RELATION OF CONFORMATIONAL AND DYNAMIC DIFFERENCES IN COLLAGEN SEQUENCES TO ALPHA2I DOMAIN INTEGRIN BINDING AFFINITY

By

JACKY WANG

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Chemistry and Chemical Biology

Written under the direction

Professor Jean Baum

And approved by

_____

_____

_____

New Brunswick, New Jersey

May, 2015

ABSTRACT OF THE THESIS

Relation of conformational and dynamic differences in collagen sequences to alpha2I domain integrin binding affinity

By JACKY WANG

Thesis Director:

Jean Baum

The α2I domain of integrin has been identified to bind to the recognition motifs in collagen, GFOGER and GAOGER, with the former having a higher binding affinity than the latter. Here we employ nuclear magnetic resonance spectroscopy and circular dichroism spectroscopy to study the structural differences between the GFOGER and GAOGER collagen model peptides (CMP) triple helices. As a consequence of the limitations from using synthetic CMPs, we have also used recombinant collagen proteins expressed from E. coli as a way to efficiently screen mutations in CMPs and provide a platform to probe binding affinities using enzyme-linked immunosorbent assay. Results from this work show that the residues in GAOGER and GFOGER triple helices have few differences in dynamics, however dynamics between strands in a collagen triple helix may play a role in modulation of binding affinity. Attempts at more efficient analysis of varying CMPs using recombinant proteins have yielded us fusion proteins that can be used with enzyme-linked immunoabsorbent assay and produce recombinant CMPs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Collagens are extracelluar matrix fibril proteins that function as both structural support and cell adhesion in multi-cellular animals. The collagen fibril structure is a supercoil of triple helices which in turn are composed of three polypeptide strands wound together as a left-handed helix and up to 28 types of collagen amino acid sequences have been described. Unique to collagen is its (Gly-Xaa-Yaa) repeating sequence, where Xaa is often proline and Yaa is frequently found as hydroxyproline. Of critical importance to the collagen triple helix structure is the presence of Gly in every third residue; the lack of a side chain in Gly fulfills steric requirements to enable the three polypeptide strands to fold together to form a triple helix [1]. Crystal structure evidence suggests that charged residue sequences in the Xaa and Yaa positions of collagen facilitate the triple helices to self-associate through charge interactions and provide binding sites for collagen binding proteins [2,3,4]. Solid phase binding assay results show that the cell surface receptor, integrin, binds to the aforementioned charged motifs in collagen in form of GXX'GEX'', with sequence dependent affinity [5,6]. As for biological significance, the event of integrin binding to collagen begins cell-substrate adhesion which in turn cascades into variations in pathologies [7].

The GFOGER collagen model peptide (CMP)-integrin α2I crystal structure shows that the majority of GFOGER's interaction with the metal ion dependent adhesion site (MIDAS) region of the integrin α2I domain stems from the middle strand of the peptide [8]. Three residues from the middle strand of the GFOGER triple helix make significant contributions to binding; the glutamate coordinates with the magnesium in the MIDAS

region, arginine salt bridges with the α2I domain's D219, and the phenylalanine lies in a surface dimple formed by the side chains of α2I's Q215 and N154 (Figure 1).

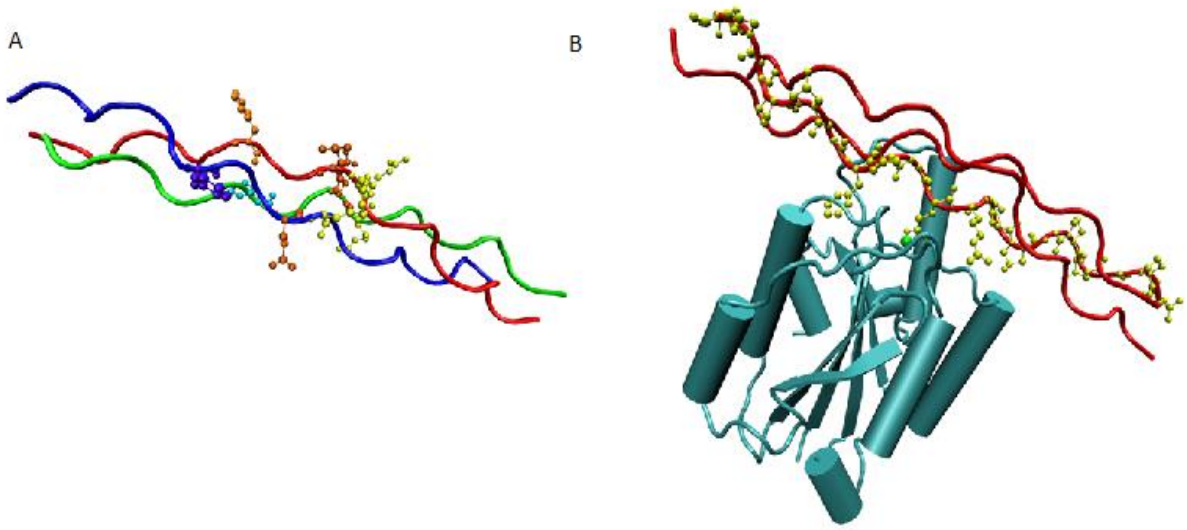


Figure 1. **(A)** Crystal structure of the triple helix GFOGER collagen model peptide. The blue strand is the leading strand, red is the middle strand and green is the trailing strand. Red CPK models represent glutamate residues, gold CPK models represent arginine residues and multi-color CPK models represent phenylalanine residues. **(B)** Crystal structure of GFOGER CMP in complex with the α2I domain. The middle strand of the triple helix is shown as a CPK model, with F14, E17, and R18 side chains extended into the MIDAS region of α2I.

However, there is a lack of evidence to explain why the α2I domain prefers one strand of the CMP peptide over other strands. Further complicating the narrative, crystal structure evidence of GFOGER in complex with the E318W α2I domain, an "activated" mutant, revealed that the α2I domain binds to two strands of the GFOGER CMP triple helix in a high affinity mode and a single strand in a low affinity mode [5]. A refined model of how collagen binds to integrin can thus provide novel therapeutics that inhibit or stimulate integrin binding to collagen.

A low resolution microfibril structure of collagen obtained from fiber diffraction has been described, yet there has been no follow up with a more detailed structure of the fibril form [9]. Consequently, characteristics of the collagen structure have not been fully explored despite decades of research due to the challenges of using full length collagen in vitro. As a compromise, collagen model peptides (CMP) are frequently used to model select regions of a full length collagen in molecular biology experiments [10,11].

Despite the availability of a high resolution crystal structures of the collagen peptide-α2I integrin complex, the exact mechanism of how integrin recognizes and binds to collagen is not fully understood [6,8]. Additionally, assay evidence shows that integrin has a higher binding affinity for GFOGER compared to other GXX'GEX" sequences, however there is a lack of evidence pertaining to why integrin has a higher affinity for one sequence over another [5,6]. Here, we employ molecular biology and biophysical techniques to probe CMP triple helices as a means to understand the conformational changes collagen undergoes before binding to integrin. Using nuclear magnetic resonance (NMR) protein and Circular Dichroism (CD) experiments we set about to test our hypothesis that differences in the conformation and dynamics of the GAOGER, low affinity for α2I, and GFOGER, high affinity, CMP triple helices dictate binding affinity.

Production of ^{15}N labeled CMPs through recombinant protein systems could be a huge boon to future attempts at modeling full length collagen through CMPs. Our previous work began by attaching the α2I binding CMP $(GPP)_4GFPGER(GPP)_5GY$ sequence to the C-terminus of the V-domain, a globular protein that facilitates trimerization; a histidine tag was attached to the N-terminus of the V-domain; and a thrombin cleavage sequence was inserted between the V-domain and the CMP [12]. The

resulting construct, V-THN-GFPGER, was inserted into the *Escherichia Coli* (E. coli) pCOLD vector to boost expression yield [13]. Initial expression attempts failed as a consequence of proline toxicity to E. coli [communication with M. Inoyue], and thus we attached the ompA signal peptide to the N-terminus of the His tag, which makes the cell secrete the protein from the cytoplasm to the periplasm to minimize the toxicity of the prolines [14]. Upon expression, the V-THN-GFPGER construct is moved to the periplasm, while the ompA peptide is cleaved, leaving only a fusion protein beginning with a His tag at the N-terminus and the GFPGER CMP at the C-terminus (Figure 2).

Figure 2. V-THN-GFPGER amino acid sequence beginning with the ompA signal peptide (purple), His tag (italicized black), V-domain (blue), thrombin cleavage site (red italicized), and GFPGER CMP (bold underline).

In this work we have successfully utilized V-THN-GFPGER construct as an enzyme-linked immunosorbent assay (ELISA) agent, however we were unable to cleave the V-domain from the GFPGER CMP. As a consequence we restructured our cleavage attempt by using Tobacco Etch Virus (TEV) protease, a very specific enzyme that can function in a variety of reaction conditions [15].

AIMS AND SIGNIFICANCE

Conformation and dynamics differences between the GFOGER and GAOGER collagen model peptides

We hypothesize that the variations in binding among integrin binding CMP triple helices are due to sequence dependent conformational shifts in the CMP that lead to a "loosening" of a single strand which precedes an integrin binding event. To investigate such dynamics, we use longitudinal relaxation (R_1), transverse relaxation (R_2), and steady state heteronuclear nuclear Overhauser effect (NOE) ^{15}N relaxation NMR experiments to measure ps to ns timescale conformational shifts in polypeptide backbone amides due to bond vibrations and sidechain rotations [16,17]. Slower dynamics, days, are investigated using hydrogen-deuterium exchange NMR experiments to measure the quantity of protection residues have from amide proton exchange with solvent [18]. We will use the aforementioned NMR experiments to examine the motion that backbone amides of individual residues undergo in collagen triple helices that bind to integrin: the high affinity $(GPO)_4GFOGER(GPO)_4GY$ sequence and the low affinity $(GPO)_4GAOGER(GPO)_4GY$ sequence.

Development of a recombinant collagen model peptide for NMR experiments

The large sizes of collagens (300-400 kD) make them unwieldy for solution NMR, since line broadening increases with increasing molecular weight due to poor tumbling in solution, which in turn complicates assignments of resonances [19]. Consequently, NMR studies of collagen are conducted through CMPs with selectively

labeled residues to model regions of collagen as seen throughout decades of research [20,21,22]. Chemical synthesis of isotopically labeled CMPs is expensive in both time and resources, thus we will attempt bio-production by expressing recombinant CMPs with unconventional biotechnology.

MATERIALS AND METHODS

NMR and CD Sample Preperation

Peptide Ac-$(GPO)_4$GFOGER$(GPO)_4$GY-$CONH_2$ with G7, G13, and F14 ^{15}N labels; designated as GFOGER; was synthesized by Biomer Technology (Pleasanton, CA). Peptide Ac-$(GPO)_4$GAOGER$(GPO)_4$GY-$CONH_2$ with G7, G13, A14, and G16 ^{15}N labels; designated as GAOGER; and peptide GSP$(GPP)_4$GFPGER$(GPP)_5$; designated as GFPGER; were synthesized by Lifetein, LLC (South Plainfield, NJ). The NMR samples of each ^{15}N labeled peptide were prepared in 10% D_2O/90% H_2O at pH 3.1 with concentrations of 3.7 mM for GFOGER and 1.0 mM for GAOGER.

NMR spectroscopy

Unless stated otherwise, all NMR experiments were performed on an Agilent Inova 600 MHz spectrometer equipped with a cryoprobe. All data was processed with NMRPipe and Sparky was used for both spectrum analysis and non-linear fits of peak intensities [23,24] .

Two Dimensional HSQCs

To examine the secondary structure of the CMPs through distinct monomer and trimer resonances, ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiments were conducted at 20 ºC [25]. For GFOGER, a 2D ^{1}H-^{15}N HSQC spectrum consisting of 256 (t_1) x 1232 (t_2) complex points was recorded using the gradient sensitivity enhanced approached with a spectral width of 7225.434 Hz and 2430.600 Hz in ^{1}H and ^{15}N dimensions respectively. For GAOGER, the 2D ^{1}H-^{15}N HSQC spectrum composed of 260 (t_1) x 1024 (t_2) complex points was collected using the gradient sensitivity enhanced approached with a spectral width of 6009.615 Hz and 1800 Hz in ^{1}H and ^{15}N dimensions respectively.

^{15}N spin relaxation experiments

To examine the ps – ns timescale dynamics of peptide backbones of the CMPs, we used ^{15}N spin relaxation measurements. Unless stated otherwise, all relaxation experiments were performed with 2 second recycle delays and the spectrometer equilibrated to 20 ºC.

R_1 experiments

Longitudinal relaxation, also known as R_1, is defined as the rate at which magnetisation is returned to an equilibrium that is parallel with the magnetic field **B** after a perturbation is applied to the magnetisation [17]. Typically, a R_1 experiment is conducted by running a series of 2D ^{1}H-^{15}N HSQCs with varying relaxation delays and fitting the resulting peak intensities of each resonance to the following non-linear decay equation:

$$A_t = A_0 e^{-R_1 t} \quad (1)$$

where A_t is the amplitude after the relaxation delay, A_0 is the amplitude at time 0 and t is the particular relaxation delay. The error propagation for each R_1 value is calculated as follows:

$$error = R_1 \sqrt{\left(\frac{SD}{T_1}\right)^2} \quad (2)$$

where SD is the standard deviation in the peak intensities and T_1 is the inverse of R_1.

