

©2009

JIANXI XIAO

ALL RIGHTS RESERVED

NMR CONFORMATIONAL AND DYNAMIC CHARACTERIZATION OF TRIPLE
HELICAL PEPTIDES

by

JIANXI XIAO

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Chemistry

written under the direction of

Professor Jean Baum

and approved by

New Brunswick, New Jersey

October, 2009

ABSTRACT OF THE DISSERTATION

NMR conformational and dynamic characterization of triple helical peptides

By JIANXI XIAO

Dissertation Director:
Professor Jean Baum

The repeating Gly-X-Y sequences and uniform rod-like structure makes collagen-like peptides a unique system for NMR studies. In this dissertation, a number of triple helical peptides modeling biologically important regions in collagen, such as mutation sites, interruption sites and collagenase cleavage sites, are investigated by a variety of NMR techniques. Structure determination strategy combining molecular modeling and NMR spectroscopy have been developed on a classic triple helical peptide. Novel approaches capable of obtaining long-range order restraints, such as residual dipolar coupling and ^{15}N relaxation measurements, are applied to obtain detailed information about the orientation of the N-H bonds, which are crucial in defining collagen structure.

Triple helical peptides modeling Gly mutations involved in Osteogenesis imperfecta (OI), a connective tissue disorder, are investigated to elucidate the structural bases of various OI phenotypes. The level of structural disruption by different Gly substitutions is found to correlate well with the lethality of OI, while

Arg/Asp causes larger disruption and is more likely to result in lethal OI. Triple helical peptides modeling natural interruptions in a heterotrimeric rather than homotrimeric environment have been successfully obtained and special features of stability, conformation, dynamics and folding at the interruption sites are detected by NMR.

Triple helical peptides, which model natural cleavage sites and potential but noncleavable sites in collagen, are explored to understand the specific recognition of collagen by matrix metalloproteinases (MMP). A single Ile at the cleavage site shows a distinct chemical shift, an unusual J-coupling value, dramatically increased dynamics and decreased local stability, suggesting that the Ile may be released from the restricted triple helical conformation and recognized by MMPs. The distribution of neighboring imino acids is also shown to be able to affect the local conformation and dynamics at the cleavage sites. This thesis correlates collagen sequence variations to the changes in structural and dynamic features of collagen-like peptides, furthering our understanding of the molecular bases for collagen-involved diseases and recognitions.

Acknowledgements

First of all, I would like to express my deepest thanks and appreciation to my advisor, Prof. Jean Baum, for her inspiration, guidance and support in the course of my Ph.D. studies. It has been challenging and sometimes frustrating, but her encouragement, patience and advice, helps me to go through it and complete my thesis research. I would also like to thank Prof. Barbara Brodsky, who is directly involved in my research and has semiofficially been my advisor for one semester. Her inspiration, enthusiasm and knowledge have greatly enriched my research experience. I would like to thank Prof. John Taylor and Prof. David Talaga for taking the time to review this dissertation and serving on the committee.

I am very grateful to all the members of Prof. Baum's group. Dr. Seho Kim, Dr. Nagarajan Murali and Dr. Yingjie Li have helped me in many aspects of NMR. Thank Kuen-phon Wu, Lijuan Kang, Rayna Addabbo, Yu-jen Chen and Chitra Narayanan for their assistance and support. I also wish to thank all the members of Prof. Brodsky's group for inspiring discussions and kind assistance, especially Dr. Balaraman Madhan, Dr. Ayumi Yoshizumi, Zhuoxin Yu, Eileen Hwang and Teresita Silva.

I would like to thank Prof. Gregg Fields for the collaboration on the collagenase cleavage project and thank Dr. Janelle Lauer-Fields for her help in the enzymatic kinetics studies. I would like to thank Prof. Masayori Inouye for the

collaboration on the bacterial expression project and thank Dr. Takeshi Yoshida, Dr. Ayumi Yoshizumi, Zhuoxin Yu, and Hiroshi Kobayashi for their helps in developing efficient expression methods for collagen peptides. I would like to thank Prof. David Case and Iwen Fu for helpful discussions on molecular dynamic simulations of triple helical peptides.

Finally, I want to express my gratitude to my family and friends for their love and support. My mother devoted her whole life to cultivating my sister and me. She would be very happy if she could know her son will be a doctor. I want to thank my wife Tingting Li, for the happiness and everything we shared and went through together.

Dedication

To my mother and my wife

Table of Contents

Abstract of the Dissertation.....	ii
Acknowledgments.....	iv
Dedication.....	vi
Table of Contents.....	vii
List of Tables.....	xii
List of Figures.....	xiii
Chapter 1 Introduction	1
1.1 Collagen: Structure, Function and Diseases	2
1.2 NMR: Structural and dynamic characterizations of proteins.....	8
1.3 Collagen peptides and NMR	10
1.4 Scope of this dissertation	12
Chapter 2 NMR methods developed on a classic triple helical peptide	13
2.1 Introduction	14
2.2 Materials and Methods.....	17
2.2.1 Sample preparation	17
2.2.2 NMR spectroscopy	18
2.2.3 Triple resonance experiments.....	18
2.2.4 NOESY and TOCSY experiments	19
2.2.5 Amide proton temperature gradients	19
2.2.6 HNHA experiments	20
2.2.7 Residual dipolar coupling experiments	20

2.2.8 ^{15}N relaxation experiments	21
2.2.8.1 Experimental procedures	21
2.2.8.2 Relaxation data analysis	22
2.3 Results and discussions.....	25
2.3.1 Sequential assignments.....	25
2.3.2 Chain assignments and inter-chain NOEs	29
2.3.3 Hydrogen bonding and backbone conformation	32
2.3.4 Model structure from NMR restraints and molecular modeling	34
2.3.5 Residual dipolar couplings	37
2.3.6 Structural insights from ^{15}N relaxation data	41
2.3.6.1 Derivation of diffusion tensor from ^{15}N relaxation data	41
2.3.6.2 Deviation of experimental and theoretical ^{15}N relaxation data	48
2.3.6.3 Hydrogen bonding leads to reorientation of Gly amide protons	49
2.3.7 Dynamics of T3-785 at fast picosecond-nanosecond timescale ...	51
2.4 Conclusions.....	53
Chapter 3 NMR studies on collagen peptides modeling Gly mutations involved in Osteogenesis Imperfecta.....	55
3.1 Introduction	56
3.2 Materials and Methods.....	60
3.2.1 Sample preparation	60
3.2.2 NMR spectroscopy	60

3.2.3 NMR diffusion experiments.....	61
3.3 Results	66
3.3.1 Design of model peptides	66
3.3.2 NMR characterization of T1-898, T1-898[G16A], and T1-898[G16S]	66
3.3.3 Effects of renucleation sequence on the incorporation of Gly to Ser substitution in triple helix.....	69
3.3.4 NMR characterization of T1-898[G16D].....	71
3.3.5 Effect of renucleation sequence and nearest neighboring sequence in Gly to Arg substitution	73
3.4 Discussion.....	77
Chapter 4 NMR Studies on heterotrimer peptides modeling a natural interruption in type IV collagen	81
4.1 Introduction	82
4.2 Materials and Methods.....	88
4.2.1 Sample preparation	88
4.2.2 NMR spectroscopy	88
4.2.3 NMR folding experiments.....	89
4.3 Results and discussion	90
4.3.1 Design of model peptides	90
4.3.2 NMR monitoring of chain-specific stability of heterotrimer A ₂ C.....	92
4.3.3 NMR shows heterotrimer formation of ABC	97

4.3.4 NMR conformation, dynamics and folding studies of heterotrimer B ₂ C	99
4.3.4.1 Sequential and chain assignments of heterotrimer B ₂ C	99
4.3.4.2 Hydrogen bonding of heterotrimer versus homotrimer	102
4.3.4.3 Model structure of heterotrimer B ₂ C	105
4.3.4.4 Dynamics of heterotrimer B ₂ C at fast picosecond timescale	107
4.3.4.5 Folding properties of heterotrimer B ₂ C	109
Chapter 5 NMR studies on triple helical peptides modeling the collagenase cleavage sites.....	111
5.1 Introduction	112
5.2 Materials and Methods.....	116
5.2.1 Sample preparation	116
5.2.2 NMR spectroscopy	116
5.2.3 Hydrogen exchange experiments	117
5.2.3.1 Experimental procedures	117
5.2.3.2 Local stabilities derived from hydrogen exchange measurements	118
5.3 Results	119
5.3.1 5-triplet model of the collagenase cleavage sites	119
5.3.2 Distribution analysis of imino acid triplets and non-imino acid triplets	123
5.3.3 Analysis of the Ile/ Leu in the X position	125

5.3.4 Peptide Design and Activities of the model peptides	128
5.3.5 Reverse activities of monomer peptides	131
5.3.6 NMR HSQC spectra of the model peptides	132
5.3.7 Conformation of the model peptides	135
5.3.8 Dynamics of the model peptides on the slow second time scale	137
5.3.9 Dynamics of the model peptides on the fast pico-second time scale	139
5.4 Discussion.....	144
References.....	150
Curriculum Vita	160

