed and the core layer, sandwiched between the surface layers, which represents the fibril interior. In this way, one SRU in each of the three layers is surrounded by all of its neighbors as in the full fibril. All analyses are presented for surface A, which places D4 and D5 on the immediate interaction surface. The alternative surface B places D1 at the

fibril forefront. From this starting structure (0 ns) to the end point (250 ns) of our MD simulation, three major conformational changes are apparent (Figure 3.3): (1) longitudinal translation of the C-telopeptide, (2) downward displacement of D5 in the overlap region, and (3) contraction of the surface layer in the gap region. From the longitudinal view (Figure 3.3a), we observe a displacement of the C-telopeptide in D5 in the N-terminal direction. This shift exposes a patch of the D4-segment (Figure 3.3a, blue) that was previously occluded by the C-telopeptide. The movement also allows the D5 segment to shift, which opens a wider range of motion for residues in the middle of the segment. In some of the MD frames, the middle of D5 is observed to fluctuate along the a-axis, bulging away from the surface and returning back (Figure 3.4). The cross-section view of a slice from the middle of the overlap region (Figure 3.3b) shows that the D5-segment also has downward movement along the b-axis, which creates a cavity, exposing much of the overlap region of D4 (blue). While fluctuations in the D5 cause outward expansion from the surface, the gap region of the surface layer contracts inward toward the fibril core, creating a much denser packing of monomers in the surface layer and exposing the previously hidden D2-segment (Figure 3.3c, red). This denser packing expels water from the surface layer (Figure 3.3, cyan). In order to characterize these motions in further detail over the time course of the simulation and address how the surface reconstruction may facilitate ligand binding, we analyzed time points of the simulation in terms of displacements,

dynamics, hydrogen bond modulation, and accessibilities.

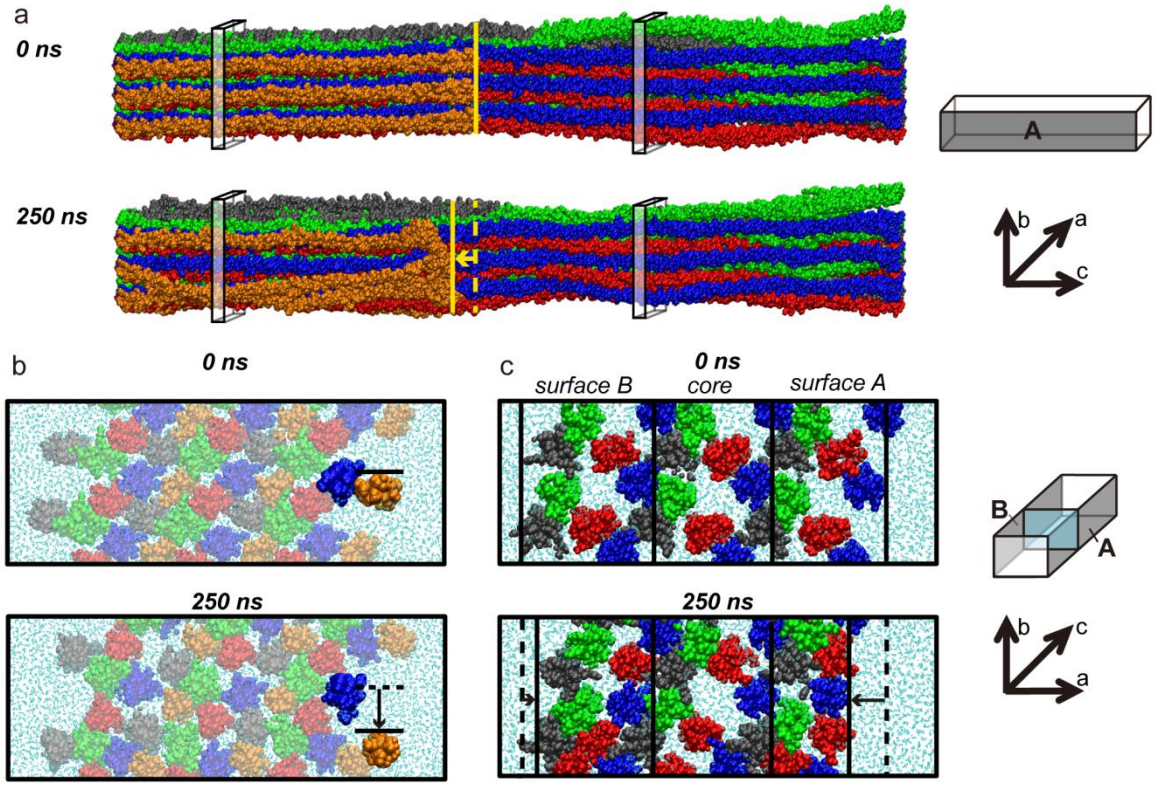

Figure 3.3 Three major movements observed during the 250-ns MD simulation. Snapshots from the starting (0 ns) and ending (250 ns) time points are shown. D-segments are color-coded: D1- gray, D2- red, D3- green, D4- blue, and D5- orange. (a) Longitudinal view of the full D-period model. The C-telopeptide on the D5-segment shifts N-terminally, exposing sites on D4 previously hidden by the C-telopeptide. The edge of D5 is indicated by the solid yellow line in both snapshots. The dashed line at 250 ns demarcates the edge of D5 at 0 ns. The four transparent slices are shown as cross-section view in panels b and c. (b) Cross-sectional views of slices taken at 10-12 nm from the N-terminus along the c-axis. The downward displacement of D5 from its starting position is indicated by black lines in the same manner of those in panel a. Motions in the

middle of D5 open a cavity that allows access to D4. (c) Cross-sectional views of slices taken at 44-46 nm from the N-terminus along the longitudinal axis in the gap region. The surface layer of the gap region contracts inward toward the core, exposing the originally partially hidden D2 and expelling waters (cyan) from within the surface layer. Solid black lines demarcate boundaries of the surface layers and the core layer at the time point indicated. Dashed black lines show the original position of the layer boundary. Boxes on the right of panels a, b and c indicate the orientation of the 3a3b fibril model, with the gray sides representing the surfaces and the blue plane representing a cross-section slice.

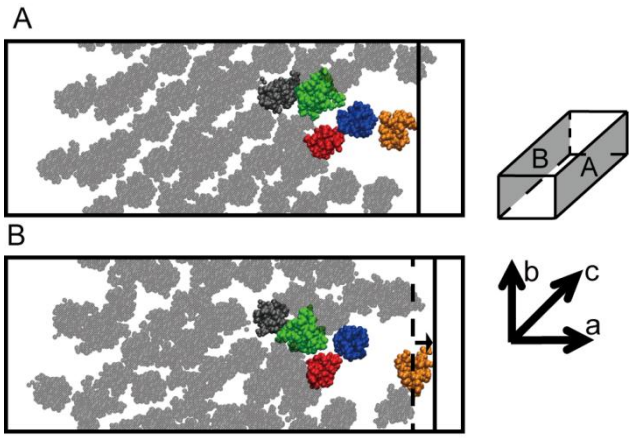

Figure 3.4 Cross-sectional views of slices taken at 10 to 12 nm from the N-terminus along the longitudinal axis (c-axis) of the collagen fibril model. The surface layer of the overlap region, highlighted by colors, expands outward (B) compared to the starting structure (A). Solid black lines show boundaries of the surface layers and dashed black line shows the original position of the layer boundary.

3.3.2 Internal dynamics of the fibril surface contribute to movements of

outermost monomers.

The internal motions of the fibril model are characterized by the root mean square deviations (RMSD) and root mean square fluctuations (RMSF) over the course of the MD simulation (Figure 3.5). To distinguish motions of the interaction surface, we have analyzed the surface and the core layers separately. We consider only the middle microfibril bundle along the b-axis since it is surrounded by its neighbors on all sides. In comparison of the RMSDs and RMSFs of the core and the surface layer, differences between the layers are interpreted as distinct motion in the surface layer relative to the interior of the fibril.

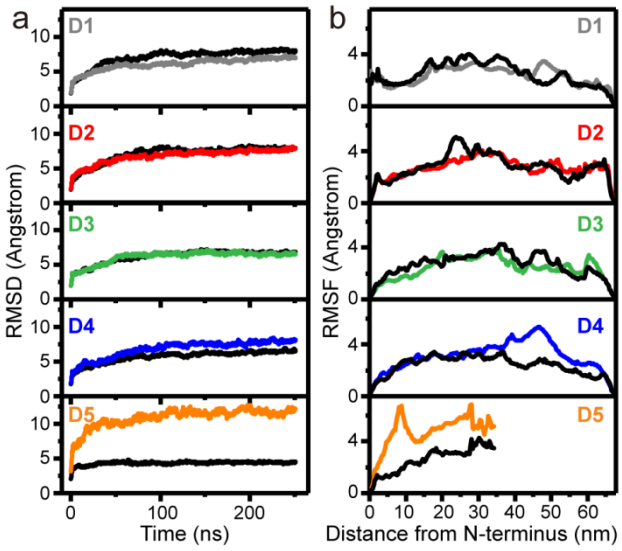

Figure 3.5 Internal motions within the fibril model. (a) Root-mean-square deviation (RMSD) during the 250-ns MD simulation of D-segments 1–5 within the core layer (black) and surface layer (color-coded as in Figs. 3.1–4). (b) Root-mean-square fluctuation (RMSF) of D-segments 1–5 over the time of the simulation. RMSF of surface layer

(colored) and core layer (black) are overlaid and aligned by distance from the N-terminus in the SRU.

The RMSD is a measure of the distance of atoms from the starting position. In Figure 3.5a, we plot RMSD vs. simulation time of the indicated D-segment in the core layer (black) and the same segment on the surface (color). The RMSDs of all segments in the core and surface layers roughly converge within about 100 ns. We observe the largest deviation in RMSD between the layers in D5 (orange), the frontline of the fibril surface. While the core layer is displaced only by ≈ 4 Å by the end of the simulation, the surface layer has a much greater displacement of ≈ 12 Å. This especially high RMSD of D5 in the surface layer is in agreement with the large movements in all directions of D5 captured in the snapshots in Figure 3.3. The D4 (blue) segment also shows a slightly higher RMSD (by ≈ 2 Å) at the surface than within the fibril core. In contrast, segments D2 and D3 do not deviate in RMSD on the surface compared to the core, and D1, the innermost D-segment from this surface, actually has a marginally lower RMSD on the surface. It should be noted that we put restraints on both termini of each D-segment, except for the N- and C-telopeptides, to mimic the covalent bonding to N- and C-terminal residues that are part of the same monomer in an adjacent D-period. These restraints introduce bias by limiting the motions of D-segments at the termini. For instance, in full collagen fibrils, the D4 C-terminus would be covalently bound

to a D5 segment in the subsequent D-period. Given that D5 shows high RMSD, its displacement could influence the bound D4 and enhance the D4 RMSD beyond what we have calculated here. These results suggest that the most substantial conformational changes to the fibril surface occur in the outermost two triple helices on the surface, D4 and D5 (Figure 3.5a).

While the RMSD is informative of the displacement of a particular region over time, the RMSF is a measure of the fluctuations of each residue during the entire MD simulation. To quantitatively characterize the most dynamic areas within the collagen fibril, we performed RMSF calculations after the system was equilibrated (Figure 3.5b) and compared the surface and core layers as in the RMSD analysis. The highest fluctuations in the surface layer relative to the core are again in D5 and in the gap region of D4, on the forefront of the interaction surface (Figure 3.5b). Given the extensive flexibility of the majority of the D5 segment, in the absence of restraints, the RMSF of its N-terminus is not likely to be static as suggested in Figure 3.5b. Instead, its dynamics could extend to the covalently bound D4 segment in the adjacent SRU, and potentially add to the fluctuations in the D4 C-terminus. No considerable differences in fluctuations are observed in the D1, D2, and D3 relative to the core layer. Together, the trends in RMSD and RMSF indicate that regions closest to the fibril surface are more dynamic and show sizeable movement over the course of the simulation compared to the core of the fibril.

3.3.3 Formation of protein–protein H-bonds supports tighter monomer packing in the surface reconstruction.

The formation of hydrogen bonds (H-bonds) is an important factor in protein folding and protein–protein interactions. By monitoring the number of protein–protein H-bonds with respect to time in the whole collagen fibril model, we observe that the protein–protein H-bonds increase by 0.169 H-bonds per residue, and protein–water H-bonds decrease by 0.142 H-bonds per residue within 250 ns of MD simulation (Figure 3.6a, b). The gain of protein–protein H-bonds with loss of protein–water H-bonds is consistent with increased monomer packing in the fibril. To determine the contribution of H-bond buildup due to monomer packing, we focused on trends of H-bonding between triple helices. Backbone atoms of different triple helices are distant enough that inter-triple helix backbone–backbone H-bonding is rare throughout the time course of the simulation (Figure 3.7). The build-up of the inter-triple helix sidechain-involved H-bonds, however, is indicative of the supermolecular packing. These H-bonds continue to increase during the entirety of the MD simulation (Figure 3.6c), not converging within 250 ns. At the same time, sidechain-involved H-bonds within triple helices remain constant. This formation of inter-triple helix protein–protein H-bonds with breakage of protein–water H-bonds supports tightening of the monomer packing within the model. Other factors, such as exposure of hydrophobic regions and distribution of

charges on the fibril surface may also accompany the reconstruction of the fibril surface.

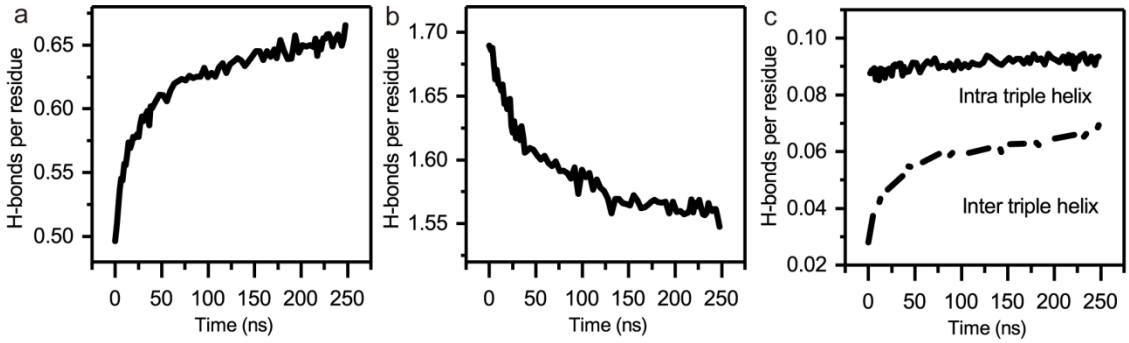

Figure 3.6 Hydrogen bond (H-bond) modulations in the MD simulation. (a) Protein–protein and (b) protein–water H-bonds per residue in the fully solvated collagen fibril model through the MD simulations. (c) Buildup of side chain involved intra- (solid) and inter- (dashed) triple helix protein–protein H-bonds per residue.

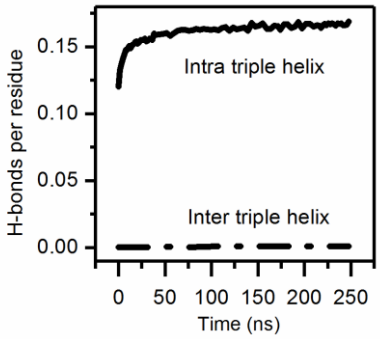

Figure 3.7 Buildup of backbone-backbone intra- (solid) and inter-triple helix (dashed) protein-protein H-bonds per residue in the 3a3b collagen fibril model through time.

3.3.4 Experimental AFM images topographic features of the type I

collagen fibril surface.

The MD simulation shows reconstruction of the type I collagen fibril that would result in changes to the surface topography. For example, the inward contraction of the gap and the outward expansion of the overlap region increase the depth between the highest point in the overlap region and lowest point in the gap region (Figures 3.3 and 3.4). These topographical features could be measured by AFM, which is an excellent tool to specifically probe physical and mechanical properties of the surfaces of materials and proteins. We imaged several type I collagen fibrils from reconstituted rat tail tendons in air adsorbed to mica. By analyzing the height profiles of the D-bands of an isolated fibril, we measured an average height difference between the overlap (peak) and the gap (valley) regions to be 4.1 ± 0.4 nm (Figure 3.8). The overall height and the relative heights of the overlap and gap regions are sensitive to environmental conditions[113,114]. Hydration induces swelling of collagen fibrils and may influence this step-height.

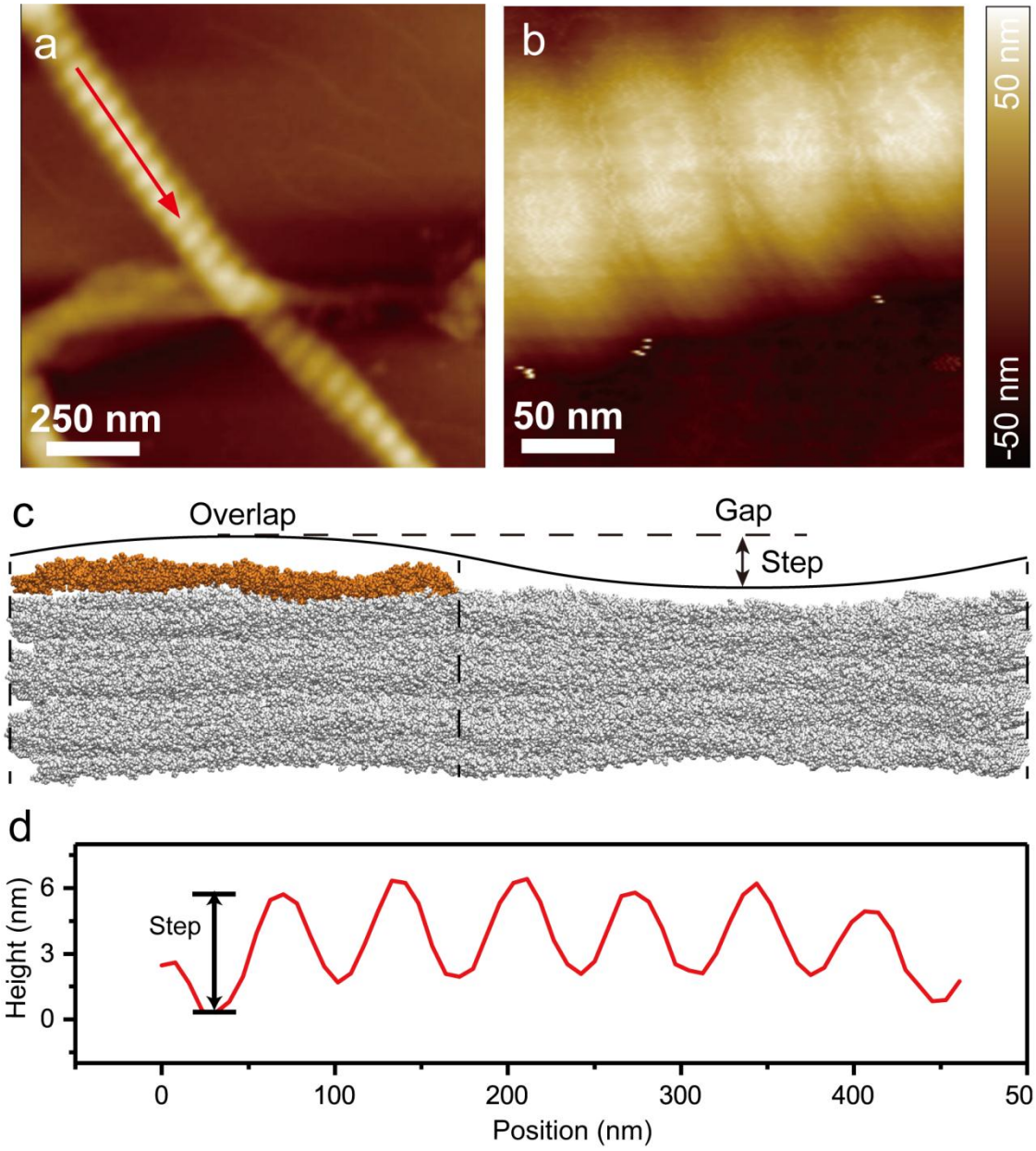

Figure 3.8 Measuring height difference between overlap and gap regions. (a-b) AFM height images of type I collagen fibrils with dimensions of (a) 2 µm × 2 µm and (b) 250 nm × 250 nm. (c) Schematic describing how the AFM height profile relates to the overlap and gap regions of the collagen fibril. D5 on the surface in the overlap region is colored orange. The step-height is the height difference between the peak of the overlap and the valley of the gap. (d) The height profile taken along the red arrow in (a). The height profile has a periodicity ≈ 67 nm, consistent with the D-period of the collagen fibril model.

3.3.5 Type I collagen fibril reconstruction has implications in the accessibility of ligand binding sites on the fibril surface.

Cell receptor and ligand binding sites on collagens have largely been determined on triple helical peptides, recombinant bacterial collagen constructs, and from imaging proteins binding to collagen monomers by microscopy methods[8-11,65,89-91]. In its monomeric form, all possible binding sites are exposed and available for interaction. However, when bundled into the supermolecular fibril, several of these binding sites become hidden from the interaction surface. To characterize accessibilities of binding motifs in the collagen fibril, we calculated solvent accessible surface area (SASA) around each residue of the fibril model, excluding surfaces on the interior of the fibril, i.e. those not accessible from the fibril interaction surface. To eliminate SASA of the interior, we used a spherical probe approximately one-half the size of the cavities within the gap region (with a radius of 8 Å). This probe size is still small enough to be sensitive to small deviations in SASA across the surface. Comparing the 8.0 Å SASA of residues in the surface layer of surface A in the starting model and at the end of the simulation, we find that fluctuations on the fibril surface substantially enhance accessibilities in parts of the fibril that are buried prior to reconstruction.

The modulation of the 8.0 Å SASA of surface A due to motions in the dynamic

surface is shown for three time points in the simulation (Figure 3.9a). In the starting structure, the 8.0 Å SASA was ≈ 0 $Å^2$ on the entire lengths of D1, D2 and D3 and in the overlap region of D4 (Figure 3.9a). Conversely, residues directly on the binding surface (D4 in the gap region and D5 in the overlap region) have considerably high 8.0 Å SASA since they are completely exposed. Consistent with the turn of the triple helix, even in the regions of high 8.0 Å SASA, the accessibilities drop to zero every three residues since the glycines of the $(G\text{-}X\text{-}X')_n$ repeating sequence point to the center of the triple helices and are not accessible from the interaction surface. Throughout the simulation, specific regions in D2 (30-50 nm) and D3 (57-65 nm) show substantial increase in SASA over time (Figure 3.9). D4 in the overlap region has fluctuating accessibilities over time, which is likely influenced by the dynamics of D5. Additionally, the N-terminal displacement of the C-telopeptide extends the accessible region on D4 at the junction of the gap and the overlap regions. Modulations in accessibility over time may provide a key as to how reconstruction of the type I collagen fibril surface may facilitate ligand binding.