In this study, the GFOGER R_1 spectra were collected with 200 (t_1) x 908 (t_2) complex points with a spectral width of 6410.256 Hz and 1800.000 Hz in the ^{1}H and ^{15}N dimensions respectively. Nine relaxation delays were used; 10, 50, 150, 200, 300, 400, 500, 600, and 700 ms. R_1 spectra for GAOGER were obtained through 200 (t_1) x 1024 (t_2) complex points with a spectral width of 7225.434 Hz and 1823.523 Hz for the ^{1}H and ^{15}N dimension respectively. Relaxation delays for GAOGER were the same as in the GFOGER experiment.

R_2 experiments

Transverse relaxation (also known as R_2) is another rate of magnetisation, after perturbation, return to the magnetic field's direction, however the magnetisation returns perpendicular to the magnetic field direction [17]. Excluding relaxation delay parameters, R_2 experiments were performed similarly to the R_1 experiment with the same fitting

equation and error propagation calculation, however the R_1 and T_1 variables are replaced by R_2 and T_2 variables respectively.

In the case of GFOGER, R_2 data was collected through spectra set to relaxation delays of; 5, 10, 10, 30, 70, 70, 90, 110 and 130 ms. Each spectra was a 2D ^{1}H-^{15}N HSQC composed of 230 (t_1) x 908 (t_2) complex points with a spectral width of 6410.256 x 1800.000 Hz in the ^{1}H and ^{15}N dimensions respectively. Relaxation delays for GAOGER were set as; 10, 10, 30, 50, 70, 70, 90, 110, and 130 ms. Spectra of GAOGER R_2 consisted of 230 (t_1) x 1024 (t_2) complex points and the spectral widths were 7225.434 x 1823.523 Hz in the ^{1}H and ^{15}N dimensions respectively.

NOE experiments

Steady state nuclear Overhauser effect of amides (also known as ^{1}H-^{15}N NOE) is the cross-relaxation between dipolar-coupled spins and as a result is dependent on the distance between ^{1}H and ^{15}N spins of an amide [17]. NMR can measure the NOE value of a particular ^{15}N labeled residue by recording two 2D ^{1}H-^{15}N HSQC's, one with proton saturation and the other without proton saturation. The peak amplitudes from both spectra are used to calculate a resonance specific NOE:

$$NOE = \frac{A_{sat}}{A_{eq}} \tag{3}$$

where A_{sat} is the peak amplitude in the presence of proton saturation and A_{eq} is the peak amplitude in the absence of proton saturation. Error propagation for each NOE calculation is obtained from the following error propagation equation:

$$error = NOE\sqrt{\left(\frac{N_{sat}}{A_{sat}}\right)^2 + \left(\frac{N_{eq}}{A_{eq}}\right)^2} \quad (4)$$

where N_{sat} is the noise level of the proton saturated peak and N_{eq} is the noise level of the peak without proton saturation.

For both GFOGER and GAOGER, the NOE values were obtained from a pair of 2D ^{1}H-^{15}N HSQC spectra, in the presence and absence of proton saturation. The GFOGER spectra contained 260 (t_1) x 1024 (t_2) complex points with a spectral width of 6410.256 x 1800.000 Hz in the ^{1}H and ^{15}N dimension respectively. Collection of NOE values for GAOGER was conducted through spectra containing 184 (t_1) x 1024 (t_2) complex points and spectral widths of 7225.434 x 1832.523 Hz in the ^{1}H and ^{15}N dimensions respectively.

Hydrogen Exchange experiments

In a hydrogen exchange experiment, the quantity of protection a residue in a polypeptide has from solvent (P) is correlated with the exchange rate the amide proton has with the solvent protons (k_{ex}) [26]. The exchange rate is derived from a two-state model from structure opening and closing as seen in equation 5:

$$\mathrm{N-H}(closed) \overset{k_{cl}k_{op}}{\longleftrightarrow} \mathrm{N-H}(open) \overset{k_{ex}}{\longrightarrow} \mathrm{N-D} \quad (5)$$

where k_{cl} and k_{op} are the rate constants for structure closing and opening respectively. Suggested from the two state model, N-H exists in an equilibrium between the closed and open state before irreversibly hydrogen exchanging into the N-D form [27]. Since NMR requires nuclei with non-zero spin to obtain a signal, we can conveniently use deuteron

(^{2}H) as a substitute for solvent proton and thus as ^{2}H replaces in polypeptide protons, the NMR signal is lost [28]. 2D ^{1}H-^{15}N HSQCs are commonly used to monitor peak intensities after a lypholized polypeptide is dissolved in concentrated D_2O. The resulting peak intensities are fitted to a non-linear fit:

$$I_t = I_0 e^{-k_{ex}t} \tag{6}$$

where I_t is the peak intensity at time t, I_0 is the peak intensity at time 0 and t is the time the spectrum was recorded after the sample was dissolved in D_2O. The P of each peak was calculated by using its k_{ex} with the following equation:

$$P = \frac{k_{int}}{k_{ex}} \tag{7}$$

where k_{int} is the theoretical monomer exchange rate for the amide proton at certain pH and temperature [26]. Error bars for the P values were calculated using the following equation

$$error = P\sqrt{\left(\frac{SD}{T_{decay}}\right)^2} \tag{8}$$

where SD is the standard deviation of the peak intensities and T_{decay} is the time constant ($1/k_{ex}$). Local stabilities of each peak were calculated using P:

$$\Delta G_{HX} = RT \ln P \tag{9}$$

where, ΔG_{HX} is the local stability, R is the molar gas constant and T is the temperature of the hydrogen exchange experiment.

In this study, both GFOGER and GAOGER hydrogen exchange experiments were conducted at 10 ºC at pD 2.54, where pD is the corrected for the glass electrode solvent isotope artifact [29]. NMR samples were equilibrated at 4 ºC for a minimum of 72 hours in water and then lyophilized. Data collection began by dissolving the lyophilized samples in 100% D_2O at pD 2.54 and quickly transferring the samples to a NMR spectrometer that was pre-emptively equilibrated to 10 ºC. Immediately after the NMR sample was placed into the spectrometer, a consecutive series of thirteen 5.85 minutes 2D ^{1}H-^{15}N HSQC spectra were taken to monitor fast-exchange protons. Upon completion of the so called "fast" HSQC scans, "normal" HSQC scans were acquired every 38.3 minutes [120 (t_1) increments with 8 scans per increment] to observe slow-exchanging amide protons.

Circular Dichroism

CD spectroscopy measures the absorbance of circularly polarized light by an optically active medium and the spectrum of a far-UV wavelength scan between 190 and 260 nm can indicate the degree of secondary structure present in the sample [30]. Raw CD data obtained from a CD spectrometer is converted to mean residue ellipicity (MRE) using the following equation:

$$[\theta] = \frac{100(signal)}{Cnl} \tag{10}$$

where θ is MRE, *signal* is the raw CD data in millidegrees, *C* is the polypeptide's concentration in millimolarity, *n* is the number of residues in the polypeptide and *l* is the pathlength in centimeters. Additionally, thermodynamic properties can be obtained by measuring the ellipicity at a fixed wavelength through a temperature gradient and the subsequent melting curve can be used to calculate the thermal stability of the sample through the melting curve's derivative.

CD experiments were performed on an Aviv Model 420 CD spectropolarimeter using 1 cm pathlength cuvettes purchased from Hellma Analytics (Mullheim, Germany). CMP samples were prepared in 0.01 mM acetic acid, pH 3.2, or Phosphate-buffered saline (PBS), pH 7.4, and peptide concentrations were measured using tyrosine (280 nm) extinction coefficients. Wavelength scans composed of 10 s averaging time for each nanometer were conducted at 5 ºC with 2 scans for each sample. Temperature scans were performed by measuring the 225 nm CD signal as a function of temperature between 0 and 70 ºC with 2 minutes of equilibration for every 0.33ºC increment.

Bacterial collagen expression and purification

All of the constructs containing the V-domain and GFPGER CMP (V-CMP) were obtained by inserting the PCR products into the pCOLD vector, which contains an N-terminal $(His)_6$ tag, using BamHI and NdeI restriction sites [13]. Oligonucleotide primers for PCR amplification of DNA fragments and mutagenesis of constructs were obtained from Integrated DNA Technologies, Inc (Coralville, IA) (Supplemental Figure 7). Mutagenesis was performed using the PfuUltra II Fusion HS DNA Polymerase kit

purchased from Agilent Technologies (Santa Clara, CA) with a modified form of the aforementioned kit's protocol where the annealing temperature was modified to match the target site's melting temperature of 55 ºC [32] (Supplemental Figure 8). The final DNA plasmids of the constructs were cloned using competent E. Coli BL21(DH5α) cells and sequencing of cloned plasmids was performed by GENEWIZ, Inc (South Plainfield, NJ) (Supplemental Figure 8).

All of the constructs that were confirmed to contain the desired V-CMP fusion proteins were transformed onto E. Coli BL21(DE3) competent cells. Cultures containing 50 mL of LB medium [20 g/L Luria Broth and 0.1 mg/mL ampicillin (Amp)] within 250 mL flasks were inoculated by the transformed cells and shaken overnight at 37 ºC. The cultures that grew overnight were diluted to 0.15-0.18 OD_{600} in 2.8 L flasks containing 1 L of LB medium and shaken at 37 ºC until the OD_{600} reached 0.5. Afterwards the cold shock step was applied by first incubating the cultures on ice-water for 5 minutes and then shaking at 15 ºC for 45-60 minutes. Protein expression was induced by inoculating the cultures with 1 mM isopropyl 1-thio-Beta-D-galacto-pyranoside (IPTG) and shaking overnight at 25 ºC. The protein content of the overnight cultures was analyzed with Tris-Tricine SDS-Page and if the cells contained the protein of interest, we harvested the cells by pelleting the 1 L cultures with centrifugation at 3,500 rpm for 40 minutes and stored the cell pellets at -80 ºC.

For purification of the constructs the following buffers were used: Buffer A (20mM NaPi, 0.5M NaCl, pH 7.4, 50 mM imidazole). Buffer B (20mM NaPi, 0.5M NaCl, pH 7.4, 0.5M imidazole).

Each 1 L culture pellet was re-suspended in 25 mL of Buffer A and then lysed with a French Press at 10,000 psi, followed by centrifugation at 5,000 rpm for 20 minutes to remove cell debris. Purification of the lysate was supplemented by ultra-centrifugation at 24,000 rpm for 1 hour to remove membrane proteins and lipids; then the lysate was filtered with a 0.22 µm filter. A 1 mL nickel-nitrilotriacetic acid-agarose (Ni-NTA) resin column was equilibrated with Buffer A and then the lysate from the ultra-centrifugation step was cycled through the Ni-NTA column for a minimum of 1 hour. Fast protein liquid chromatography (FPLC) was used with Buffer B as the elution buffer and each FPLC fraction was analyzed with Tris-Tricine SDS-Page. Fractions containing the V-CMP fusion protein were dialyzed into PBS and stored at 4 ºC until further use.

Enzyme-Linked Immunosorbent Assay

Immulon-2 96-Well Microtitier EIA plates were coated with 100 µL of PBS, pH 7.4, containing 1mg/mL V-THN-GFPGER, type I collagen (rat tail extract), or bovine serum albumin (BSA) in each well and incubated overnight at 4 ºC. The next day began by blocking each well with 200 µL Tris-buffered saline (TBS), pH 7.4 with 50 mg/mL BSA for 1 hour at room temperature and unless stated otherwise each subsequent step of the experiment was conducted at room temperature. The wells were then washed three times with 200 uL of wash buffer [TBS, pH 7.4, 1 mg/mL BSA with 5mM (Ethylenediaminetetraacetic acid) EDTA or 5 mM $MgCl_2$] before the addition of BSA or recombinant α2I domain variants; wild type α2I, α2T (α2I with G337 to T339 deleted), and L/O (G172C and L322C α2I mutant). Each protein sample was diluted to 10 µg/mL in appropriate wash buffers and then incubated in the wells for 1 hour. Another wash step was performed three times and then 100 uL anti- α2 antibody (mouse extract), diluted to a

1:2000 volume ratio, was added to each well with a 1 hour incubation. The plate was washed three times again, then 100 uL of anti-mouse anibody (goat extract), diluted to a 1:15000 volume ratio, was added to each well with 30 minutes of incubation and afterwards each well was treated to a wash step four times. The next step was to add 100 uL of a 1:1 peroxidase-substrate/peroxide mixture to each well, incubate for 20 minutes and then quench each reaction with 100 uL of 1M H_2SO_4. Colorimetry was used to analyze color of the wells with absorbance measurements at 405 nm.