List of Tables

2.1	X-ray structure and RDCs.....	40
2.2	The ^{15}N longitudinal and transverse relaxation rates R_1 and R_2 and heteronuclear NOEs for labeled residues in T3-785 at 20°C at 500 MHz..	43
2.3	The ^{15}N longitudinal and transverse relaxation rates R_1 and R_2 and heteronuclear NOEs for labeled residues in T3-785 at 20°C at 600 MHz..	44
2.4	The ^{15}N longitudinal and transverse relaxation rates R_1 and R_2 and heteronuclear NOEs for labeled residues in T3-785 at 20°C at 800 MHz..	45
2.5	Hydrogen bond information of the labeled Glycines before and after the modification of the amide protons.....	51
3.1	Sequences and thermal stability of model peptides.....	64
3.2	Monomer diffusion coefficients of model peptides.	65
4.1	Sequences of polypeptides A, B and C.	91
5.1	Triple helical peptides modeling the collagenase cleavage sites and their activities.....	130
5.2	Monomer peptides incorporating the sequences at the collagenase cleavage sites and their activities	132
5.3	Hydrogen Exchange data of model peptides.....	138

List of Figures

1.1	Triple helical structure of collagen.	7
2.1	^1H - ^{15}N HSQC spectra of T3-785 at 20°C.	27
2.2	Strips from the HNCA spectra of T3-785. Sequential connections in individual chains are traced out as shown in colored lines.	28
2.3	Strips from the HN(CA)NH and HN(CACO)NH spectra of T3-785.	28
2.4	Comparison of predicted NOEs from the X-ray structure and experimental NOEs from the NOESY-HSQC experiment for T3-785.	31
2.5	Amide proton temperature gradients ($\Delta\delta/\Delta T$) of T3-785.	33
2.6	$^3J_{\text{HNH}\alpha}$ coupling constants of T3-785.	34
2.7	Model structure of T3-785 aligned with its X-ray structure.	37
2.8	The relative position of triple helical peptide in the diffusion tensor frame...46	
2.9	S^2 of labeled residues in T3-785.	53
3.1	Effect of Gly to Ala/Ser substitution on equilibrium conformations of peptide T1-898.	70
3.2	Effect of Gly to Asp substitution.	72
3.3	Effect of Gly to Arg substitution in different environments.	75
3.4	Effect of different Gly substitutions on the equilibrium conformations.	76
4.1	Heterotrimer peptide models of interruptions in type IV collagen.	87
4.2	Overlapped ^1H - ^{15}N HSQC spectra of peptide C and a 2A:1C peptide mixture	95

4.3	Comparison of trimer resonances of labeled Ile, Ser, Leu and Ala residues in peptide C and in a 2A:1C peptide mixture at different temperatures.	96
4.4	^1H - ^{15}N HSQC spectra of homotrimer B at 3°C and heterotrimer ABC at 0°C	98
4.5	^1H - ^{15}N HSQC spectrum of heterotrimer peptide B2C.	101
4.6	Strips from the HNCA spectra of heterotrimer B2C.	102
4.7	Amide proton temperature gradients of peptide B and C and B2C.	104
4.8	Model structure of heterotrimer B2C.	106
4.9	Heteronuclear ^1H - ^{15}N NOE values of heterotrimer peptide B2C.	108
4.10	Residue-specific folding of chain A (Left) and chain B (right) in heterotrimer peptide B2C at 10°C.	110
5.1	Natural collagenase cleavage sites in human types I-III collagens.	121
5.2	Alignments of the sequences at the collagenase cleavage sites in types I-III collagens across different species.	122
5.3	Cumulative frequencies of imino acid triplets in CO1A1, CO1A2, CO2A1 and CO3A1.	123
5.4	Frequency of m+n triplet pattern in human type III collagen.	125
5.5	Analysis of Ile/Leu in the X-position of the $\alpha 1$ chain of types I-III collagen.	127
5.6	HSQC spectra of active peptide T3-778 and inactive peptide T3-785.	134
5.7	HSQC spectra of peptide T3-778I-L, T3-778P-A and T3-778IT-PO.	135
5.8	J-coupling values of model peptides.	136
5.9	Local stabilities of model peptides.	139

5.10 Heteronuclear ^1H - ^{15}N NOE values of heterotrimer peptide T3-778.	141
5.11 R_2/R_1 values of labeled Ile17/leu17 at the cleavage sites in T3-778 series peptides.....	142
5.12 R_2 values of labeled residues in T3-778I-L at different concentrations.	143
5.13 Effect of neighboring Pro on Ile at the cleavage site.....	149

Chapter 1

Introduction

1.1 Collagen: Structure, Function and Diseases

Collagen is the most abundant protein in mammals, making up about one quarter of the total weight of proteins (Miller & Gay 1987; Shoulders & Raines 2009). As the major structural protein, collagen is found in all connective tissues, such as skin, bone, tendon, ligament, cartilage and blood vessels (van der Rest & Garrone 1991; Myllyharju & Kivirikko 2004). Collagen forms a molecular scaffold that provides mechanical strength and structural integrity to the human body, supporting and protecting the softer tissues and internal organs. There are 28 different types of collagens that have so far been identified (Kielty & Grant 2002; Thiagarajan et al. 2008). Types I, II and III collagens, also called interstitial collagens, are most abundant (Fields 1991). Type I collagen, a heterotrimer of two $\alpha 1$ and one $\alpha 2$ chains, represents over 90% of all collagens and provides tensile strength to skin, bone, tendon and ligament (Di Lullo et al. 2002). Types II collagen, a homotrimer of three $\alpha 1$ chains, is found predominantly in cartilage and together with proteoglycans provides compressive properties to joints (Eyre 2002). Type III collagen, another homotrimer, however, often occurs in elastic tissues including skin, blood vessels and various internal organs, and is more prevalent in newborns than in adults (Kuivaniemi et al. 1997).

Collagen contains a characteristic triple-helical conformation, consisting of three left-handed polyproline II-like helices which are right-handed supercoiled around a common axis (Figure 1.1) (Rich & Crick 1961; Ramachandran 1967; Bella et al. 1994). The three monomer chains can be distinguished structurally as a

leading chain, a middle chain and a lagging chain, as there is one residue staggering between the chains. The close packing of the three chains can only accommodate Gly as every third residue, generating the repetitive (Gly-X-Y)_n sequence pattern. The X and Y positions can be almost any amino acid, but they are frequently Pro and Hydroxyproline (Hyp or O), respectively. It is found that both Pro and Hyp are important for triple helix stability and Gly-Pro-Hyp is the strongest nucleating sequence to force the triple helix formation (Persikov et al. 2000; Hyde et al. 2006). All Gly are deeply buried inside the triple helix and their amide protons are pointed toward the center, while X and Y residues are located outside and their amide protons are exposed to solvent. The amide protons of Gly form hydrogen bonds with the carbonyl oxygen of the X residues in the neighboring chain, which are crucial in defining collagen structure (Bella et al. 1994; Kramer et al. 1999).

Interstitial collagens maintain a perfect Gly-X-Y repeating pattern throughout their ~1000 residue long triple helix domains (van der Rest et al. 1991; Kielty & Grant 2002). Numerous studies have shown that even a single Gly substitution in the triple helix results in diseases (Myllyharju & Kivirikko 2001; Byers & Cole 2002; Di Lullo et al. 2002; Marini et al. 2007). For instance, more than 600 different Gly to X mutations in type I collagen have been found to be associated with a connective tissue disorder osteogenesis imperfecta (OI), which is distinguished by fragile bones and easy susceptibility to fracture (Marini et al. 2007). Studies show that Gly to X mutations in type II collagen are associated with chondrodysplasias, a disease that is characterized by abnormal growth or

development of cartilage (Tiller et al. 1995). Gly substitutions in type III collagen leading to Ehlers Danlos Syndrome type IV, which may cause sudden death when aortic ruptures occur, have also been identified (Prockop & Kivirikko 1995; Kuivaniemi et al. 1997). However, the molecular mechanism of how Gly to X mutations lead to various clinical phenotypes remains poorly understood.

Despite the strict requirement of Gly as every third residue in the triple helix domain of fibrillar collagens, non-fibrillar collagens, such as type IV collagen in basement membranes, can tolerate interruptions in the Gly-X-Y pattern (Kielty & Grant 2002; Mohs et al. 2006; Thiagarajan et al. 2008). More than 20 interruptions are found in type IV collagen, while studies suggest that those interruptions may play important functions such as serving as binding sites (Brazel et al. 1987; Mohs et al. 2006). However, one extra Gly to X mutation other than the natural interruptions, results in pathological conditions (Hudson et al. 2003; Kalluri 2003). In the $\alpha 5$ chain of type IV collagen, more than 150 Gly substitutions have been identified to be associated with Alport Syndrome, a genetic disorder characterized by gradual kidney failure (Lemmink et al. 1997; Hudson et al. 2003). Both interruptions and mutations disrupt the repeating Gly-X-Y sequences, however, they have totally contrary functional consequences. It is not known if they share any similarities in structural or dynamic features.

As the major structural protein in extracellular matrix, collagen also plays a critical role in many aspects of cellular processes, including cell development, cell

adhesion and cell migration, through interactions with other matrix molecules and cellular receptors (van der Rest et al. 1991; Di Lullo et al. 2002; Myllyharju & Kivirikko 2004). Despite the relative uniform rod-like conformation and the repeating Gly-X-Y sequence, collagen can bind specifically to a plethora of molecules such as integrins, fibronectin, proteoglycans, glycoprotein VI, phospholipid vesicles, discoidin domain receptor, the mannose receptor family, and matrix metalloproteinases (Fields 1991; Kadler 1994; Knight et al. 2000; Lauer-Fields et al. 2002; Myllyharju & Kivirikko 2004). For instance, a critical GXOGER hexapeptide motif has been identified within the collagen sequence for binding to integrin $\alpha 2\beta 1$, which is involved in numerous cellular processes such as platelet activation and aggregation (Siljander et al. 2004). The comparison of X-ray crystal structures of free triple helical peptide $(\text{GPO})_2\text{GFOGER}(\text{GPO})_3$ and the peptide complexed with the I-domain of integrin $\alpha 2\beta 1$ indicates that the peptide undergoes conformational changes during the binding process (Emsley et al. 2000; Emsley et al. 2004). Studies on the binding of discoidin domain receptor to collagen suggest that the binding sequences on collagen span only two or three triplets including some charged residues (Konitsiotis et al. 2008).