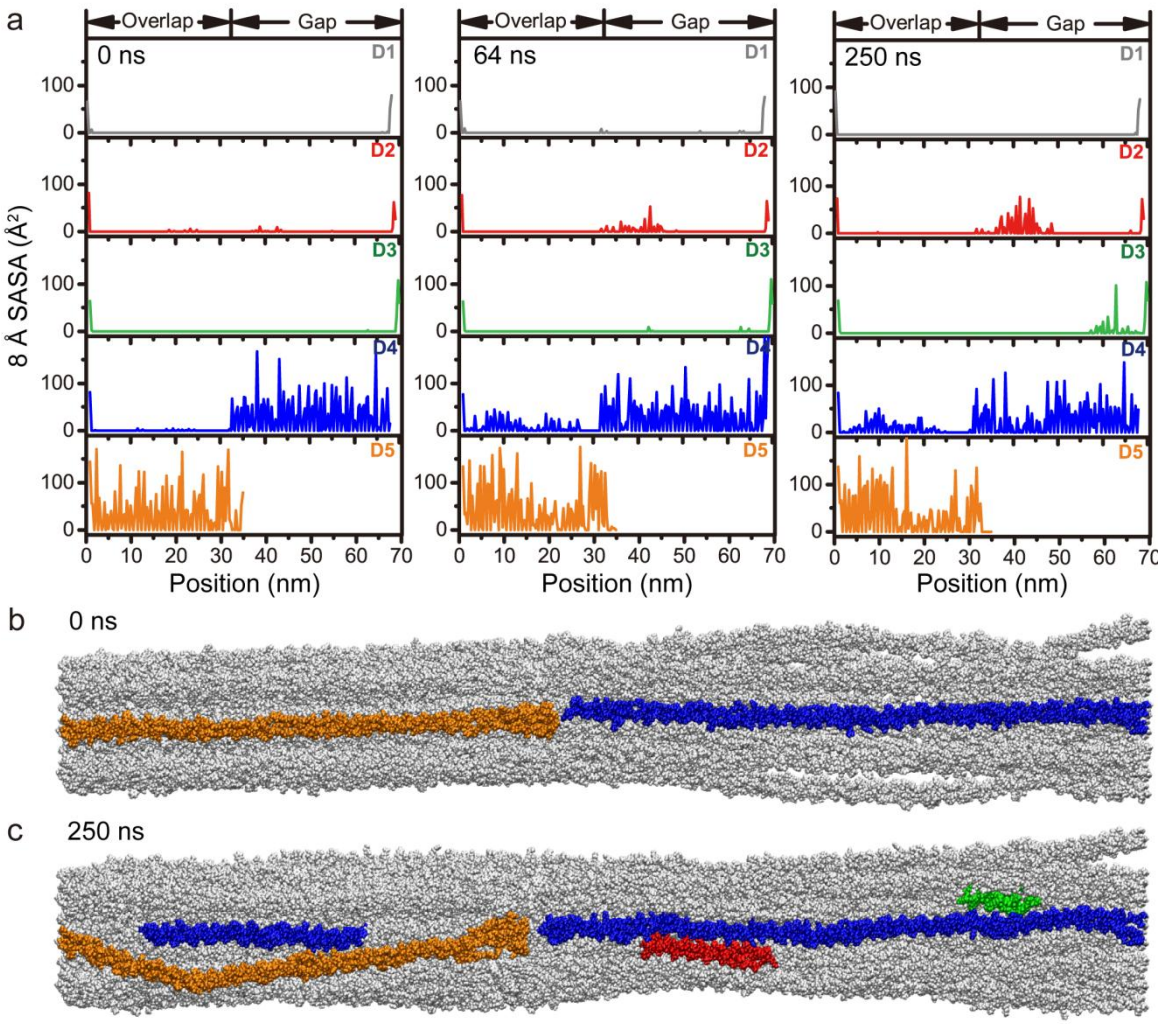

Figure 3.9 Accessibilities of the collagen fibril model. 8.0 Å SASA of "surface A" of the fibril model at 0 ns, 64 ns, and 250 ns time points of the MD simulation (a). (b-c) Longitudinal view of the fibril model highlighting residues with 8.0 Å SASA higher than 15 $Å^2$ in (b) the starting model and (c) at 250 ns. Although invisible in the starting model, D4 (blue) in the overlap region and parts of D2 (red) and D3 (green) in the gap region become accessible due to fibril surface reconstruction.

Through analysis of these SASA calculations, we find that fluctuations of the fibril surface reveal cryptic regions important for ligand interaction. In the overlap region, dynamics of the fibril surface open access to at least one of the

hidden major ligand binding regions on D4[1,86] (Figure 3.10a). This is a concentrated zone of molecular binding partners, including MMPs, collagen-binding integrins, discoidin domain receptors, heat shock protein 47, and fibronectin[1,86]. In this region, the SASA is not only seen to increase with longer simulation time, but has fluctuating accessibilities over time, especially in the region between 18-28 nm. We specifically probed an integrin αI domain binding site within this region at the edge of the overlap with the sequence GQRGER (Figure 3.10). In the starting model, this site is hidden by the C-telopeptide in D5 and has an 8.0 Å SASA value of ≈ 0 $Å^2$, and as a static structure, it would seem to be unavailable for binding. Monitoring the 8.0 Å SASA over time, we observe variability of accessibility at this site. By 120 ns, the binding site has maximum accessibility (Figure 3.10a). At this point, the C-telopeptide has translated longitudinally toward the N-terminus and unveils the integrin αI domain binding site, as shown in the cross-sectional view of Figure 3.10b. Although more open, dynamics within the fibril cause the 8.0 Å SASA to fluctuate, as seen by the lower accessibility at 250 ns (Figure 3.10a). Variable accessibility of specific regions of the fibril surface through time is consistent with a dynamic surface that is flexible to assume a multitude of conformations.

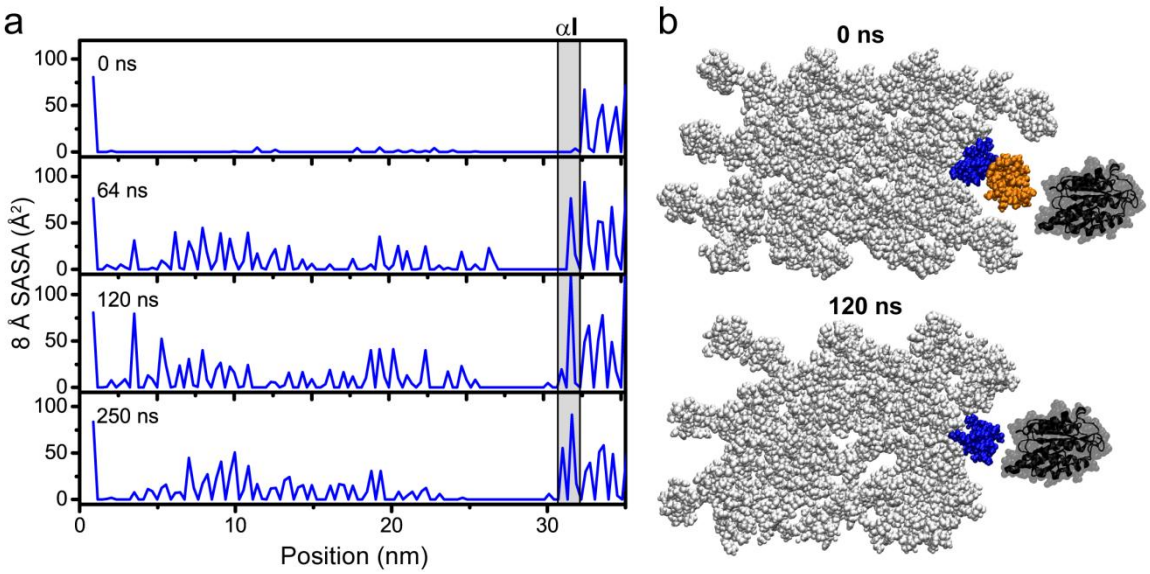

Figure 3.10 Measuring accessibility of a major ligand binding region on the D4-segment that contains an integrin αI domain binding site. (a) 8.0 Å SASA at 0 ns, 64 ns, 120 ns, and 250 ns MD simulation time of the overlap region of D4. The integrin αI domain binding site, GQRGER, is highlighted by the gray box. (b) Cross-sectional view of the GQRGER αI domain binding site on D4, nine residues deep along the longitudinal axis (gray box in panel A), at 0 ns (starting model) and 120 ns (maximal accessibility of this site). GQRGER on the D4-segment is blue and the neighboring C-telopeptide on the D5-segment is orange. The integrin α2I domain is shown in black (PDB ID: 1aox[57]). In the starting structure, the C-telopeptide immediately on the fibril surface is obstructing αI access to GQRGER. However, after 120 ns, the C-telopeptide is translated longitudinally, and out of the cross-sectional slice, allowing αI access to the GQRGER binding motif in the fibril.

3.4 Discussions

In this report, we present an early stage atomic model of the type I collagen

fibril surface. This is the first all-atom MD simulation of a collagen fibril model that allows discrimination of the binding surface from the interior. Although our simulation is certainly not fully equilibrated (and hence the description of surface reconstruction is incomplete), we have found that fluctuations at the interaction surface of the type I collagen fibril allow sampling of rare events on the hundreds of nanoseconds timescale. These fluctuations involve three major movements (Figure 3.3): (1) downward movement of D5, (2) inward contraction of the gap region, and (3) longitudinal displacement toward the N-terminus of the C-telopeptide. We analyzed the RMSD, RMSF, and H-bond perturbations throughout the 250 ns MD simulation. The RMSD and RMSF show that the outermost monomers on the surface of the fibril have greater spatial deviation from the starting structure and are more flexible than analogous regions in the fibril interior (Figure 3.5). We find that the fluctuations are accompanied by formation of new protein–protein H-bonds and disruption of protein–water H-bonds (Figure 3.6a, b). The number of protein–protein H-bonds on the surface increases over the timeframe of the MD simulation replacing protein–water H-bonds and expelling water from the surface in the process. This is similar to the collagen fibrillogenesis process, which is driven by the loss of water molecules from the protein surface[115-118]. The formation of protein–protein H-bonds optimizes supermolecular packing of monomers, as we observe an increase in sidechain–sidechain inter-triple helix hydrogen bonds (Figure 3.6c). The quantification of the H-bonding is consistent with our

observation that the surface layer of the gap region becomes more compact, tightening monomer interactions while displacing water in the surface reconstruction. This difference in monomer packing between the fibril surface and interior is consistent with an inhomogeneous assembly proposed by several researchers[119-122]. In particular, Gutsmann et al. suggested that the collagen fibril has a harder, denser surface layer and softer core[120]. They proposed that this inhomogeneous structure might be more resistant to bending and reduce deformation of collagen fibrils.

The reconstruction changes the surface topography of the fibril. AFM allows us to experimentally characterize the fibril surface in terms of height profiles, electrostatics, and mechanical properties; and it provides a complimentary tool to probe the MD fibril surface model. It should be noted however, that these computational and experimental methods provide information on different timescales, on the order of hundreds of nanoseconds in MD and seconds in AFM. At this stage, we have measured the overlap–gap step-height on the surface to be 4.1 ± 0.4 nm, as a starting point. It is interesting to note that this step-height is somewhat larger than would be expected based on the starting model for the MD simulation. Further work is being pursued to better understand the relationship between the MD simulations and AFM measurements. The topography of the collagen fibril is sensitive to environmental conditions, such as hydration, pH, and salt concentrations. Soft,

biological materials can now be imaged to sub-nanometer resolution in physiological buffers[123-125]. Current studies in our laboratory are working toward characterizing the collagen surface topography and mechanical properties in different environmental conditions. More recently, the evolution of high speed AFM technology has provided a way to study protein dynamics on millisecond timescales[126,127]. With these advancements, future studies will allow us to address biologically interesting questions about physical, electrical, mechanical, and dynamical properties of the collagen fibril surface.

We present the analysis of our MD simulation assuming that segments D4 and D5 are exposed to the surface. The exposure of this surface is based on fitting the type I collagen microfibril X-ray diffraction model to the corrugated profile of the fibril observed by SEM and AFM[99,100] and access to certain ligand binding sites, especially those of decoron and MMPs[20,22,98,101]. We would expect that in order for ligand binding to occur, the binding site should be accessible from the binding surface[20,98]. However, this does not necessitate that the binding site is open at all times given the now observed dynamics of the fibril surface. In our time-dependent SASA analysis, we find that cryptic sites that are unavailable in the static structure, become exposed due to the observed conformational fluctuations of the type I collagen fibril. As shown in the 8.0 Å SASA (Figure 3.9), the flexibility of D5 allows access to parts of D4 in the overlap region, and conformational changes in the gap region expose parts of D2 and D3. The

ability of the collagen fibril surface to sample these conformations, and not maintain a single, rigid conformation may enable or inhibit cellular processes through exposure of cryptic sites, such as those in the major ligand binding region in the overlap of D4, which houses binding sites for MMPs, secreted protein acidic and rich in cysteine (SPARC), discoidin domain receptor 2 (DDR2), phosphophoryn, fibronectin, and integrins. In this way, these ligand binding sites are cryptic, and the dynamics of the monomer on the collagen fibril surface provide a means of exposure. Additionally, previous MD simulations have shown that dynamics of the individual chains within the triple helices themselves play an important role in facilitating binding processes[128,129].

Our early model of the type I collagen fibril surface provides a new framework upon which future studies can build and that now allows us to address important biological questions, such as: What regulates the conformational transitions that provide access to cryptic binding sites? How are the surface dynamics perturbed by environmental conditions, such as pH and salt concentration? Are there slower timescale motions that may affect ligand binding? How might the reorganizations that occur on this timescale affect biological activity? Our simulation considers the surface layer of an isolated fibril. Further studies on the influence of environmental conditions or the impact of interacting molecules on the fibril dynamics may help to gain a better

understanding of the regulation of these dynamic processes. Our current study is limited to fast, nanosecond timescales accessible by MD. From this, we cannot deduce events that occur on longer timescales. However time resolved AFM may be instrumental in observing collagen fibril surface rearrangements and ligand binding in action.

3.5 Conclusions

Collagen fibrils are often presented as long, rigid rods that provide tensile strength to connective tissues, such as bones and tendons. However, within the ECM, they are also very biologically active, interacting with numerous cell receptors, enzymes, and ECM components to carry out critical cellular functions. Here, we ask how collagen binding partners are able to access their recognition motifs that are seemingly sequestered by the complex collagen fibril architecture of bundled triple helices. MD and AFM are excellent techniques by which we can characterize the collagen fibril surface. Through an all-atom MD simulation of a type I collagen fibril surface model, we show that the fibril surface is not merely a rigid rod, but is actually dynamic on the nanosecond timescale and samples conformations not observed in static models. Through reconstruction of the fibril surface, cryptic binding sites are unveiled for several collagen binding partners. Fluctuations of the C-telopeptide and D5 especially expose the major ligand binding region of D4, including an integrin αI domain binding motif, GQRGER. The observed

dynamics and reconstruction of the fibril surface promote its role as a “smart fibril” to keep certain binding sites cryptic, and to allow accessibility of recognition domains when appropriate. This suggests that through the transient availability of binding sites, collagen binding partners are able to interact with the collagen fibril to uphold their cellular functions. In addition, alternate conformations at the fibril surface expand possible drug targets against fatal collagen diseases.

Chapter 4 Intrinsic local destabilization of the C-terminus predisposes integrin α1 I domain to a conformational switch induced by collagen binding

The article cited in this chapter is: Nunes, A. M.; Zhu, J.; Jezioro, J.; Minetti, C. A.; Remeta, D. P.; Farndale, R. W.; Hamaia, S. W.; Baum, J. Intrinsic local destabilization of the C-terminus predisposes integrin α1 I domain to a conformational switch induced by collagen binding. @Protein science : a publication of the Protein Society 2016, 25, 1672.[130]

This work is available online at:

https://onlinelibrary.wiley.com/doi/full/10.1002/pro.2972

I am the second author on this paper. I performed the protein expressions and

purifications, NMR resonance assignments, protein secondary structure predictions and hydrogen-deuterium exchange experiments and analyzed the data. I partially wrote and edited this paper.

4.1 Introduction

Collagen interactions with α1β1 integrin receptors play a key role in numerous cellular processes, spanning cell development to differentiation and hemostasis to immune responses.[88,131,132] Integrin α1β1 is widely expressed in mesenchyme cells, the immune system, and a minority of epithelial tissues. [132] Functionally, α1 is one of four collagen binding I domains containing β1 partners that includes α2, α10, and α11. Upon binding to a collagen peptide, the I domain of integrin α1β1 (α1 I) undergoes conformational changes[57,59,60,62] similar to those observed in the α2 I domain, [57,59,60,62] in which coordination of a collagen glutamate to the metal ion-dependent adhesion site (MIDAS) at the top of α1 I induces a shift from the “closed” (unliganded) to “open” (liganded) conformation. This switch induces allosteric changes in the C-terminus of the Rossmann-fold structure, with helix αC unfolding and helix α7 displacing 12 Å downwards. The resultant movement permits coordination to a second allosteric site within helix α7 to the β-subunit and propagates these structural changes throughout the entire integrin macromolecule.[131]

Despite the critical functional importance of collagen-integrin interactions *in vivo*, the mechanism of integrin binding to collagen requires elaboration at the structural, dynamic, and thermodynamic levels. In an attempt to gain mechanistic insights into the I domain activation process induced by collagen, several “gain-of-function” mutants that result in enhanced binding to collagen have been studied.[56,64,65,67,133] Specifically, the α1 I mutant E317A has been characterized by high resolution X-ray crystallography[134] and numerous binding assays report an increase of binding to collagen. [133,134] The residue E317 is located at the top of helix α7 and is linked to helix αC (R287) via a salt bridge, a stabilizing feature of the closed unliganded form. The E317A mutation mutation eliminates this salt bridge and adopts a transition state between the α1 I closed-unbound form and the open-bound form. Indeed, an X-ray structure structure of the E317A mutant has revealed a novel conformation in which helix helix α7 is in the closed upward position, helix αC is unstructured (as observed in the open-bound form), and the metal ion is reported as adopting a unique penta-coordination. [134] Although the gain in function is primarily ascribed to structural changes that are implicated in reducing the ligand binding barrier relative to wild type protein, dynamics must be considered when describing the mechanism by which a gain of activity occurs.

There is increasing evidence that conformational flexibility assumes a central role in the molecular mechanisms of protein-protein interactions.[135-138] NMR is

well suited for studying the conformation and dynamics of proteins and provide information over a broad spectrum of timescales ranging from picoseconds to seconds and hours.[135,136] Conformational fluctuations on the micro to millisecond timescales have been shown to be critical determinants of biological processes in protein recognition and allosteric events.[135-139] In this study, we explore the role of slower timescale dynamic events that occur in the unbound free protein to unravel specific mechanisms that precede collagen binding. Hydrogen–deuterium exchange (HDX) kinetics measured via NMR affords the advantage of characterizing slow conformational fluctuations in denaturant free environments and thereby provides site-specific information on local stability in the native state.[140-144]

In the current investigation, NMR HDX experiments demonstrate the importance of slow motions in terms of predisposing integrin to conformational changes upon binding. Our NMR HDX results suggest a relationship between regions exhibiting a reduced local stability in the unbound I domain and those that undergo significant conformational changes upon binding. The αC and α7 helices of α1 I are at the center of such major perturbations and have reduced local stabilities in the unbound state, relative to other structural elements. Significantly, a combined energetic and structural characterization suggests that E317A activation and enhancement of collagen binding affinity are primarily dynamic in origin. The latter includes intermediary timescale motions

in helix αC and MIDAS, as well as propagation of slow conformational fluctuations to additional structural elements within the C-terminus. Our findings findings underscore the relevance of slow conformational dynamics, intrinsic to the free α1 I, and the concomitant reduction of local stability within regions of a conformational switch. The latter assumes a critical role in allosteric regulation presumably by decreasing the overall energetic penalty associated with ligand-binding interactions.

4.2 Materials and Methods

4.2.1 I domains expression and purification

The recombinant α1 I from human integrin α1β1 used for these studies corresponds to residues T141-E335. Protein purification was conducted as previously described,[64] but cells were grown in M9 media supplemented with $^{15}NH_4Cl_2$, and [$^{13}C_6$]-D-glucose and deuterated water was used when required to obtain isotopically ^{2}H, ^{13}C, ^{15}N labeled proteins. Protein concentration was determined via measurement of the absorption at 280 nm employing the respective molar extinction coefficients.

4.2.2 NMR spectroscopy

Spectra were acquired on a 700-MHz Bruker spectrometer equipped with a cryoprobe. Triple resonance experiments allowed assignments of 98 and 90% of the complete backbone resonances (195 residues) for wild type α1 I and

E317A/α1 I, respectively, and agree with previous NMR studies of wild type α1 I.[145] Chemical shifts of all assigned resonances of E317A/α1 I were deposited in the BMRB under accession number 26822. TALOS+[79] was used to estimate the secondary structure in solution based on the ^{13}C resonances. The analysis of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO resonances[146] reveals a comparable α1 I and E317A/α1 I secondary structure as in the crystal structures. [59,134,147,148] The chemical shift perturbation (Δ*δ*) of the backbone amides caused by mutation of E317A/α1 I relative to wild type α1 I was calculated by the equation: $\Delta\delta = \sqrt{((0.154\Delta N^2) + \Delta H^2)/2}$ where Δ*N* and Δ*H* correspond the chemical shifts difference between the E317A's and the wild type's α1 I amide nitrogens and protons, respectively.[149] All spectra were processed using nmr-Pipe[150] and Sparky.

4.2.3 Hydrogen–deuterium exchange (HDX)

The amide exchange experiments were performed at 20ºC for the ^{15}N-α1 I sample with a pD of 7.10 and at 25ºC for ^{15}N-E317A/α1 I with a pD of 7.4. Samples were lyophilized in 10 mM PIPES buffer containing 140 mM NaCl, 5 mM or 25 mM $MgCl_2$ and 1 mM DSS, with concentrations spanning the range 0.3-0.5 mM. A series of ^{1}H-^{15}N HSQC spectra of the D_2O sample were acquired every 10 min for 24 hours, followed by several spectra of 1 hour duration up to 2 months for α1 I and 1 month for E317A/α1 I. Considering the time required to setup and acquire the NMR spectra, the first time points for α1

I and E317A/α1 I following resuspension in D_2O were 15, and 20 min, respectively. HDX reaction of the amide proton is generally described by a two-step model between the folded (NH_{closed}) and unfolded (NH_{open}) states versus the exchanged ($NH_{exchanged}$) state. [151] The protection factor (P_f) of each amide proton was determined by a ratio of individual intrinsic rate constant (k_{int}) for the intrinsic chemical HDX reaction of the freely exposed amide group and observed rate constant of exchange (k_{obs}). The amide proton decays were monitored by plotting the peak intensities against the incubation times in order to obtain the k_{obs}.[80] Experimental uncertainties for k_{obs} were obtained from fitting errors. For residues with P_f too low or too high to be quantified under our experimental conditions, the k_{obs} values were estimated to be faster than k_{obs}= (-Ln (0.05)/t_{min}) s^{-1} for the lower limit residues, where t_{min} is the first time point of each experiment considering that more than 95% of the signal intensity change has occurred. The upper limit residues were estimated to be slower than k_{obs}= (-Ln (0.95)/t_{max}) s^{-1}, where t_{max} is the time point of the last spectrum acquired. Thus, HDX kinetics were estimated to be faster than 3.0×10^{-3} s^{-1}, and 2.3×10^{-3} s^{-1} for α1 I, and E317A/α1 I, respectively. The exchange in these residues is too fast for observation in the first NMR spectrum and too slow within the time frame of the experiments, thereby retaining their initial peak intensity. In contrast, the k_{obs} value for residues that do not exchange after the last experimental point were estimated to be slower than 1×10^{-8} s^{-1} and 2.1×10^{-7} s^{-1} for α1 I and E317A/α1 I, respectively. The k_{int} values were

calculated from the amino acid sequence utilizing the methods of Bai *et al* [81] and Connelly *et al* [82] and the program SPHERE (http://landing.foxchase.org/research/labs/roder/sphere/).[152] For α1 I, 5% of the residues were excluded from kinetic analysis due to unreliable data caused by severe resonance overlap. The exchange free energy of the amide protons was calculated from the equation $\Delta G_{HX} = RT \ln (P_f)$, where *R* is the gas constant and *T*is the absolute temperature at which exchange was monitored. Under extreme conditions where the exchange rate is much faster than the refolding rate, ΔG_{HX} estimated on the basis of EX2 limit would be larger than the actual value. Solvent accessibility was predicted using the program ASAView[153] inputting the crystal structure of the unbound I domain. $\Delta rC\alpha_{open/closed\ form}$ was calculated by measuring the Cα distance between the closed X-ray structure (PDB # 1pt6) and the averaged coordinates of each conformer of the open-bound NMR structure (PDB # 2m32).

4.2.4 Isothermal titration calorimetry (ITC)

Thermodynamic binding parameters for the association of Ac-$(GPO)_4$GLOGEN$(GPO)_4$GY-NH_2 collagen peptide (synthesized by Dominique Bihan at Cambridge University and LifeTein) with wild type α1 I and E317A/α1 I were determined via isothermal titration calorimetry employing a VP-ITC (MicroCal, Northampton, MA). Protein stock solutions were dialyzed exhaustively against a buffer comprised of 5 mM PIPES, 140 mM NaCl, and

100 mM $MgCl_2$ (pH = 7.3). The standard solutions were filtered using a 0.22 µm pore size membrane, and thoroughly degassed for 10 min. The calorimetric sample cell (1.4 mL) was filled with a 50 µM GLOGEN standard prepared in the final protein dialysate and the titration syringe (300 µL) contained a 500 µM solution of wild type α1 I or E317A/α1 I. Each titration experiment consisted of a 2.0 µL pre-injection followed by 30 consecutive 10.0 µL injections during which the reaction heats are monitored and integrated for 5.0 min. Binding isotherms were generated by recording the integrated heats normalized for α1 I concentration versus the protein:peptide ratio. A nonlinear least squares fit of the resultant profile to a single site binding model facilitates characterization of thermodynamic parameters for the protein:peptide. complex including the affinity (K_a), Gibbs free energy (ΔG), enthalpy (ΔH), entropy (ΔS), and stoichiometric ratio (n). Experimental uncertainties are expressed as fitting errors determined from least squares minimization using the Origin software program.