TEV Protease expression and purification

E. coli BL21 CodonPlus(DE3)-RIL cells containing the pRK793 plasmid were obtained from American Type Culture Collection (Manassas, VA) and the pRK793 plasmid contains the N-terminal His_7 tagged TEV protease fused with the Maltose Binding Protein (MBP) which is attached at the N terminus of the $(His)_7$ tag [15].

E. coli cells with the TEV protease were streaked onto a LB agar plate with Amp and chloramphenicol (Cm) and incubated overnight at 37 ºC in an oven. Isolated colonies from the overnight culture plates were used to inoculate 20 mL LB medium (20 g/L Luria Broth, 0.1 mg/mL Amp and 0.25 mg/mL Cm) cultures contained in 125 mL flasks and shaken overnight at 37 ºC. Each pre-culture was completely diluted into a 1 L LB medium, grown to 0.5 OD_{600} with shaking at 37 ºC, inoculated with 0.5 mM IPTG to induce protein expression and then shaken overnight at 30 ºC. The protein content of the overnight cultures was analyzed with Tris-Glycine SDS-Page and cultures containing the TEV protease were pelleted at 3,500 rpm for 40 minutes and stored at -80 ºC.

For purification of the TEV protease the following buffers were used: Buffer A (50 mM Na-Pi, pH 8.0, 0.5 M NaCl, 1% Triton X-100, 20 mM imidazole). Wash Buffer (50 mM NaPi, pH 8.0, 20 mM imidazole). Buffer B (50 mM NaPi, pH 8.0, 0.3 M imidazole).

Each 1 L culture pellet was re-suspended in 25 mL of Buffer A, next lysed with a French Press at 10,000 psi and then centrifuged at 8,000 rpm for 30 minutes to remove cell debris. The lysate was filtered with a 0.45 µm pore membrane filter and then cycled on a pre-equilibrated Ni-NTA resin column overnight. On the next day, the column was washed with Wash Buffer and TEV protease was eluted Buffer B. After analyzing each purification step with Tris-Glycine SDS-Page, we dialyzed the eluted TEV protease into 50 mM Tris-HCl, phH 7.5, 1mM EDTA, 5mM dithiothreitol (DTT). For storage, the TEV protease was diluted into glycerol, with a final glcyerol concentration of 50% v/v, and stored at -80 ºC.

Cleavage of the V-TEV-GFPGER

Cleavage of the V-TEV-GFPGER was conducted based on a suggested protocol provided by D.S. Waugh, 2010 [33]. The protein solution containing V-TEV-GFPGER was heated at >50 ºC for 30 minutes and then cooled on ice to 37 ºC. TEV protease was added to the heated V-TEV-GFPGER solution with 1 A_{280} TEV protease per 50 A_{280} substrate and the reaction mixture was incubated at room temperature overnight. The reaction mixture was analyzed with Tris-Tricine SDS-Page and MALDI TOF to check if the resulting CMP, GFPGER, was successfully removed from the V-domain; if the

cleavage was successful, we cycled the reaction mixture through a Ni-NTA resin column overnight and dialyzed the GFPGER containing flow through into a buffer of interest.

RESULTS

Secondary structure of integrin binding collagen model peptides

Labeled residues in or near the integrin binding sequence were investigated by biophysical characterization methods to examine conformational differences among the GFOGER and GAOGER CMP triple helices. Two dimensional ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) spectroscopy of CMPs can detect peaks from a population of trimerized peptides and monomer peptide strands, both of which manifest as distinct peaks due to differences in their local chemical environment [34]. Assignments of the resulting 2D HSQC spectra peaks were based on a staggered triple helix model obtained from x-ray diffraction data [35] (Figure 3). G7 and G13 trimer resonances of both peptides appeared with similar chemical shifts in both peptides, with G7 trimer resonances being overlapped in both peptides since G7 is within the GPO repeat region. The GAOGER peptide additionally had G16 labeled, which showed overlap resonances for two of the three strands. F14 trimer resonances from the GFOGER peptide and A14 trimer resonances from the GAOGER peptide lied further downfield from the glycine trimer resonances in the ^{15}N dimension. In both spectra, the downfield hydrophobic residues (alanine and phenylalanine), revealed that two of the three trimer resonances were adjacent to each other, while the third resonance was upfield from the peak pair (Figure 4).

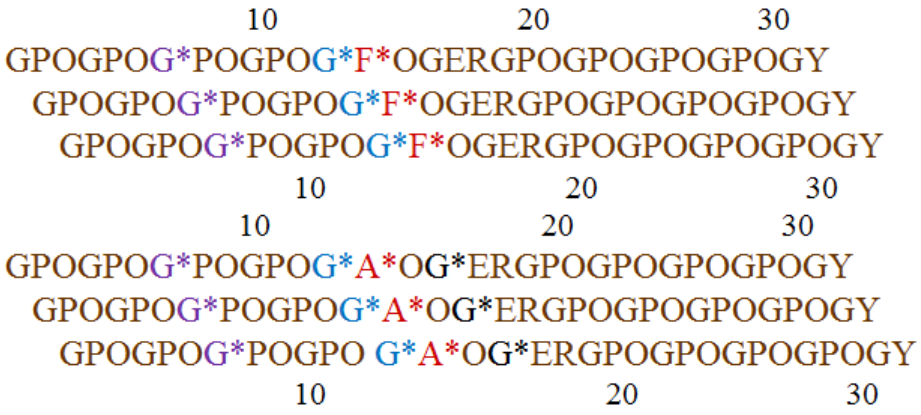


Figure 3. (A) One residue staggered triple helix model applied to the GFOGER CMP used in this study. Asterisk represent ^{15}N labeled residues. **(B)** Staggered triple helix model for GAOGER. Asterisk represent ^{15}N labeled residues.

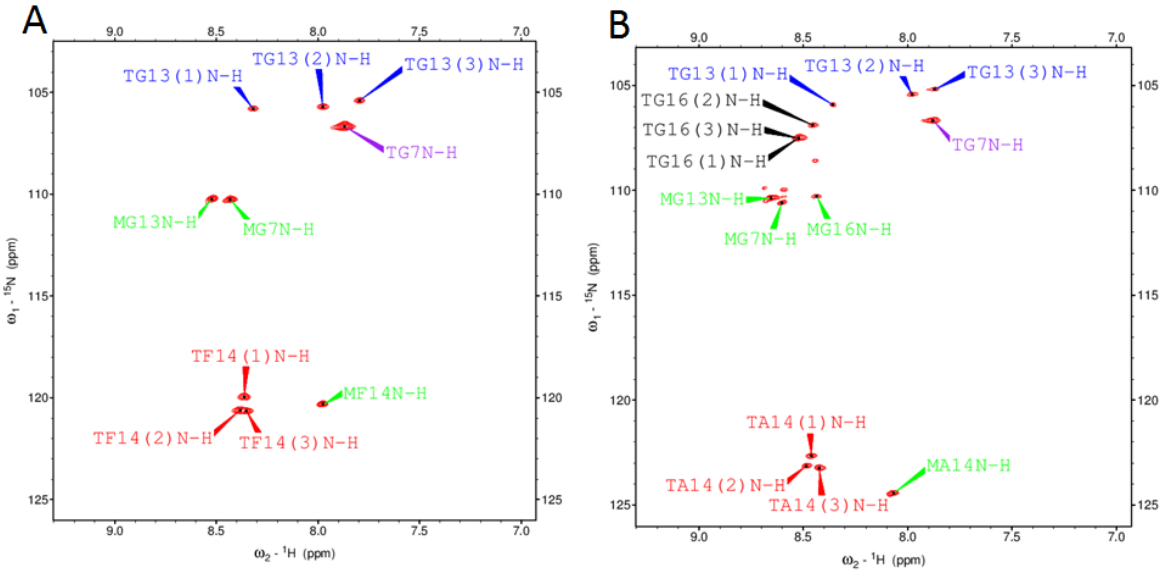


Figure 4. (A) ^{1}H-^{15}N HSQC of the GFOGER CMP triple helix and monomer. **(B)** ^{1}H-^{15}N HSQC of the GAOGER CMP triple helix and monomer. In both spectra, Gly trimer resonances clustered upfield in the nitrogen dimension, Gly monomer resonances clustered downfield from the Gly trimers, and Phe/Ala resonances were the most downfield in the nitrogen dimension.

Spectra from the CD wavelength scans of the GFOGER and GAOGER showed that they had a very similar triple helix character with a maximum near 225 nm and a

minimum near 200 nm (Figure 5). Upon characterizing the secondary structures through 2D ^{1}H-^{15}N HSQC NMR and CD wavelength scans, we set out to investigate the thermal stability of said secondary structures through CD temperature scans (Figure 6) [31, 36]. Thermal stabilities of both CMP triple helices obtained from the temperature scans are in agreement with collagen stability calculator thermal stabilities [37].

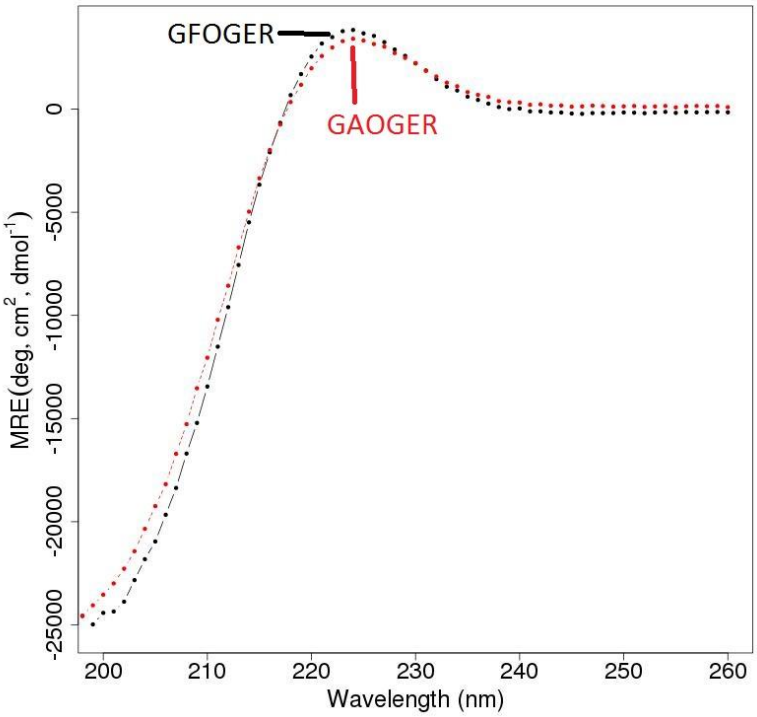


Figure 5. CD wavelength scan showing the triple helical character of GFOGER (colored black) and GAOGER (colored red)

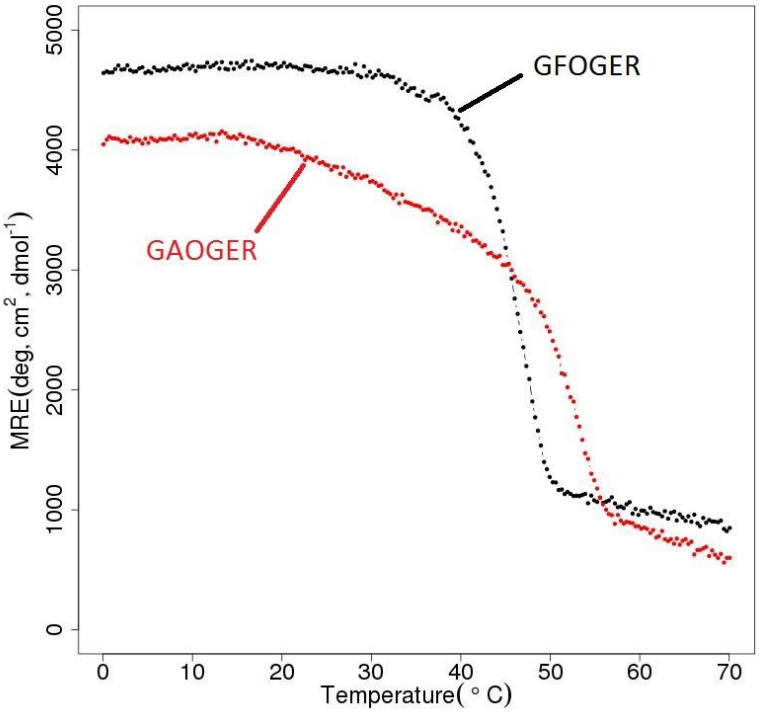


Figure 6. Temperature scan of GFOGER (black line) and GAOGER (red line). T_m of GFOGER is 46ºC (theoretical T_m is 43.1ºC) and 51ºC for GAOGER (theoretical T_m is 51.3ºC).

Conformational dynamics of integrin binding collagen model peptides measured by NMR

With the 2D ^{1}H-^{15}N resonance assignments, we set out to investigate the fast time scale dynamics of each CMP trimer resonance using ^{15}N relaxation experiments [38, 39]. NOE measurements of the GFOGER and GAOGER CMPs were mostly uniform within and among the two peptides (Figure 7). Deviation from the levelness of the NOE measures stemmed from G7, which had the lowest NOE values for both peptides, and

also from G13(3) of GFOGER which showed the lowest NOE values among residues located within the integrin binding motifs.