Though the interactions of collagen and other proteins have been extensively investigated, the structural basis for their recognition remains poorly understood. One typical example is the binding and cleaving of collagen by matrix metalloproteinases (MMP), which are critical in the process of collagen degradation (Cawston et al. 1999; Brinckerhoff & Matrisian 2002). The

degradation of collagen is an integral part in many biological processes such as wound healing, cell migration, tissue remodeling and organ morphogenesis (Baragi et al. 1997; Visse & Nagase 2003). However, accelerated breakdown of collagen may result in many diseases such as arthritis, tumor cell invasion, glomerulonephritis and metastasis (Riley et al. 1995; Chung et al. 2004). The stability and geometry of triple helix makes collagen resistant to most proteases. It can be degraded by only a few proteinases including a group of matrix metalloproteinases (MMPs). These MMPs can bind and cleave interstitial collagens at a unique locus approximately three-fourths away from the N-terminus of the collagens. The cleavage site is after one Gly residue in the sequence of Gly-[Ile/Leu]-[Ala/leu] (Miller et al. 1976). Interestingly, there are dozens of other sites in the collagens that contain the cleavage scissile bond Gly-Ile/Leu but are not hydrolyzed (Fields 1991). It has been proposed that the distribution of 'weak helix' triplets, imino acids, hydroxyproline and charged residues as well as the side-chain volume concentration could provide recognition information for MMPs (Fields 1991). The variations in the triple helical conformation is also suggested to play a key role in the high specificity of the collagen cleavage, however, the details are largely unknown (Ryhanen et al. 1983).

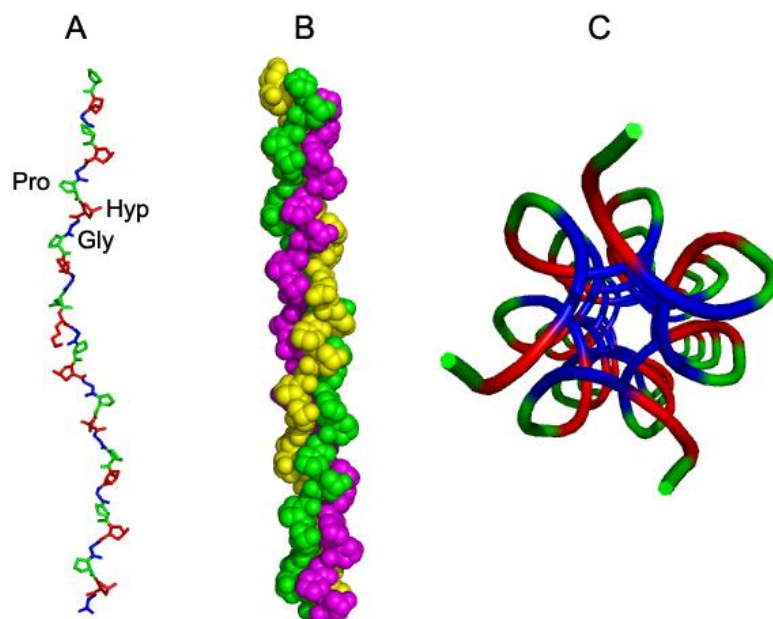


Figure 1.1 Triple helical structure of collagen.

A. The stick model of the left-handed polyproline II conformation of a monomer chain of collagen. Collagen has a repetitive Gly-X-Y sequence pattern (Gly colored in blue, X colored in green, Y colored in red), in which X and Y residues are frequently Proline and Hydroxyproline.

B. Vertical view of the sphere model of collagen triple helical structure. Three monomer chains are right-handed supercoiled about a central axis, with one-residue staggering between different chains (leading chain colored in yellow, middle chain colored in green and lagging chain colored in pink).

C. Cross-section view of the cartoon model of collagen triple helical structure. Gly are closely buried inside the triple helix (Gly colored in blue, X colored in green, Y colored in red). All the models are generated on the X-ray structure of peptide (POG)₄EKG(POG)₅ (PDB # 1QSU) by PyMol.

1.2 NMR: Structural and dynamic characterizations of proteins

NMR has emerged as a powerful tool for the study of protein structure and dynamics (Sutcliffe & Dobson 1991; Cavanagh et al. 1996; Tjandra et al. 1997; Palmer 2001; Wuthrich 2001; Kay 2005). NMR and X-ray crystallography are two major techniques for protein structure determination at the atomic level. X-ray crystallography relies on the availability of crystallized protein samples, while NMR investigates small to medium-size proteins in solution. Advances in solution NMR techniques in the past decade extend the capacity of NMR structure determination to proteins with molecular weights on the order of 100 kDa (Palmer 2004). Compared with X-ray crystallography, NMR has the unique ability to elucidate dynamic properties of biomolecules over a wide range of timescales from picosecond to second, which is crucial to understanding the molecular basis of biological processes (Kay 2005).

NMR structure determination for normal globular proteins has become a standard multistep process (Wuthrich 2003). Starting from the resonance assignments by pairs of triple-resonance experiments, conformational constraints such as proton-proton distances from NOEs and the torsion angle constraints from coupling constants will be derived. With a starting structure generated from the known amino acid sequence, a group of structures can be calculated with the inputs of experimentally determined NMR restraints. This traditional approach of NMR structure determination, however, is severely limited for multidomain proteins and linear nucleic acids where there are not enough NOEs to define the global

conformation (Hennig et al. 2001; Ryabov & Fushman 2007). Therefore, the restraints which contain long range order information are crucial for accurate NMR structure determination for those biomolecules. Novel techniques such as residual dipolar coupling and paramagnetic enhancement have been fast developing to obtain the long range restraints and are widely utilized in structure refinement, structure validation and determination of interdomain orientations (Prestegard et al. 2004; Bax & Grishaev 2005; Clore & Iwahara 2009).

As a probe of internal mobility and dynamic features of proteins, NMR spectroscopy can give insights into the relation of motion and function (Kern et al. 2005; Valentine & Palmer 2005). NMR studies have shown that conformational fluctuations in enzymes are relevant to the catalytic process (Cole & Loria 2002). It is particularly useful that dynamics of proteins spanning a wide range of timescales can be evaluated by various NMR methods (Palmer 2004). ^{15}N relaxation measurements and the derived order parameter S^2 provide useful information on the fast, picosecond-nanosecond timescale motion such as bond libration, vibration and sidechain rotation (Palmer 2001). The hydrogen exchange experiments, on the other hand, probe the slow, second timescale events such as local unfolding and the breathing of a helix (Englander et al. 1997). The exchange rates of the amide protons depend on their flexibility and capacity to be exposed to the solvents. Therefore, alterations in the exchange rates obtained from hydrogen exchange measurements can be used to identify residues with significant dynamic fluctuations, which could be involved in the interactions with other proteins.

1.3 Collagen peptides and NMR

Due to its large size and rod-like shape, collagen is too difficult to be directly investigated by solution state NMR. Instead, Triple helical peptides modeling various biologically important regions in collagen have been designed for NMR studies (Baum & Brodsky 1997; Buevich & Baum 2001). Peptides that not only satisfy the strict Gly-X-Y sequence requirement but also contain a high content of imino acids have been shown to be capable of forming a stable triple helix (Persikov et al. 2000; Hyde et al. 2006; Brodsky et al. 2008). The most stabilizing Gly-Pro-Hyp tripeptide sequences are frequently added at one or two ends to force the homotrimer formation. Alternative strategies, such as covalent linkage to force the selection of three chains and electrostatic interactions within the $(\text{Gly-X-Y})_n$ sequences to direct desired self-assembly, have been created for heterotrimer peptides (Fields et al. 1993; Ottil et al. 1996; Fiori et al. 2002; Sacca & Moroder 2002; Gauba & Hartgerink 2008; Madhan et al. 2008). These triple helical model peptides provide valuable opportunities to investigate the effects of sequence variations on stability, structure, dynamics, folding, binding, aggregation and collagenolytic protease activity (Baum & Brodsky 1999; Kramer et al. 1999; Lauer-Fields & Fields 2002; Renner et al. 2004; Brodsky & Baum 2008; Brodsky et al. 2008).

Though much simpler than collagen, triple helical peptides are still a challenging system for NMR studies as a result of their repetitive Gly-X-Y sequences, rod-like shapes and identical chains. The rod-like shape of collagen

peptides causes a high molecular tumbling rate, which results in broadened peaks with decreased intensity (Li et al. 1993; Long et al. 1993). The linear shape also means very few long-range contacts to be expected. The repetitive sequences and identical chains give rise to serious problems of degeneracy in chemical shifts. The overlapping resonances and insufficient data make the NMR structure determination of a full peptide unfeasible. Instead, the peptides are ^{15}N and/or ^{13}C selectively labeled at specific positions and residue-specific information is obtained by NMR measurements (Buevich & Baum 2001; Buevich et al. 2004; Li et al. 2009).

Triple helical peptides in the native state, in the unfolded state and in the mutant forms have been extensively investigated by NMR in our lab (Xu et al. 2003; Li et al. 2005; Li et al. 2007). Real time NMR folding studies were utilized to explore the mechanism of collagen triple helix folding and the influences of Gly substitutions and neighboring imino acids on folding (Baum & Brodsky 1998; Buevich & Baum 2001; Hyde et al. 2006). The effects of sequence variations on dynamics were also examined and model structures of triple helical peptides containing breaks or mutations in the Gly-X-Y sequences were recently solved (Long et al. 1993; Liu et al. 1998; Li et al. 2007; Li et al. 2009). These studies advanced our understandings of the relation between structure, dynamics, function and sequence variations in triple helix.