4.2.5 Resonance assignments

NMR peak assignments to obtain backbone $^{1}H_N$, ^{15}N, $^{13}C_O$, $^{13}C_\alpha$, $^{13}C_\beta$ chemical shifts were achieved using uniformly $^{2}H/^{13}C/^{15}N$-labeled samples and/or $^{1}H/^{13}C/^{15}N$-labeled (0.5-0.8 mM) in 50 mM phosphate (NaPi) buffer with 140 mM NaCl at pH 6.7 for α1 I and pH 7.0 for E317A/α1 I, containing 10 mM BME and 5 mM $MgCl_2$. The AutoLink program incorporated in CARA was used to

connect most of the intra-residue and sequential correlations of the peaks derived from TROSY versions of 3D HNCO, HN(CA)CO, HNCACB, and CBCA(CO)NH or HN(COCA)CB spectra at 25ºC. Spectra were acquired on a 700 Bruker spectrometer equipped with a cryoprobe. Chemical shifts were referenced with external DSS. Spectra were processed using nmr-Pipe and Sparky.

4.2.6 Characterization of GLOGEN by CD spectroscopy

The collagen-like triple-helical peptide (THP), Ac-$(GPO)_4$GLOGEN$(GPO)_4$GY-NH_2 (GLOGEN) was synthesized by LifeTein and by Dominique Bihan at Cambridge University and purified as previously described with its identity verified by mass spectroscopy. The peptide concentration was determined by monitoring absorption of the Tyr residue at 280 nm using an extinction coefficient of 1280 $M^{-1}cm^{-1}$ for GLOGEN in its monomeric form. The lyophilized peptide was dissolved in water and equilibrated at 4°C for a minimum of 24 hrs to ensure complete trimerization. The characteristic collagen triple helix maximum ellipticity monitored at 224 nm on an AVIV Model 400 circular dichroism spectrophotometer (Aviv Biomedical, Lakewood, NJ) confirmed the THP conformation. GLOGEN standards were prepared in a buffer system comprised of 5 mM PIPES and 5 mM $MgCl_2$ (pH 7.4) at a concentration of 100 μM monomer, Wavelength scans were recorded over the range of 190 – 260 nm at 5 °C employing an averaging time of 10 sec

and 0.5 nm intervals. GLOGEN stability was assessed via temperature-dependent measurements recorded at 224 nm employing a constant heating rate of 1.0 ºC·min^{-1} and 0.1 ºC increments. .Analysis of the resultant temperature-dependent profile yielded a transition midpoint of ~50 ºC for GLOGEN secondary structure stability.

4.2.7 α1 I-GLOGEN complex by NMR

The α1 I-GLOGEN complex was performed using two different constructs of human α1 I, a longer construct corresponding to residues T141-E335 (wild type α1 I) and a shorter construct comprising the T141-I331 residues (α1IΔC). The longer construct was used for the NMR titration experiments, by monitoring the ^{15}N-labeled wild type α1 I peaks on the ^{15}N-^{1}H-TROSY spectra with the incremental addition of unlabeled GLOGEN THP, reaching a THP:α1I molar ratio of 2:1. Spectra were acquired on a Bruker 600 MHz at 25ºC with 0.71-0.32 mM ^{15}N-labeled wild type α1 I in TBS buffer, containing 5 mM $MgCl_2$ and 20 mM BME at pH 7.4.

Chemical shift perturbation of α1 I backbone amides caused by the conformational switch, closed and open-bound state, was determined on a 1 mM ^{15}N-labeled α1IΔC sample in 5 mM PIPES buffer containing 25 mM $MgCl_2$ and 140 mM NaCl, monitored in the absence and presence of 2 mM GLOGEN THP. ^{15}N-^{1}H-TROSY spectra was acquired on a Bruker 700 MHz at 30ºC.

4.2.8 I domain adhesion assay

Wild type α1 I and E317A/α1 I adhesion to type I collagen from rat tail (BD Biosciences) and type IV collagen from human placenta (Advanced BioMatrix) was determined colorimetrically as described previously. Immulon 2HB 96-well plates (Thermo Scientific) were coated with 10 μg/ml collagen overnight at 4 ºC and blocked for 1 hr with 5 % BSA solution. The washing and adhesion buffer consisted of 5 mM PIPES and 140 mM NaCl in the presence of 5 mM $MgCl_2$ or 5 mM EDTA. Following three washings, binding was achieved by adding 10 μg/ml wild type α1 I or E317A/α1 I to the wells and incubating for 1 hr at room temperature (RT). The binding was detected by first incubating the mouse anti-α1 I monoclonal antibody (Millipore) with a 1:2000 dilution for 1 hr at RT, followed by incubation for 30 min with a 1:5000 diluted goat HRP-conjugated anti-mouse IgG antibody (GenScript) at RT, incorporating washing steps between each antibody addition. Color was developed using a TMB Substrate Kit (Pierce) according to manufacturer's instructions. Experiments were performed in triplicate with the data corresponding to mean values ± standard deviation.

4.3 Results

4.3.1 HDX NMR experiments reveal conformational fluctuations in the I domain of human α1β1 integrin

HDX measurements of α1 I integrin provide detailed information regarding conformational fluctuations based on the amide NH protection factors (P_f). [140,144] Well-dispersed resonances in the [^{1}H-^{15}N]-TROSY spectrum (Figs. 4.2A and 4.6A) are characteristic of a well-folded protein.[59] HDX measurements for 186 non-overlapping residues in α1 I reveal 32% fast-exchanging amide protons with low P_f, and 23% slowly exchanging residues with high P_f (Fig. 4.1A). The remaining 45% exhibit a time-dependent decrease in peak intensity and their observed exchange rate constants (k_{obs}), P_f, and free exchange energies (ΔG_{HX}) are summarized in Appendix Table A.1.

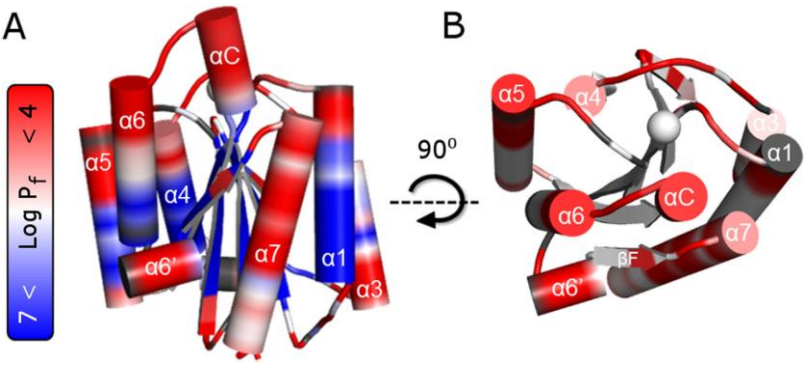

Figure 4.1 Hydrogen–deuterium exchange of α1 I integrin. The logarithmic value of protection factors (P_f) obtained from the HDX exchange rates, are mapped in the representation of **(A)** wild type α1 I (PDB #1pt6) in the closed-unbound form. The residues that do not exchange with solvent for over two months exhibit high Pf (Log $P_f > 7$) and are colored in blue, whereas residues that are highly dynamic (i.e., exchange with solvent so fast that these disappear from the NMR spectra after 20 min) exhibit reduced Pf (Log $P_f < 4$) and are colored in red (refer to color bar). **(B)** Top view highlighting the flexibility of residues located at the top of collagen binding site in α1 I. Note that the blue color is removed for ease of

visualization. Unassigned or overlapped peaks are colored in gray and the metal is represented as a sphere.

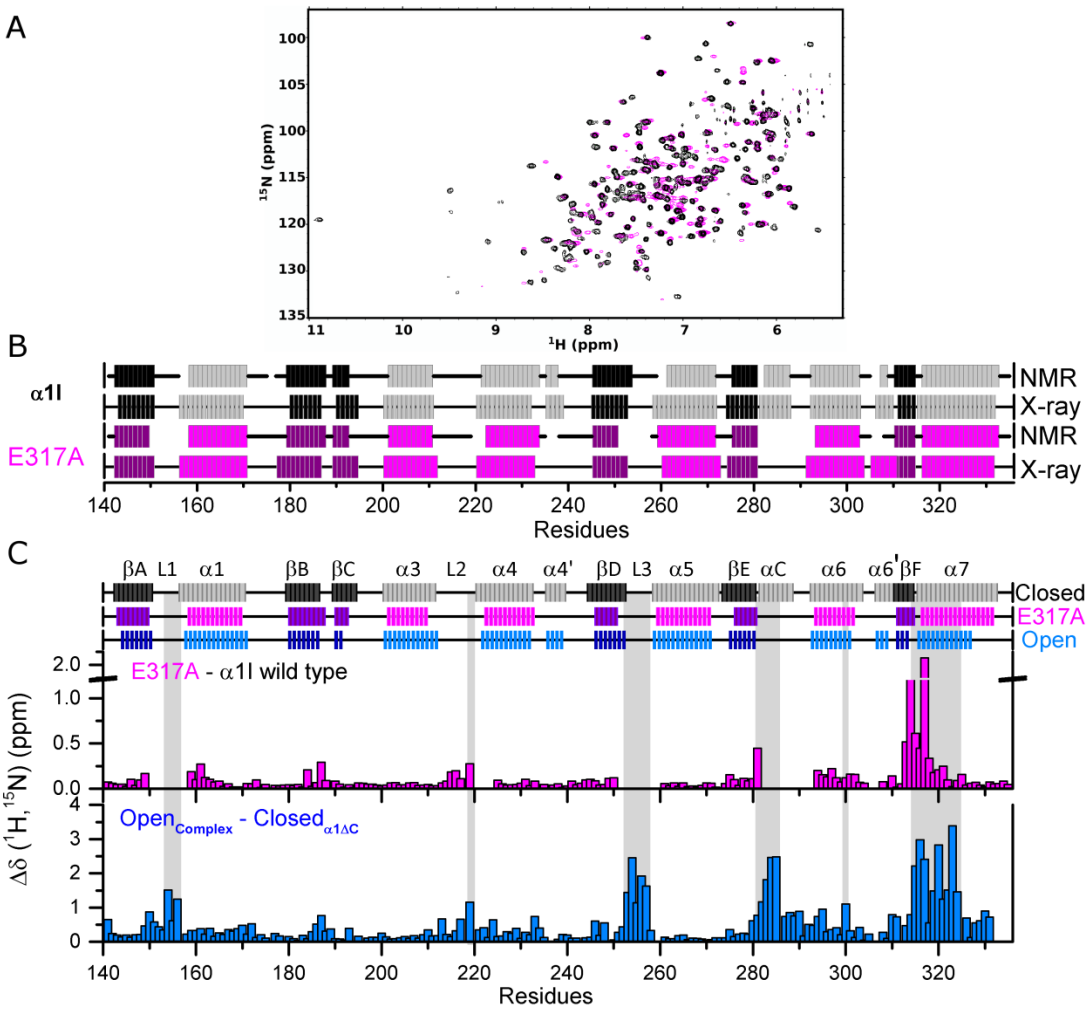

Figure 4.2 Impact of an E317A/α1 I activating mutation and GLOGEN complexation on α1 I peptide backbone. (A) Overlaid ^{1}H-^{15}N-TROSY spectra of wild type α1 I (black) and activating mutant E317A/α1 I (magenta). Spectra were acquired on a Varian 600 MHz at 20 ºC. Differences in the resonance distribution of ^{1}H-^{15}N-TROSY spectrum between E317A mutant and wild type α1 I correlate with unique conformations observed in the E317A/α1 I X-ray crystal structure. (B) Comparison of wild type α1 I and E317A/α1 I secondary structure in solution and crystal. Secondary structure elements predicted via TALOS+ based on the ^{13}C

resonance assignments of wild type α1 I (black/gray) and E317A/α1 I (purple/magenta) with the secondary structure deduced from X-ray studies reveal similar solution and crystal structures. Unassigned residues are represented by the lack of NMR data. (C) Chemical shift perturbation (Δδ) of the wild type α1 I backbone $^{1}H_{N}$ and ^{15}N atoms caused by mutation of the E317A residue (top plot) and by conformational switch (bottom plot). The secondary elements of X-ray closed α1 I structure (black, pbd #1pt6), X-ray of E317A transitional structure (magenta, PDB #4a0q) and NMR α1 IΔC-GLOGEN open-bound structure (blue, PDB # 2m32) are represented at the top of the plot with β-strands and α-helices in darker and lighter colors bars, respectively. Please note, that the plots present different scales on the y axis. Shaded gray regions highlight the residues with larger chemical shift perturbations (Δδ > 1 ppm) between the closed and open forms. The large structural rearrangement occurs in the MIDAS loops (L1, L2, and L3) and in the C-terminus residues (βF-strand and αC, α6, and α7 helices). The chemical shift perturbation caused by an E317A mutation is much smaller than that observed during the conformational switch and primarily surrounds the mutation site (βF-strand and top of helix α7). The small changes for most residues of helix α7 indicate a similar chemical environment to wild type α1 I in the unbound closed form. The MIDAS loops and helix αC resonances are not observed in the NMR spectra, presumably due to resonance broadening caused by microsecond dynamics.

There are four major observations that can be deduced from the HDX kinetics data of α1 I (Fig. 4.1A, Table A.1). First, α1 I contains a highly protected β-sheet core with amide protons that exhibit P_f values of 10^6−10^7 [Log P_f = 6–7

in Fig. 4.1A] comparable in magnitude to those observed for typical 10–20 kDa proteins.[154,155] Second, residues located in the upper face possess lower P_f values with faster HDX exchange rates than residues in the lower face (Fig. 4.1B). Most of these fast-exchanging residues comprise unstructured loops and α-helical structures with the exception of helix α3. Third, highly protected loops of α3-βC and α6-βF are located in the bottom face (P_f of 10^5−10^6 compared to 10^2−10^4 in other loops), consistent with the existence of a hydrophobic intramolecular pocket observed in other I domains.[156-159] Fourth, fast HDX rates are observed in helices αC and α7. Indeed, these residues present remarkably low P_f values for a helical conformation (10^4−10^5). In fact, only G283 in helix αC ($G^{283}SYNR^{287}$) exhibit measurable P_f values on the order of 10^5. These two helices are known to undergo a major structural rearrangement upon complexation (Fig. 4.3A).[160,161] Thus, the data suggest a relationship between fast HDX rates in the unbound α1 I and residues that undergo significant conformational change upon complexation. HDX can be interpreted in terms of the free exchange energies or local stability (ΔG_{HX}) by assuming an EX2 limit,[151] whereby the conformational equilibrium between unfolded and folded states is much faster than the intrinsic exchange rate. HDX will be discussed within the context of local stability (ΔG_{HX}) in subsequent sections of this manuscript.

Low local stabilities derived from HDX experiments can arise from increased

solvent accessibility. We have therefore investigated the relationship between local stabilities derived from HDX experiments, solvent accessible surface areas (SASA), and conformational changes in the α1 I structure (ΔrCα) induced by collagen binding (Fig. 4.3B). High Δ*r*Cα values correspond to large conformational differences between the unbound-closed and bound-open forms of α1 I. In the overlay of both α1 I forms (Fig. 4.3), helix α7 translocates Å downward upon ligand binding and helix αC becomes an unstructured loop.[160] Solvent accessible surface areas (SASA) derived from analysis of the three-dimensional structures measure solvent accessibility to an amide within the protein. Residues characterized by large SASA values are highly accessible to solvent and usually exhibit low P_f values (e.g., loops), while residues with low SASA values are buried from solvent. The most striking feature illustrated in this figure is the presence of residues (highlighted by drop lines) with low SASA that present low P_f values represented by reduced local stabilities (ΔG_{HX}), all of which correspond to residues with the highest displacement (high ΔrCα). These residues are confined to the C-terminus of α1 I (Fig. 4.3B) and are primarily located in helices αC and α7.

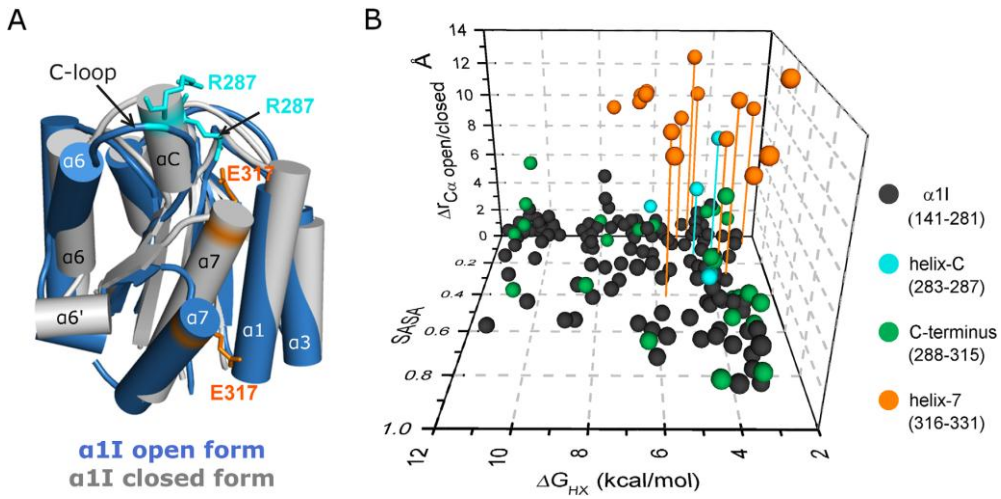

Figure 4.3 Relationship between conformational change upon complexation for α1 I residues, local stability, and solvent accessibility (SASA). (A) The X-ray α1 I closed-unbound structure (gray, PDB #1pt6) is overlaid with the first conformer of the GLOGEN bound open α1 I NMR structure (blue, PDB #2m32). The stick representation of E317 (orange) and R287 (cyan) residues highlight their distinct positions before and after collagen binding. (B) Correlation between α1 I free energy of exchange (ΔG_{HX}) obtained from the protection factors, solvent accessible surface area (SASA), and distance between the alpha-carbon ($\Delta rC\alpha_{open/closed\ form}$) in the closed and open structures of α1 I taken from the X-ray[6] and NMR[7] structures. The C-terminal residues are colored in green except for residues assigned to helix-7 (orange) and C-helix (cyan). Drop lines highlight residues that have $\Delta rC\alpha > 5$ Å, $\Delta G_{HX} < 6$ kcal/mol and SASA < 0.5. Helices that have unusually low local stabilities (i.e., αC and α7 helices in the closed form) undergo large structural changes upon binding to GLOGEN.

4.3.2 Activated E317A α1 I mutant is more dynamic than its wild type counterpart

Our NMR data, using an E317A/α1 I sequence construct that is similar to the X-ray study, reveals 90% of the expected resonances (independent of the experimental conditions). The majority of missing resonances correspond to residues located in the L1 and L3 loops of MIDAS and helix αC and are attributed to conformational exchange dynamics on the micro-millisecond timescale (Figs. 4.2B, C. For the resonances that are observed, ^{13}C chemical shift analysis[146] suggests that E317A/α1 I retains a secondary structure in solution which is identical to the crystal structure[134] (Fig. 4.2B) with the exception of helix αC and the MIDAS loops for which the NMR resonances are missing. Chemical shift difference spectra between E317A and the wild type protein support the notion that helix α7 is positioned upward as it is in the closed closed state (Fig. 4.2C). These results do not agree with previous NMR studies[160,162] that reported α7 in a downward position as in the open-bound form. This discrepancy may arise from the absence of reported peak assignments or from a different protein construct used in previous NMR studies studies relative to the crystal structure[134]and our present NMR studies. The Scanlon group uses a construct exhibiting a truncation of the last three residues residues in helix α7 with the sequence terminating at I331.[160]

E317A/α1 I HDX experiments indicate that 31 and 20% of the residues are in fast and slow exchange, respectively and explicit kinetic parameters have been been determined for 34% of these residues (Table A.1). Comparison of the

HDX parameters for E317A/α1 I and α1 I (Figs. 4.1A and 4.4A) indicate similar tendencies overall to those observed for α1 I; however, there are local regions in E317A that exhibit a decrease in P_f values. A more quantitative comparison of ΔG_{HX} reveals that the HDX of the beta-sheet core residues in α1 I and E317A E317A do not change appreciably, yet regions important for conformational rearrangement from the closed to open form exhibit both a lower average and narrower distribution of ΔG_{HX} (Fig. 4.4B). This is particularly significant in helix α6, strand βF, and helix α7, with the activating mutant in helix α7 exhibiting a lower and narrower distribution of ΔG_{HX}. Indeed, the C-terminus secondary-structure elements present an overall destabilization of 5.6 kcal·mol^{-1} relative to wild type α1 I. The βF-strand ΔG_{HX} median value is 4.0 kcal·mol^{-1} less stable in the mutant, while helices α6 and α7 are destabilized by by 0.7 and 0.9 kcal·mol^{-1}, respectively.

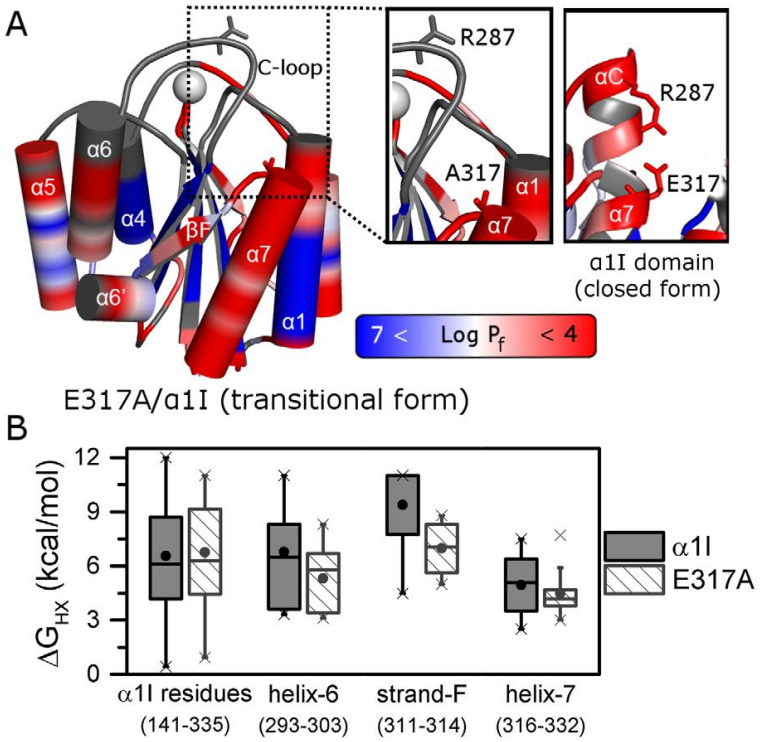

Figure 4.4 Hydrogen–deuterium exchange of α1 I activating mutant, E317A. (A) The logarithmic value of protection factors (Log P_f, refer to color bar) is mapped onto the crystal structure of the activated E317A/α1 I (PDB #4a0q) mutant. Unassigned or overlapped peaks are colored in gray, and the metal is represented as a sphere. A close-up of E317 and R287 residues comprising the salt-bridge in the wild type (right panel) appears on the left for E317A. (B) Boxplot diagram representing the dispersion of local free energy of exchange for amide protons (ΔG_{HX}) of α1 I wild type (gray) and E317A (hatch white). This box represents the statistical distribution between 25% and 75% of the full sequence and the C-terminus with helices α6, α7, and βF strand plotted separately. The mean value is depicted by a ball and the whiskers represent maximum/minimum values of all data.

4.3.3 Energetics of wild type α1 I and E317A interactions with a collagen model peptide

Full-length collagen sequences contain multiple binding sites exhibiting variable integrin affinities[11,56,163] that effectively precludes an accurate energetic characterization. Synthetic collagen-like triple-helical peptides (THP) represent a suitable model for biophysical [60,161] Chin, 2013 #201,164-167 and functional studies[11,56,163] of collagen. In an effort to elucidate the forces that promote α1 I binding to collagen and the origins of E317A enhanced activity observed in adhesion assays toward collagens I and IV (Fig. 4.5), we have employed ITC to derive complete thermodynamic binding profiles for association of Ac-$(GPO)_4$GLOGEN$(GPO)_4$GY-NH_2 (GLOGEN) THP with wild type α1 I and the E317A mutant. We specifically selected this high affinity GLOGEN binding sequence[163] to provide structural, dynamic, and thermodynamic insight on the α1 I-collagen interaction.