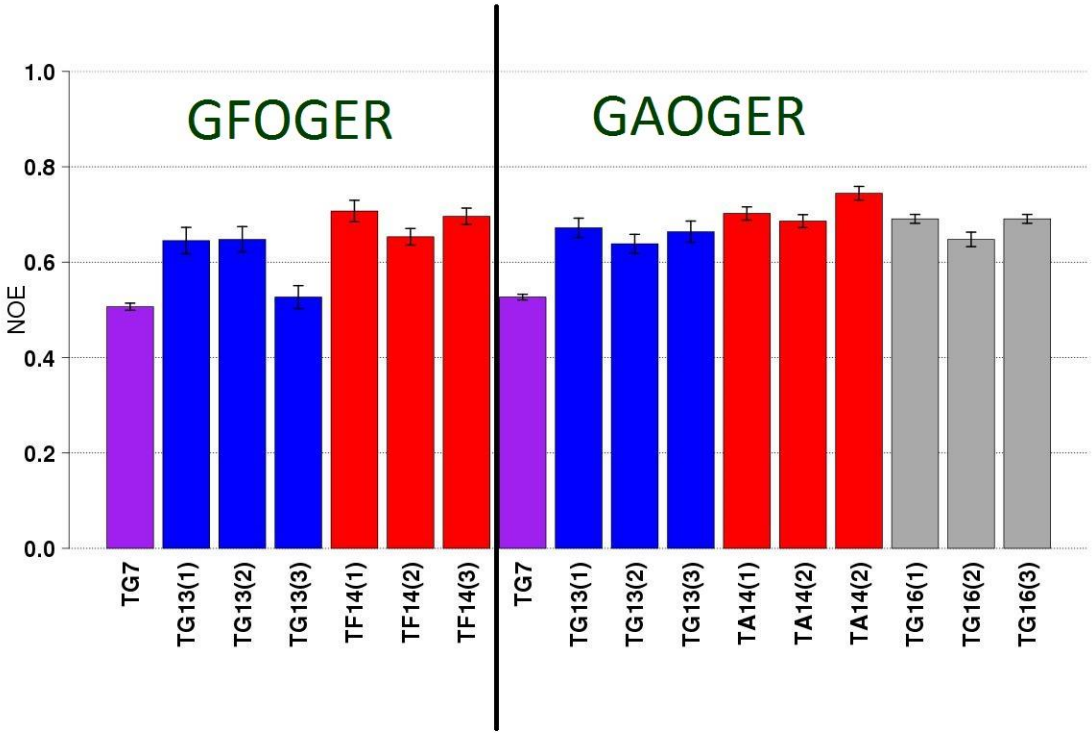


Figure 7. NOE values of GFOGER and GAOGER CMPs. In this figure and subsequent NMR dynamics figures, bars are color coded to their respective positions in the CMPs.

^{1}H-^{15}N R_1 and R_2 values for both peptide sets showed that the X14 resonances had lower spin relaxation values compared to the Gly resonances and there were no conclusive differences among Gly spin relaxation values (Figure 8).

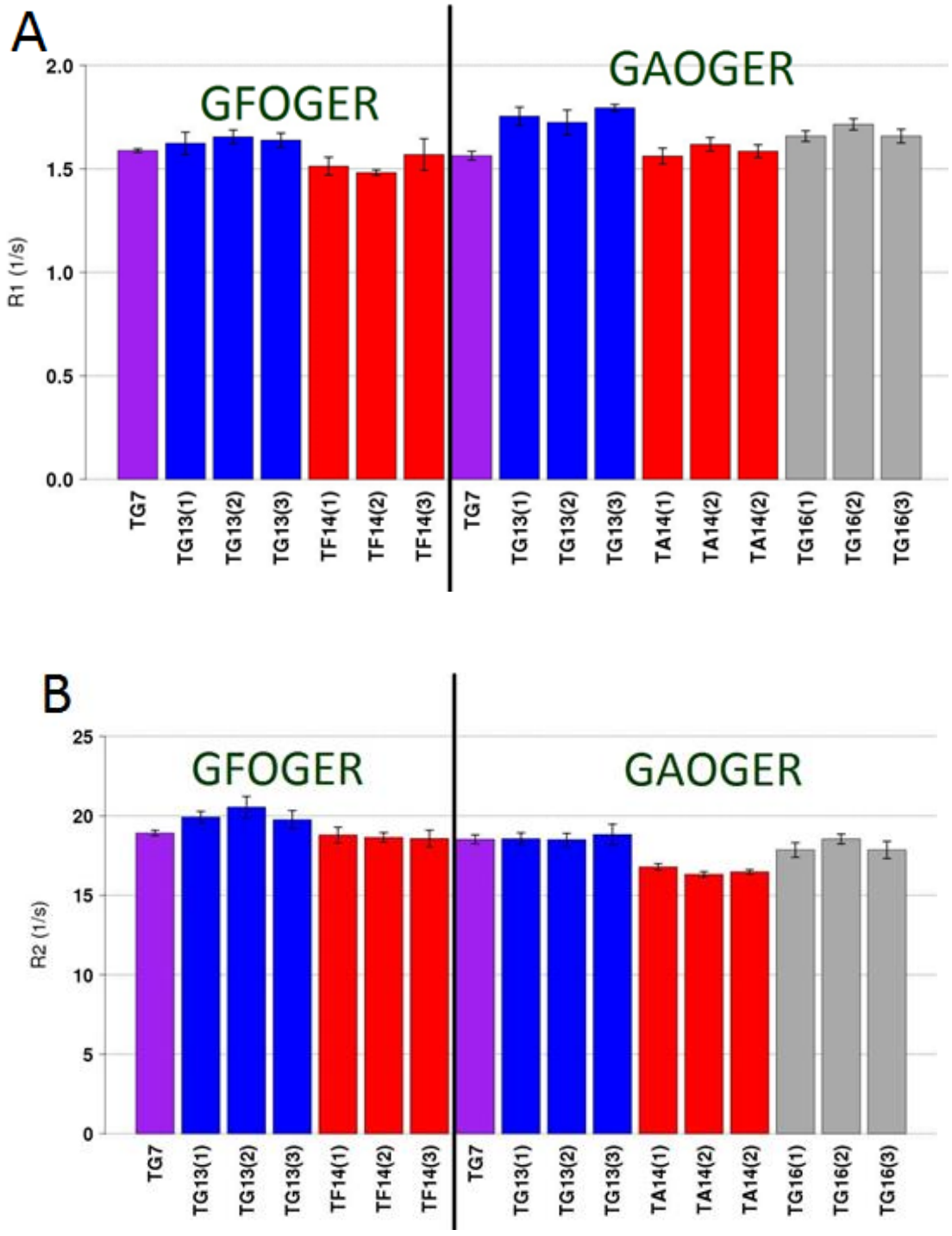


Figure 8. Relaxation values of the α2I binding CMP triple helices. **(A)** R1 values of the CMPs. **(B)** R2 values of the CMPs. The Phe and Ala residues seem to have relaxation values that were often lower than the Gly residues.

Table 1. Parameters obtained from ^{15}N relaxation experiments for select labeled residues of GAOGER and GFOGER.

GFOGER			
Resonance	**NOE**	**R1** (1/s)	**R2** (1/s)
TG7	0.51 ± 0.008	1.59 ± 0.009	18.9 ± 0.17
TG13(1)	0.65 ± 0.028	1.62 ± 0.054	19.9 ± 0.36
TG13(2)	0.65 ± 0.027	1.65 ± 0.033	20.5 ± 0.67
TG13(3)	0.53 ± 0.024	1.64 ± 0.034	19.8 ± 0.57
TF14(1)	0.71 ± 0.023	1.51 ± 0.044	18.8 ± 0.49
TF14(2)	0.65 ± 0.017	1.48 ± 0.014	18.6 ± 0.3
TF14(3)	0.7 ± 0.017	1.57 ± 0.076	18.6 ± 0.52

GAOGER			
Resonance	**NOE**	**R1** (1/s)	**R2** (1/s)
TG7	0.53 ± 0.006	1.56 ± 0.022	18.5 ± 0.28
TG13(1)	0.67 ± 0.020	1.75 ± 0.044	18.6 ± 0.36
TG13(2)	0.64 ± 0.02	1.72 ± 0.06	18.5 ± 0.39
TG13(3)	0.66 ± 0.022	1.8 ± 0.017	18.8 ± 0.65
TA14(1)	0.70 ± 0.014	1.56 ± 0.039	16.8 ± 0.21
TA14(2)	0.69 ± 0.014	1.62 ± 0.033	16.3 ± 0.17
TA14(3)	0.74 ± 0.014	1.59 ± 0.031	16.5 ± 0.14
TG16(1)	0.69 ± 0.009	1.66 ± 0.025	17.9 ± 0.46
TG16(2)	0.65 ± 0.015	1.72 ± 0.027	18.5 ± 0.31
TG16(3)	0.69 ± 0.009	1.66 ± 0.033	17.9 ± 0.55

The lack of differences among fast dynamics led us to investigate the slow dynamics (days) through hydrogen exchange experiments, which measure the amount of protection amide protons had from solvent. Relatively high protection factors within a peptide indicates strong intra-helix hydrogen bonding within the triple helix that hinders

the amide proton from exchanging with solvent protons, meaning that the particular strand is tightly wound to the collagen triple helix (Figure 9). G7 in both peptides showed the high protection factors relative to their respective triple helix from solvent since hydrogen bonding contributes to a tightly wound triple helix in the GPO repeat region [8]. After accounting for the error bars, there are no major differences between the G13 resonances in both CMPs. Among the 14th residues, F14(1) and A14(1) showed the lowest protection factors. The GAOGER G16(2) resonance showed a much lower protection factor compared to its other two counterpart G16 resonances which, themselves, showed protection from solvent at a level similar to G7, however the significance of that result is challenged by the overlapping of those peaks.

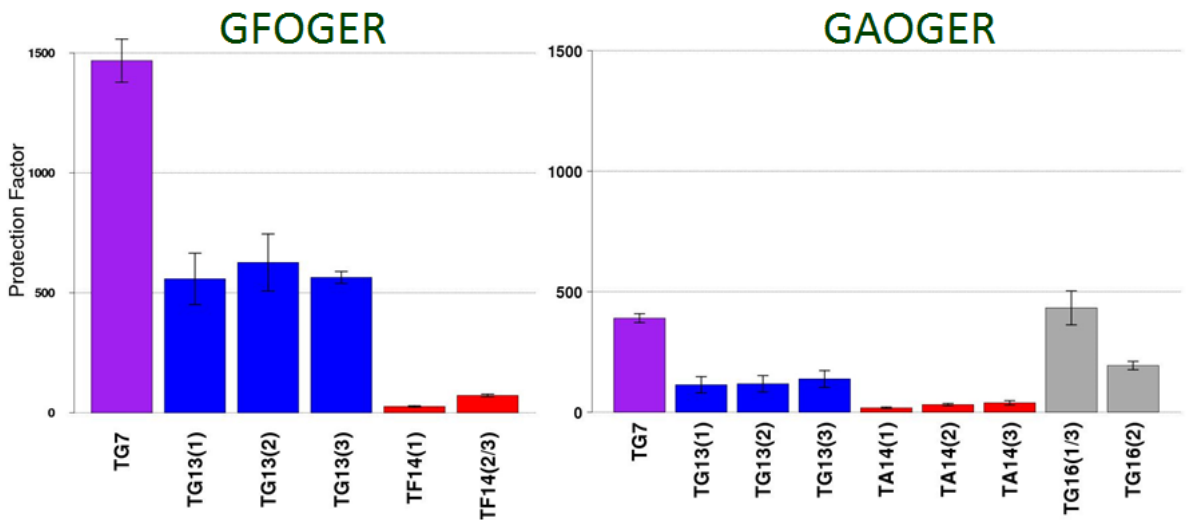


Figure 9. Protection factors calculated from hydrogen exchange rates for GFOGER and GAOGER. Peak overlap was observed to occur with TF14(2) and TF14(3)

Table 2. Stabilities of select labeled residues modeled through hydrogen exchange experiments

GFOGER			GAOGER		
Resonance	**P_f**	**ΔG_{HX} (kcal/mol)**	**Resonance**	**P_f**	**ΔG_{HX} (kcal/mol)**
TG7	1468 ± 90	4.1	TG7	390 ± 18	3.4
TG13(1)	559 ± 107	3.6	TG13(1)	114 ± 34	2.7
TG13(2)	627 ± 118	3.6	TG13(2)	118 ± 34	2.7
TG13(3)	564 ± 25	3.6	TG13(3)	138 ± 35	2.8
TF14(1)	26 ± 2.8	1.8	TA14(1)	18 ± 3.3	1.6
TF14(2)	72 ± 5.6	2.4	TA14(2)	32 ± 4.8	1.9
			TA14(3)	39 ± 9.2	2.1
			TG16(1)	433 ± 70	3.4
			TG16(2)	193 ± 17	3

Expression of recombinant collagen model peptide

We designed recombinant fusion proteins to express CMPs in E. coli that can function as an ELISA agent and as an isotopically labeled protein for NMR experiments. Expression of the original V-THN-GFPGER construct was attempted first to test its ELISA capabilities (Figure 10).