1.4 Scope of this dissertation

The repeating Gly-X-Y sequences and uniform rod-like structure makes collagen-like peptides a unique system for NMR studies. In this dissertation, a number of triple helical peptides modeling biologically important regions in collagen, such as mutation sites, interruption sites and collagenase cleavage sites, are investigated by a variety of NMR techniques. The global objective is to correlate collagen sequence variations to the changes in structural and dynamic features of collagen-like peptides, furthering our understanding of the molecular bases for collagen-involved diseases and recognitions. Chapter 2 describes the development of NMR structure determination strategies on a classic triple helical peptide T3-785 and novel approaches to obtain additional orientational restraints from ^{15}N relaxation and residual dipolar coupling measurements. Chapter 3 focuses on the characterization of the consequences of a Gly residue substituted by different bulkier residues. Chapter 4 extends our NMR studies to triple helical peptides modeling natural interruptions in the Gly-X-Y sequences in a heterotrimeric rather than a homotrimeric environment. Chapter 5 describes the application of NMR spectroscopy to identify altered conformation and dynamics at the native collagenase cleavage sites, giving insights into the mechanism of collagen triple helix cleaved by matrix metalloproteinases. All the studies focus on the structural and dynamic consequences of sequence variations in collagen and ultimately the mechanism of collagen diseases and recognition.

Chapter 2

**NMR methods developed on a classic triple
helical peptide**

2.1 Introduction

Collagen, the major structural protein in connective tissues, possesses a unique triple helical conformation, which consists of three supercoiled polypyrroline II-like helices with repetitive sequences of Gly-X-Y (Miller & Gay 1987; Shoulders & Raines 2009). Collagen is also crucial for multiple interactions with cells and other matrix molecules, and is involved in various diseases such as arthritis, cancer and metastasis (Chung et al. 2004; Myllyharju & Kivirikko 2004; Heino 2007). The mechanism by which enzymes and proteins recognize, bind to or cleave a particular site in the uniform triple helix, however, is not understood. Small variations of structure and flexibility in collagen are crucial to understanding the specificity of its recognition (Stultz 2002; Emsley et al. 2004).

NMR has been developed as a powerful tool for the study of protein structure and dynamics in the past decades (Palmer 2001; Wuthrich 2001; Kay 2005). As a huge molecule with rod-like shape, collagen, however, is too difficult to be directly investigated by NMR. Therefore, triple helical peptides modeling various biologically important regions in collagen have been designed for NMR studies (Buevich & Baum 2001; Fiori & Moroder 2001). Peptides that include several Gly-X-Y triplets from the site of interest and three or more Gly-Pro-Hyp triplets at either or both ends have been shown to be able to form a stable triple helix (Persikov et al. 2000; Brodsky et al. 2008). These triple helical model peptides provide valuable opportunities to investigate the effects of sequence variations on stability, structure, dynamics, folding, binding, aggregation and

collagenolytic protease activity (Baum & Brodsky 1999; Kramer et al. 1999; Renner et al. 2004; Brodsky et al. 2008; Li et al. 2009).

NMR structure determination of peptide models of the triple helix, however, is still very challenging due to its linear rod-like shape and repeating sequence. The anisotropic tumbling of collagen peptides results in broader and weaker trimer resonances, and the repetitive sequences and homotrimer or heterotrimer chains cause seriously overlapped resonances (Li et al. 1993; Long et al. 1993). Moreover, standard approaches of using NOEs and J-couplings to obtain structures of model triple helical peptides are limited because only a small number of distances can be obtained via NOE experiments and they tend to be very short range (Li et al. 2007; Li et al. 2009). Therefore, the restraints which contain long range order information are crucial for accurate NMR structure determination for triple helical peptides.

NMR structure techniques capable of obtaining long range restraints such as residual dipolar coupling have been rapidly developing and are widely utilized in structure refinement and validation (Bax et al. 2001; Prestegard et al. 2004; Clore & Iwahara 2009). Dipolar couplings contain orientational information of internuclear vectors relative to the magnetic field, which could be used to define both local and global geometry. However, dipolar interactions are generally averaged to zero for biomolecules in isotropic solution as a result of Brownian diffusion. Only after the discovery of liquid-crystalline media such as lipid bicelles,

filamentous bacteriophage and polyacrylamide gels, successful partial alignment for biomolecules could be achieved and meaningful residual dipolar coupling measurements became possible (Prestegard et al. 2000; Bax & Grishaev 2005).

Similar to residual dipolar couplings, NMR relaxation, in principle, also provides restraints for the orientations of internuclear vectors relative to the external magnetic field for non-spherical molecules (Tjandra et al. 1997; Hashimoto et al. 2000; Ryabov & Fushman 2007). This relaxation approach doesn't require any special sample conditions such as liquid-crystalline media; instead, it takes advantage of the inherent shape of the molecule itself. The dependence of NMR relaxation times on the rotational diffusion anisotropy gives rise to the angle information for the internuclear vectors such as N-H bonds. However, to derive structural information from relaxation data, it requires anisotropic molecules with D_{\parallel}/D_{\perp} (the ratio of diffusion tensor constants parallel and perpendicular) greater than ~ 1.5 , and the absence of large amplitude internal motions and conformational exchange in the molecules (Tjandra et al. 1997). Up till now, there have been only a few examples of the use of ^{15}N relaxation rates for global structure refinement and for determination of inter-domain orientations for globular proteins with intermediate anisotropy (Fushman et al. 1999; de Alba & Tjandra 2000; Wu et al. 2003). The molecules with large anisotropy, such as collagen peptides, which could be particularly suitable for this approach, have not been explored yet.

Aiming to develop NMR structural characterization strategies for the triple helix, we will start from a classic peptide T3-785 (with sequence (POG)₃ITGARGLAG(POG)₄Y where O stands for Hydroxyproline), which models an imino acid poor region occurring one triplet C-terminal to the unique collagenase cleavage site in type III collagen. The crystal structure of T3-785 has been solved and it shows that the peptide is a long straight rod (Kramer et al. 1999). Here, we will combine molecular modeling and NMR spectroscopy to obtain a model structure of this peptide and utilize residual dipolar couplings to probe the triple helix for the first time. Particularly, we will also demonstrate that ¹⁵N relaxation measurements, and their dependence on rotational diffusion anisotropy, can be used to determine the orientation of the N-H bonds relative to the protein backbone in the very anisotropic triple helical peptide model of collagen.

2.2 Materials and Methods

2.2.1 Sample preparation

Peptide Ac-(POG)₃ITGARGLAG(POG)₄Y-CONH₂, designated as T3-785, was synthesized by Tufts University Core Facility (Boston, MA). The peptide was selectively ¹⁵N/¹³C doubly labeled at positions G15, L16, A17, G18 and G24. The NMR sample of the peptide was prepared in 10% D₂O/90% H₂O at pH 3.1 with a concentration of 5.9 mM.

2.2.2 NMR spectroscopy

NMR experiments were performed on a Varian Inova 500MHz spectrometer, a Varian Inova 600MHz spectrometer equipped with a cryoprobe or a Varian VNMRS 800MHz spectrometer. All data were processed using the Felix 2004 software package (MSI, San Diego, CA) or NMRPipe (Delaglio et al. 1995), and then were analyzed by Felix or NMRView (Johnson & Blevins 1994). A solvent suppression filter, apodization with a 90° sine-bell window function and zero filling to double complex points were utilized. Backward linear prediction was used to extend the indirect dimensions to 1.5 times.

2.2.3 Triple resonance experiments

To accomplish sequential assignments, ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) (Kay et al. 1992) and a couple of triple resonance experiments including HNCA (Ikura et al. 1990) and HA(CA)NH/HA(CACO)NH (Kay et al. 1991) were carried out at 10°C. The 2D ^1H - ^{15}N HSQC spectrum composed of $128(t_1) \times 1024(t_2)$ complex points was recorded using the gradient sensitivity enhanced approach. The spectral widths were 1500 and 5000 Hz in the ^{15}N and ^1H dimensions, respectively. 3D HNCA experiment consisted of $80(t_1) \times 32(t_2) \times 1024(t_3)$ complex points, and was recorded with spectral widths of 4000(F_1), 1500(F_2) and 5000(F_3) Hz. 3D HA(CA)NH and HA(CACO)NH experiments comprised $80(t_1) \times 28(t_2) \times 1024(t_3)$ and $80(t_1) \times 32(t_2) \times 1024(t_3)$ complex points, respectively; and both were recorded with spectral widths of 2200(F_1), 1600(F_2) and 6000(F_3) Hz.

2.2.4 NOESY and TOCSY experiments

3D ^{15}N edited NOESY-HSQC (Fesik & Zuiderweg 1988; Marion et al. 1989) experiments were performed with a mixing time of 30ms at 20°C and another mixing time of 50ms at 10°C. A short mixing time of 30ms was employed to wipe out spin diffusion, while different temperatures were utilized to reduce overlapped resonances. The NOESY-HSQC experiments consisted of $120(t_1) \times 32(t_2) \times 512(t_3)$ or $112(t_1) \times 32(t_2) \times 512(t_3)$ complex points, and were recorded with spectral widths of 5000(F_1), 1500(F_2) and 5000(F_3) Hz. 3D ^{15}N edited TOCSY-HSQC (Fesik & Zuiderweg 1988; Messerle et al. 1989) experiments were carried out with a mixing time of 50ms at 20°C at 10°C. The TOCSY-HSQC experiments comprised $120(t_1) \times 32(t_2) \times 512(t_3)$ or $128(t_1) \times 32(t_2) \times 512(t_3)$ complex points, and were recorded with spectral widths of 5000(F_1), 1500(F_2) and 5000(F_3) Hz.