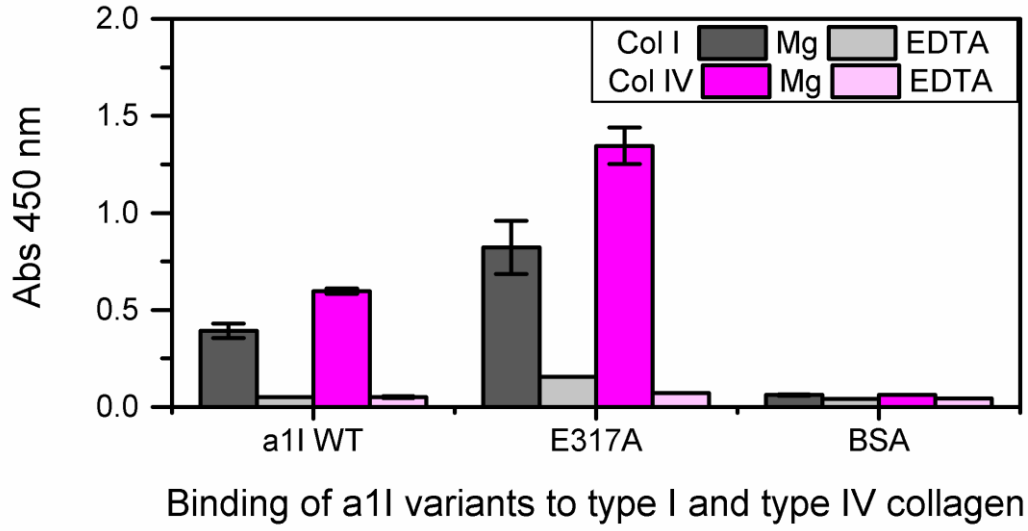

Figure 4.5 Adhesion of wild-type α1 I and gain-of-function E317A/α1 I mutant to type I and IV collagen. The recombinant wild type α1 I and E317A/α1 I was incubated in wells coated with type I collagen extracted from rat tail and type IV collagen extracted from human

placenta, and the adhesion was measured in the presence of 5 mM $MgCl_2$ (dark bars) and 5 mM EDTA (lighter bars). BSA served as a control surface coating. Experiments were performed in triplicate with the data corresponding to mean values, and error bars representing the standard deviation. E317A/α1 I mutation increases the affinity towards collagen relative to the wild type α1 I retaining the specificity of α1 I to type IV over type I collagen.

Association of both I-domains with THP is characterized by formation of a 1:1 stoichiometric complex with favorable enthalpic and entropic contributions (Table 4.1). The Gibbs free energy of - 5.98 kcal·mol^{-1}determined for the α1 I-GLOGEN interaction yields a dissociation constant of ~ 20 μM that corroborates solid phase binding studies.[163] While wild type interactions are predominantly entropy-driven, peptide association with the mutant is characterized by nearly identical enthalpic and entropic components. The E317A activation of α1 I is accompanied by an approximately threefold enhancement in binding affinity, which is the result of a more favorable enthalpy (i.e., $\Delta\Delta H = -1.88$ kcal·mol^{-1}) and a concomitant loss in entropy (i.e., $\Delta T\Delta S = -1.32$ kcal·mol^{-1}). This characteristic enthalpy–entropy compensation yields a small net gain in the overall Gibbs binding free energy ($\Delta\Delta G$) of −0.56 kcal·mol^{-1}. Direct HDX characterization by NMR of the wild type α1 I-GLOGEN complex is not possible as the 20 μM dissociation constant determined for α1 I-GLOGEN interaction positions the interconversion between bound–unbound

forms in an intermediate timescale, which results in resonance broadening within the NMR spectra of the complex (Fig. 4.6). Nevertheless, the regions of α1 I most broadened by the addition of GLOGEN include the binding interface (particularly MIDAS, Fig. 4.6C,D) and the regions of structural rearrangement (primarily the C-terminus, Fig. 4.6C,E).

Table 4.1 Thermodynamic binding parameters derived from calorimetric measurements of the wild type and activating mutant E317A/α1 I titrated into GLOGEN at 5.0°C

Integrin	K_d (μM)	K_a 10^5 (M^{-1})	ΔG (kcal·mol^{-1})	ΔH (kcal·mol^{-1})	TΔS (kcal·mol^{-1})
α1I	19.9±0.50	0.50±0.01	-5.98±0.15	-1.32±0.02	+4.66±0.10
E317A/α1I	7.25±0.13	1.38±0.02	-6.54±0.02	-3.20±0.02	+3.34±0.05

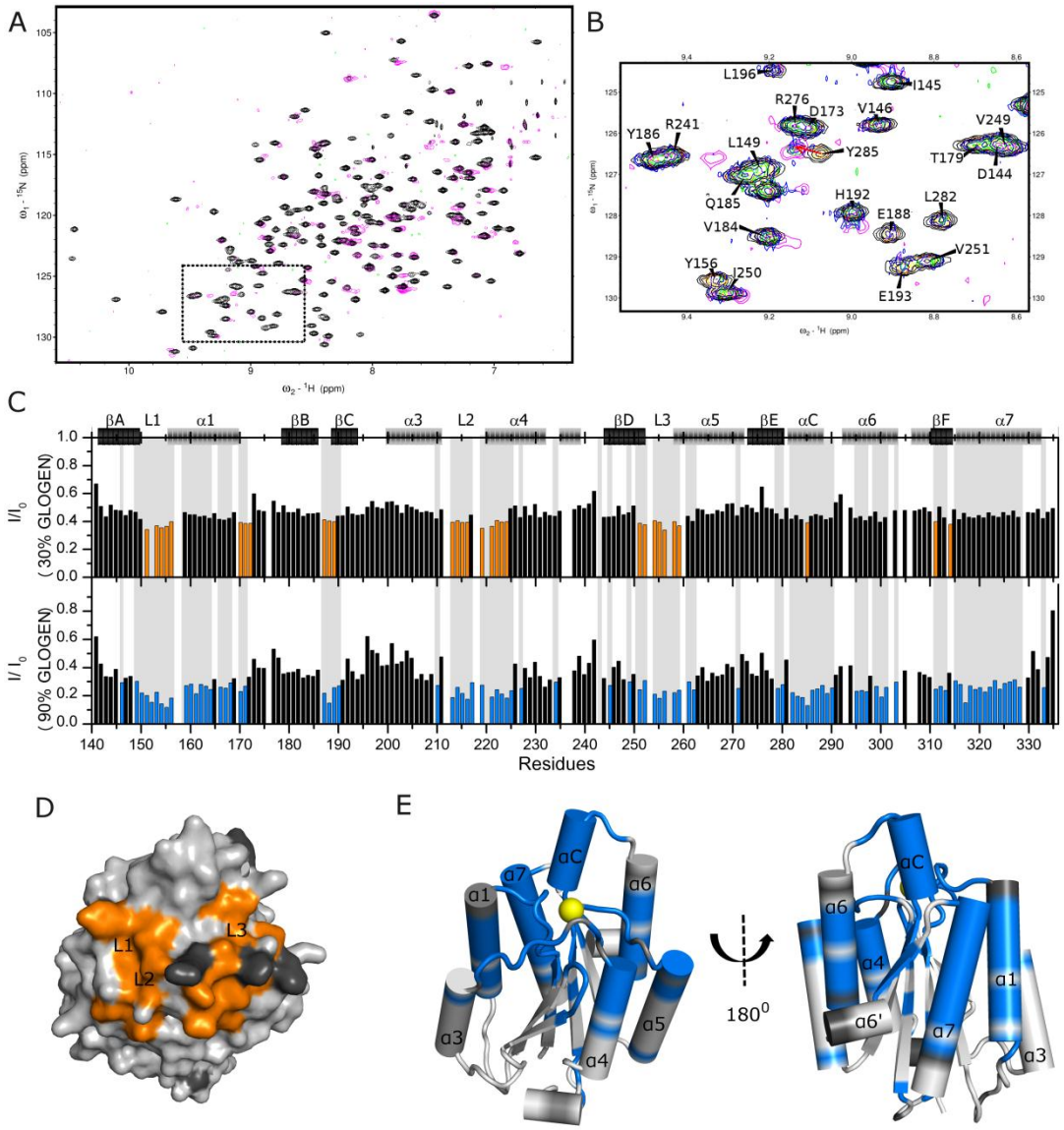

Figure 4.6 α1 I-GLOGEN complex by NMR. Complex formation between wild type α1 I and GLOGEN collagen model peptide leads to an overall and selective peak intensity loss as observed in: (A) The overlaid ^{1}H-^{15}N TROSY spectra of free α1 I (black) and a 1:2 complex of α1 I bound to GLOGEN (magenta) in TBS buffer at pH 7.4. (B) The ^{1}H-^{15}N TROSY overlay of the boxed region in (A) without GLOGEN (black), and in the presence of 30 % (orange), 90 % (blue) and 200 % GLOGEN highlights the residue specific peak intensity loss as a function of peptide concentration. The arrow highlights the chemical shift change of Y285 over 30 % GLOGEN. All spectra were acquired on a Bruker 600 MHz, at 25 ºC. (C) Intensity loss plot

upon addition of 30 % and 90 % GLOGEN to α1 I. The most affected residues exhibiting an intensity drop greater than 80 % (I/I_0 lower than 0.4) upon addition of 30 % GLOGEN are colored orange and mapped in the surface representation of the closed form deduced from the α1 I X-ray structure (PBD #1pt6) (D), while the 70 % (I/I_0 lower than 0.3) intensity drop caused by the addition of 90 % GLOGEN is highlighted by the blue color of the residues in the plot and in the cartoon representation of α1 I (E). These residues represent the binding interface and undergo conformational rearrangement. At high THP concentrations, the MIDAS loops and αC, α1, α4, α6, and α7 helices are regions most affected by GLOGEN binding. Regions colored in gray represent residues for which it is not possible to determine the loss of intensity due to extensive peak broadening or lack of assignment.

4.4 Discussion

The mechanisms that regulate α1 I ligand affinity of integrin and its intrinsic ability to undergo structural rearrangements upon binding to the rigid rod-shaped collagen remain unclear. Conformational fluctuations have been proposed as a driving force for several binding and allosteric events.[135-138] The majority of these studies have focused on dynamic fluctuations occurring on the micro- to millisecond timescale[135,139,168] in contrast to the impact of slower timescales on conformational changes. Much of what is known about allosteric activation of integrins caused by collagen binding has been derived from structural studies of both α1 and α2 I domains.[59,60,134,147,160] These studies have captured static conformational “snapshots” of the collagen binding α1 I in

three different conformations: a closed-unbound structure determined by X-ray,[59,147] a transitional conformation adopted by the gain-of-function E317A/α1 I mutant determined by X-ray, [134] and an open-bound form determined by solution NMR with a complexed GLOGEN THP. [160] The first glimpse of α1 I slow dynamics originated from a HDX mass spectroscopy study on a α1 I rat-human chimera and provided important insights on the impact of different ligands on integrin activation.[169] To assess the role of intrinsic conformational fluctuations on the structural switch, our studies focus on elucidating the dynamics of unbound human α1 I integrin and its activating E317A mutant at the atomic level by HDX NMR spectroscopy in combination with ITC to evaluate the binding energetics. The resultant data suggest that slow timescale motions may be an integral determinant of αI-domain propensity to undergo significant structural changes upon collagen binding.

4.4.1 Intrinsic local destabilization of a1 I in theunbound-closed form facilitates a conformational switch to the open form

The difference in local stability as seen by HDX experiments between the upper upper and bottom face of the α1 I is quite significant, with the region that binds collagen,[160,169] at the top of α1 I, undergoing much faster exchange rates (Fig. 4.1B). We propose that this increase in plasticity of the collagen binding interface might be required to allow the globular I domain to incorporate the rigid rod-like shape of the collagen structure. These slow timescale motions

probed by HDX experiments are important for collagen recognition mechanisms in matrix metalloproteinases[165] and may thereby assist in optimizing α1 I-collagen interactions.

The unusually low local stability of αC and α7 in wild type α1 I reveals that these these helices undergo a local breathing or accordion-like motion. Notably, these helices are at the center of major structural rearrangements in the conformational switch of α1 I. In the complex, helix αC becomes unfolded and helix α7 exhibits a major displacement of 12 Å. The highly dynamic character of of helix α7 is consistent with NMR and molecular dynamics studies of other I domain integrins.[170-173] The data suggest a correlation between residues prone for allosteric movement and reduced local stability despite low solvent accessibility. We hypothesize that local destabilization of structural elements in the unbound I domain facilitates the conformational rearrangement induced by collagen.

4.4.2 Enhanced dynamics of unbound E317A/a1 I contributes to gain of functionality

To test our hypothesis we used an activated mutant of α1 I, E317A, that has been characterized by X-ray crystallography[134] and affinity assays.[133,134] It provides mechanistic insight into the α1 I activation process due to its increase of affinity toward collagen and its transitional conformation.[134] Here, we

hypothesize that the increased binding affinity is also the result of higher local destabilization of the regions of conformational rearrangement.

NMR HDX measurements described in this study reveal that the E317A dynamics are more complex than wild type α1 I, particularly in the region of αC and MIDAS loops. Helix αC is a unique structural element of collagen-binding integrins that is proposed to be a determinant of selectivity for collagen[60,174,175] by functioning via steric hindrance.[133] In wild type α1 I, NMR resonances are observed for helix αC, whereas the latter are not detected in E317A. Signal loss provides valuable dynamic information, as peak broadening observed within the activating mutant indicates that helix αC and MIDAS loops are in conformational exchange.[135,139] In addition, the absence of peaks for helix αC, rather than the existence of sharp resonances expected for unfolded regions, implies that these residues are in intermediate exchange on the NMR timescale rather than fully unfolded as observed in the X-ray structure (Fig. 4.3). The E317A chemical shifts surrounding the MIDAS loops (Fig. 4.2C) support a chemical environment that is distinct from wild type α1 I. Thus, we propose that the conformation of helix αC and MIDAS loops are in conformational exchange in solution, and that helix αC may be interconverting between folded-unfolded conformations on a fast micro-second to millisecond timescale, reducing the steric hindrance to collagen.

On the slower timescale motions, our HDX results on the activating mutant E317A/α1 I reveal a propagation of the trends already observed in wild type α1 I, and reflect an even lower and narrower range of local stabilities for the C-terminus residues involved in allosteric movement. The slow conformational dynamics are propagated from helix α7 to helix α6 and strand βF in the gain-of-function mutant, disturbing to a higher degree (than in wild type α1 I) the contacts within the hydrophobic intramolecular pocket.[156-159] Indeed, destabilization of the hydrophobic core has been linked with an increase of binding toward collagen.[158,176] Moreover, deletion of three residues within helix α7 enhances the binding of α1 I to GLOGEN and allows visualization of the complex by NMR (Fig. 4.6C).[160] Higher destabilization of the E317A C-terminus coupled with the increase in binding affinity supports the hypothesis that these slow transient fluctuations observed in the unbound E317A/α1 I contribute to the gain-of-function by facilitating conformational rearrangement.

4.4.3 Thermodynamic basis of integrin-collagen interactions

Further corroboration of the forces driving association of α1 I with the THP may be deduced via comparisons of the respective thermodynamic profiles for both wild type α1 I and E317A mutant (refer to Table 4.1). The E317A mutant exhibits approximately a three-fold increase in affinity relative to wild type α1 I. The net enthalpic improvement of E317A association (i.e., $\Delta\Delta H = -1.9$

kcal·mol^{-1}) compared to its wild type counterpart presumably results from abrogation of an unfavorable enthalpy that arises from local unfolding of the αC helix, and salt-bridge disruption that occurs upon collagen binding to the wild type. We suggest that hydrogen bonds are more transient in E317A due to absence of the salt bridge, faster dynamics of the αC helix and MIDAS loops, and the more destabilized C-terminus. Collectively, our findings on the E317A mutant suggest that the unbound form populates an ensemble of states that are conformationally more suited to ligand association, thereby reducing the overall energy (enthalpy) penalty imposed by major structural changes that must occur in the wild type.

4.5 Conclusions

Elucidating the forces that drive allostery is critical to understanding the complex transformations of biomolecules. The unique heterogeneous shape of a rod-like extracellular matrix protein associated with a globular cellular receptor prompts us to explore new timescales. HDX experiments on the wild type and activated α subunit I domains of integrin suggest that collagen binding and the induced conformational change are facilitated by destabilization of the C-terminus secondary-structure elements and residues comprising the integrin–collagen interface. Nature designed this domain with regions prone for conformational rearrangement given their inherent dynamics and intrinsic destabilization. Collectively, our characterization of slow dynamics

in the integrin α1 I domain and the underlying binding energetics provides an instructive example on the relationship between local destabilization and propensity for allosteric structural changes. In an era where NMR microsecond timescale motions are critical for elucidating allosteric mechanisms, this study highlights the importance of exploring different timescales as part of a comprehensive experimental strategy to delineate allosteric and binding events.

Chapter 5 Effects of integrin on collagen fibril assembly kinetics, morphologies and viscosities

5.1 Introduction

The extra-cellular matrix (ECM) is a non-cellular component within tissues for all cells to make close contact with. The ECM not only provides load-bearing structure for tissues to ensure their integrities but also initiates biochemical interactions and signal transductions to carry out crucial biological functions. [2] The ECM contains a number of proteins, among which type I collagen is the most abundant constituent.

Type I collagen molecules are composed of three polyproline type II (PPII) helices. [177]. Each chain contains a collagenous domain composed of repeating Gly-X-Y sequences, where X is often proline and Y is often hydroxyproline. Each collagen molecule is ~300 nm in length and ~1.5 nm in diameter. Five triple-helical collagen molecules are staggered with each other by a distance of 67 nm (a D period) and arranged in a quasi-hexagonal packing [6,97]. The D period, the repeating unit of the collagen fibril, consists of an “overlap region” and a “gap region”. These D-banded patterns are represented as periodic patterns in atomic force microscopy (AFM) and electron microscopy (EM).

Collagen is first synthesized as procollagen, which contains the triple helical part in the middle, flanked by non-helical N- and C- propeptides to help three collagen α chains to twist into a triple helix. The soluble procollagen is then enzymatically processed by cleavage of N- and C- propeptides. After that, the

tropocollagens spontaneously aggregate into elongated staggered arrays to build the higher order collagen fibrils. [178-181] Collagen fibril assembly involves three main processes: the initiation of microfibril formation, the linear elongation of homogeneous fibrils and the lateral association of fibrils into fiber networks. [182-184]

The structural organizations of type I collagen vary greatly in different tissues. In tendons and ligaments, collagen fibrils bundle to form dense, straight cables; in corneas and bones, collagen fibrils organize into regular layers; in dermis and areolar tissues, collagen is arranged loosely as irregular layers; and in skin, collagen fibrils form basket weaves. [29-31] The diameters of type I collagen in various tissues diverge notably as well. Thin fibrils (~20 nm) are found in corneas where optical transparency is important, while thick collagen fibrils (~500 nm) are present in mature tendons where high tensile strength is essential. [32] Additionally, collagen fibers in various tissues differ in mechanical properties. They are rigid in bones, compliant in skin and elastic in cartilage. [33-35]

Collagen molecules can be obtained by dissolving animal tissues in acidic solutions. *In vitro*, they reconstitute into elongated fibrils with the characteristic alternating light and dark bands under TEM as *ex vivo* collagen fibrils. [185] The self-assembly of collagen suggests that all the information required to form

collagen fibrils is recorded in the collagen sequence. However, *in vivo*, the immediate self-assembly of collagen is not sufficient to explain the diverse appearance of collagen fibrils in different ECMs. The *in vivo* regulation of collagen fibril formation has been studied for over a century and a number of biomolecules have been suggested to regulate the assembly process of collagen, including proteoglycans in the ECM such as decorin, fibromodulin, lumican, osteoglycin, keratocan and biglycan; [38-42] enzymes required for post-translational modification of collagen after secretion such as 4-hydroxylase, lysyl hydroxylases and oxidases [43,44]; proteins involved in the translation of collagen in the secretory pathway such as HSP47 [45]; matrix metalloproteinases (MMPs) responsible for the degradation of extracellular matrix proteins [46] and cell membrane proteins, which are involved in the cellular controls for collagen assembly such as integrins. [47-49] Although enormous progress has been made, the mechanism of how type I collagen fibrils assemble, how length and diameter are regulated, how the multiscale organization is achieved, and how fibril formation kinetics is controlled *in vivo* remain elusive.

It has been proposed that collagen fibril formation is under close cellular control and integrin is possibly one of the proteins that participate in the regulation process. [47-49] Collagen binding integrins constitute a major group of cell receptors for ECM, mediating cell-collagen interactions to carry out

functions including cell spreading, migration, proliferation and differentiation. In the integrin family, four integrins are recognized to bind to collagens: α1β1, α2β1, α10β1, α11β1 integrins. [186-188] Here we focus on the more highly characterized α1β1 and α2β1 integrins.

Integrin α1β1 and α2β1 bind to type I collagen, mediated by the α-inserted domain (αI domain), by the sequence GXX'GEX'', where X'' is often R or N as determined by exploiting the collagen model peptide toolkit. [56,66] When unligated, the α1β1 integrin I domain remains in a "closed form", in which a salt bridge between Arg^{287} and Glu^{317} (Arg^{288} and Glu^{318} for α2 I domain) is thought to keep the I domain in a low-affinity conformation. Two integrin I domain mutants were reported to mimic the activated form of integrin I domains (α1I E317A [134] and α2I E318W [64]) and enhance their binding abilities to type I collagen. Here we use these two mutants to represent the activated integrin I domains and study the influence on collagen fibrillogenesis by integrin states.

Integrin has been proposed to regulate collagen fibril formation and has been studied for more than ten years. Li et al. found that vascular smooth muscle cells coordinate the assembly of type I collagen fibrils through a dynamic relationship between collagen, integrin, fibronectin and the cytoskeleton. [48,49] In addition, integrin is proposed to guide the migration of cells to produce new collagen and regulate the expressions of MMPs for collagen cleavage during

wound-healing. [46] Heino et al. [47] studied collagen fibril recognition by integrins. Integrin binding, integrin involved cell spreading and collagen contractions of monomeric and fibrillar collagens were studied and compared. They also suggested that integrin might guide or promote pericellular collagen fibrillogenesis. Similarly, Schiro et al. suggested that integrin might be used by the cell to modify the tertiary structure of the matrix. [71] De Fougerolles et al. reported that α2 integrin null mutation is embryonic lethal and both α1β1 and α2β1 both play roles in the reorganization and contraction of collagen matrices. [189]

Although enormous progress has been made, the mechanism of how integrin regulates collagen fibril formation in the ECM remains unclear. Here, we aim to elucidate the roles integrin plays in regulating type I collagen fibrillogenesis from the molecular point of view.

5.2 Materials and Methods

5.2.1 Integrin I domain expression and purification

The recombinant α1 I from human integrin α1β1 and gain-of-function mutant E317A used for these studies correspond to residues T141-E335. α2 I from human integrin α2β1 and gain-of-function mutant E318W correspond to residues S142-E336. All proteins contain a histag in the N-terminus. Protein purifications were conducted as previously described [64]. Protein

concentrations were determined via measurements of the absorptions at 280 nm employing the respective molar extinction coefficients: 12950 $M^{-1} \times cm^{-1}$ for α1I and α1I E317A, 20400 $M^{-1} \times cm^{-1}$ for α2I, and 25900 $M^{-1} \times cm^{-1}$ for α2I E318W.

5.2.2 Collagen-integrin I domain binding assays by biolayer interferometry (BLI)

Binding studies were performed on an Octet RED96 system (Pall Forte Bio LLC, Fremont, CA). All experiments were carried out at 25 °C on 96-well plates containing 200 µL of sample in each well. Amine reactive second-generation (AR2G) biosensors (Pall Forte Bio LLC, Fremont, CA) were activated by EDC/sulfo NHS and loaded with 0.03 mg/mL type I collagen from rat tail (Discovery Labware Inc., Bedford, MA) in 10 mM sodium acetate (pH 5.0) for 60 s, followed by a 300 s quench by 1.0 M Ethanolamine (pH 8.5). TBS buffer (pH 7.4) with 10 mg/mL bovine serum albumin (BSA) and 0.05% (v/v) Tween 20 was used as the working buffer for biosensor equilibrations, analyte dilutions and dissociations. In order to ensure the proper metal-mediated binding between integrin I domains and collagen, 5.0 mM of $MnCl_2$ was added to the buffer. TBS buffer without $MnCl_2$ was used as a negative control to represent non-specific binding. The sensors were equilibrated in working buffer for 120 s to acquire the baseline and submerged in working buffer containing integrin I domains with concentrations ranging from 0.1 to 10.0 µM

for 200 s to measure associations. Integrin I domains were then dissociated in working buffer for 600 s. The biosensors were regenerated in 10.0 mM glycine (pH 1.7) three times between runs. All steps were performed with shaking at 1000 rpm.