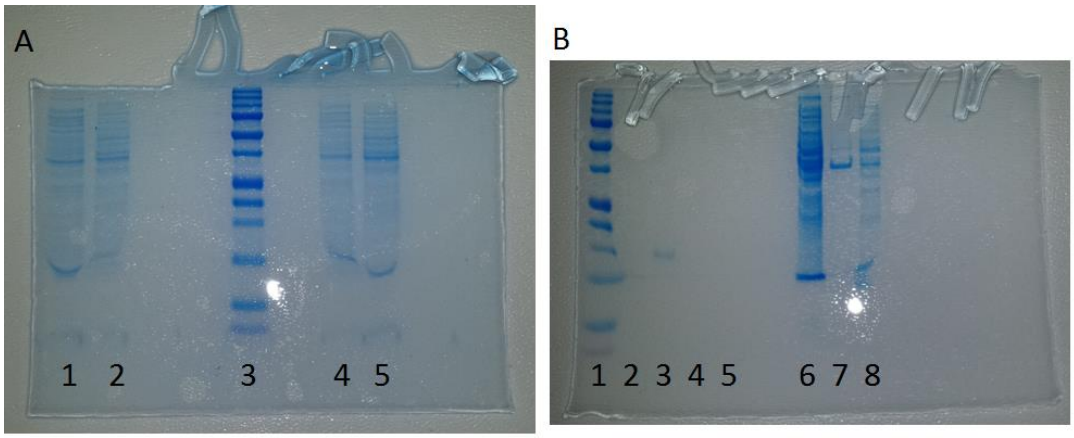


Figure 10. (A) SDS-Page gel of lysed cells during expression. Lanes 2 and 4 are before IPTG inoculation. Lanes 1 and 5 are after IPTG inoculation. (B) SDS-PAGE of samples during stages of purification. Lane 2 is Fast Protein Liquid Chromatography (FPLC) fraction 1, lane 3 is FPLC fraction 2, lane 4 is FPLC fraction 3, lane 5 is FPLC fraction 5, lane 6 is the insoluable portion of the cell lysate from French press, lane 7 is the lysate after cycling through the Ni-NTA column and lane 8 is the flow through from the wash step

The α2I domain recognizes the GFOGER sequence as a binding substrate, however our recombinant peptide does not have hydroxyprolines, since E. coli lack prolyl hydroxyprolases [40]. Nevertheless, we pressed on to investigate the binding affinity of V-THN-GFPGER to α2I variants using ELISA (Figure 11).

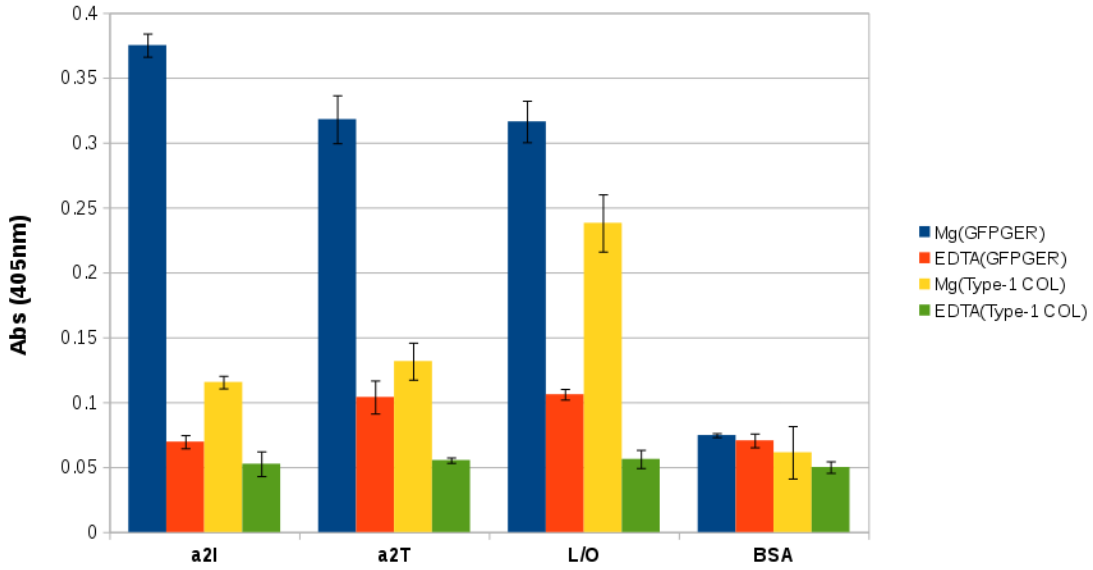


Figure 11. ELISA used to measure binding affinity of α2I, α2T (truncated form of α2I), L/O (G172C/L322C α2I mutant), and BSA (bovine serum albumin, non-binding control) to collagen. GFPGER (V-THN-GFPGER) shows a specificity higher than Type I collagen (rat tail extract)

ELISA results indicate that α2I was able to bind to the V-THN-GFPGER with a higher affinity than type I collagen. Type I collagen showed a higher binding affinity for L/O α2I compared to wild type α2I, in contrast to V-THN-GFPGER's binding trend, however the L/O α2I mutation is such that a disulfide bond is formed between a pair of cysteine mutant residues to maintain a constant "open" MIDAS region.

In an attempt to cleave the GFPGER CMP from the V-domain, we mutated the thrombin cleavage site to a TEV protease cleavage site (Figure 12). The new construct, V-TEV-GFPGER, displayed a similar expression yield to the original construct with no apparent increase or decrease of yield (Figure 13).

YEKTAIAIAVALAGFATXAQA*HHH*
*HHH*MDEQEEKAKVRTELIQELAQ
GLGGIEKKNFPTLGDEDLDHTYM
TKLLTYLQEREQAENSWRKRLLK
GIQDHALD*ENLYFQG*SPGPPGP
PGPPGPPGFPGERGPPGPPGPP
GPPGPPGY

Figure 12. Amino acid sequence of V-TEV-GFPGER starting with the ompA signal peptide (purple), His tag (italicized black), V domain (blue), TEV cleavage site (italicized orange), and finally GFPGER (bold underline).

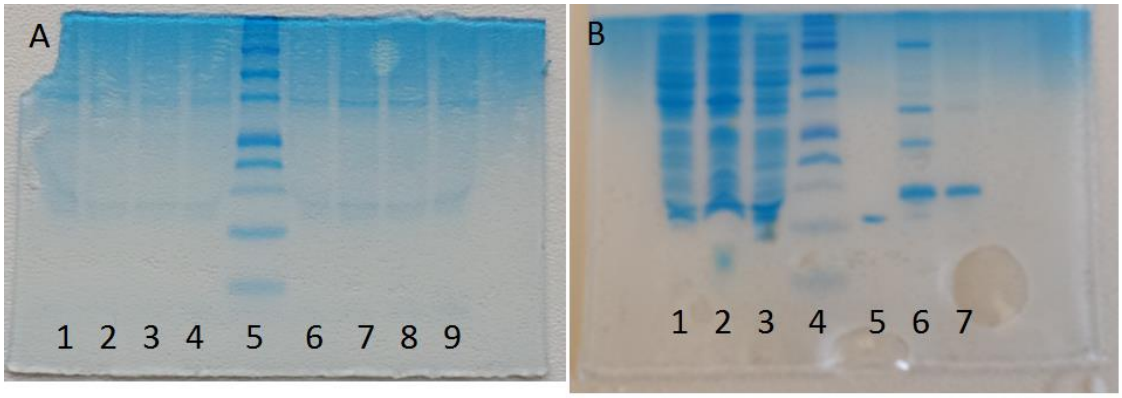


Figure 13. (A) SDS-Page of cell lysate during expression. Lanes 1 through 4 are the cell contents before IPTG inoculation. Lanes 6 to 9 are after IPTG inoculation. **(B)** SDS-Page of samples during the steps of purification. Lane 1 is insoluable material from cell lysate after French press, lane 2 is lysate that did not bind to the Ni-NTA column, lane 3 is flow through from wash step, lane 5 is FPLC fraction 1 containing V-domain that expressed without GFPGER peptide, lane 6 is FPLC fraction 2 with V-TEV-GFPGER and lane 7 is FPLC fraction 3 with V-TEV-GFPGER

Initially cleavage was unsuccessful with a 4ºC overnight incubation. We amended the reaction procedure by heating the V-TEV-GFPGER to a minimum of 50ºC to unfold it, followed up by quickly cooling the solution to 37ºC with ice and finally adding the TEV protease for an overnight incubation. A Matrix-assisted laser desorption/ionization-

time of flight (MALDI-TOF) mass spectrum revealed that the V-TEV-GFPGER was efficiently cleaved, yielding V-domain and GFPGER CMP with no remaining starting V-TEV-GFPGER (Figure 14). Both the TEV protease and V-domain contain a C-terminal His tag, which prompted us to purify the reaction mixture with a Ni-NTA resin. The CD wavelength scan spectrum of the flow through from the purification procedure showed a minimum near 205 nm, however it lacked the maximum near 220 nm that its synthetic copy contained (Figure 15).

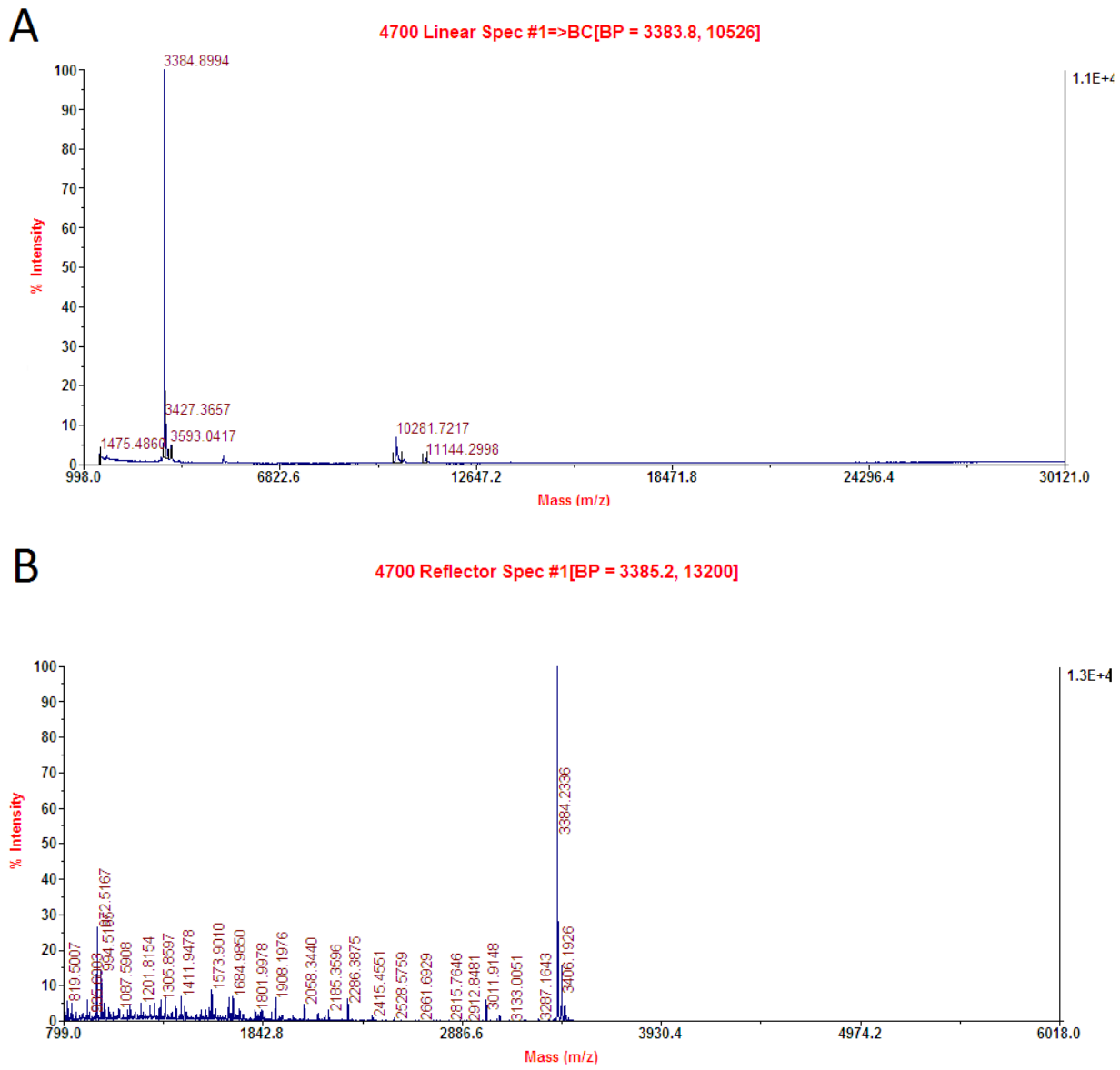


Figure 14. (A) MALDI-TOF spectra of V-TEV-GFPGER after TEV protease cleavage. **(B)** Inset showing only 800-6,000 m/z. The 3,384.8894 m/z peak is the GFPGER peptide free from the 10,281.7217 m/z V-domain peak.