2.2.5 Amide proton temperature gradients

To obtain hydrogen bonding information, amide proton temperature gradients were measured by taking a series of ^1H - ^{15}N HSQC spectra at a range of temperatures from 0°C to 25°C with an interval of 5°C. The sample was equilibrated for >3 hours before the spectrum was taken at each temperature. Chemical shifts of the amide protons were measured at each temperature and the temperature gradient was obtained for each residue by linear regression of its chemical shifts versus temperature.

2.2.6 HNHA experiments

3D HNHA experiment (Vuister & Bax 1993) was carried out with an H-H coupling period of 25ms at 20°C to measure $^3J_{\text{HNH}\alpha}$ coupling constants, which provided information about phi angles. The time domain data sets consisted of 512 complex points in the t_3 dimension, and 100 and 28 increments in the t_1 and t_2 dimensions, respectively. The spectral widths were 1600 and 6000 Hz in the ^{15}N and ^1H dimensions, respectively. A correction factor for calculating $^3J_{\text{HNH}\alpha}$ coupling constants was utilized as previously described (Li et al. 2007).

2.2.7 Residual dipolar coupling experiments

$^1D_{\text{NH}}$, one-bond ^{15}N - ^1H residual dipolar coupling, was derived by calculating the difference in the corresponding J couplings ($^1J_{\text{NH}}$) measured in both aligned and isotropic states (Andrec et al. 2001; Bax et al. 2001). The $^1J_{\text{NH}}$ couplings were extracted from 2D IPAP- $[^1\text{H}, ^{15}\text{N}]$ -HSQC spectra, recorded as two interleaved data matrices (Skrynnikov & Kay 2000). The differences of the chemical shifts for each residue in the two matrices were determined as the values for $^1J_{\text{NH}}$ couplings. The sample in the isotropic state was prepared as described above. The samples in the aligned state were prepared using polyacrylamide gels.

A 10 % polyacrylamide gel was prepared following protocols previously described (Sass et al. 2000). Immediately after the addition of 30 μL 5 % APS (ammonium persulfate) and 30 μL 5 % TEMED (tetramethylethylenediamine) to the solution of 333 μL 30 % Bis-Acrylamide and 607 μL water, the mixture was

carefully pipetted into a 3" Tygon tube (with an inner diameter of 5/32"), which had been capped by parafilm on one end. After ~2 minutes for polymerization, the gel was pushed out from the tube, washed for >15 hours in deionized water, and cut to a length of 23 mm. After incubation in the peptide solution (350 μ L) >24 hours at room temperature, the cut gel was gently inserted into a Shigemi tube and compressed with the tube insert.

2.2.8 ^{15}N relaxation experiments

2.2.8.1 Experimental procedures

^{15}N R_1 , R_2 and heteronuclear NOE experiments (Fan et al. 1993; Palmer 1993; Farrow et al. 1994) were performed at least two times at magnetic fields 500 and 600 MHz and one time at 800 MHz. For 500 MHz NMR, the spectral widths were 1500 and 5000 Hz in the ^{15}N and ^1H dimensions, respectively. Nine relaxation delays (50, 150, 250, 400, 500, 600, 700, 800, 900 ms) were used in the R_1 measurement. Seven relaxation delays (4.8, 14.4, 24, 33.6, 43.2, 52.8 and 62.4 ms) were used in the R_2 measurement. For the NOE measurements, spectra were recorded in the presence and absence of ^1H saturation. A recycle decay of 1.5s, 3s and 2s were used in the R_1 , R_2 and NOE experiments, respectively. The R_1 spectra composed of 128 or 100 (t_1) \times 512(t_2) complex points; the R_2 spectra composed of 128 or 100 (t_1) \times 1024(t_2) complex points; the NOE spectra consisted of 128 or 88 (t_1) \times 512(t_2) complex points, while 128 or 96 scans per t_1 increment were used. The parameters were setup similarly for the R_1 , R_2 and NOE experiments at 600 and 800 MHz magnetic fields. The R_1 , R_2 and NOE values were obtained as the

average of all data sets at each magnetic field and the error was determined as the larger value of the experimental error or the fitting error. The error of R_2/R_1 ratio was calculated as follows:

$$\left(\frac{R_2}{R_1}\right)_{\text{error}} = \left(\frac{R_2}{R_1}\right) * \sqrt{\left(\frac{R_{1, \text{error}}}{R_1}\right)^2 + \left(\frac{R_{2, \text{error}}}{R_2}\right)^2} \quad (1)$$

2.2.8.2 Relaxation data analysis

The relaxation parameters R_1 , R_2 and NOE are defined by the spectral density function as follows (Lipari & Szabo 1982; Fan et al. 1993):

$$R_1 = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2J(\omega_N) \quad (2)$$

$$R_2 = \frac{d^2}{2}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{3}[4J(0) + 3J(\omega_N)] + R_{\text{ex}} \quad (3)$$

$$\text{NOE} = 1 + \frac{\gamma_H}{\gamma_N} (d^2) \frac{6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)}{R_1} \quad (4)$$

where

$$d^2 = \frac{\hbar^2 \gamma_N^2 \gamma_H^2}{4r_{\text{NH}}^6} \quad \text{and} \quad c^2 = \frac{\omega_N^2 (\delta_{\parallel} - \delta_{\perp})^2}{3} ;$$

in which \hbar is Planck's constant divided by 2π ; γ_N and γ_H are the gyromagnetic ratios for the ^{15}N nuclei and amide proton, respectively; r_{NH} is the NH bond length; ω_H and ω_N are the angular resonance frequencies for the ^1H and ^{15}N nuclei, respectively.; δ_{\parallel} and δ_{\perp} are the parallel and perpendicular components of the chemical shift tensor; R_{ex} represents the contribution of conformational exchange to the R_2 relaxation; $J(\omega_i)$ is the spectral density function. Two factors contributing to R_1 and R_2 are dipolar interaction between ^{15}N - ^1H (the d^2 part in equations 1 and 2) and ^{15}N chemical shift anisotropy (the c^2 part in equations 1 and 2).

The spectral density function depends on the type of molecular motion. Protein dynamics at fast picosecond-to-nanosecond timescale can be described by overall motion (the global reorientation for the whole molecule) and internal motion (local orientation of a bond within its own molecular frame). Using Lipari and Szabo's model-free formalism (Lipari & Szabo 1982), which assumes independent overall motion versus internal motion, the spectral density function for globular proteins with isotropic motion can be expressed as follows:

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{(1 - S^2) \tau}{1 + (\omega \tau)^2} \right] \quad (5)$$

where

$$\frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e};$$

in which τ_c and τ_e are the overall correlation time and effective correlation time for the internal motion, respectively; S^2 is the generalized order parameter.

For axially symmetric molecules such as triple helical peptides, assuming much faster internal motion than overall motion ($\tau_e \ll \tau_1, \tau_2$ and τ_3) and the internal motion within the extreme limit ($(\omega \tau_e)^2 \ll 1$), the spectral density function can be approximated as follows (Farrow et al. 1994; Tjandra et al. 1995):

$$J(\omega) = \frac{2}{5} \left\{ S^2 \left[\frac{A \tau_1}{1 + (\omega \tau_1)^2} + \frac{B \tau_2}{1 + (\omega \tau_2)^2} + \frac{C \tau_3}{1 + (\omega \tau_3)^2} \right] + (1 - S^2) \tau_e \right\} \quad (6)$$

where

$$1/\tau_1 = 6D_{\parallel}, \quad 1/\tau_2 = 6D_{\parallel} + D_{\perp} \text{ and } 1/\tau_3 = 2D_{\parallel} + 4D_{\perp};$$

(τ_1 , τ_2 and τ_3 are the three correlation times; D_{\parallel} and D_{\perp} are the parallel and perpendicular components of the diffusion tensor.)

and

$$A = \frac{1}{4}(1 - 3 \cos^2 \theta)^2,$$

$$B = 3 \sin^2 \theta \cos^2 \theta,$$

$$C = \frac{3}{4} \sin^4 \theta;$$

(A, B and C are coefficients dependent on the θ angle, which is the angle between the N-H vector and the principal axis of diffusion tensor as illustrated in Figure 2.1b.)

As shown above, dynamic parameters can be derived from the R_1 , R_2 and NOE measurements to evaluate protein motions, including τ_c (or τ_1 , τ_2 and τ_3 for anisotropic molecules) for overall motion as well as τ_e and S^2 for internal motion. For anisotropic molecules, those measurements also contain structural information, the θ angle between the N-H vector and the principal axis of diffusion tensor. In the absence of significant amplitude internal motion ($\tau_e = 0$) and conformational exchange ($R_{ex} = 0$), the R_1 , R_2 and NOE values depend on the diffusion tensor, the θ angle, and S^2 , while the R_2/R_1 ratio depends only on the diffusion tensor (specifically the parallel and perpendicular components D_{\parallel} and D_{\perp}) and the θ angle, as S^2 is proportionally included in both R_1 and R_2 , and thus can be crossed out in the R_2/R_1 ratio. Based on the experimental R_2/R_1 values and the

known structure, the diffusion tensor could be derived using a fitting program R_2R_1 _diffusion (Tjandra et al. 1995).