5.2.3 *In vitro* collagen kinetics of fibrillogenesis

Type I collagen from rat tail (Discovery Labware Inc., Bedford, MA) was diluted in 10 mM phosphate buffered saline (PBS) pH 7.4 with 5.0 mM $MgCl_2$ to a final concentration of 0.5 mg/mL and mixed with integrin I domains at concentrations ranging from 6.25 to 100.00 μg/mL. Aliquots (100 μL) of mixed samples were transferred to 96 well UV transparent plates (Corning Incorporated, Corning, NY) with three replicates for each condition. The plates were incubated at 37 °C in a POLAR Star Omega plate reader (BMG Labtech, Cary, NC). Changes in absorbance at 313 nm were recorded every 3 minutes until all turbidity curves reached plateaus. The turbidity curves were fit by the four-parameter sigmoidal models in GraphPad Prism 6 to obtain the half time of fibrillogenesis ($t_{1/2}$), hill slopes and final plateau levels by the following equation:

$$Y = Bottom + \frac{Top - Bottom}{1 + 10^{(log\ t_{1/2} - log\ t)\cdot\ HillSlope}}$$

where *Y* is the 313 nm absorbance, *Bottom* is the absorbance of baseline in lag phase, *Top* is the absorbance after fibrillogenesis curve reaches plateau, *t* is the time during the fibrillogenesis, $t_{1/2}$ is the time that gives an absorbance

halfway between *Bottom* and *Top*, and *HillSlope* describes the steepness in the middle phase of the curve.

5.2.4 Collagen fibrils morphology characterizations by atomic force microscopy (AFM)

Collagen fibril samples were taken from the end products of the fibrillogenesis described in 5.2.3. For each sample, a sample volume of 20 μL was deposited on a 1 cm x 1 cm square of freshly cleaved mica (Ted Pella Inc.,Redding, CA) and incubated at room temperature for 15 min. Then the surface of the sample was washed with 1 mL of deionized water and left to dry at room temperature for 1 h before being imaged. The samples were imaged by an NX-10 instrument (Park Systems, Suwon, South Korea) in non-contact mode with PPP-NCHR tips (nominal force constant 42 N/m; 330 kHz frequency; Nanosensors, Neuchatel, Switzerland). Image processing was conducted using XEI (Park Systems, Suwon, South Korea).

5.2.5 Collagen hydro-gel viscosity characterizations by rheology

The mechanical properties of collagen hydrogels were measured with Rheometrics SR-2000 parallel plate rheometer with a temperature-controlled incubation chamber (TA instruments, New Castle, DE). Aliquots of 200 μL solutions from the final products of the fibrillogenesis described in 5.2.3 were transferred to the bottom plate of the rheometer. A top plate of 20 mm diameter

was lowered down to a height of 0.6 mm, spreading the fibril solution throughout the plate. The viscosities of the gels were measured at rotation frequencies ranging from 0.1 to 10.0 Hz. All measurements were performed at 25 °C.

5.3 Results

5.3.1 Integrin I domains bind monomeric type I collagen

To determine whether integrin I domains bind to collagen, acid-extracted type I collagen from rat tail tendon was coated onto the biosensors and recombinant wild type (WT) integrin I domains α1I, α2I and gain-of-function variants α1I E317A and α2I E318W were allowed to interact. All integrin I domains showed significant binding to type I collagen in the presence of metal and negligible non-specific bindings when metal was absent. Both gain-of-function mutants showed increased binding capabilities relative to wild type. (Fig. 5.1)

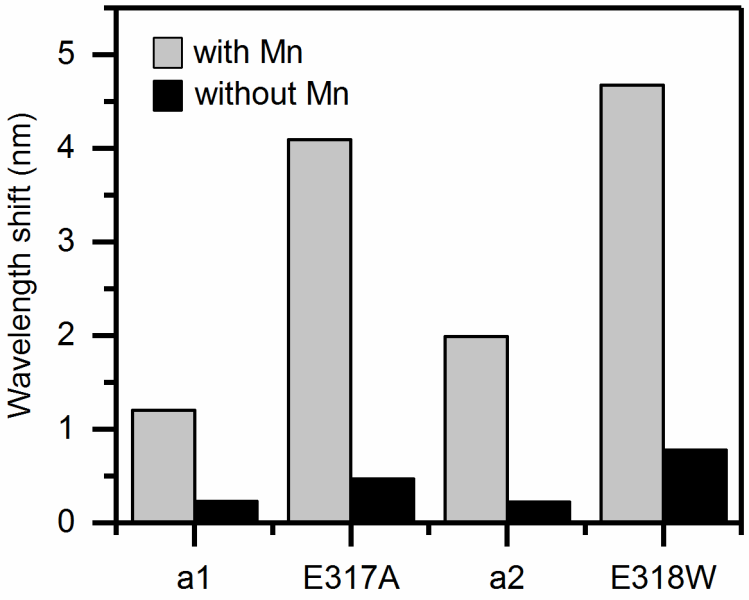

Figure 5.1 Biolayer interferometry (BLI) assays of integrins and gain-of-function mutants binding capabilities to type I collagen. The biosensors were loaded with type I collagen and allowed to bind to 10.0 μM of the indicated WT integrin I domain and variant. The maximum wavelength shifts are represented by gray bars. Bindings in the absence of metals represent non-specific bindings of integrins and are shown as black bars.

To further characterize the interactions in details, net binding of each WT I domain or variant was calculated by subtracting non-specific binding from gross binding signals. All WT I domains and variants showed strong dose dependencies (Fig. 5.2).

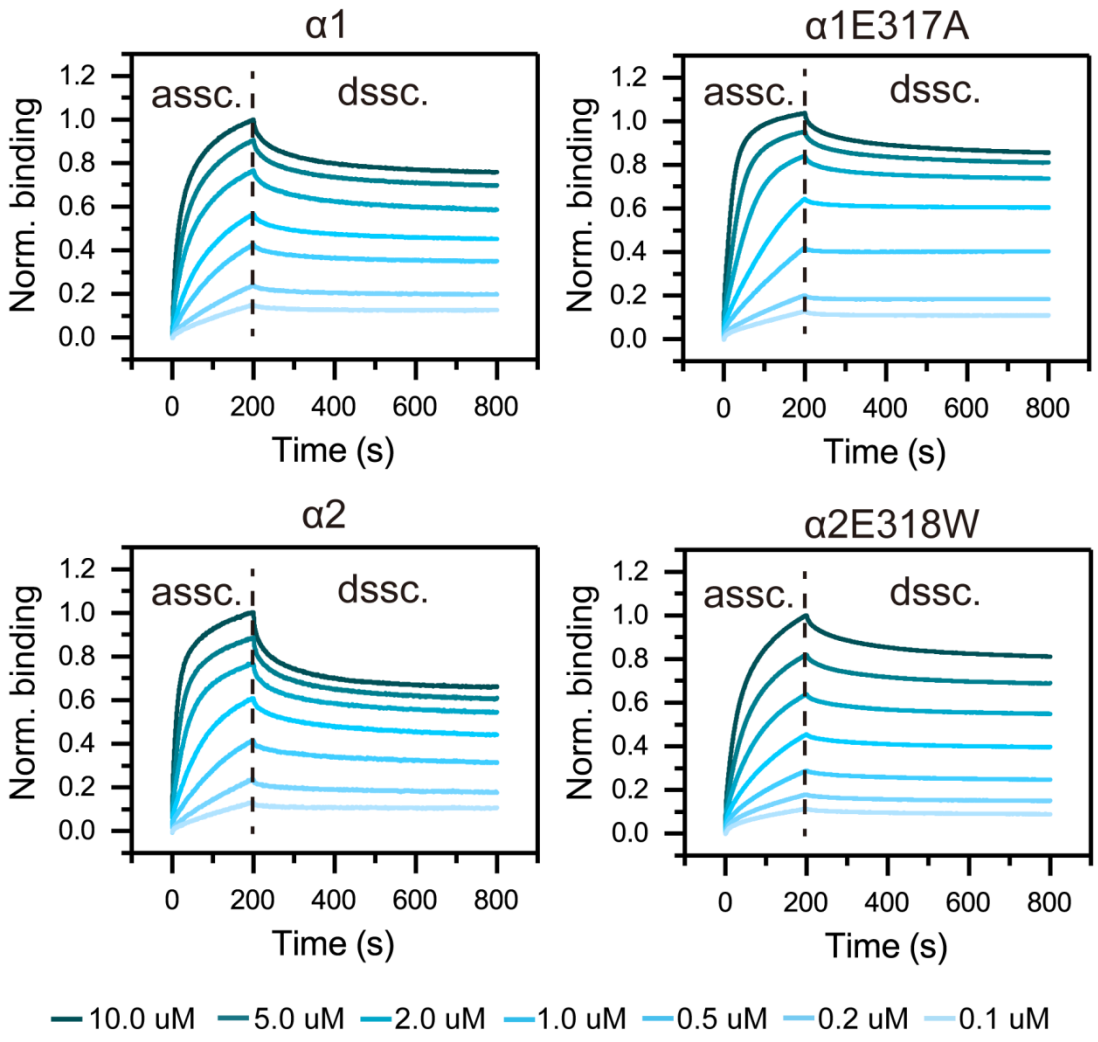

Figure 5.2 Biolayer interferometry (BLI) sensorgrams showing associations and dissociations of integrin I domain variants to immobilized type I collagen. 0.03 mg/mL type I collagen was loaded onto the biosensors. The sensors were equilibrated in working buffer to acquire the baseline, and allowed to bind to 0.1 to 10.0 μM integrin I domain. A negative control representing non-specific binding (without metal) was subtracted from the gross sensorgram signals so that only specific binding is shown.

5.3.2 Wild type Integrin I domains and gain-of-function mutants modulate type I collagen fibril formation

The fibril structures might be relevant to the assembly kinetics. The kinetics of type I collagen fibrillogenesis co-mixed with WT integrin I domains and variants were examined by measuring the solution turbidity (absorbance at 313 nm). The solutions contain the same concentration of collagen (0.5 mg/ml) and WT integrin I domains and mutants (20ug/ml). As shown in Fig. 5.3A, all curves show sigmoidal characteristics, consistent with the two-phase cooperative mechanism of collagen fibrillogenesis suggested by Wood et al. In the first phase, collagen monomers aggregate and stack to form microfibrils. This process is proved by AFM images captured within 5 min of fibril formation, which show thin rod-like species of 1.5 nm (collagen monomer) to 6.0 nm (collagen microfibril) in diameter and 300 to 800 nm in length. The second phase is characterized by a fast increase in the turbidity curves. In this phase, the microfibrils further assemble in parallel and longitudinally into homogeneous thicker fibrils. [179-181] This mechanism is also supported by real-time AFM observing growth steps of collagen self-assembly. [190]

Our results show that integrin WT I domains and their gain-of-function mutants slowed down the initial phase. WT I domains did not influence the second phase dramatically; gain-of-function mutants slowed down the second phase of collagen fibril formation (Fig. 5.3A). Negligible delays of fibril formation were observed when metal was not present (to abolish integrin binding to collagen) or BSA replaced integrin I domains (non-specific binding). Halftime of

fibrillogenesis ($t_{1/2}$), hill slopes and final plateau levels were fit by four-parameter sigmoidal models in GraphPad Prism 6 (Chapter 5.2.3). All WT integrin I domains increased $t_{1/2}$, similar hill slopes and final plateau levels of collagen fibril formation (Fig. 5.3C-E). The gain-of-function mutants, which mimic the activated states of integrin I domains, showed more significant increased $t_{1/2}$, decreased hill slopes and final plateau levels of collagen fibril formation, relative to their WT variants.

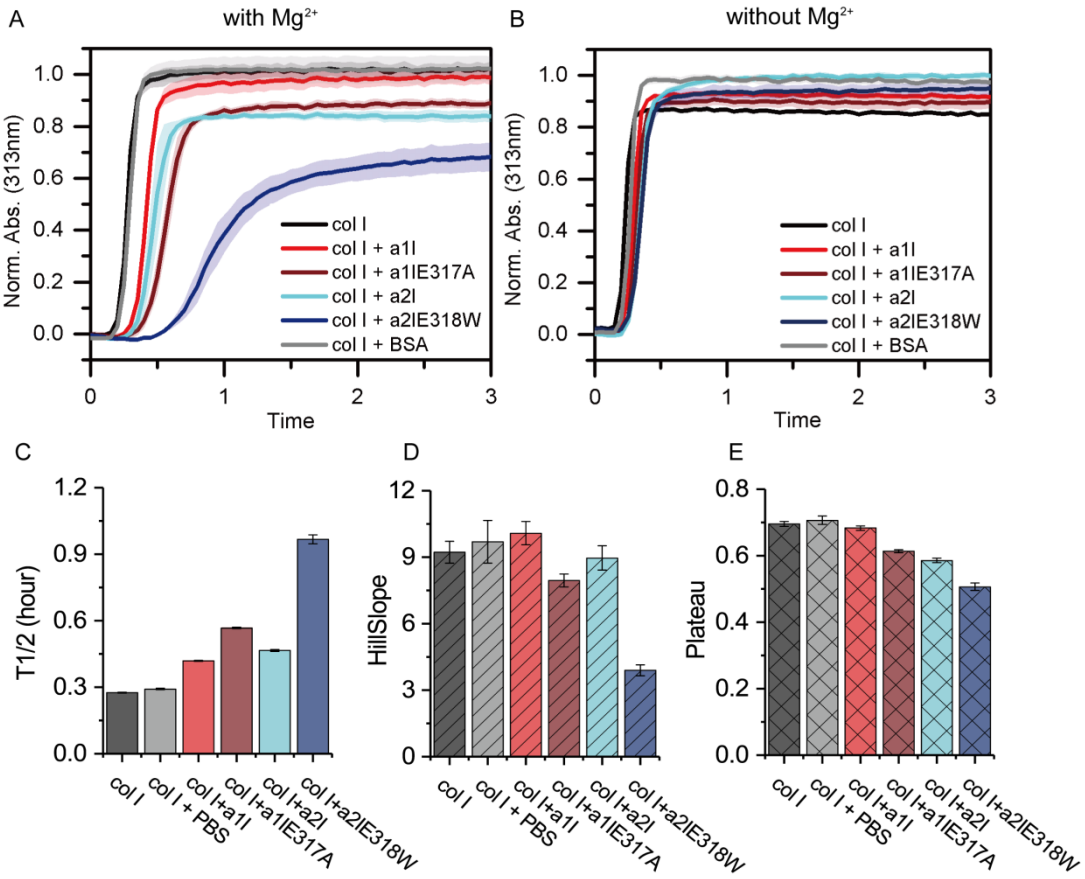

Figure 5.3 Effects of WT integrin I domains and gain-of-function variants on type I collagen fibrillogenesis. Fibril assembly kinetics of type I collagen alone and in the presence of integrin I domains and BSA with (A) and without (B) $MgCl_2$. Type I collagen in

PBS was incubated at 37 ºC for fibril formation, and turbidity was measured at 313 nm over time. Metal ions were added in the buffer to facilitate metal mediated integrin-collagen interactions. Fibrillogenesis kinetics with integrin I domains replaced by BSA or in the absence of metal act as negative controls for non-specific integrin I domain binding.

To further show that specific integrin-collagen interactions slowed down collagen fibril formation, we tested the dose-dependencies. α1, α1E317A, α2, and α2E318W integrin I domains showed clear dose dependencies in $t_{1/2}$ (Fig. 5.4). This might be attributed to a higher amount of integrin I domains interacting with collagen, thus decreasing the probability of collisions between collagen molecules and resulting in fewer nucleation centers. The hill slopes showed less dose dependency than $t_{1/2}$. The plateau levels were random, suggesting that integrins might not affect the final product amounts of collagen fibril formation.

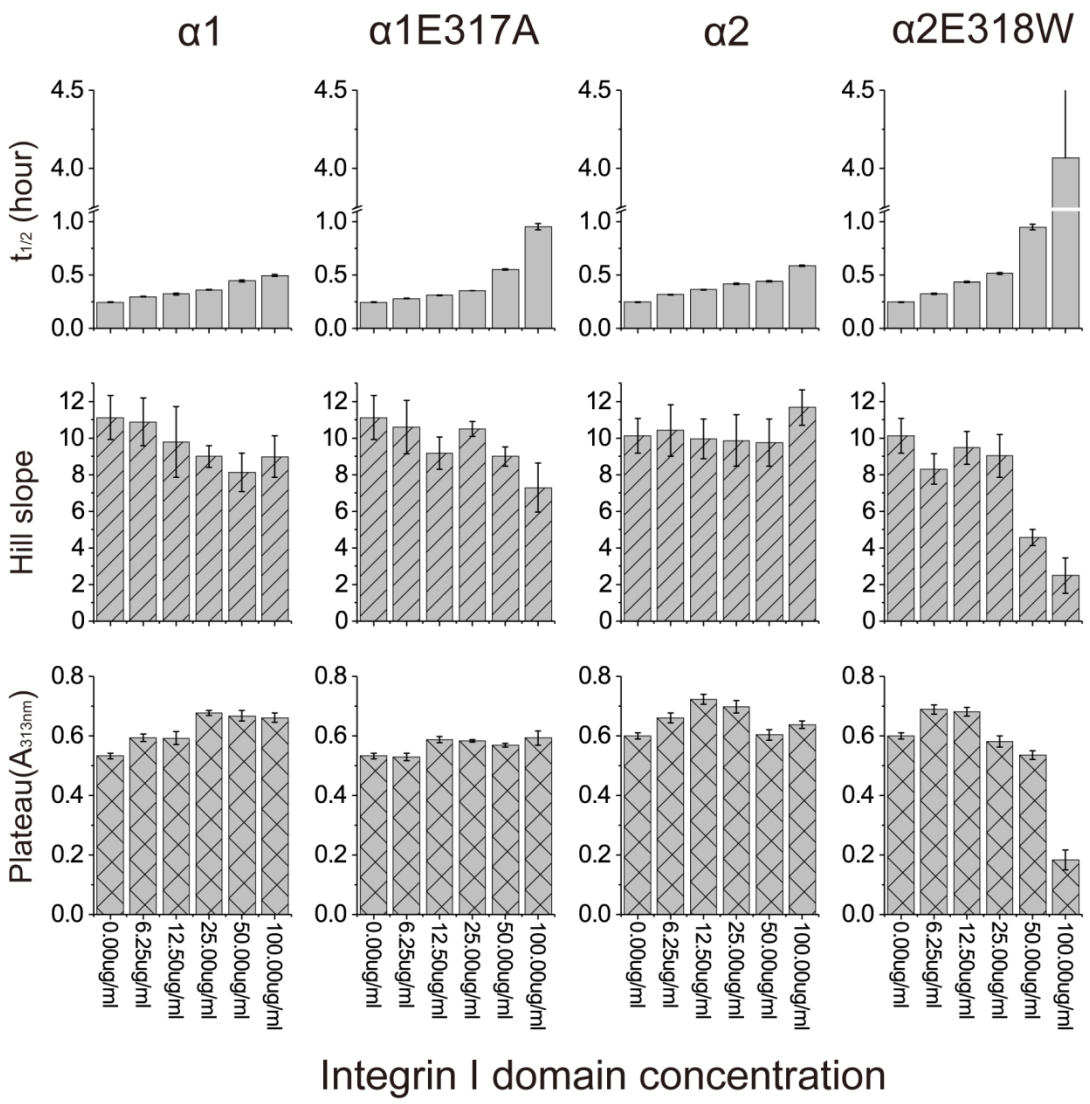

Figure 5.4 The effects of integrin I domains and their gain-of-function mutants on type I collagen fibril formation are dose-dependent. Type I collagen turbidity measured by absorbance at 313 nm in the absence and presence of α1, α1E317A, α2, and α2E318W integrin I domains at concentrations ranging from 6.25 μg/ml to 100.00 μg/ml. Halftime ($t_{1/2}$), hill slopes and final plateau levels from four-parameter sigmoidal curve fits are expressed in bar plots. Data are shown as mean ± S.E. of three replicates.

5.3.3 Collagen fibril morphology with and without integrin I domain co-incubations.

In order to characterize the morphologies of the reconstituted collagen fibrils, we used atomic force microscopy (AFM) to look at the final products of the collagen fibril formation kinetic studies. Fig. 5.5 shows the comparative AFM images of reconstituted type I collagen with or without integrin I domain co-incubation. Amplitude images were shown to describe the textures of collagen more clearly. The selected fibrils with the largest heights in each condition are shown. In the case of pure collagen, the image exhibits D-banded straight and thick collagen fibrils. After introducing integrin α1I and α2I, compared to pure collagen, the collagen fibrils have similar patterns. When gain-of-function integrin I domain mutants are involved, only thin and pliant fibrils are observed.

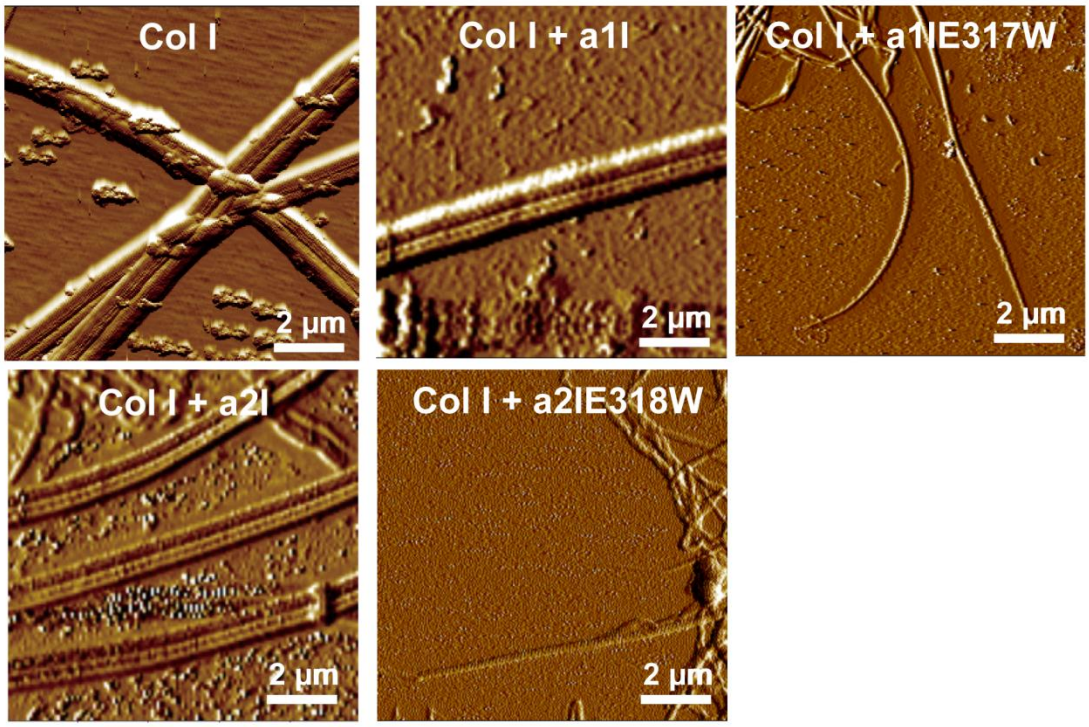

Figure 5.5 Representative AFM images of the greatest fibril heights found in each

condition. All images were taken in non-contact amplitude mode. In each incubation condition (collagen alone, collagen with α1I, α1IE317A, α2I, and α2IE318W), the fibril heights were measured. The images with the greatest fibril heights are shown.

In the AFM images taken on pure collagen and collagen co-incubated with integrin I domains, two distinct species of collagen fibrils were observed. Shown in Fig. 5.6A, one species is thick (~ 0.5 μm) and straight. This species is more mature, with clear D-bands. The other species is much thinner, is often curved, less smooth, less D-banded and form networks. Fig. 5.6B illustrates the relationships between the two collagen species. Thin and curved fibrils fuse in parallel and form higher order thicker fibrils, and the thicker fibrils fuse further to form even higher order species. As shown in Fig. 5.5, pure type I collagen self-assembly products contain both thin and thick collagen fibrils. The presence of α1I and α2I did not seem to affect the sizes and shapes of the thicker species collagen fibrils. However, when collagen was incubated with gain-of-function integrin I domain mutants, only thin species were found, implying that these mutants stopped thin fibrils from fusing into higher order thick fibrils.