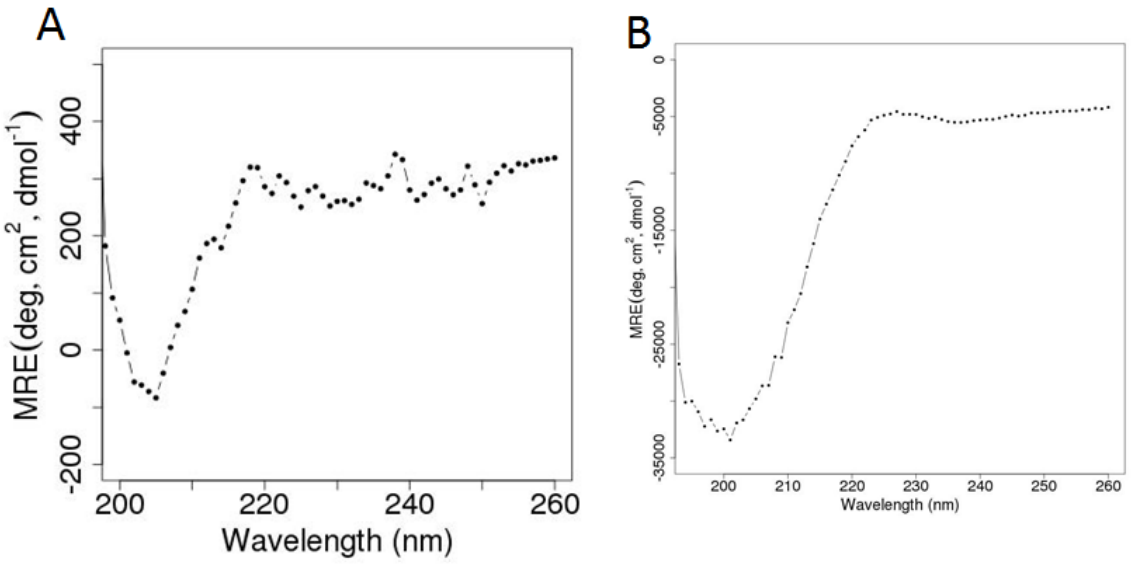


Figure 15. CD wavelength scan of (A) recombinant GFPGER CMP and (B) synthetic analog of GFPGER CMP triple helix. The minimum is near 205 nm which is shifted from the synthetic copy's minimum and is also less steep. The maximum that should appear near 220 nm is not observed in the recombinant form.

DISCUSSION

Integrin binding collagen model peptide secondary structure characterization

Biophysical spectroscopy techniques in this study have yielded information on the secondary structure of the GFOGER and GAOGER CMP triple helices. Differences in 2D HSQC chemical shifts arise due to local chemical environments around atomic nuclei, which suggests that the NMR samples of GFOGER and GAOGER have both monomer and trimer populations due to presence of both monomer and trimer resonances. Comparison of GFOGER and GAOGER trimer resonances showed that both peptides had similar trimer Gly chemical shifts, which suggests that both peptides adopt similar triple helix structures in solution. Additionally the TA14(1) and TF14(1) trimer

resonances appeared upfield from their two counterpart intra-helix resonances in the ^{15}N dimension, which could mean that TA14(1) and TF14(1) belong to a single strand of their respective triple helix with a conformation that is very different from the other two strands. Further support for a unique strand conformation in the triple helix arises from the G16 residue in the GAOGER triple helix, as the G16(2) does not overlap with its sister G16 resonances in the triple helix. Data from the CD temperature scans showed that both peptides have similar thermal stabilities which were in agreement with the collagen stability calculator values [37]. Between the atomic details from NMR and overall secondary structure data from CD; both methods conclude that GFOGER and GAOGER triple helices have similar secondary structure with a single strand in the triple helix in a conformation different from the two other strands.

Solution dynamics from ^{15}N spin relaxation experiments

R_1, R_2, and NOE values for most glycine trimer resonances are very similar within and among the α2I binding triple helices. The 14^{th} residues, Phe in GFOGER and Ala in GAOGER, showed lower R_1 and R_2 values compared to their intra-helix Gly neighbors, which supports the notion that non-Pro and non-charged residues in the XX' locations create local flexibilities. The implication of these results is that a minimal amount of dynamic differences exists between backbone Gly amides of GFOGER and GAOGER. From these polypeptide backbone fast dynamics results, which are in agreement with literature values, we can conclude that both CMP triple helices are mostly rigid in picosecond to nanosecond timescales [18].

Slow exchange measured by Hydrogen Exchange

Our hydrogen exchange results provide insight into the backbone amide slow dynamics of α2I binding CMP triple helices and GFOGER protection factors are in agreement with literature values [18, 37]. Among the 14th residues, F14(1) and A14(1) showed the lowest protection factors, which provides further evidence that one strand of the CMP triple helix contributes to α2I binding more than the other strands. From GAOGER's additional labeled residue, G16, G16(2) resonance showed a much lower protection factor compared to its other two counterpart G16 resonances, further supporting that one strand is more flexible that the other two strands in the α2I binding triple helices. Additionally, GAOGER's protection factors were a fraction of counterpart GFOGER protection factors, which may indicate that the GAOGER triple helix was unfolding and refolding during some time of the experiment. NMR results in this study have produced evidence of local conformational differences between intra-helix backbone amides within the GFOGER triple helix which exist as events that persist for days and thus lend support to literature evidence that α2I selectively binds one strand of a triple helix. Despite the details obtained about the GFOGER triple helix, there is a lack of evidence in this study to support the hypothesis that binding affinity is dictated by differences in dynamics among GXX'GEX" sequences in collagen triple helices. With the hypothesis seemingly nulled by data from this study, it is likely that GFOGER has a higher α2I affinity than GAOGER, due to the Phe in GFOGER fitting into the hydrophobic pocket in the α2I MIDAS region.

Progress in development of recombinant collagen model peptides

ELISA results show that α2I was able to bind to the recombinant V-THN-GFPGER construct, which indicates that V-THN-GFPGER was able to form a CMP triple helix and the V-domain was able to adhere to the mictrotiter plate. The GFPGER peptide also showed a higher binding affinity than type I collagen, however it is actually an increase of specificity since the truncation of the collagen through a model peptide increases the accessibility of the GFPGER binding sequence to α2I [8].

In an attempt isolate the recombinant GFPGER CMP, we mutated the thrombin cleavage site to a TEV protease cleavage site and expressed the new construct. MALDI-TOF spectra of the cleavage reaction mixture showed that cleavage attempt was successful, however attempts at characterizing the peptide product have not yielded concrete results. The CD wavelength scan spectrum of the flow through from the purification procedure showed a minimum near 205 nm. The spectrum indicates some degree of a triple helix formed, however the minimum is shifted away from the literature minimum of 200 nm and the maximum value that should appear near 220 nm is not present. Potentially, the deviation from literature values from the recombinant GFPGER CD wavelength spectrum could be the consequence of a low sample concentration since the minimum of the recombinant peptide is less steep than the synthetic GFPGER's minimum. Results from this study have thus provided evidence that E. coli can be indeed be used to produce CMPs as an alternative to chemical synthesis.

CONCLUSIONS

A detailed understanding of the molecular mechanisms that lead to the integrin α2I domain selectively binding to different collagen sequences is of vital importance, yet there is a lack of insight into as how these molecular mechanisms proceed. Hydrogen exchange experiment evidence in this study indicates that one phenylalanine in GFOGER triple helix is more flexible than the other two phenylalanines, which supports the theory that one strand of the GFOGER triple helix is more flexible than the other two. Accounting for all of the NMR results in this study, however, there is lack of detail showing that differences in the dynamics of the GXX'GEX" collagen triple helices guide the binding of one strand to the α2I domain. Dynamic and the conformation of the integrin α2I domain are more likely to be the driving forces for recognition of the collagen triple helix.

There is a great amount of potential in using recombinant CMPs as a platform to efficiently screen variations in integrin binding sequences along with mutations in said binding sequences. Our molecular biology work has provided us novel fusion proteins that can be used as ELISA reagents and cleaved in-vitro to yield recombinant CMPs. The combination of molecular biology and biophysics can pave the way towards furthering our understanding of how integrin binds to collagen, thus leading to novel therapeutics to treat diseases.