2.3 Results and discussions

2.3.1 Sequential assignments

The HSQC spectrum of peptide T3-785 was well resolved (Figure 2.1a). All labeled residues in the central region had three trimer resonances due to the non-equivalent environments among the three chains, while G24 in the repeating GPO environment showed a single trimer resonance. One monomer resonance was also observed for each residue, supporting a trimer form of the peptide in equilibrium with monomer. Residues with the same superscripted number (eg. 1T G15, 1T L16) in the HSQC spectrum were assigned to the same chain. These chain-specific sequential assignments were based on the triple resonance experiments HNCA, HA(CA)NH and HA(CACO)NH.

As the HNCA spectrum correlates the N and amide proton NH in residue i (N_i and NH_i) to the alpha carbons CA in residue i and $i-1$ (CA_{i-1} and CA_i), two peaks for residue i , one originated from CA_i and the other from CA_{i-1} , can be observed in one HNCA strip if residue $i-1$ exists and is doubly labeled (Figure 2.2). Gly has distinct chemical shifts in the ^{15}N dimension and Gly15 doesn't have any preceding labeled residue, therefore, Gly15 was easily determined by its characteristic single peak in each strip. By matching the alpha carbon chemical shifts, three groups of sequentially connected residues were identified, as

indicated by different colors for the corresponding peaks (red vs black vs green). A pair of HA(CA)NH/HA(CACO)NH experiments, correlating NH in residue *i* to HA in residue *i* and/or HA in residue *i*-1, were performed to remove any ambiguity in the assignments from HNCA spectra. Similarly, sequential connections in the three individual chains were traced out as shown in colored lines (Figure 2.3). These results lead to the sequential assignments of the resonances in the HSQC spectrum (Figure 2.1).

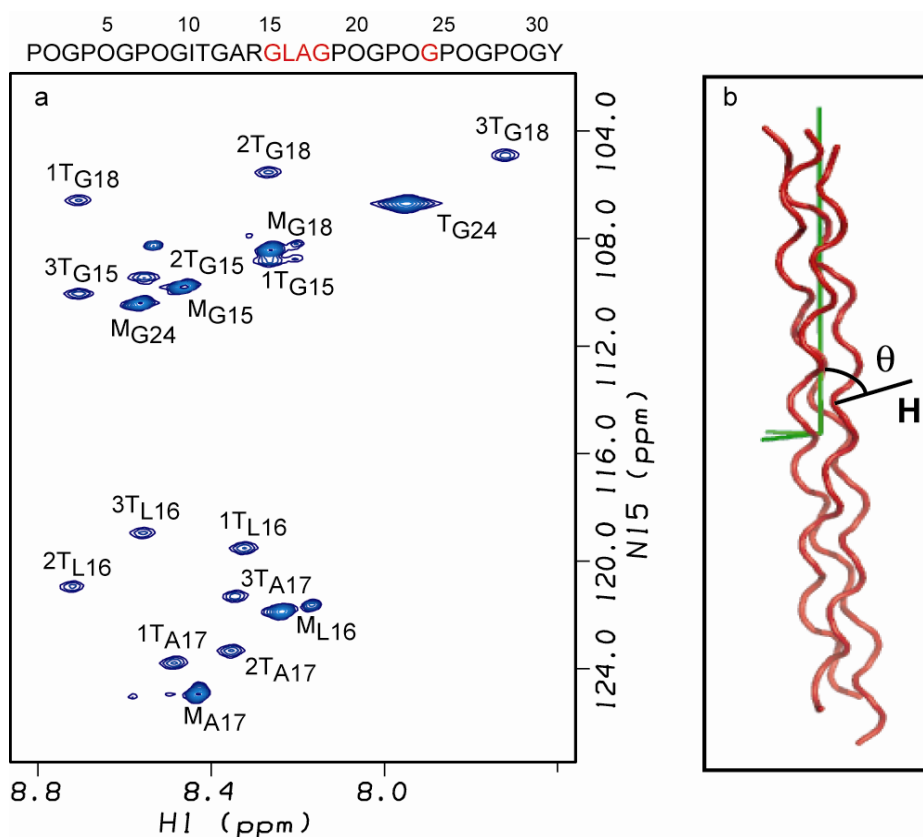


Figure 2.1 (a) ^1H - ^{15}N HSQC spectra of T3-785 at 20°C. $^{15}\text{N}/^{13}\text{C}$ doubly labeled residues are colored red in the sequence. The monomer and trimer peaks are denoted with a superscript M or T, respectively. The trimer resonances are further assigned to individual chains indicated by a 1, 2 or 3 referring to the leading, middle or lagging chain. (b) The structure of T3-785 positioned in the diffusion frame (in green). The θ angle between a representative N-H vector and the principal axis of the diffusion tensor is illustrated.

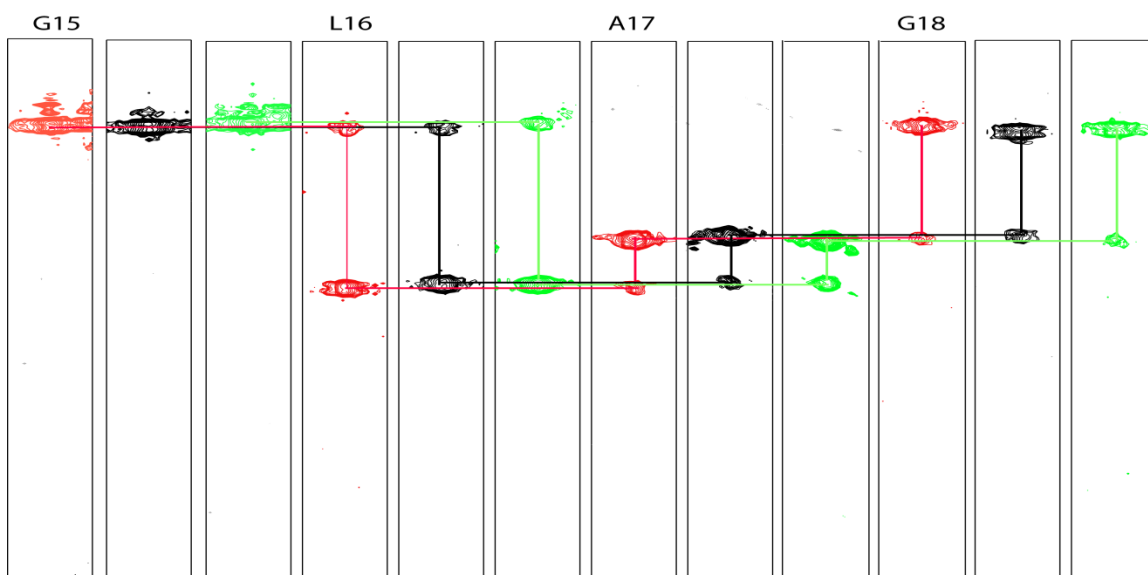


Figure 2.2 Strips from the HNCA spectra of T3-785. Sequential connections in individual chains are traced out as shown in colored lines.

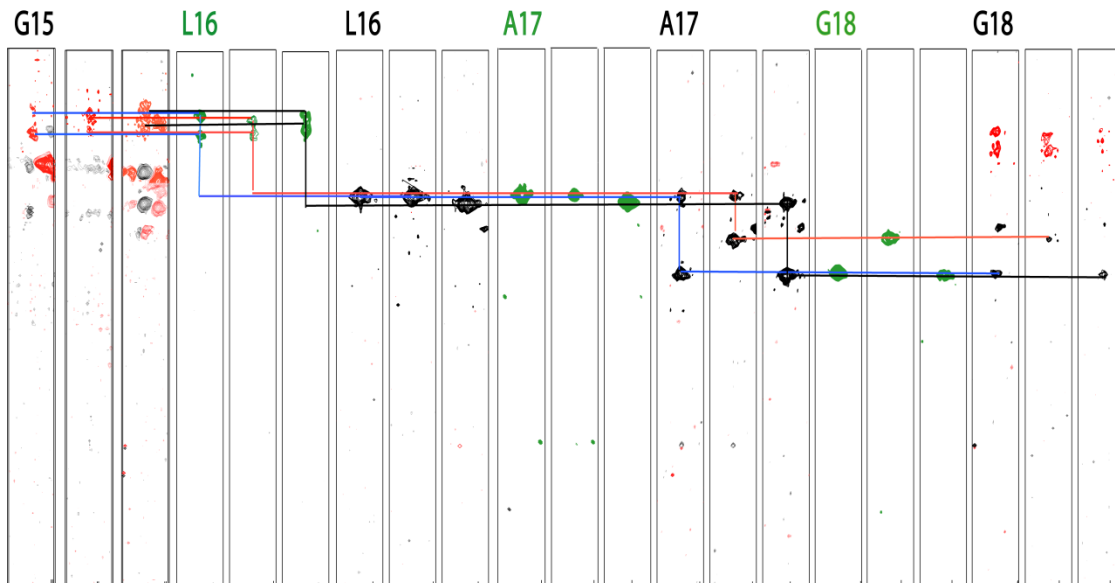


Figure 2.3 Strips from the HN(CA)NH (black) and HN(CACO)NH (green) spectra of T3-785. Sequential connections in individual chains are traced out as shown in colored lines.

2.3.2 Chain assignments and inter-chain NOEs

From triple resonance experiments, the trimer resonances were assigned to specific residues in different chains numbered 1, 2 and 3. However, the three chains in triple helix were not the same because of one-residue staggering, and they could be distinguished as the leading chain, the middle chain and the lagging chain. As a result, residues in different chains (e.g. leading chain vs lagging chain) would have different patterns of interactions with neighboring chains and inter-chain NOEs could be utilized to accomplish the chain assignments.