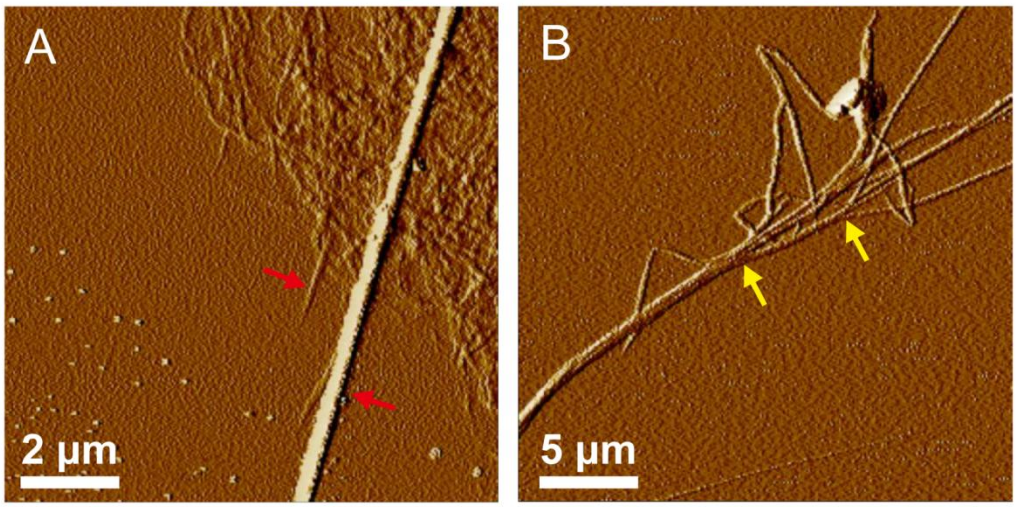

Figure 5.6 Type I collagen fibrillogenesis products have two main species. (A) One species is thick and straight, while the other species is thin and curvy (marked by red arrows). (B) Thin and curved species fuse in parallel and form thicker higher order collagen fibrils. Two fusion sites are marked by yellow arrows. Once fused, thicker collagen fibrils become more rigid and straight. Both AFM images were taken under amplitude mode to show clear views of collagen fibril textures.

To further characterize the thinner species of collagen fibril products incubated alone or with integrin I domains, heights of the collagen fibril thin species were measured and are shown in Fig. 5.7. The distributions of the fibril heights are not symmetric bell shapes, but having trailing effects that fit lognormal distributions. This is possibly because fibrils have the tendency to fuse into higher order fibrils. In pure collagen, the trailing effects are more pronounced than in the presence of integrin I domains (Fig. 5.7 black curves). Wild type integrin I domains decrease the collagen fibril heights and the gain-of-function mutants reduce collagen fibril heights to an even greater extent.

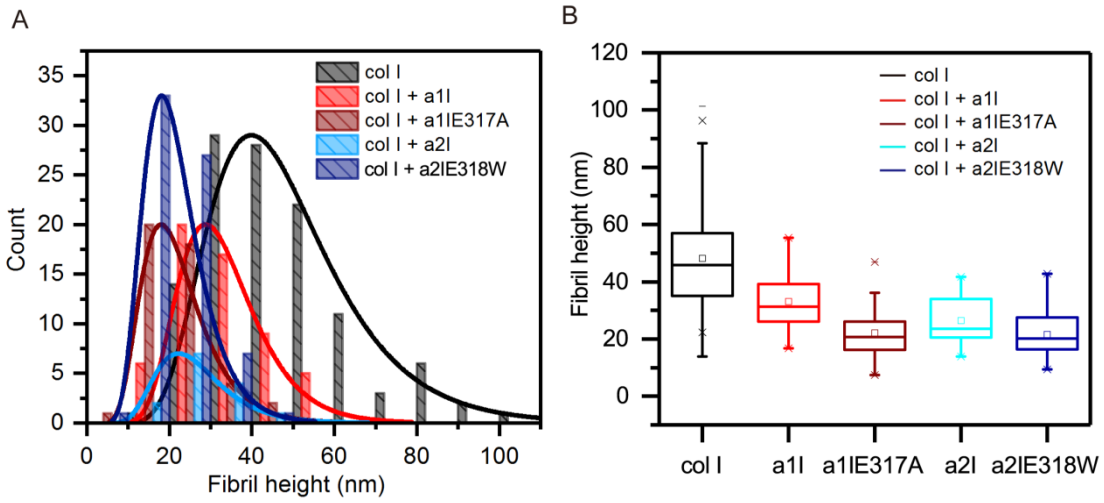

Figure 5.7 Type I collagen fibril height distributions affected by WT integrin I domains and their gain-of-function mutants. (A) Type I collagen fibril heights were taken from multiple collagen fibrils in multiple samples. The heights were plotted with bin sizes of 5 nm. The distribution curves were fit to a lognormal distribution (solid lines). (B) Boxplot diagram representing the collagen fibril height distributions. The boxes represent the statistical distribution between 25% and 75% of the measurements in each condition. The mean value is depicted by a circle, and the median is represented by a line inside the box. The whiskers represent maximum and minimum values of all height measurements or the calculated confidence ranges when some measurements are suspect outliers (denoted by ×). The sample sizes are 117, 57, 45, 14 and 69, respectively. A few outliers of collagen fibrils formed by co-incubation with a1I or a2I that have heights greater than 110 nm are not included.

5.3.4 Mechanical properties

Collagen forms elastic networks of fibrils, which confer tensile strength.

Reconstituted collagen gels are used extensively for cell-matrix mechanical interaction studies. Here we used rheology to characterize viscosities of collagen gels, by measuring stress relative to the change in shear strain rates. Viscosity reveals the interactions between molecules resisting flowing in a fluid. As shown in Fig. 5.8, collagen gels are non-Newtonian materials, showing reduced viscosities with higher shear rates, consistent with reports by earlier studies [191]. The incorporation of integrin I domains in type I collagen gel substantially increase the apparent gel viscosities. Integrin gain-of-function mutants have a greater effect. Co-incubation with α2I E318W increases the collagen gel viscosity by 10 fold relative pure collagen at all shear rates. The increased viscosities of collagen when integrin I domains are involved suggests that more stress is needed to overcome the interactions between molecules to keep the collagen gel moving.

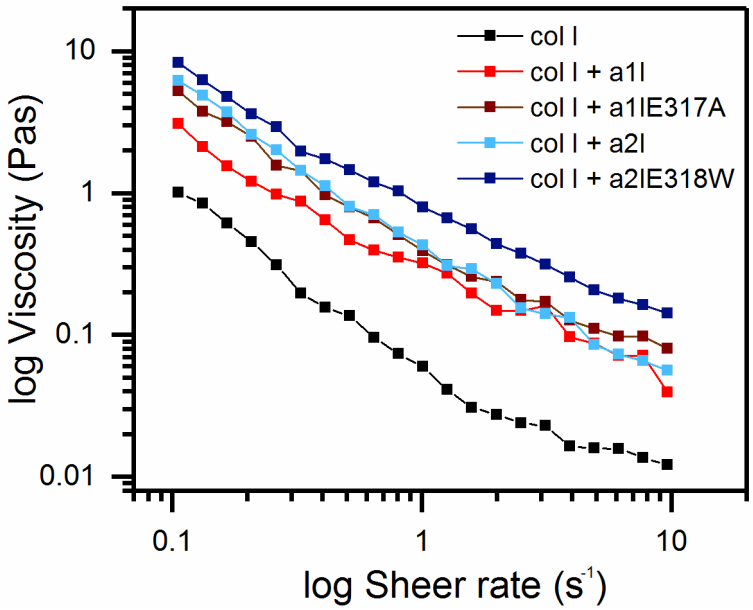

Figure 5.8 Apparent viscosity behavior of pure collagen and collagen

co-incubated with integrin I domains vs. shear rate. Aliquots of 200 μL solutions from the final products of the fibrillogenesis were transferred to the bottom plate of the rheometer and pressed by a top plate of 20 mm diameter to form a flat cylinder of 0.6 mm thick. The temperature of the plates was kept at 25 °C.

5.4 Discussion

Collagen is the main component of the ECM, playing dual roles of providing load-bearing structure to uphold tissue integrities and initiating biochemical interactions for crucial biological functions. *In vivo*, the diversities in collagen fibril shape and organizations in different tissues cannot be merely explained by the self-assembly of type I collagen and must be regulated by some collagen binding partners. Integrins have been shown to regulate collagen fibril synthesis, turnover and collagen gel contractions. [47,192-197] These data clearly indicate that integrin has a crucial role in the formation of proper collagen fibrils in tissues. Here we studied the interactions between collagen and integrin I domains of different activation states and carefully investigated the changes in collagen fibrils formed in the presence of integrin I domains.

Recombinant integrin I domains bound collagen type I immobilized on BLI biosensors in a dose-dependent manner, as did the gain-of-function mutants which mimic the activated states of integrin I domains. Integrin α2I showed higher binding than α1I and their gain-of-function mutants showed higher

binding than the WT I domains, consistent with previous findings. [198,199] In the present study, collagen fibrillogenesis was assayed in the presence of WT integrin I domains and gain-of-function mutants. The presence of integrin I domains elongates the $t_{1/2}$ in a strict dose-dependent manner and decreases the hill slope in a rough dose-dependent manner (Fig. 5.4). Together with the two-phase cooperative mechanism of collagen fibrillogenesis suggested by Wood et al. [179-181], our data suggests that the presence of integrin I domains elongate the first stage, in which collagen monomers assemble into microfibrils. However, our data is not consistent with the observations that α1I accelerated fibrillogenesis in the beginning and inhibited later process *in vitro*, reported by Heino et al. [47] The discrepancies might result from differences in the integrin I domain constructs. The integrin I domain used by Heino et al. contains a 211 amino acid N-terminal GST tag, while our construct instead contains a more robust ten amino acid N-terminal His tag. The gain-of-function integrin I domains mutants slow down the first stage, in which collagen monomers assemble into microfibrils, and also the second stage, in which microfibrils further assemble in parallel and longitudinally into thicker fibrils. The gain-of-function mutants had the same inhibitory effects on collagen fibril formation as wild type I domains, but with greater magnitudes, in agreement with the binding affinities shown in Fig. 5.1. This suggests a mechanism for regulating collagen fibrillogenesis by varying the activation state of integrin I domains.

We next investigated morphologies of collagen fibrils formed in the presence of integrin I domains. Pure collagen self-assembly products showed fibrils in two species: (1) straight and thick fibrils with clear D-bands and (2) thin and curved fibrils that usually form dense networks. WT Integrin I domains reduced the size of the thin species, but not affect the sizes of thick fibrils. Gain-of-function integrin I domain mutants reduced the sizes of thin species fibrils to a larger extent than WT I domains and completely inhibited the thick fibrils (Fig. 5.5, 5.7). We found that thick collagen fibrils were formed by the fusion of thin fibrils (Fig. 5.6B). A possible mechanism could be that wild type integrins slow down the formation of microfibrils and the lateral fusion by interacting with collagen and competing with other collagen molecules. This would result in thinner bulk collagen fibrils. The wild type integrin I domains did not stop the thin collagen fibrils from forming higher order thick fibrils, probably due to the weak binding affinities before activation. The activated states are able to bind weak integrin binding sites and therefore might bind to thin collagen fibrils strongly and shield the binding sites involved in collagen association and protect thin collagen fibrils from further associations.

Heino et al [47] have studied integrin-mediated cell adhesion to type I collagen in some detail and provide a valuable basis for our study. They have concluded that integrins bind to type I collagen, enable cell spreading on collagen and

reorganize collagen fibrils. They proposed a mechanism of cellular regulation of fibrillogenesis: due to the higher binding affinity to collagen monomers than to fibrils, integrins on cell surfaces grab a newly synthesized collagen monomer and guide fibril formation to the pericellular area. Once the grabbed collagen monomer assembles into a fibril, the new collagen monomers might compete for integrin binding with the fibril. Canty et al. proposed “fibripositors” on fibroblasts are the cellular regulators of initial collagen fibril assembly occurring near the tendon fibroblast surface. [200] α2β1 integrins are located on the surfaces of fibroblasts and are possibly involved in this regulation process. Our study is focused on the mechanism of collagen self-assembly in the molecular level and might be the basis for the more complicated mechanism i*n vivo*.

.

In addition to the single fibril morphology, the organization of fibrils and their mechanical properties define the diverse architecture of ECMs. Therefore we studied the viscosities of bulk collagen gel, the final products of fibrillogenesis. The increased viscosities of bulk collagen gel upon addition of integrin I domains suggest that the collagen fibrils have a higher tendency to tangle and bind together in the presence of I domains, resisting flow and making the liquid stickier. The viscosity behavior of a collagen gel results from the interactions between two component phases: collagen fibril network and interstitial solution. When a collagen gel is subjected to shear or tension, the collagen fibril

network supports virtually all of the load; and under compression, the collagen network transfers most of the load to the interstitial solution, which is incompressible, to prevent the network from collapsing. [201] In our case, when a collagen gel is under sheer, the interactions between fibrils in the network were enhanced when integrin I domains were present. It is likely that the integrin I domains increased the gelation network associations of collagen fibrils in the bulk solutions.

5.5 Conclusions

We report on the formation of type I collagen fibrils in the presence of recombinant integrin I domains and gain-of-function mutants. The conducted experiments show that WT integrin I domains slowed down the kinetics of type I collagen in the first phase and the gain-of-function integrin I domains mutants mimicking the activated states slowed down both phases. Mature collagen fibrils (> 100 nm in diameter) and immature, thin and floppy ones (~ 20 nm in diameter) were observed in the fibrillogenesis final products, and integrin I domains were found to reduce the heights of the immature fibrils. The mature species were observed by AFM to be fused to the immature species and the gain-of-function mutants of integrin I domains inhibited mature species formation. Integrin I domains were found to enhance the viscosities of collagen gels, implying stronger intermolecular interactions between collagen fibrils. We propose that i*n vivo*, by controlling its activation state, integrin might play a role

in the cellular regulation of type I collagen self-assembly kinetics, equilibration, morphologies and mechanical properties accounting for the various architectures of ECMs.

Appendix

Table A.1 Kinetic parameters (individual intrinsic rate constant (k_{int}), observed exchange rate constants (k_{obs}), logarithm of protection factors (P_f), and free exchange energies (ΔG_{HX}) for wild type α1 I and E317A/α1 I residues obtained by the NMR HDX experiments.

α1 I					E317A/ α1 I				
AA	k_{int} (s^{-1})	k_{obs} (s^{-1})	Log P_f	ΔG_{HX} (kcal/mol)	AA	k_{int} (s^{-1})	k_{obs} (s^{-1})	Log P_f	ΔG_{HX} (kcal/mol)
T141	2.05	3.33E-03	2.79	3.74	T141	3.23	2.27E-03	3.15	4.23
Q142	3.82	3.33E-03	3.06	4.10	Q142	6.02	2.27E-03	3.42	4.59
L143	0.88	4.86±0.14E-05	4.26±0.01	5.71±0.02	L143	1.38	6.08±0.26E-05	4.36±0.02	5.84±0.03
D144	0.65	1.04E-08	7.80	10.46	D144	1.05	2.60E-08	7.61	10.20
I145	0.26	1.04E-08	7.40	9.92	I145	0.42	2.60E-08	7.20	9.66
V146	0.25	1.04E-08	7.38	9.89	V146	0.39	2.60E-08	7.17	9.62
I147	0.28	1.04E-08	7.44	9.97	I147	0.45	2.60E-08	7.23	9.70
V148	0.25	1.04E-08	7.38	9.89	V148	0.39	2.60E-08	7.17	9.62
L149	0.40	1.04E-08	7.58	10.17	L149	0.63	2.60E-08	7.38	9.90
D150	0.65	1.04E-08	7.80	10.46	D150*	1.05			
G151	2.59	1.04E-08	8.40	11.26	G151*	4.15			
S152	7.28	3.33E-03	3.34	4.48	S152*	11.50			
N153	12.90	3.33E-03	3.59	4.81	N153*	20.40			

S154	10.30	3.33E-03	3.49	4.68	S154*	16.20			
I155	0.78	1.90±0.20E-04	3.61±0.05	4.85±0.06	I155*	1.23			
Y156	0.66	1.77±0.08E-05	4.57±0.02	6.13±0.03	Y156*	1.05			
P157*					P157*				
W158	0.47	3.33E-03	2.15	2.88	W158	5.13	2.27E-03	3. 35	4.50
D159	0.82	3.33E-03	2.39	3.21	D159	1.32	2.27E-03	2.76	3.71
S160	3.26	8.80±0.59E-04	3.57±0.03	4.79±0.04	S160	5.22	2.27E-03	3.36	4.51
V161	0.84	3.93±0.96E-08	6.33±0.11	8.49±0.14	V161	1.32	6.02±0.62E-05	4.34±0.04	5.82±0.06
T162	1.29	8.50±0.66E-07	6.18±0.03	8.29±0.04	T162	2.04	1.76±0.14E-05	5.06±0.03	6.79±0.05
A163	3.33	1.27±0.03E-05	5.42±0.01	7.27±0.01	A163	5.25	2.44±0.20E-04	4.33±0.04	5.81±0.05
F164	1.21	1.04E-08	8.07	10.82	F164	1.90	2.60E-08	7.86	10.55
L165	0.63	1.04E-08	7.78	10.44	L165	1.00	2.60E-08	7.59	10.17
N166	4.00	1.04E-08	8.58	11.51	N166	6.31	2.60E-08	8.39	11.25
D167	2.21	1.04E-08	8.33	11.17	D167	3.55	2.60E-08	8.14	10.91
L168	0.37	1.04E-08	7.55	10.12	L168	0.59	2.60E-08	7.35	9.86
L169	0.34	1.04E-08	7.52	10.08	L169	0.54	2.60E-08	7.32	9.81
E170	0.40	1.04E-08	7.59	10.18	E170	0.64	2.60E-08	7.39	9.92
R171	1.79	3.33E-03	2.73	3.66	R171	2.85	2.27E-03	3.10	4.16
M172	3.40	1.70±0.10E-05	5.30±0.02	7.11±0.03	M172	5.37	1.66±0.02E-05	5.51±0.01	7.39±0.01
D173	1.36	3.33E-03	2.61	3.50	D173	2.19	2.39±0.10E-06	5.96±0.02	8.00±0.20
I174	0.26	1.88±0.09E-06	3.14±0.02	4.21±0.03	I174	0.42	2.52±0.16E-04	3.22±0.03	4.31±0.04
G175	2.30	3.21±0.22E-06	5.86±0.03	7.85±0.04	G175	3.63	5.47±0.20E-06	5.82±0.02	7.81±0.02
P176*					P176*				

K177	1.10	3.33E-03	2.52	3.38	K177	1.74	2.27E-03	2.88	3.87
Q178	3.18	2.19±0.32E-03	3.16±0.06	4.24±0.08	Q178	5.01	2.46±0.19E-03	3.31±0.03	4.44±0.05
T179	2.83	2.38±0.60E-05	5.08±0.11	6.81±0.15	T179	4.47	2.60E-08	8.24	11.05
Q180	3.82	2.44±0.39E-07	7.19±0.07	9.65±0.09	Q180	6.02	2.60E-08	8.36	11.22
V181	0.66	1.92±0.22E-07	6.54±0.05	8.77±0.06	V181	1.05	2.60E-08	7.61	10.20
G182	2.83	1.04E-08	8.43	11.31	G182	4.47	2.60E-08	8.24	11.05
I183	0.58	1.04E-08	7.74	10.39	I183	0.91	2.60E-08	7.55	10.12
V184	0.25	1.04E-08	7.38	9.89	V184	0.39	2.60E-08	7.17	9.62
Q185	1.75	1.04E-08	8.23	11.03	Q185	2.75	2.60E-08	8.02	10.76
Y186	1.79	1.04E-08	8.24	11.05	Y186*	2.82			
G187	4.38	4.26±0.25E-05	5.01±0.03	6.72±0.03	G187	6.92	2.27E-03	3.48	4.67
E188	0.96	3.33E-03	2.46	3.30	E188	1.54	2.27E-03	2.83	3.80
N189	4.61	3.33E-03	3.14	4.21	N189	7.33	2.27E-03	3.51	4.71
V190	0.88	3.33E-03	2.42	3.25	V190	1.38	2.27E-03	2.78	3.73
T191	1.29	8.70±1.23E-07	6.17±0.06	8.28±0.08	T191	2.04	4.72±0.21E-06	5.64±0.02	7.56±0.03
H192	5.01	3.33E-03	3.18	4.26	H192	15.70	2.27E-03	3.84	5.15
E193	1.47	3.07±0.15E-05	4.68±0.02	6.28±0.03	E193	3.97	4.15±0.33E-05	4.98±0.03	6.68±0.05
F194	0.86	2.23±0.09E-05	4.58±0.02	6.15±0.02	F194	1.37	2.70±0.05E-06	4.71±0.01	6.31±0.01
N195	7.45	3.33E-03	3.35	4.49	N195	11.70	2.27E-03	3.71	4.98
L196	1.15	1.49±0.06E-05	4.89±0.02	6.55±0.02	L196	1.82	4.95±0.52E-06	5.57±0.05	7.46±0.06
N197	4.00	1.42±0.16E-06	6.45±0.05	8.65±0.06	N197	6.31	1.87±0.08E-06	6.53±0.02	8.76±0.03
K198	4.00	8.68±1.35E-07	6.66±0.07	8.94±0.09	K198	6.31	2.15±0.08E-06	6.47±0.02	8.67±0.02
Y199	1.49	1.13±0.06E-05	5.12±0.02	6.87±0.03	Y199	2.34	1.42±0.02E-05	5.22±0.01	7.00±0.01

S200	5.52	3.33E-03	3.22	4.32	S200	8.71	2.27E-03	3.58	4.81±0.02
S201	9.81	2.85±0.21E-04	4.54±0.03	6.08±0.04	S201	15.50	4.08±0.16E-04	4.58±0.02	6.14±0.02
T202	3.56	3.33E-03	3.03	4.06	T202	5.62	2.27E-03	3.39	4.55
E203	1.03	3.33E-03	2.49	3.34	E203	1.65	2.27E-03	2.86	3.84
E204	0.46	5.35±0.55E-04	2.94±0.04	3.94±0.06	E204	0.75	2.27E-03	2.52	3.38
V205	0.30	4.10±0.11E-06	4.86±0.01	6.52±0.02	V205	0.47	5.10±0.16E-06	4.97±0.01	6.66±0.02
L206	0.40	1.04E-08	7.58	10.17	L206	0.63	2.39±0.15E-07	6.42±0.03	8.61±0.06
V207	0.26	1.03±0.05E-05	4.40±0.02	5.90±0.03	V207	0.41	1.25±0.04E-05	4.51±0.01	6.05±0.02
A208	1.52	3.16±0.13E-06	5.68±0.02	7.62±0.02	A208	2.40	4.07±0.19E-06	5.77±0.02	7.74±0.03
A209	2.10	7.14±0.28E-06	5.47±0.02	7.34±0.02	A209	3.31	7.61±0.28E-06	5.64±0.02	7.56±0.02
K210	1.91	2.11±0.04E-05	4.96±0.01	6.65±0.01	K210	3.02	2.46±0.06E-05	5.09±0.01	6.83±0.01
K211	2.52	5.62±0.13E-05	4.65±0.01	6.24±0.01	K211	3.98	8.66±0.26E-05	4.66±0.01	6.25±0.02
I212*	0.52				I212	0.81	2.45±0.08E-05	4.52±0.01	6.06±0.02
V213	0.25	8.11±0.44E-04	2.48±0.02	3.33±0.03	V213	0.39	2.27E-03	2.23	3.00
Q214	1.75	2.02±0.14E-03	2.94±0.29	3.94±0.39	Q214	2.75	2.27E-03	3.08	4.14
R215	4.00	4.18±0.36E-05	4.98±0.04	6.68±0.05	R215	6.31	5.87±0.45E-05	5.03±0.03	6.75±0.04
G216	6.48	3.33E-03	3.29	4.41	G216	10.20	2.27E-03	3.65	4.90
G217	5.78	3.33E-03	3.24	4.46	G217	9.12	2.27E-03	3.60	4.83
R218	3.73	3.33E-03	3.05	4.09	R218*	5.89			
Q219	4.00	3.33E-03	3.08	4.13	Q219	6.31	2.27E-03	3.44	4.62
T220	2.83	3.33E-03	2.93	3.93	T220*	4.47			
M221	3.25	2.67±0.45E-04	4.08±0.07	5.48±0.10	M221	5.13	2.27E-03	3.35	4.50
T222	2.30	9.61±0.28E-05	4.38±0.01	5.87±0.02	T222	3.63	2.27E-03	3.20	4.30