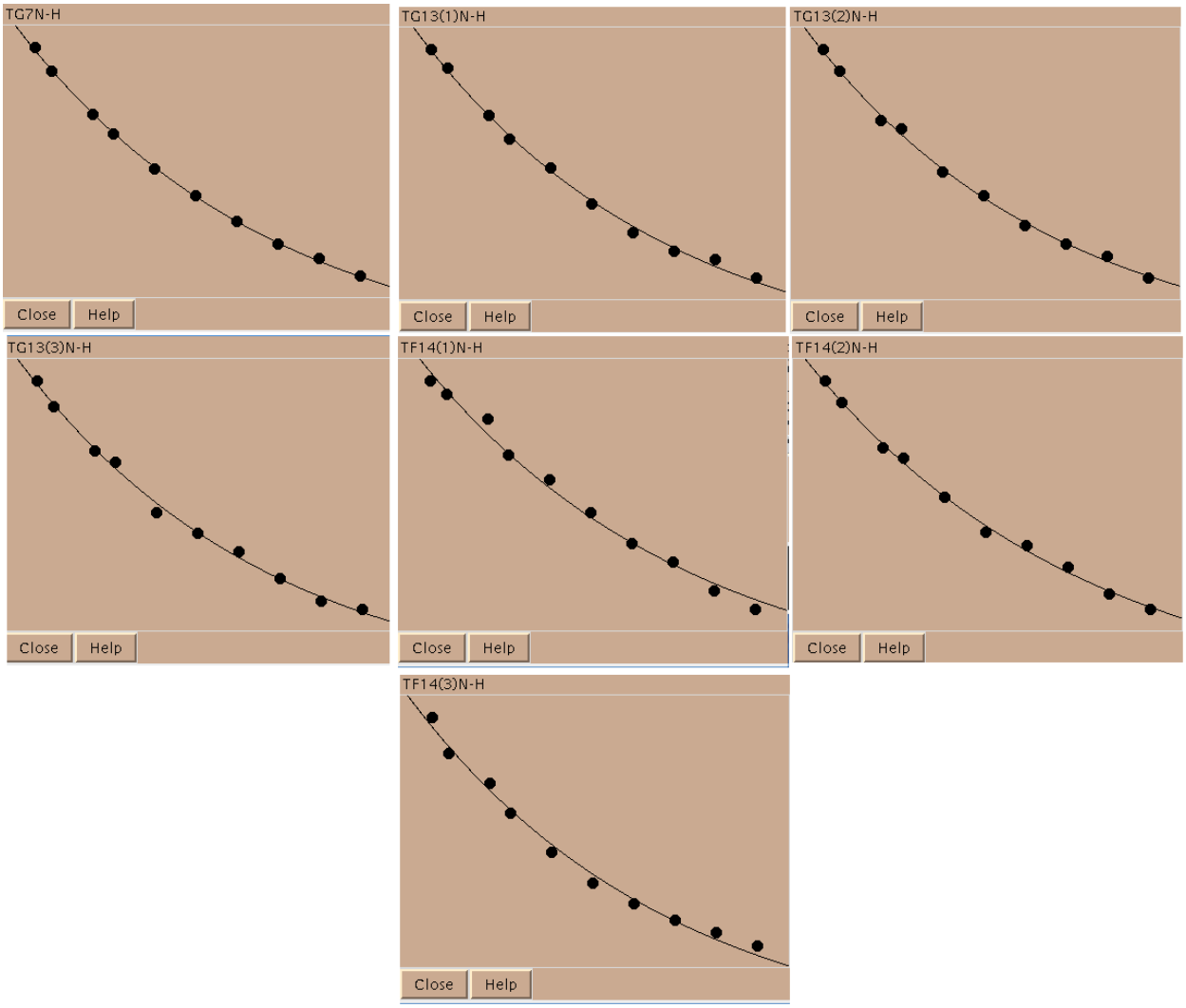
REFERENCES

1. Emsley, J., Knight, G., Farndale, R.W., and Barnes, M.J. Structure of the Integrin α2β1-binding Collagen Peptide. J. Mol. Biol. 2004. 335(4): 1019-1028
2. Kielty, C.M., Sheratt, M.J., and Shuttleworth, C.A., Elastic Fibres. Cell. 2002. 115: 2817-2828
3. Miles, A.J., Knutson, J.R., Skubtiz, A.P.N., Furcht, L.T., McCarthy, J.B., and Fields, G.B., A Peptide Model of Basement Membrane Collagen α1(IV) 531- 543 Binds the α3β1 Integrin. J. Biol. Chem. 1995. 270: 29047-29050
4. Kim, J.K., Xu, Y., Xu, X., Keene, D.R., Gurusiddapa, S., Liang, X., Wary, K.K., and Hook, M. A. Novel Binding Site in Collagen Type III for Integrins α1β1 and α2β1. J. Mol. Biol. 2005. 280: 32512-32520
5. Hamaia, S.W., Pugh, N., Raynal N., Némoz B., Stone R., Gullberg D., Bihan D., and Farndale R.W. Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin α1β1 using collagen toolkits II and III. J. Biol. Chem. 2012. 287: 26019-26028
6. Carafoli, F., Hamaia, S.W., Bihan. D., Hohenester, E., and Farndale, R.W. An activating mutation reveals a second binding mode of the integrin α2 I domain to the GFOGER motif in collagens. PLoS ONE. 2013. 8(7): e69833- e69833
7. Hynes, R.O. Integrins: Bidirectional, Allosteric Signaling Machines. Cell. 2002. 110(6): 673-687
8. Emsley, J., Knight, G., Farndale, R.W., and Barnes, M.J. Structural Basis of Collagen Recognition by Integrin α2β1. Cell. 2000. 101: 47-56
9. Orgel, J.P., Irving, T.C., Miller, A., and Wess, T.J. Microfibrillar structure of type I collagen in situ. Proc. Natl. Acad. Sci. U.S.A. 2006. 103(24): 9001
10. Long, C.G., Braswell, E., Zhu, D, Apigo, J., Baum, J., and Brodsky, B. Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. Biochemistry. 1993. 32(43): 11688-11695
11. Bella, J., Eaton, M., Brodsky, B., and Berman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. Science. 1994. 266: 75-81
12. Xu, Y., Keene, D.R, Bujnicki, J.M., Hook, M., and Lukomski, S. Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices. J. Biol. Chem. 2002. 277: 27312 – 27318.
13. Qing, G., Ma, L.C., Khorchid, A., Swapna, G.V., Mal, T.K., Takayama, M.M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G.T., Ikura, M., and Inoyue, M. Cold-shock induced high-yield protein production in Escherichia coli. Nat. Biotechnol. 2004. 22(7): 877-882.
14. Takahara, M., Hibler, D.W., Barr, P.J., Gerlt, J.A., and Inouye, M. The ompA signal peptide directed secretion of Staphylococcal nuclease A by Escherichia coli. J. Biol. Chem. 1985. 260. 2670-2674.
15. Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh D.S. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng. 2001 14(12): 993-1000
16. Palmer, A.G., NMR Characterization of the Dynamics of Biomacromolecules. Chem. Rev. 2004. 3623-3640

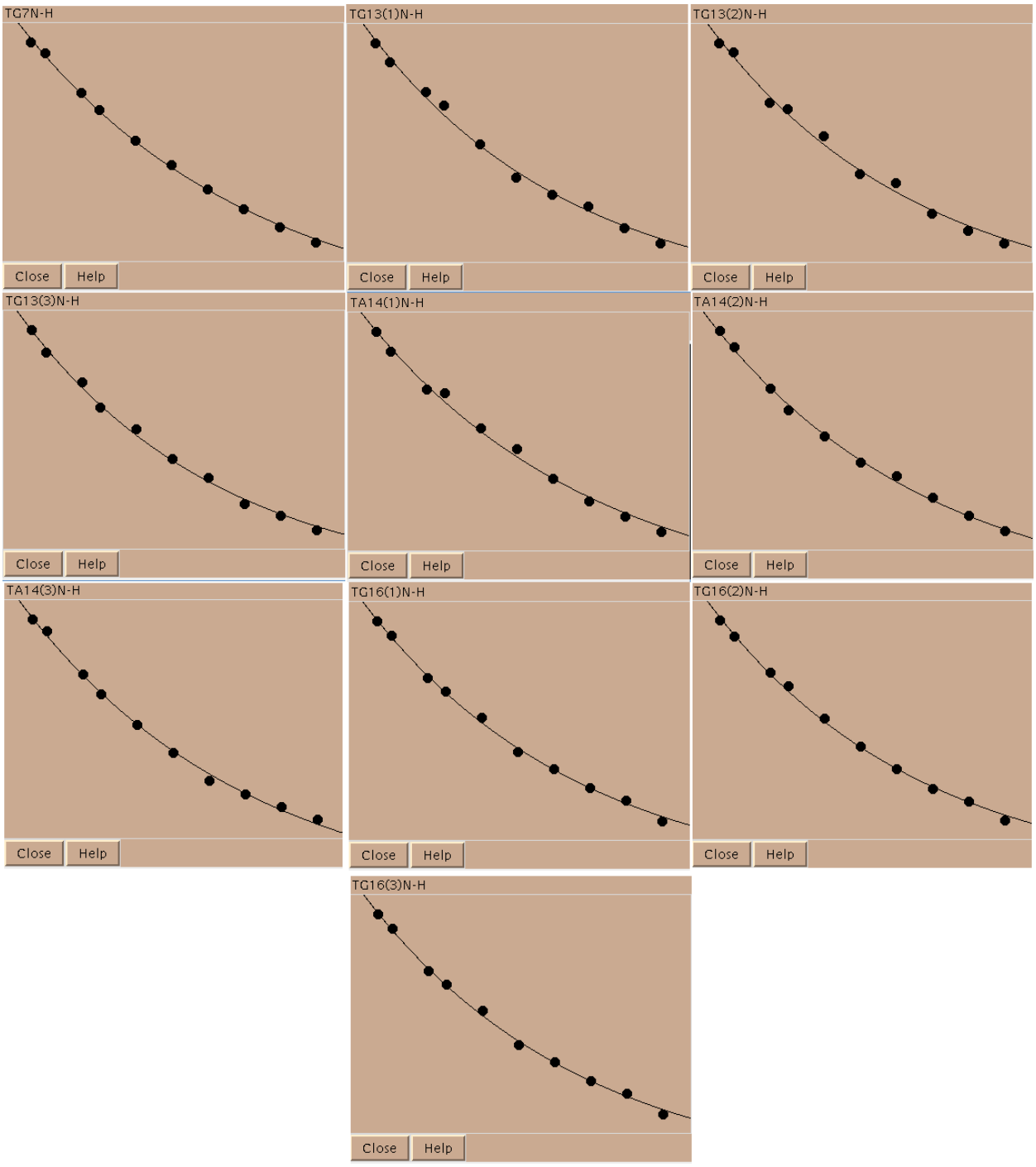
17. Morin, S., A practical guide to protein dynamics from ^{15}N spin relaxation in solution. Progress in Nuclear Magnetic Resonance Spectroscopy. 2011. 59(3): 245-262
18. Fan, P., Li. M.H., Brodsky, B., and Baum. J. Backbone Dynamics of (Pro-Hyp-Gly)$_{10}$ and a Designed Collagen-like Triple Helical Peptide by ^{15}N NMR Relaxation and Hydrogen-Exchange Measurements. Biochemistry. 1993. 32(48): 13299-13309.
19. Jelinski, L.W., Sullivan, C.E., and Torchia, D.A. ^{2}H NMR study molecular motion in collagen fibrils. Nature. 1980. 284. 531-534
20. Li, M.H., Fan, P., Brodsky, B., and Baum, J. Two-dimensional NMR assignments and conformation of (Pro-Hyp-Gly)$_{10}$ and a designed collagen triple-helical peptide. Biochemistry. 1993. 32(29): 7377-7387
21. Mayo, K.H., Parra-Diaz, D., McCarthy, J.B., and Chelberg, M. Cell adhesion promoting peptide GVKGDKGNPGWPGAP from the collagen type IV triple helix: cis/trans proline-induced multiple ^{1}H NMR conformations and evidence for a KG/PG multiple turn repeat motif in the all-trans proline state. Biochemistry. 1991. 30(33): 8251-8267
22. Fields, G.B. and Prockop, D.J. Perspectives on the synthesis and application of triple-helical, collagen-model peptides. Biopolymers. 1996. 40(4): 345-357
23. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR. 6, 277-293 (1995).
24. Goddard, T.D, and Kneller, D.G. SPARKY 3, University of California, San Francisco
25. Kay, L., Keifter, P., and Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J. Am. Chem. Soc. 1992. 114(26): 10663-10665
26. Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. Primary structure effects on peptide group hydrogen exchange. Biochemistry. 1972. 11(2): 150-158
27. Huyghues-Despointes, B.M.P., Pace, C.N., Englander, S.W., and Scholtz, J.M. Measuring the Conformational Stability of a Protein by Hydrogen Exchange. Methods Mol. Biol. 2001. 168: 68-92
28. Rabi, I.I., Zacharias, J.R., Millman, S., and Kusch, P. A New Method of Measuring Nuclear Magnetic Moment. Physical Review. 1938. 53(4): 318
29. Glasoe, P.K, and Long, F.A. Use of glass electrodes to measure acidities in deuterium oxide. J. Phys. Chem. 1960. 64: 188-190
30. Myers, J.K., Pace, C.N., and Scholtz, J.M. Helix propensities are identical in proteins and peptides. Biochemistry. 1997. 36(36). 10923-10929
31. Greenfield, N.J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat. Protoc. 2006. 1: 2527-2535
32. Kibbe, W.A., OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. 2007. 35(webserver issue)
33. Waugh, D.S. TEV Protease FAQ. National Cancer Institute. 2010.

34. Madhan, B., Xiao, J., Thiagarajan, G., Baum, J., and Brodsky, B. NMR Monitoring of Chain-Specific Stability in Heterotrimeric Collagen Peptides. J. Am. Chem. Soc. 2008. 130(41): 13520 – 13521
35. Fraser, R.D.B., MacRae, T.P., and Suzuki, E. Chain conformation in the collagen molecule. J. Mol. Biol. 1979. 129(3): 463-481
36. Reyes, C.D. and Garcia, A.J. Engineering integrin-specific surfaces with a triple-helical collagen-mimetic peptide. J. Biomed. Mater. Res. A. 2003. 65(4): 511-523
37. Persikov, A.V., Ramshaw, J.A., and Brodsky, B. Prediction of collagen stability from amino acid sequence. J. Biol. Chem. 2005. 280(19): 19343-19349
38. Xiao, J., Addabbo, R.M., Lauer, J.L., Fields, G.B., and Baum, J. Local Conformation and Dynamics of Isoleucine in the Collagenase Cleavage Site Provide a Recognition Signal for Matrix Metalloproteinases. J. Biol. Chem. 2010. 285: 34181-34190
39. Xiao, J., Madhan, B., Li, Y., Brodsky, B., and Baum, J. Osteogenesis Imperfecta Model Peptides: Incorporation of Residues Replacing Gly within a Triple Helix Achieved by Renucleation and Local Flexibility. Biophysical. 2011. 101(2): 449-458
40. Rutschmann, C., Baumann, S., Cabalzar, J., Luther, K.B., and Hennet, T. Recombinant expression of hydroxylated human collagen in Escherichia coli. Appl. Microbiol. Biotechnol. 2013. 98(10): 4545-4455