All NOEs observed from the NOESY-HSQC experiments were summarized in the experimental contact map as described previously (Figure 2.4b) (Li et al. 2007). The colored circles were utilized to represent experimentally observed NOEs. The shaded squares in the background denote NOEs expected from the X-ray structure of T3-785 (yellow for inter-chain NOEs and gray for intra-chain NOEs) (Kramer et al. 1999). A predicted contact map was made to summarize the information of those expected NOEs or contacts within 5Å (Figure 2.4a). Those contacts within 5Å were categorized into NH-NH (●), NH-HA (●) and NH-side chain proton (●) contacts. The squares were shaded gray if the predicted contact was intra-chain and yellow if it was inter-chain. Comparing the predicted contact map with the experimental contact map, it was noted that much less NOEs were observed than predicted. Particularly, only a few inter-chain NOEs including ¹G15 NH- ²G15 NH, ²G15NH- ³G15 NH, ¹G18 NH- ²G18 NH, ²G18 NH- ³G18 NH, ¹L16 NH-³R14 αH, ²L16 NH-¹A17 αH and ³L16 NH-²A17 αH could be observed. The

inter-chain NOEs between Gly15 in the 2nd chain with the Gly in the other two chains, suggested that the 2nd chain was the middle chain. The inter-chain NOE between ¹L16 NH-³R14 αH indicated that the 1st chain was the leading chain, while the third chain was the lagging chain. All other inter-chain NOEs supported these chain assignments. The inter-chain NOEs played a key role in the chain assignment, and it also provided crucial distance restraints for obtaining model structures.

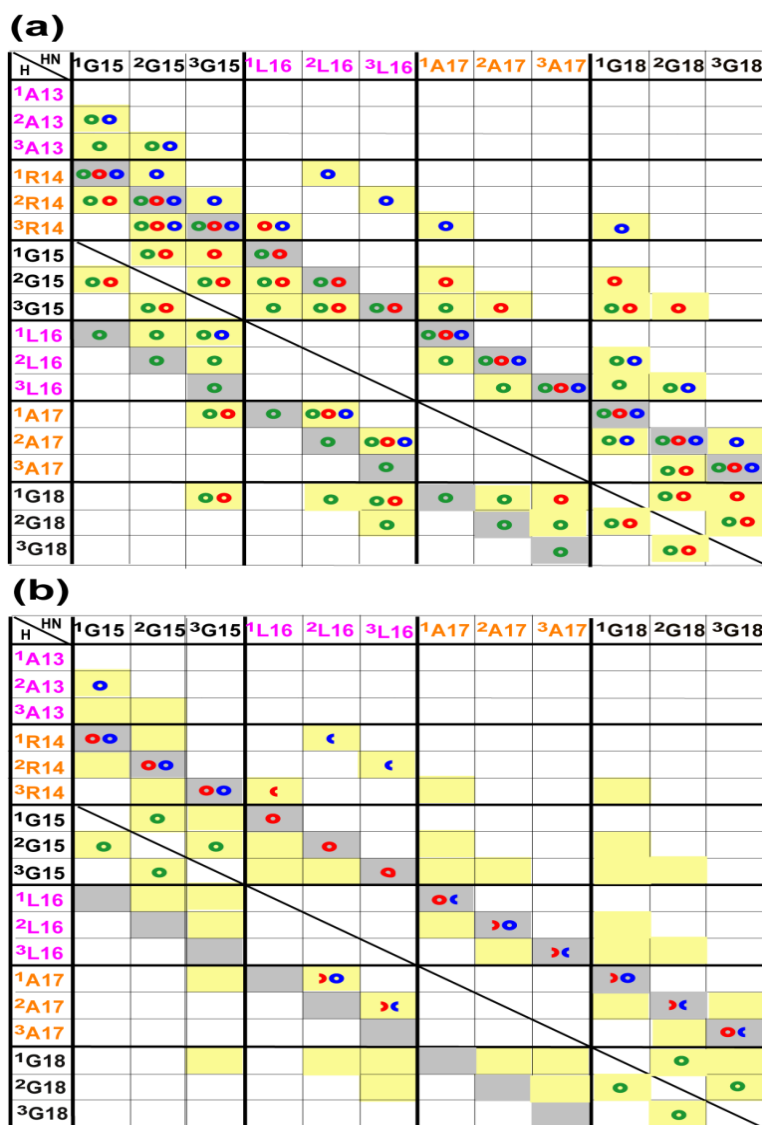


Figure 2.4 Comparison of predicted NOEs from the X-ray structure and experimental NOEs from the NOESY-HSQC experiment for T3-785. (a) Predicted contact map from the X-ray structure of T3-785. Contacts within 5Å are shaded in gray if the contact is intra-chain and in yellow if it is inter-chain. Those contacts within 5Å are further categorized into NH-NH (○), NH-HA (○) and NH-side chain proton (○) contacts. The colored circles are overlaid on the shading. (b) Experimental contact map based on observed NOEs for T3-785. The colored full circles indicate observed NOEs for NH-NH (○), NH-HA (○) and NH-side chain proton (○), while the half circles (◐) represent overlapped NOEs consistent with prediction. The experimental contacts are overlaid on the shaded contacts for easy comparison.

2.3.3 Hydrogen bonding and backbone conformation

Amide proton temperature gradients ($\text{NH } \Delta\delta/\Delta T$) were measured for peptide T3-785 to obtain hydrogen bonding information (Figure 2.5). Early studies suggested that amide protons showing the temperature gradients with a more positive value than -4.6 supported the existence of hydrogen bonds (Baxter & Williamson 1997; Cierpicki & Otlewski 2001). From the experiments, Gly15, Gly18 and Gly24, showed their $\text{NH } \Delta\delta/\Delta T$ to be in the range of -4 to -1, indicating the formation of hydrogen bonds (Figure 2.5). These results were consistent with the X-ray structure and early hydrogen exchange studies of the peptide, which both suggest the formation of hydrogen bonds for Gly. Residues Leu16 and Ala17 in the triple helix, as well as the residues in the monomer state, all had temperature gradients more negative than -4.6, indicating the absence of hydrogen bonds. In all, the hydrogen bonding information of the Glycines obtained from temperature gradient measurements provided useful restraints for obtaining model structures of the peptide.

The dihedral angle Φ is a critical parameter for defining the backbone conformation and it can be calculated from $^3J_{\text{HNH}\alpha}$ coupling constants by the parameterized Karplus equation: $^3J_{\text{HNH}\alpha} = 6.51\cos^2(\phi-60^\circ) - 1.76\cos(\phi-60^\circ) + 1.6$ (Vuister & Bax 1993). $^3J_{\text{HNH}\alpha}$ coupling constant is usually obtained using the 3D HNHA experiment, in which the volume ratio of the cross peak to diagonal peak is measured. Due to the different relaxation rates of cross peak and diagonal peak in the HNHA experiment, the cross peak intensity is artificially reduced relative to the

diagonal peak intensity (Wang & Bax 1996). Therefore, a correction factor of 1.16 was utilized to calculate the final coupling constants (Li et al. 2007). From the measurements, all residues contain a J coupling value around 4 or 5, which is normal for a triple helix polyproline II conformation. Gly showed two measured J coupling values, as it had two alpha protons.

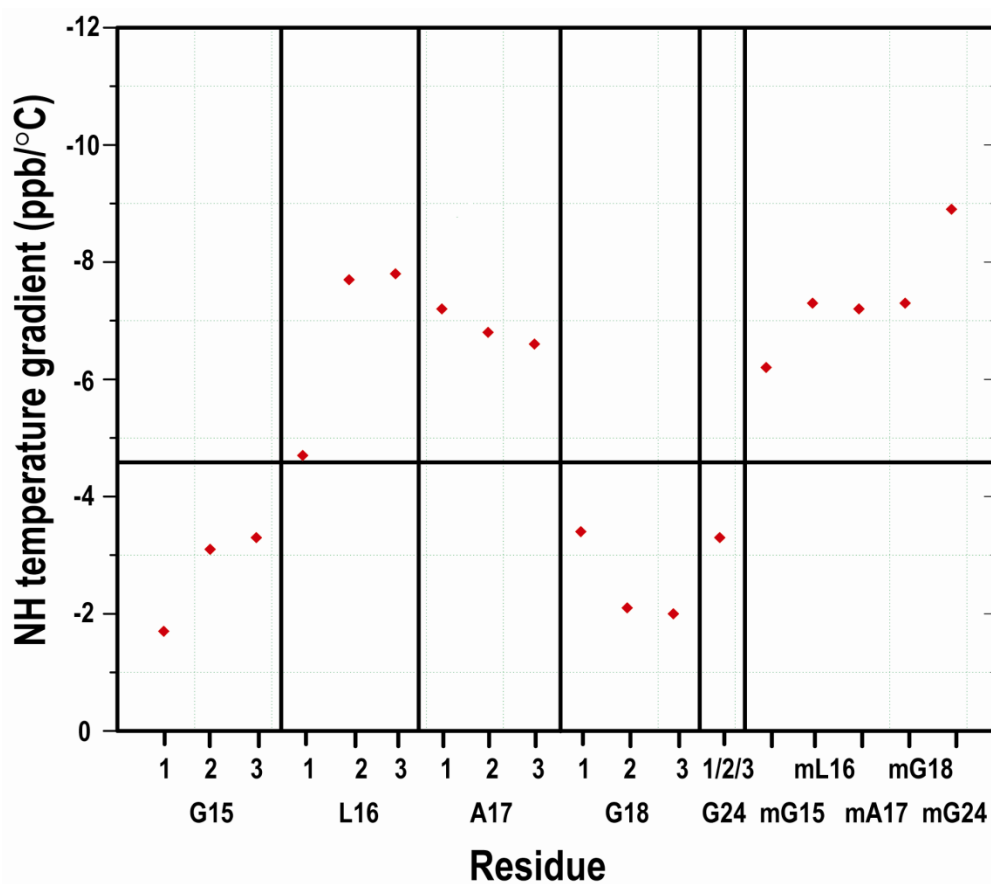


Figure 2.5 Amide proton temperature gradients ($\Delta\delta/\Delta T$) of T3-785. The black horizontal line corresponds to a value of -4.6, a cut-off for hydrogen bonding, with less negative values indicative of hydrogen bonding.