A223	3.33	3.33E-03	3.00	4.02	A223*	5.25			
L224	0.55	2.66±0.10E-05	4.32±0.02	5.79±0.02	L224	0.87	2.60E-08	7.53	10.09
G225	2.41	1.90±0.10E-06	6.34±0.06	8.51±0.08	G225*	3.80			
I226	0.58	1.04E-08	7.74	10.39	I226	0.91	2.60E-08	7.55	10.12
D227	0.62	1.04E-08	7.78	10.43	D227	1.00	2.60E-08	7.58	10.17
T228	1.19	5.49±0.26E-07	6.34±0.02	8.50±0.03	T228	1.89	2.60E-08	7.86	10.54
A229	3.33	1.04E-08	8.51	11.41	A229	5.25	2.60E-08	8.31	11.14
R230	2.52	1.04E-08	8.38	11.25	R230	3.98	2.60E-08	8.18	10.98
K231	3.18	1.04E-08	8.49	11.38	K231	5.01	2.60E-08	8.28	11.11
E232	0.86	7.64±0.93E-07	6.05±0.05	8.12±0.07	E232	1.37	2.60E-08	7.72	10.36
A233	1.49	4.76±0.24E-05	4.50±0.02	6.03±0.03	A233	2.37	3.72±0.26E-07	6.80±0.03	9.13±0.04
F234	1.21	6.85±0.57E-07	6.25±0.04	8.38±0.05	F234	1.90	2.03±6.90E-05	4.97±1.47	6.67±1.98
T235	2.05	8.66±0.24E-06	5.37±0.01	7.21±0.02	T235	3.23	1.16±0.02E-05	5.45±0.01	7.30±0.01
E236*	1.03				E236	1.65	1.23±0.04E-05	5.46±0.01	7.32±0.02
A237*	1.49				A237*	2.37			
R238	2.52	2.08±0.19E-04	4.08±0.04	5.48±0.05	R238	3.98	1.66±0.15E-04	4.38±0.04	5.88±0.05
G239	6.48	4.49±0.40E-04	4.16±0.04	5.58±0.05	G239	10.20	3.49±0.79E-04	4.47±0.10	5.99±0.13
A240	3.10	3.36±0.33E-04	3.97±0.04	5.32±0.06	A240	4.90	3.81±0.16E-04	4.11±0.02	5.51±0.02
R241*	2.52				R241	3.98	2.15±1.16E-04	4.27±0.23	5.72±0.31
R242	4.19	3.33E-03	3.10	4.16	R242	6.60	2.27E-03	3.46	4.65
G243*	6.48				G243	10.20	2.27E-03	3.65	4.90
V244	0.62	3.33E-03	2.27	3.04	V244	0.98	2.27E-03	2.63	3.53
K245	1.39	3.33E-03	2.62	3.52	K245	2.19	2.27E-03	2.98	4.00

K246	2.52	7.83±1.08E-04	3.51±0.06	4.70±0.08	K246	3.98	2.27E-03	3.24	4.35
V247	0.55	1.04E-08	7.72	10.36	V247	0.87	2.60E-08	7.52	10.09
M248	1.49	1.04E-08	8.16	10.94	M248	2.34	2.60E-08	7.95	10.67
V249	0.54	1.04E-08	7.72	10.35	V249	0.85	2.60E-08	7.51	10.08
I250	0.28	1.04E-08	7.44	9.97	I250	0.45	2.60E-08	7.23	9.70
V251	0.25	1.04E-08	7.38	9.89	V251	0.39	2.19±1.56E-07	6.25±0.31	8.38±0.41
T252	1.29	1.04E-08	8.09	10.86	T252*	2.04			
D253*	1.68				D253*	2.69			
G254	2.59	2.10±0.05E-05	5.09±0.01	6.83±0.01	G254*	4.15			
E255	0.96	5.73±1.31E-04	3.23±0.10	4.33±0.13	E255*	1.54			
S256	3.49	2.11±0.33E-04	3.22±0.07	4.32±0.09	S256	5.56	2.27E-03	3.39	4.55
H257	6.31	3.33E-03	3.28	4.40	H257*	19.80			
D258	2.38	3.33E-03	2.85	3.83	D258	6.48	3.39±0.21E-06	6.28±0.03	8.43±0.04
N259	4.30	3.33E-03	3.11	4.17	N259	6.88	2.27E-03	3.48	4.67
H260	6.60	3.33E-03	3.30	4.42	H260	20.70	2.69±0.12E-06	6.89±0.02	9.24±0.02
R261	5.69	3.33E-03	3.23	4.34	R261	15.20	2.27E-03	3.83	5.13
L262	0.92	3.33E-03	2.44	3.27	L262	1.45	5.24±0.16E-06	5.44±0.01	7.30±0.02
K263	1.18	3.33E-03	2.55	3.42	K263	1.86	2.27E-03	2.91	3.91
K264	2.52	3.33E-03	2.88	3.86	K264	3.98	2.27E-03	3.24	4.35
V265	0.55	4.58±0.83E-07	6.08±0.08	8.16±0.1	V265	0.87	2.69±0.11E-06	5.51±0.02	7.39±0.02
I266	0.28	1.04E-08	7.44	9.97	I266	0.45	2.60E-08	7.24	9.70
Q267	1.42	3.60±0.10E-06	5.60±0.01	7.51±0.02	Q267	2.24	1.40±0.02E-05	5.21±0.01	6.98±0.01
D268	1.68	1.06±0.09E-06	6.20±0.04	8.32±0.05	D268	2.69	4.00±0.20E-06	5.83±0.02	7.82±0.03

C269	4.94	1.47±0.11E-07	7.53±0.03	10.09±0.04	C269	7.90	1.48±0.08E-06	6.73±0.02	9.02±0.03
E270	1.88	7.49±0.56E-07	6.40±0.03	8.58±0.04	E270	3.00	3.70±0.08E-06	5.91±0.01	7.93±0.01
D271	0.75	1.10±0.07E-04	3.84±0.03	5.14±0.04	D271	1.22	1.55±0.14E-04	3.90±0.04	5.23±0.05
E272	0.43	1.15±0.10E-05	4.58±0.04	6.14±0.05	E272	0.70	2.99±0.7E-05	4.37±0.01	5.86±0.01
N273	4.61	1.97±0.14E-04	4.37±0.03	5.86±0.04	N273	7.33	2.27E-03	3.51	4.71
I274	0.82	9.23±0.71E-06	4.95±0.03	6.64±0.05	I274	1.29	2.48±0.05E-05	4.72±0.01	6.33±0.01
Q275	1.42	2.23±0.45E-06	5.80±0.09	7.79±0.12	Q275	2.24	2.22±0.20E-06	6.00±0.04	8.05±0.05
R276	4.00	1.58±0.08E-06	6.40±0.02	8.59±0.03	R276	6.31	2.60E-08	8.38	11.25
F277	2.00	1.04E-08	8.28	11.11	F277	3.16	2.60E-08	8.08	10.84
S278	5.65	1.04E-08	8.73	11.72	S278	8.91	2.60E-08	8.53	11.45
I279	0.78	1.04E-08	7.87	10.56	I279*	1.23			
A280	1.24	1.04E-08	8.08	10.83	A280	1.95	2.60E-08	7.87	10.56
I281	0.39	6.90±1.37E-07	5.75±0.09	7.72±0.12	I281	0.62	2.60E-08	7.37	9.89
L282	0.33	4.82±0.38E-07	5.83±0.03	7.82±0.05	L282*	0.51			
G283	1.56	4.44±0.28E-05	4.55±0.03	6.10±0.04	G283*	3.80			
S284	8.16	3.33E-03	3.39	4.55	S284*	11.50			
Y285	2.25	3.33E-03	2.83	3.80	Y285*	3.55			
N286	7.28	3.33E-03	3.34	4.48	N286*	11.50			
R287	8.16	3.33E-03	3.39	4.55	R287*	8.31			
G288	5.78	3.33E-03	3.24	4.35	G288*	10.20			
N289	9.59	3.33E-03	3.46	4.64	N289*	15.10			
L290	1.15	3.33E-03	2.54	3.40	L290*	1.82			
S291	3.03	3.33E-03	2.96	3.97	S291	4.78	2.27E-03	3.32	4.46

T292	3.56	3.33E-03	3.03	4.06	T292	5.62	2.27E-03	3.39	4.55
E293	1.03	3.33E-03	2.49	3.34	E293*	1.65			
K294	1.36	3.33E-03	2.61	3.50	K294	2.16	2.27E-03	2.98	4.00
F295	1.59	3.33E-03	2.68	3.59	F295	2.51	4.77±0.58E-05	4.72±0.05	6.33±0.07
V296	0.48	1.63±0.71E-05	4.47±0.02	6.00±0.03	V296	0.76	2.27E-03	2.52	3.39
E297	0.47	8.23±0.33E-06	4.76±0.02	6.38±0.02	E297	0.75	2.27E-03	2.52	3.38
E298	0.46	5.96±0.21E-06	4.89±0.02	6.56±0.02	E298	0.75	3.69±0.45E-05	4.31±0.05	5.78±0.07
I299	0.28	2.76±0.27E-08	6.00±0.04	8.05±0.06	I299	0.44	2.27E-03	2.29	3.07
K300	1.13	1.04E-08	8.04	10.78	K300	1.78	1.14±0.07E-05	5.19±0.03	6.97±0.04
S301	6.48	4.23±0.10E-06	6.19±0.01	8.30±0.01	S301	10.20	3.62±0.50E-04	4.45±0.06	5.97±0.08
I302*	0.78				I302	1.23	1.19±0.03E-05	5.02±0.01	6.73±0.02
A303	1.24	1.04E-08	8.08	10.83	A303	1.95	1.21±0.08E-06	6.21±0.03	8.32±0.03
S304*	4.92				S304*	7.76			
E305	1.30	3.33E-03	2.59	3.48	E305	2.08	2.27E-03	2.96	3.97
P306*					P306*				
T307	1.03	3.33E-03	2.49	3.34	T307*	11.20			
E308	1.03	3.33E-03	2.49	3.34	E308	1.65	2.27E-03	2.86	3.84
K309	1.36	3.33E-03	2.61	3.50	K309	2.16	1.14±0.03E-05	5.28±0.01	7.08±0.01
H310	4.17	6.73±0.14E-06	5.79±0.01	7.77±0.01	H310	13.10	1.13±0.03E-05	6.07±0.01	8.14±0.02
F311	2.72	1.04E-08	8.42	11.29	F311	7.27	2.06±0.04E-06	6.55±0.01	8.78±0.01
F312	1.39	1.04E-08	8.13	10.90	F312	2.19	4.32±0.29E-05	4.71±0.03	6.31±0.04
N313	7.45	3.33E-03	3.35	4.49	N313	11.70	2.27E-03	3.71	4.98
V314	0.88	1.04E-08	7.92	10.63	V314	1.38	2.20±0.22E-06	5.80±0.04	7.78±0.06

S315	3.56	3.33E-03	3.03	4.06	S315	5.62	2.27E-03	3.39	4.55
D316	2.11	1.27±0.03E-04	4.22±0.01	5.66±0.01	D316	3.39	2.27E-03	3.17	4.26
E317	0.43	1.48±0.06E-04	3.47±0.02	4.65±0.02	A317	2.23	2.27E-03	2.99	4.01
L318	0.39	3.33E-03	2.07	2.78	L318	0.87	2.27E-03	2.58	3.47
A319	1.29	2.35±0.06E-05	4.74±0.01	6.36±0.01	A319	2.04	6.22±0.52E-04	3.52±0.04	4.72±0.05
L320	0.55	2.63±0.80E-04	3.32±0.13	4.46±0.18	L320	0.87	2.27E-03	2.58	3.47
V321	0.26	6.10±0.76E-04	2.63±0.05	3.53±0.07	V321	0.41	2.27E-03	2.26	3.02
T322	1.29	3.33E-03	2.59	3.47	T322	2.04	2.27E-03	2.95	3.96
I323	0.62	9.74±0.57E-05	3.80±0.03	5.10±0.03	I323	0.98	1.39±0.21E-04	3.85±0.07	5.16±0.09
V324	0.25	3.33E-03	1.87	2.51	V324	0.39	4.97±0.28E-06	4.89±0.02	6.56±0.03
K325	1.39	3.33E-03	2.62	3.52	K325	2.19	2.27E-03	2.98	4.00
T326	2.35	3.33E-03	2.85	3.82	T326	3.71	2.27E-03	3.21	4.31
L327	0.88	2.30±0.07E-06	5.58±0.01	7.49±0.02	L327	1.38	5.33±0.18E-05	4.41±0.01	5.92±0.02
G328	2.41	2.98±0.09E-05	4.91±0.01	6.58±0.02	G328	3.80	2.27E-03	3.22	4.32
E329	0.96	5.38±0.11E-05	4.25±0.01	5.71±0.01	E329	1.54	2.27E-03	2.83	3.80
R330	1.79	2.92±0.12E-05	4.79±0.02	6.42±0.02	R330	2.85	2.27E-03	3.10	4.16
I331	0.65	9.00±0.45E-06	4.86±0.02	6.52±0.03	I331	1.02	2.27E-03	2.65	3.56
F332	0.71	3.96±0.05E-05	4.25±0.01	5.71±0.01	F332	1.12	1.96±0.21E-06	5.76±0.05	7.72±0.06
A333	2.41	1.96±0.15E-03	3.09±0.03	4.14±0.04	A333*	3.80			
L334	0.55	3.33E-03	2.22	2.98	L334*	0.87			
E335	0.01	3.33E-03	0.32	0.42	E335*	0.01			

* Kinetic-undetermined residue

References

1. Di Lullo, G.A., Sweeney, S.M., Korkko, J., Ala-Kokko, L. & San Antonio, J.D. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *J Biol Chem* **277**, 4223-31 (2002).
2. Prockop, D.J. & Kivirikko, K.I. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* **64**, 403-34 (1995).
3. Kadler, K.E., Baldock, C., Bella, J. & Boot-Handford, R.P. Collagens at a glance. *J Cell Sci* **120**, 1955-8 (2007).
4. Kadler, K. Extracellular matrix. 1: fibril-forming collagens. *Protein Profile* **1**, 519-638 (1994).
5. Kadler, K. Extracellular matrix 1: Fibril-forming collagens. *Protein Profile* **2**, 491-619 (1995).
6. Petruska, J.A. & Hodge, A.J. A Subunit Model for the Tropocollagen Macromolecule. *Proc Natl Acad Sci U S A* **51**, 871-6 (1964).
7. Hulmes, D.J., Wess, T.J., Prockop, D.J. & Fratzl, P. Radial packing, order, and disorder in collagen fibrils. *Biophys J* **68**, 1661-70 (1995).
8. Giudici, C. et al. Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. *J Biol Chem* **283**, 19551-60 (2008).
9. San Antonio, J.D., Lander, A.D., Karnovsky, M.J. & Slayter, H.S. Mapping the heparin-binding sites on type I collagen monomers and fibrils. *J Cell Biol* **125**, 1179-88 (1994).
10. Wang, H., Fertala, A., Ratner, B.D., Sage, E.H. & Jiang, S. Identifying the SPARC binding sites on collagen I and procollagen I by atomic force microscopy. *Anal Chem* **77**, 6765-71 (2005).
11. Farndale, R.W. et al. Cell-collagen interactions: the use of peptide Toolkits to investigate collagen-receptor interactions. *Biochem Soc Trans* **36**, 241-50 (2008).
12. Bella, J., Eaton, M., Brodsky, B. & Berman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **266**, 75-81 (1994).
13. Bella, J. Collagen structure: new tricks from a very old dog. *Biochem J* **473**, 1001-25 (2016).
14. Okuyama, K., Wu, G., Jiravanichanun, N., Hongo, C. & Noguchi, K. Helical twists of collagen model peptides. *Biopolymers* **84**, 421-32 (2006).
15. Orgel, J.P., Irving, T.C., Miller, A. & Wess, T.J. Microfibrillar structure of type I collagen in situ. *Proc Natl Acad Sci U S A* **103**, 9001-5 (2006).
16. Streeter, I. & de Leeuw, N.H. Atomistic modeling of collagen proteins in their fibrillar environment. *J Phys Chem B* **114**, 13263-70 (2010).
17. Streeter, I. & de Leeuw, N.H. A molecular dynamics study of the interprotein interactions in collagen fibrils. *Soft Matter* **7**, 3373-3382 (2011).
18. Varma, S. et al. Effect of intrinsic and extrinsic factors on the simulated D-band length of type I collagen. *Proteins* **83**, 1800-12 (2015).
19. Nair, A.K., Gautieri, A., Chang, S.W. & Buehler, M.J. Molecular mechanics of mineralized collagen fibrils in bone. *Nat Commun* **4**, 1724 (2013).

20. Orgel, J.P., San Antonio, J.D. & Antipova, O. Molecular and structural mapping of collagen fibril interactions. *Connect Tissue Res* **52**, 2-17 (2011).
21. Herr, A.B. & Farndale, R.W. Structural insights into the interactions between platelet receptors and fibrillar collagen. *J Biol Chem* **284**, 19781-5 (2009).
22. Hoop, C.L., Zhu, J., Nunes, A.M., Case, D.A. & Baum, J. Revealing Accessibility of Cryptic Protein Binding Sites within the Functional Collagen Fibril. *Biomolecules* **7**(2017).
23. Sampath, T.K., Wientroub, S. & Reddi, A.H. Extracellular matrix proteins involved in bone induction are vitamin D dependent. *Biochem Biophys Res Commun* **124**, 829-35 (1984).
24. Siljander, P. & Lassila, R. Studies of adhesion-dependent platelet activation: distinct roles for different participating receptors can be dissociated by proteolysis of collagen. *Arterioscler Thromb Vasc Biol* **19**, 3033-43 (1999).
25. Zaman, M.H. et al. Migration of tumor cells in 3D matrices is governed by matrix stiffness along with cell-matrix adhesion and proteolysis. *Proc Natl Acad Sci U S A* **103**, 10889-94 (2006).
26. Vogel, W.F. Collagen-receptor signaling in health and disease. *Eur J Dermatol* **11**, 506-14 (2001).
27. Heino, J. The collagen family members as cell adhesion proteins. *Bioessays* **29**, 1001-10 (2007).
28. Jin, H. & Varner, J. Integrins: roles in cancer development and as treatment targets. *Br J Cancer* **90**, 561-5 (2004).
29. Ushiki, T. Collagen fibers, reticular fibers and elastic fibers. A comprehensive understanding from a morphological viewpoint. *Arch Histol Cytol* **65**, 109-26 (2002).
30. Mouw, J.K., Ou, G. & Weaver, V.M. Extracellular matrix assembly: a multiscale deconstruction. *Nat Rev Mol Cell Biol* **15**, 771-85 (2014).
31. Kadler, K.E. Fell Muir Lecture: Collagen fibril formation in vitro and in vivo. *Int J Exp Pathol* **98**, 4-16 (2017).
32. Hulmes, D.J.S. Building Collagen Molecules, Fibrils, and Suprafibrillar Structures. *Journal of Structural Biology* **137**, 2-10 (2002).
33. Silver, F.H., Christiansen, D.L., Snowhill, P.B. & Chen, Y. Role of storage on changes in the mechanical properties of tendon and self-assembled collagen fibers. *Connect Tissue Res* **41**, 155-64 (2000).
34. Kato, Y.P. et al. Mechanical properties of collagen fibres: a comparison of reconstituted and rat tail tendon fibres. *Biomaterials* **10**, 38-42 (1989).
35. Rho, J.Y., Kuhn-Spearing, L. & Zioupos, P. Mechanical properties and the hierarchical structure of bone. *Med Eng Phys* **20**, 92-102 (1998).
36. Rosenblatt, J., Devereux, B. & Wallace, D.G. Injectable collagen as a pH-sensitive hydrogel. *Biomaterials* **15**, 985-95 (1994).
37. Achilli, M., Lagueux, J. & Mantovani, D. On the effects of UV-C and pH on the mechanical behavior, molecular conformation and cell viability of collagen-based scaffold for vascular tissue engineering. *Macromol Biosci* **10**, 307-16 (2010).
38. Mihai, C., Iscru, D.F., Druhan, L.J., Elton, T.S. & Agarwal, G. Discoidin domain

receptor 2 inhibits fibrillogenesis of collagen type 1. *J Mol Biol* **361**, 864-76 (2006).
39. Svensson, L. et al. Fibromodulin-null mice have abnormal collagen fibrils, tissue organization, and altered lumican deposition in tendon. *J Biol Chem* **274**, 9636-47 (1999).
40. Chakravarti, S. et al. Lumican regulates collagen fibril assembly: skin fragility and corneal opacity in the absence of lumican. *J Cell Biol* **141**, 1277-86 (1998).
41. Reed, C.C. & Iozzo, R.V. The role of decorin in collagen fibrillogenesis and skin homeostasis. *Glycoconj J* **19**, 249-55 (2002).
42. Kalamajski, S. & Oldberg, A. The role of small leucine-rich proteoglycans in collagen fibrillogenesis. *Matrix Biology* **29**, 248-253 (2010).
43. Holster, T. et al. Loss of assembly of the main basement membrane collagen, type IV, but not fibril-forming collagens and embryonic death in collagen prolyl 4-hydroxylase I null mice. *J Biol Chem* **282**, 2512-9 (2007).
44. Pornprasertsuk, S., Duarte, W.R., Mochida, Y. & Yamauchi, M. Overexpression of lysyl hydroxylase-2b leads to defective collagen fibrillogenesis and matrix mineralization. *J Bone Miner Res* **20**, 81-7 (2005).
45. Thomson, C.A. & Ananthanarayanan, V.S. Structure-function studies on hsp47: pH-dependent inhibition of collagen fibril formation in vitro. *Biochem J* **349 Pt 3**, 877-83 (2000).
46. Steffensen, B., Hakkinen, L. & Larjava, H. Proteolytic events of wound-healing--coordinated interactions among matrix metalloproteinases (MMPs), integrins, and extracellular matrix molecules. *Crit Rev Oral Biol Med* **12**, 373-98 (2001).
47. Jokinen, J. et al. Integrin-mediated cell adhesion to type I collagen fibrils. *J Biol Chem* **279**, 31956-63 (2004).
48. Li, S., Van Den Diepstraten, C., D'Souza, S.J., Chan, B.M. & Pickering, J.G. Vascular smooth muscle cells orchestrate the assembly of type I collagen via alpha2beta1 integrin, RhoA, and fibronectin polymerization. *Am J Pathol* **163**, 1045-56 (2003).
49. Kadler, K.E., Hill, A. & Canty-Laird, E.G. Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators. *Curr Opin Cell Biol* **20**, 495-501 (2008).
50. Hynes, R.O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673-87 (2002).
51. Tashiro, K. et al. An IKLLI-containing peptide derived from the laminin alpha1 chain mediating heparin-binding, cell adhesion, neurite outgrowth and proliferation, represents a binding site for integrin alpha3beta1 and heparan sulphate proteoglycan. *Biochem J* **340 (Pt 1)**, 119-26 (1999).
52. van der Flier, A. & Sonnenberg, A. Function and interactions of integrins. *Cell Tissue Res* **305**, 285-98 (2001).
53. Loftus, J.C. & Liddington, R.C. Cell adhesion in vascular biology. New insights into integrin-ligand interaction. *J Clin Invest* **99**, 2302-6 (1997).
54. Loftus, J.C. & Liddington, R.C. New insights into integrin-ligand interaction. *J Clin Invest* **100**, S77-81 (1997).