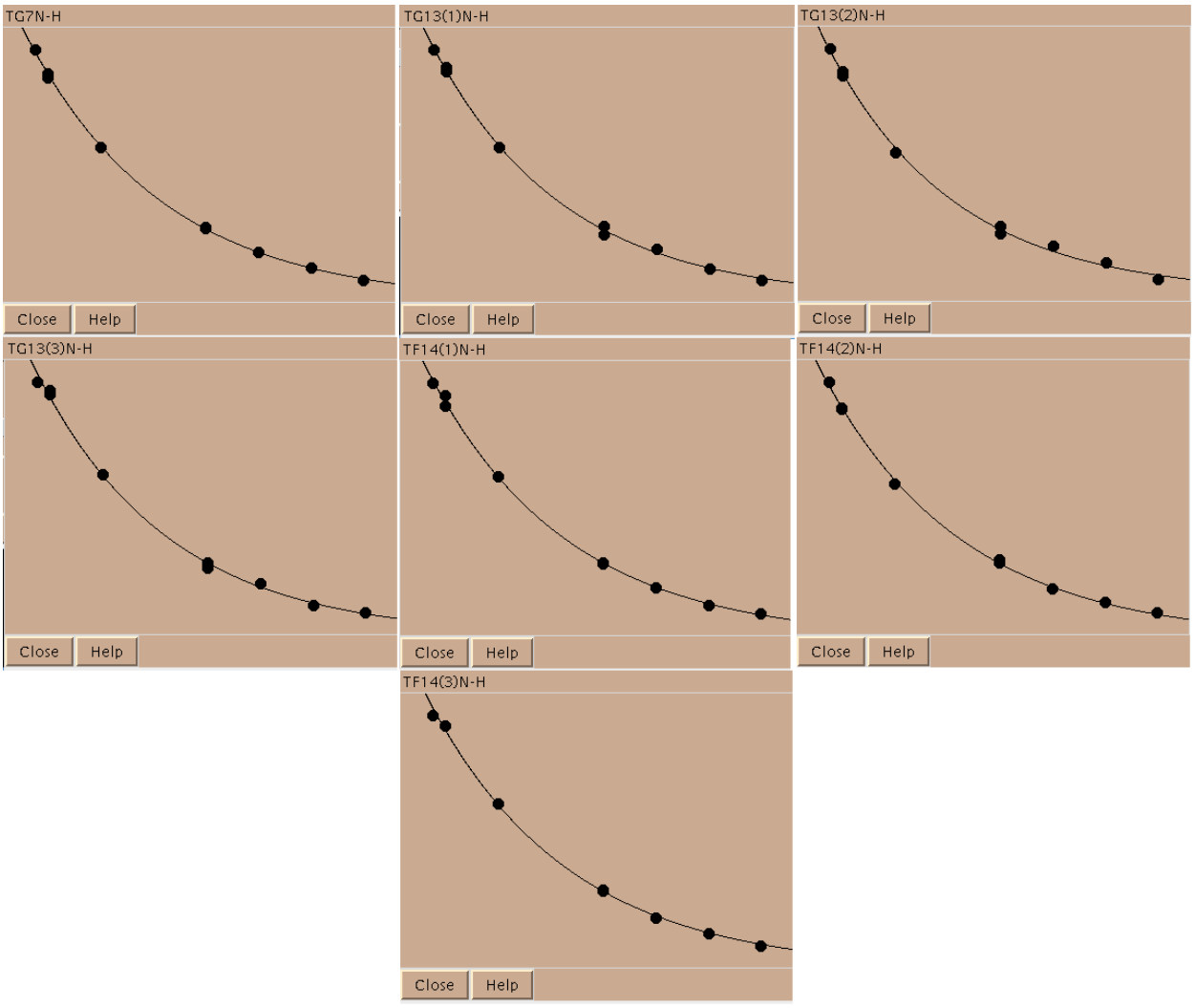
SUPPLEMENTAL INFORMATION



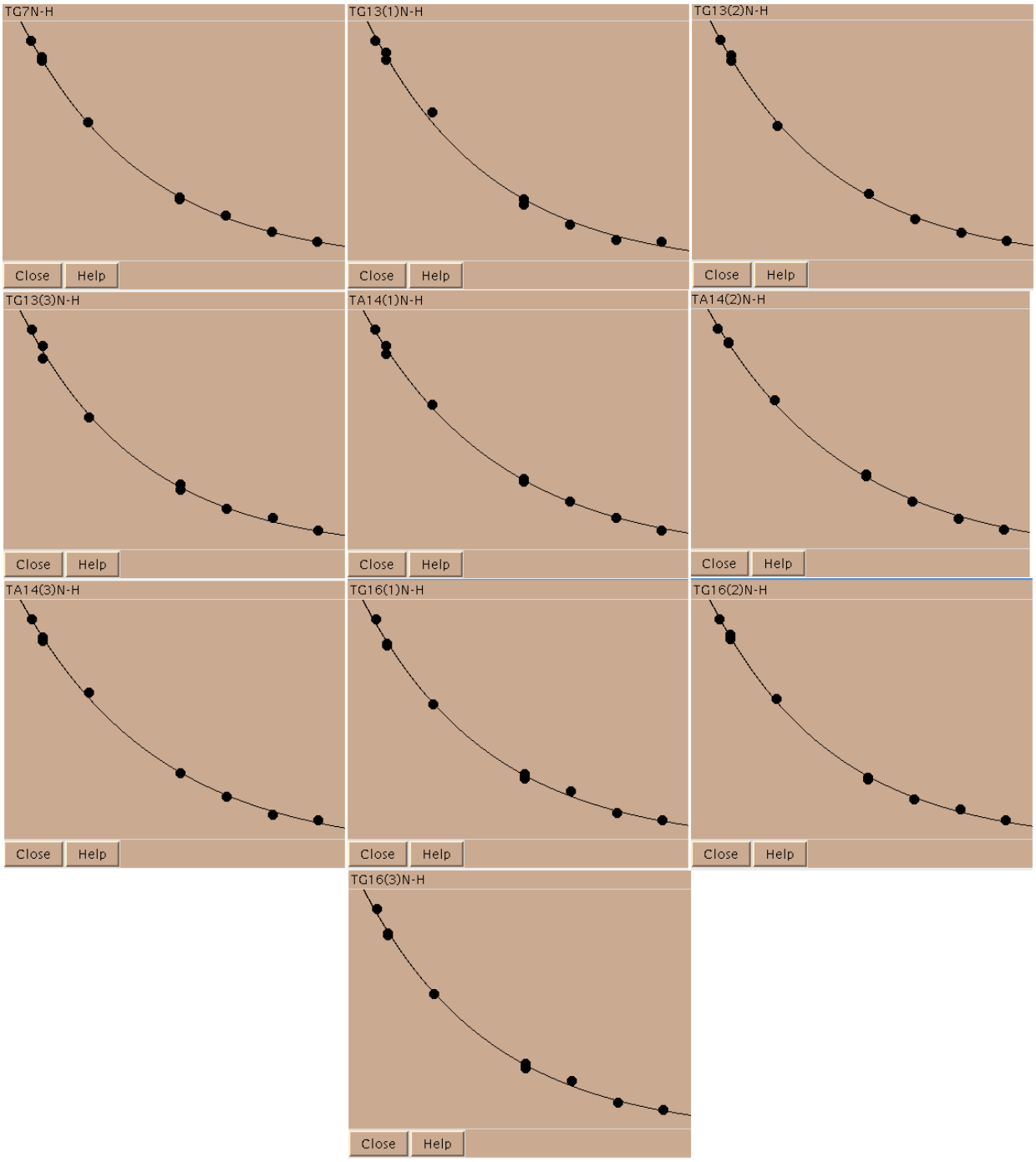
Supplemental Figure 1. GFOGER R_1 peak decay fittings calculated with Sparky



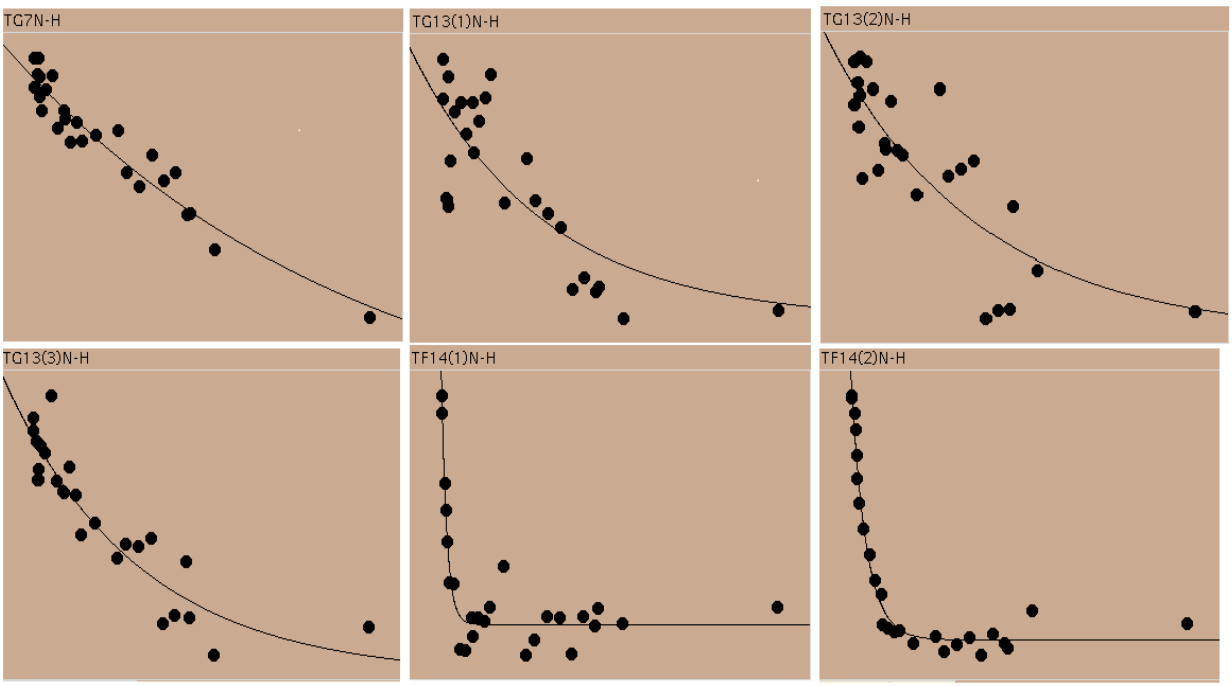
Supplemental Figure 2. GAOGER R_1 fittings from Sparky



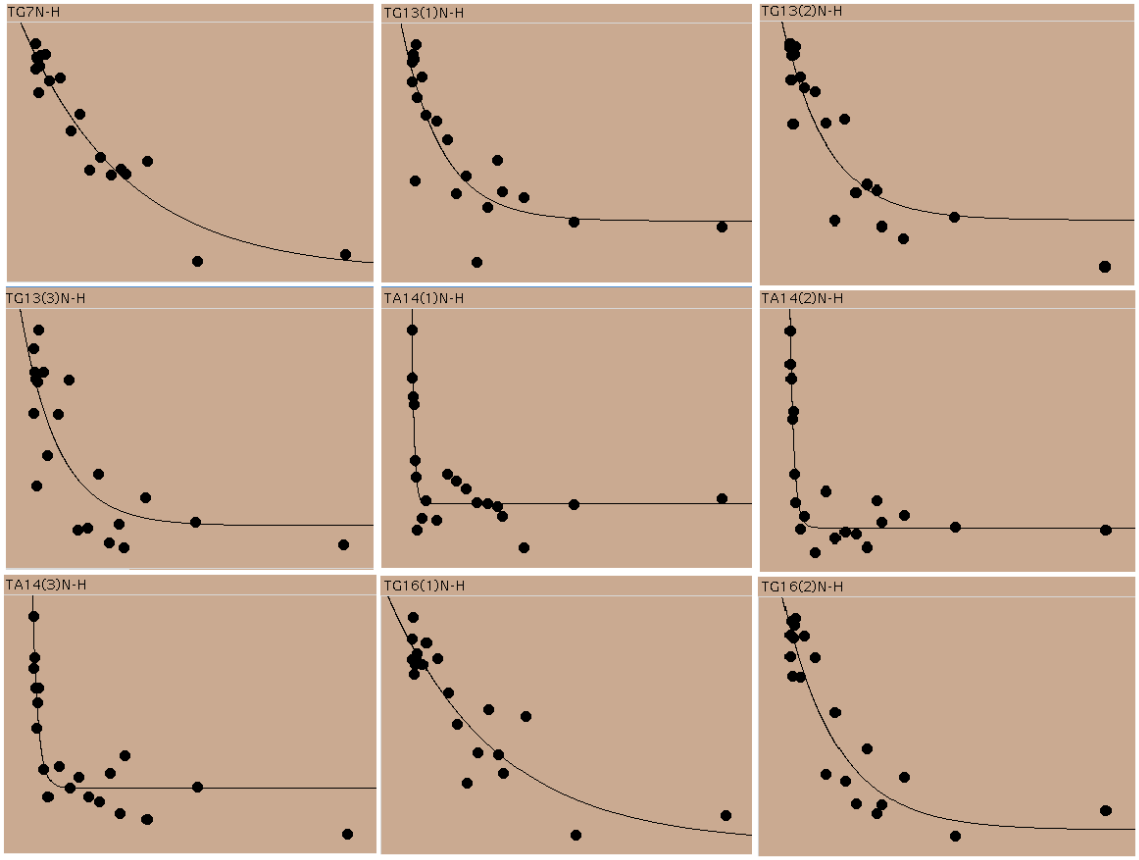
Supplemental Figure 3. GFOGER R_2 fittings provided by Sparky



Supplemental Figure 4. GAOGER R_2 fittings provided by Sparky



Supplemental Figure 5. GFOGER H_{EX} fittings provided by Sparky



Supplemental Figure 6. GAOGER H_{EX} fittings calculated with Sparky

TEV-CL

5’ – GGG TAT TCA AGA TCA TGC CCT TGA TGA AAA CCT GTA CTT CCA GGG

TAG TCC CGG GCC GC – 3’

TEV-CL-complement

5’ – GCG GCC CGG GAC TAC CCT GGA AGT ACA GGT TTT CAT CAA GGG CAT

GAT CTT GAA TAC CC – 3’

Supplemental Figure 7. DNA oligonucleotide used for mutating the thrombin cleavage site to a TEV cleavage site.

```
tatgaaaaaaccgcgattgcgattgcggtggcgctggcgggctttgcgaccgcgcaggcg
catcatcatcatcatcatatggatgaacaggaagaaaaagcgaaagtgcgcaccgaactg
attcaggaactggcgcagggcctgggcggcattgaaaaaaaaaactttccgaccctgggc
gatgaagatctggatcatacctatatgaccaaactgctgacctatctgcaggaacgcgaa
caggcggaaaacagctggcgcaaacgcctgctgaaaggcattcaggatcatgcgctggat
gaaaacctgtattttcagggcagcccgggcccgccgggcccgccgggcccgccgggcccg
ccgggctttccgagcgaacgcggcccgccgggcccgccgggcccgccgggcccgccgggc
ccgccgggctat
```

Supplemental Figure 8. DNA sequence of the V-TEV-GFPGER construct with the ompA signal peptide attached at the N-terminus of the His_6 tag