55. Cheah, M. & Andrews, M.R. Integrin Activation: Implications for Axon Regeneration. *Cells* **7**(2018).
56. Siljander, P.R. et al. Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. *J Biol Chem* **279**, 47763-72 (2004).
57. Emsley, J., King, S.L., Bergelson, J.M. & Liddington, R.C. Crystal structure of the I domain from integrin alpha2beta1. *J Biol Chem* **272**, 28512-7 (1997).
58. Lee, J.O., Rieu, P., Arnaout, M.A. & Liddington, R. Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell* **80**, 631-8 (1995).
59. Nymalm, Y. et al. Jararhagin-derived RKKH peptides induce structural changes in alpha1I domain of human integrin alpha1beta1. *J Biol Chem* **279**, 7962-70 (2004).
60. Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J. & Liddington, R.C. Structural basis of collagen recognition by integrin alpha2beta1. *Cell* **101**, 47-56 (2000).
61. Kamata, T., Liddington, R.C. & Takada, Y. Interaction between collagen and the alpha(2) I-domain of integrin alpha(2)beta(1). Critical role of conserved residues in the metal ion-dependent adhesion site (MIDAS) region. *J Biol Chem* **274**, 32108-11 (1999).
62. Chin, Y.K. et al. The structure of integrin alpha1I domain in complex with a collagen-mimetic peptide. *J Biol Chem* **288**, 36796-809 (2013).
63. Yang, W., Shimaoka, M., Salas, A., Takagi, J. & Springer, T.A. Intersubunit signal transmission in integrins by a receptor-like interaction with a pull spring. *Proc Natl Acad Sci U S A* **101**, 2906-11 (2004).
64. Carafoli, F., Hamaia, S.W., Bihan, D., Hohenester, E. & Farndale, R.W. An activating mutation reveals a second binding mode of the integrin alpha2 I domain to the GFOGER motif in collagens. *PLoS One* **8**, e69833 (2013).
65. Raynal, N. et al. Use of synthetic peptides to locate novel integrin alpha2beta1-binding motifs in human collagen III. *J Biol Chem* **281**, 3821-31 (2006).
66. Knight, C.G. et al. The collagen-binding A-domains of integrins alpha(1)beta(1) and alpha(2)beta(1) recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. *J Biol Chem* **275**, 35-40 (2000).
67. Aquilina, A. et al. A novel gain-of-function mutation of the integrin alpha2 VWFA domain. *Eur J Biochem* **269**, 1136-44 (2002).
68. Lahti, M. et al. Structure of Collagen Receptor Integrin alpha1I Domain Carrying the Activating Mutation E317A. *J Biol Chem* **286**, 43343-51 (2011).
69. Racine-Samson, L., Rockey, D.C. & Bissell, D.M. The role of alpha1beta1 integrin in wound contraction. A quantitative analysis of liver myofibroblasts in vivo and in primary culture. *J Biol Chem* **272**, 30911-7 (1997).
70. Jenkins, G., Redwood, K.L., Meadows, L. & Green, M.R. Effect of gel re-organization and tensional forces on alpha2beta1 integrin levels in dermal fibroblasts. *Eur J Biochem* **263**, 93-103 (1999).
71. Schiro, J.A. et al. Integrin alpha 2 beta 1 (VLA-2) mediates reorganization and contraction of collagen matrices by human cells. *Cell* **67**, 403-10 (1991).

72. Fleischmajer, R. et al. There is binding of collagen IV to beta 1 integrin during early skin basement membrane assembly. *Ann N Y Acad Sci* **857**, 212-27 (1998).
73. Gardner, H., Kreidberg, J., Koteliansky, V. & Jaenisch, R. Deletion of integrin alpha 1 by homologous recombination permits normal murine development but gives rise to a specific deficit in cell adhesion. *Dev Biol* **175**, 301-13 (1996).
74. Gardner, H., Broberg, A., Pozzi, A., Laato, M. & Heino, J. Absence of integrin alpha1beta1 in the mouse causes loss of feedback regulation of collagen synthesis in normal and wounded dermis. *J Cell Sci* **112 (Pt 3)**, 263-72 (1999).
75. Connolly, M. Analytical molecular surface calculation. *Journal of Applied Crystallography* **16**, 548-558 (1983).
76. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-13 (1983).
77. Rosano, G.L. & Ceccarelli, E.A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172 (2014).
78. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302 (1999).
79. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-23 (2009).
80. Molday, R.S., Englander, S.W. & Kallen, R.G. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150-8 (1972).
81. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86 (1993).
82. Connelly, G.P., Bai, Y., Jeng, M.F. & Englander, S.W. Isotope effects in peptide group hydrogen exchange. *Proteins* **17**, 87-92 (1993).
83. Fitzkee, N.C., Torchia, D.A. & Bax, A. Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain. *Protein Sci* **20**, 500-12 (2011).
84. Greenfield, N.J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols* **1**, 2527 - 2535 (2006).
85. Zhu, J. & Kaufman, L.J. Collagen I self-assembly: revealing the developing structures that generate turbidity. *Biophys J* **106**, 1822-31 (2014).
86. Sweeney, S.M. et al. Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J Biol Chem* **283**, 21187-97 (2008).
87. Springer, T.A. & Wang, J.H. The three-dimensional structure of integrins and their ligands, and conformational regulation of cell adhesion. *Adv Protein Chem* **68**, 29-63 (2004).
88. Barczyk, M., Carracedo, S. & Gullberg, D. Integrins. *Cell Tissue Res* **339**, 269-80 (2010).
89. Keene, D.R. et al. Decorin binds near the C terminus of type I collagen. *J Biol Chem* **275**, 21801-4 (2000).

90. Xu, Y. et al. Multiple binding sites in collagen type I for the integrins alpha1beta1 and alpha2beta1. *J Biol Chem* **275**, 38981-9 (2000).
91. Lisman, T. et al. A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* **108**, 3753-6 (2006).
92. Yu, Z., Visse, R., Inouye, M., Nagase, H. & Brodsky, B. Defining requirements for collagenase cleavage in collagen type III using a bacterial collagen system. *J Biol Chem* **287**, 22988-97 (2012).
93. An, B. et al. Definition of the native and denatured type II collagen binding site for fibronectin using a recombinant collagen system. *J Biol Chem* **289**, 4941-51 (2014).
94. Peng, Y.Y., Stoichevska, V., Schacht, K., Werkmeister, J.A. & Ramshaw, J.A. Engineering multiple biological functional motifs into a blank collagen-like protein template from Streptococcus pyogenes. *J Biomed Mater Res A* **102**, 2189-96 (2014).
95. Knight, C.G. et al. Identification in collagen type I of an integrin alpha2 beta1-binding site containing an essential GER sequence. *J Biol Chem* **273**, 33287-94 (1998).
96. Wess, T.J. & Orgel, J.P. Changes in collagen structure: drying, dehydrothermal treatment and relation to long term deterioration. *Thermochimica Acta* **365**, 119-128 (2000).
97. Hulmes, D.J. & Miller, A. Quasi-hexagonal molecular packing in collagen fibrils. *Nature* **282**, 878-80 (1979).
98. Perumal, S., Antipova, O. & Orgel, J.P. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proc Natl Acad Sci U S A* **105**, 2824-9 (2008).
99. Raspanti, M., Alessandrini, A., Gobbi, P. & Ruggeri, A. Collagen fibril surface: TMAFM, FEG-SEM and freeze-etching observations. *Microsc Res Tech* **35**, 87-93 (1996).
100. Hulmes, D.J., Jesior, J.C., Miller, A., Berthet-Colominas, C. & Wolff, C. Electron microscopy shows periodic structure in collagen fibril cross sections. *Proc Natl Acad Sci U S A* **78**, 3567-71 (1981).
101. Orgel, J.P., Eid, A., Antipova, O., Bella, J. & Scott, J.E. Decorin core protein (decoron) shape complements collagen fibril surface structure and mediates its binding. *PLoS One* **4**, e7028 (2009).
102. Xu, Z., Zhao, W., Wang, Z., Yang, Y. & Sahai, N. Structure analysis of collagen fibril at atomic-level resolution and its implications for intra-fibrillar transport in bone biomineralization. *Phys Chem Chem Phys* **20**, 1513-1523 (2018).
103. Rainey, J.K. & Goh, M.C. An interactive triple-helical collagen builder. *Bioinformatics* **20**, 2458-9 (2004).
104. Krivov, G.G., Shapovalov, M.V. & Dunbrack, R.L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778-95 (2009).
105. Case, D.A. et al. *AMBER 2017* (2017).
106. Maier, J.A. et al. ff14SB: Improving the Accuracy of Protein Side Chain and

Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-713 (2015).
107. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).
108. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. & Haak, J.R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690 (1984).
109. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
110. Hopkins, C.W., Le Grand, S., Walker, R.C. & Roitberg, A.E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **11**, 1864-74 (2015).
111. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-8, 27-8 (1996).
112. Roe, D.R. & Cheatham, T.E., 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**, 3084-95 (2013).
113. Grant, C.A., Brockwell, D.J., Radford, S.E. & Thomson, N.H. Effects of hydration on the mechanical response of individual collagen fibrils. *Applied Physics Letters* **92**, 233902 (2008).
114. Spitzner, E.C., Roper, S., Zerson, M., Bernstein, A. & Magerle, R. Nanoscale Swelling Heterogeneities in Type I Collagen Fibrils. *ACS Nano* **9**, 5683-94 (2015).
115. Engel, J. & Prockop, D.J. Does bound water contribute to the stability of collagen? *Matrix Biol* **17**, 679-80 (1998).
116. Kadler, K.E., Hojima, Y. & Prockop, D.J. Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen self-assembly is a classical example of an entropy-driven process. *J Biol Chem* **262**, 15696-701 (1987).
117. Kadler, K.E., Holmes, D.F., Trotter, J.A. & Chapman, J.A. Collagen fibril formation. *Biochem J* **316 (Pt 1)**, 1-11 (1996).
118. McBride, D.J., Jr., Kadler, K.E., Hojima, Y. & Prockop, D.J. Self-assembly into fibrils of a homotrimer of type I collagen. *Matrix* **12**, 256-63 (1992).
119. Strasser, S., Zink, A., Janko, M., Heckl, W.M. & Thalhammer, S. Structural investigations on native collagen type I fibrils using AFM. *Biochem Biophys Res Commun* **354**, 27-32 (2007).
120. Gutsmann, T. et al. Evidence that collagen fibrils in tendons are inhomogeneously structured in a tubelike manner. *Biophys J* **84**, 2593-8 (2003).
121. Mosler, E. et al. Stress-induced molecular rearrangement in tendon collagen. *J Mol Biol* **182**, 589-96 (1985).
122. Sarkar, S.K. et al. Molecular dynamics of collagen side chains in hard and soft tissues. A multinuclear magnetic resonance study. *Biochemistry* **26**, 6793-800 (1987).
123. Drake, B. et al. Imaging crystals, polymers, and processes in water with the

atomic force microscope. *Science* **243**, 1586-9 (1989).
124. Muller, D.J. & Dufrene, Y.F. Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. *Nat Nanotechnol* **3**, 261-9 (2008).
125. Radmacher, M., Tillamnn, R.W., Fritz, M. & Gaub, H.E. From molecules to cells: imaging soft samples with the atomic force microscope. *Science* **257**, 1900-5 (1992).
126. Kodera, N., Yamamoto, D., Ishikawa, R. & Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature* **468**, 72-6 (2010).
127. Ando, T., Uchihashi, T. & Kodera, N. High-speed AFM and applications to biomolecular systems. *Annu Rev Biophys* **42**, 393-414 (2013).
128. Nerenberg, P.S. & Stultz, C.M. Differential unfolding of alpha1 and alpha2 chains in type I collagen and collagenolysis. *J Mol Biol* **382**, 246-56 (2008).
129. Siebert, H.C. et al. Interaction of the alpha2A domain of integrin with small collagen fragments. *Protein Cell* **1**, 393-405 (2010).
130. Nunes, A.M. et al. Intrinsic local destabilization of the C-terminus predisposes integrin alpha1 I domain to a conformational switch induced by collagen binding. *Protein Sci* **25**, 1672-81 (2016).
131. Luo, B.H., Carman, C.V. & Springer, T.A. Structural basis of integrin regulation and signaling. *Annu Rev Immunol* **25**, 619-47 (2007).
132. Gardner, H. Integrin alpha1beta1. *Adv Exp Med Biol* **819**, 21-39 (2014).
133. Tulla, M. et al. Effects of conformational activation of integrin alpha 1I and alpha 2I domains on selective recognition of laminin and collagen subtypes. *Exp Cell Res* **314**, 1734-43 (2008).
134. Lahti, M. et al. Structure of collagen receptor integrin alpha(1)I domain carrying the activating mutation E317A. *J Biol Chem* **286**, 43343-51 (2011).
135. Tzeng, S.R. & Kalodimos, C.G. Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* **21**, 62-7 (2011).
136. Palmer, A.G., 3rd. Enzyme dynamics from NMR spectroscopy. *Acc Chem Res* **48**, 457-65 (2015).
137. Kay, L.E. New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. *J Mol Biol* **428**, 323-331 (2016).
138. Lisi, G.P. & Loria, J.P. Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem Rev* **116**, 6323-69 (2016).
139. Nussinov, R., Ma, B. & Tsai, C.J. Multiple conformational selection and induced fit events take place in allosteric propagation. *Biophys Chem* **186**, 22-30 (2014).
140. Englander, S.W. & Mayne, L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu Rev Biophys Biomol Struct* **21**, 243-65 (1992).
141. Bai, Y. & Englander, S.W. Hydrogen bond strength and beta-sheet propensities: the role of a side chain blocking effect. *Proteins* **18**, 262-6 (1994).
142. Berjanskii, M. & Wishart, D.S. NMR: prediction of protein flexibility. *Nat Protoc* **1**, 683-8 (2006).
143. Mittermaier, A.K. & Kay, L.E. Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* **34**, 601-11 (2009).

144. Skinner, J.J., Lim, W.K., Bedard, S., Black, B.E. & Englander, S.W. Protein dynamics viewed by hydrogen exchange. *Protein Sci* **21**, 996-1005 (2012).
145. Chin, Y.K. et al. Assignments of human integrin alpha1I domain in the apo and Mg(2)(+) bound states. *Biomol NMR Assign* **8**, 117-21 (2014).
146. Wishart, D.S. & Sykes, B.D. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* **4**, 171-80 (1994).
147. Rich, R.L. et al. Trench-shaped binding sites promote multiple classes of interactions between collagen and the adherence receptors, alpha(1)beta(1) integrin and Staphylococcus aureus cna MSCRAMM. *J Biol Chem* **274**, 24906-13 (1999).
148. Salminen, T.A. et al. Production, crystallization and preliminary X-ray analysis of the human integrin alpha1 I domain. *Acta Crystallogr D Biol Crystallogr* **55**, 1365-7 (1999).
149. Mulder, F.A., Schipper, D., Bott, R. & Boelens, R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J Mol Biol* **292**, 111-23 (1999).
150. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).
151. Englander, S.W., Mayne, L. & Krishna, M.M. Protein folding and misfolding: mechanism and principles. *Q Rev Biophys* **40**, 287-326 (2007).
152. Zhang, Y.-Z. Protein and peptide structure and interactions studied by hydrogen exchange and NMR. *Thesis, University of Pennsylvania, PA, USA* (1995).
153. Ahmad, S., Gromiha, M., Fawareh, H. & Sarai, A. ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics* **5**, 51 (2004).
154. Huyghues-Despointes, B.M., Scholtz, J.M. & Pace, C.N. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol* **6**, 910-2 (1999).
155. Huyghues-Despointes, B.M., Pace, C.N., Englander, S.W. & Scholtz, J.M. Measuring the conformational stability of a protein by hydrogen exchange. *Methods Mol Biol* **168**, 69-92 (2001).
156. Qu, A. & Leahy, D.J. Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha L beta 2) integrin. *Proc Natl Acad Sci U S A* **92**, 10277-81 (1995).
157. Huth, J.R. et al. NMR and mutagenesis evidence for an I domain allosteric site that regulates lymphocyte function-associated antigen 1 ligand binding. *Proc Natl Acad Sci U S A* **97**, 5231-6 (2000).
158. Xiong, J.P., Li, R., Essafi, M., Stehle, T. & Arnaout, M.A. An isoleucine-based allosteric switch controls affinity and shape shifting in integrin CD11b A-domain. *J Biol Chem* **275**, 38762-7 (2000).
159. Valdramidou, D., Humphries, M.J. & Mould, A.P. Distinct roles of beta1 metal ion-dependent adhesion site (MIDAS), adjacent to MIDAS (ADMIDAS), and ligand-associated metal-binding site (LIMBS) cation-binding sites in ligand recognition by integrin alpha2beta1. *J Biol Chem* **283**, 32704-14 (2008).
160. Chin, Y.K. et al. The Structure of Integrin alpha1I Domain in Complex with a

Collagen Mimetic Peptide. *J Biol Chem* (2013).

161. Emsley, J., Knight, C.G., Farndale, R.W. & Barnes, M.J. Structure of the integrin alpha2beta1-binding collagen peptide. *J Mol Biol* **335**, 1019-28 (2004).
162. Shi, M. et al. Enhancing integrin alpha1 inserted (I) domain affinity to ligand potentiates integrin alpha1beta1-mediated down-regulation of collagen synthesis. *J Biol Chem* **287**, 35139-52 (2012).
163. Hamaia, S.W. et al. Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin alpha1beta1 using collagen toolkits II and III. *J Biol Chem* **287**, 26019-28 (2012).
164. Li, Y., Brodsky, B. & Baum, J. NMR conformational and dynamic consequences of a gly to ser substitution in an osteogenesis imperfecta collagen model peptide. *J Biol Chem* **284**, 20660-7 (2009).
165. Xiao, J., Addabbo, R.M., Lauer, J.L., Fields, G.B. & Baum, J. Local conformation and dynamics of isoleucine in the collagenase cleavage site provide a recognition signal for matrix metalloproteinases. *J Biol Chem* **285**, 34181-90 (2010).
166. Fu, I., Case, D.A. & Baum, J. Dynamic Water-Mediated Hydrogen Bonding in a Collagen Model Peptide. *Biochemistry* **54**, 6029-37 (2015).
167. Xiao, J., Yang, Z., Sun, X., Addabbo, R. & Baum, J. Local amino acid sequence patterns dominate the heterogeneous phenotype for the collagen connective tissue disease Osteogenesis Imperfecta resulting from Gly mutations. *J Struct Biol* **192**, 127-37 (2015).
168. Popovych, N., Sun, S., Ebright, R.H. & Kalodimos, C.G. Dynamically driven protein allostery. *Nat Struct Mol Biol* **13**, 831-8 (2006).
169. Weinreb, P.H. et al. Dynamic structural changes are observed upon collagen and metal ion binding to the integrin alpha1 I domain. *J Biol Chem* **287**, 32897-912 (2012).
170. Legge, G.B. et al. NMR solution structure of the inserted domain of human leukocyte function associated antigen-1. *J Mol Biol* **295**, 1251-64 (2000).
171. Nam, K., Maiorov, V., Feuston, B. & Kearsley, S. Dynamic control of allosteric antagonism of leukocyte function antigen-1 and intercellular adhesion molecule-1 interaction. *Proteins* **64**, 376-84 (2006).
172. Zhang, H. et al. An unusual allosteric mobility of the C-terminal helix of a high-affinity alphaL integrin I domain variant bound to ICAM-5. *Mol Cell* **31**, 432-7 (2008).
173. Leung, H.T. et al. NMR characterization of the conformational fluctuations of the human lymphocyte function-associated antigen-1 I-domain. *Protein Sci* **23**, 1596-606 (2014).
174. Dickeson, S.K., Mathis, N.L., Rahman, M., Bergelson, J.M. & Santoro, S.A. Determinants of ligand binding specificity of the alpha(1)beta(1) and alpha(2)beta(1) integrins. *J Biol Chem* **274**, 32182-91 (1999).
175. Kapyla, J. et al. Integrin alpha(2)I domain recognizes type I and type IV collagens by different mechanisms. *J Biol Chem* **275**, 3348-54 (2000).
176. Bajic, G., Yatime, L., Sim, R.B., Vorup-Jensen, T. & Andersen, G.R. Structural insight on the recognition of surface-bound opsonins by the integrin I domain of

complement receptor 3. *Proc Natl Acad Sci U S A* **110**, 16426-31 (2013).
177. Adzhubei, A.A., Sternberg, M.J. & Makarov, A.A. Polyproline-II helix in proteins: structure and function. *J Mol Biol* **425**, 2100-32 (2013).
178. Williams, B.R., Gelman, R.A., Poppke, D.C. & Piez, K.A. Collagen fibril formation. Optimal in vitro conditions and preliminary kinetic results. *J Biol Chem* **253**, 6578-85 (1978).
179. Wood, G.C. & Keech, M.K. The formation of fibrils from collagen solutions. 1. The effect of experimental conditions: kinetic and electron-microscope studies. *Biochem J* **75**, 588-98 (1960).
180. Wood, G.C. The formation of fibrils from collagen solutions. 2. A mechanism of collagen-fibril formation. *Biochem J* **75**, 598-605 (1960).
181. Wood, G.C. The formation of fibrils from collagen solutions. 3. Effect of chondroitin sulphate and some other naturally occurring polyanions on the rate of formation. *Biochem J* **75**, 605-12 (1960).
182. Djabourov, M., Lechaire, J.P. & Gaill, F. Structure and rheology of gelatin and collagen gels. *Biorheology* **30**, 191-205 (1993).
183. Gelman, R.A., Poppke, D.C. & Piez, K.A. Collagen fibril formation in vitro. The role of the nonhelical terminal regions. *J Biol Chem* **254**, 11741-5 (1979).
184. Pederson, A.W., Ruberti, J.W. & Messersmith, P.B. Thermal assembly of a biomimetic mineral/collagen composite. *Biomaterials* **24**, 4881-90 (2003).
185. Holmes, D.F. & Chapman, J.A. Axial mass distributions of collagen fibrils grown in vitro: results for the end regions of early fibrils. *Biochem Biophys Res Commun* **87**, 993-9 (1979).
186. Velling, T., Kusche-Gullberg, M., Sejersen, T. & Gullberg, D. cDNA cloning and chromosomal localization of human alpha(11) integrin. A collagen-binding, I domain-containing, beta(1)-associated integrin alpha-chain present in muscle tissues. *J Biol Chem* **274**, 25735-42 (1999).
187. Camper, L., Hellman, U. & Lundgren-Akerlund, E. Isolation, cloning, and sequence analysis of the integrin subunit alpha10, a beta1-associated collagen binding integrin expressed on chondrocytes. *J Biol Chem* **273**, 20383-9 (1998).
188. Takada, Y. & Hemler, M.E. The primary structure of the VLA-2/collagen receptor alpha 2 subunit (platelet GPIa): homology to other integrins and the presence of a possible collagen-binding domain. *J Cell Biol* **109**, 397-407 (1989).
189. de Fougerolles, A.R. et al. Regulation of inflammation by collagen-binding integrins alpha1beta1 and alpha2beta1 in models of hypersensitivity and arthritis. *J Clin Invest* **105**, 721-9 (2000).
190. Cisneros, D.A., Hung, C., Franz, C.M. & Muller, D.J. Observing growth steps of collagen self-assembly by time-lapse high-resolution atomic force microscopy. *J Struct Biol* **154**, 232-45 (2006).
191. Lai, G., Li, Y. & Li, G. Effect of concentration and temperature on the rheological behavior of collagen solution. *Int J Biol Macromol* **42**, 285-91 (2008).
192. Ravanti, L., Heino, J., Lopez-Otin, C. & Kahari, V.M. Induction of collagenase-3 (MMP-13) expression in human skin fibroblasts by three-dimensional collagen is mediated by p38 mitogen-activated protein kinase. *J Biol Chem* **274**, 2446-55

(1999).

193. Ivaska, J. et al. Integrin alpha2beta1 mediates isoform-specific activation of p38 and upregulation of collagen gene transcription by a mechanism involving the alpha2 cytoplasmic tail. *J Cell Biol* **147**, 401-16 (1999).
194. Langholz, O. et al. Collagen and collagenase gene expression in three-dimensional collagen lattices are differentially regulated by alpha 1 beta 1 and alpha 2 beta 1 integrins. *J Cell Biol* **131**, 1903-15 (1995).
195. Riikonen, T. et al. Integrin alpha 2 beta 1 is a positive regulator of collagenase (MMP-1) and collagen alpha 1(I) gene expression. *J Biol Chem* **270**, 13548-52 (1995).
196. Ekholm, E. et al. Diminished callus size and cartilage synthesis in alpha 1 beta 1 integrin-deficient mice during bone fracture healing. *Am J Pathol* **160**, 1779-85 (2002).
197. Pozzi, A., Wary, K.K., Giancotti, F.G. & Gardner, H.A. Integrin alpha1beta1 mediates a unique collagen-dependent proliferation pathway in vivo. *J Cell Biol* **142**, 587-94 (1998).
198. Tulla, M. et al. Selective binding of collagen subtypes by integrin alpha 1I, alpha 2I, and alpha 10I domains. *J Biol Chem* **276**, 48206-12 (2001).
199. Nykvist, P. et al. Distinct recognition of collagen subtypes by alpha(1)beta(1) and alpha(2)beta(1) integrins. Alpha(1)beta(1) mediates cell adhesion to type XIII collagen. *J Biol Chem* **275**, 8255-61 (2000).
200. Canty, E.G. et al. Coalignment of plasma membrane channels and protrusions (fibripositors) specifies the parallelism of tendon. *J Cell Biol* **165**, 553-63 (2004).
201. Allen, T.D., Schor, S.L. & Schor, A.M. An ultrastructural review of collagen gels, a model system for cell-matrix, cell-basement membrane and cell-cell interactions. *Scan Electron Microsc*, 375-90 (1984).