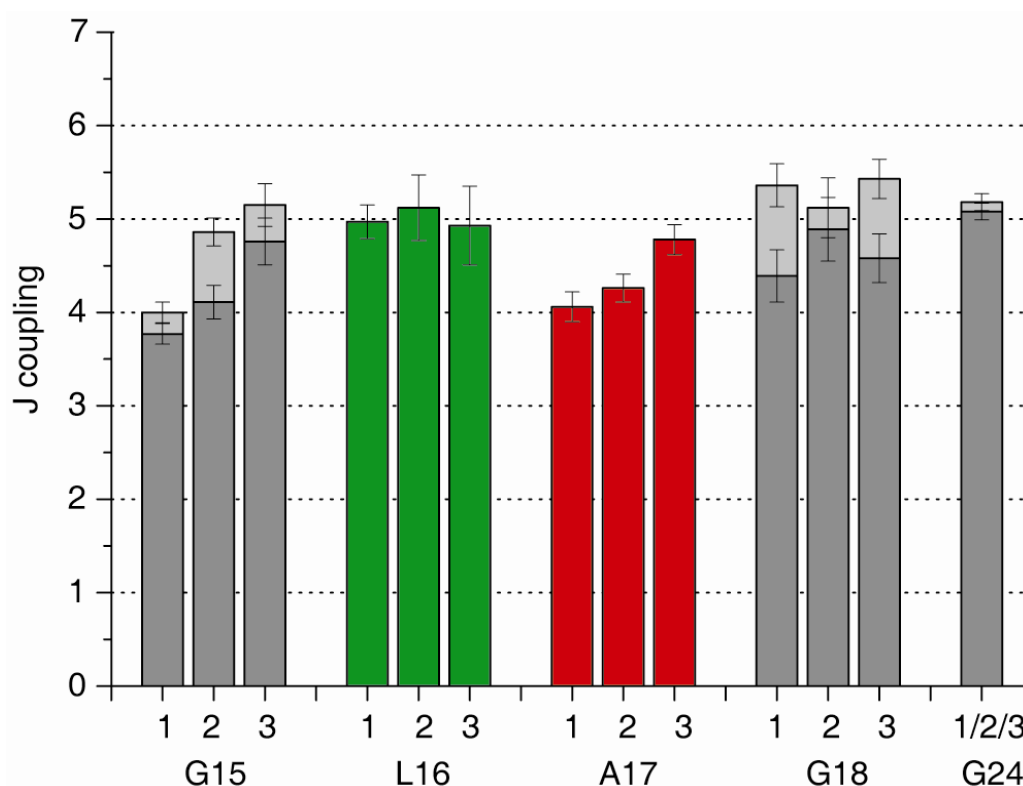


Figure 2.6 $^3J_{\text{HNH}\alpha}$ coupling constants of T3-785. All the residues show a J coupling value between 4-5 Hz, which is typical for triple helix.

2.3.4 Model structure from NMR restraints and molecular modeling

Compared with globular proteins, the triple helical peptide T3-785 showed much fewer NOEs, which were insufficient for NMR structure determination. Alternatively, a model structure was built by molecular modeling strategy incorporating NMR restraints using Molecular Operating Environment (MOE) (Chemical Computing Group Inc., Montreal, Canada) following previously described protocols (Li et al. 2007). First, the X-ray structure of another triple helical peptide (POG)₃-**POGEKGPOG**-(POG)₄ was chosen as the starting structure (Kramer et al. 2000), trying to see if a model structure similar to the X-ray

structure of T3-785 can be obtained by this method. The middle residues POGEKGPOG were mutated to ITGARGLAG, making the sequence for peptide T3-785: (POG)₃-ITGARGLAG-(POG)₄. Then energy minimization was performed on the 9 central residues with NMR restraints. The repeating POG ends were not expected to deviate from the idealized triple helix, so they were fixed in all the simulations.

The NMR restraints include distance restraints from NOEs, hydrogen bonding restraints and Φ angles. The distance limits from 1.8 to 4.2 Å were set for inter-chain NOEs shown in the experimental contact map (Figure 2.4b). The hydrogen bonding restraints for Gly15 and Gly18 include hydrogen bonding distance limits of 1.8-2.2 Å and hydrogen bonding N-H-O angle limits of 145-180 degrees. The Φ angle restraints were more complicated, as one J coupling value of ~4.5 gave rise to four possible Φ angles: roughly -180, -60, 20, 100 degrees. As positive Φ angles were generally unfavorable for any amino acid except Gly, a range of -180 to -60 degrees was set as Φ angle limits for Leu and Ala, while a larger range of -180 to 100 degrees was set as Φ angle limits for Gly. Exact Φ angle limits were determined by precisely calculating the four solutions of the J couplings of each residue. The Φ angle limits could be further narrowed to around -180 degree or around -60 degree, if the energy minimized structure didn't produce proper Φ angle consistent with the experimental J-coupling value.

After energy minimization, the model structures were then validated using the program Molprobity, which evaluates clashes, rotamer and ramachandran plot (Lovell et al. 2003). A validated model structure of T3-785 was compared with the X-ray structure of T3-785. The backbone RMSD of 1.21 Å indicated that the model structure was similar to the X-ray structure (Figure 2.7a). The Ramachandran plot for central residues in T3-785 indicated that all the residues were in the nice triple helical polyproline II conformation (Figure 2.7b). In all, molecular modeling incorporating NMR restraints was demonstrated to be a useful approach to obtain model structures for triple helical peptide T3-785.

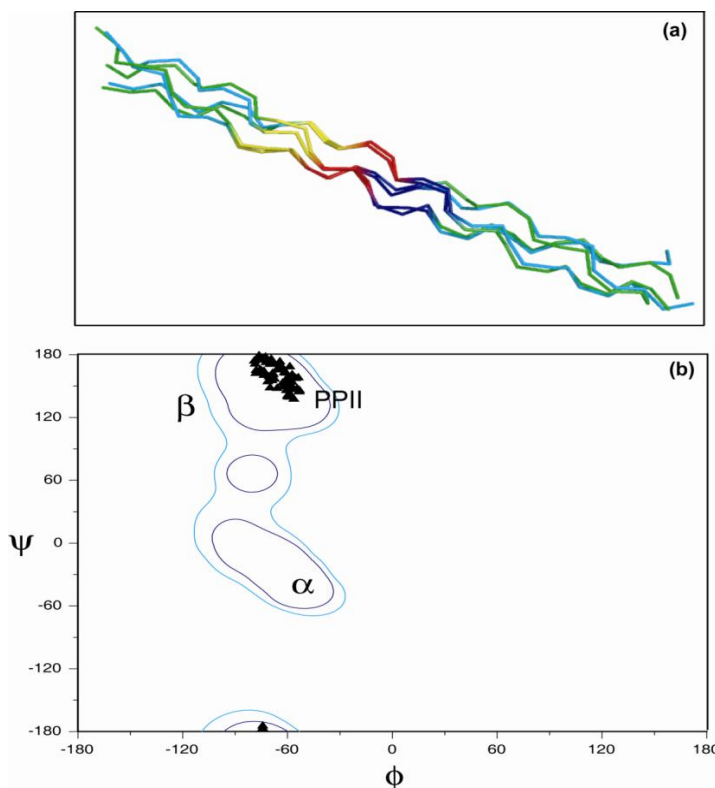


Figure 2.7 (a) Model structure of T3-785 (cyan) aligned with its X-ray structure (green). Triplets ITG, ARG and LAG are colored in yellow, red and blue, respectively. (b) Ramachandran plot for central residues in T3-785. Favorable regions for α -helix, β -sheet and polyproline II are indicated, while favorable regions for Pro are circled.

2.3.5 Residual dipolar couplings

Residual dipolar couplings (RDCs) have been widely utilized to obtain local as well as global structural information in biomolecules (Lipsitz & Tjandra 2004; Prestegard et al. 2004). It has been demonstrated that RDCs provide particularly useful information for the biomolecules which lack enough NOEs to define the overall structure, such as DNA and multi-domain proteins (Zhou et al. 1999;

Tolman 2001; MacDonald & Lu 2002). Triple helical model peptides, containing linear shapes like DNA, also suffer from lack of long-range NOEs. Thus, the RDC measurements were first developed on T3-785, and then applied to another triple helical peptide (POG)₁₀Gly→Ala, to investigate how RDCs could help to define normal triple helical conformation as well as mutations in the triple helix.



X-ray crystallography studies have shown that some triple helical model peptides do not display idealized triple helix structure (Bella et al. 1994; Kramer et al. 1998; Kramer et al. 1999; Bella et al. 2006). For peptide T3-785, it shows different helical symmetry among different zones: a 7₅ symmetry for the repeating POG terminal, while a 10₇ symmetry for the imino acid poor central region (Kramer et al. 1999). The two ends also appear to be bent relative to the straight central region (Table 2.1). For peptide (POG)₁₀Gly→Ala, it displays a single helical symmetry of 10₇ for the whole peptide, but a local bulge and untwisting is also observed at the Gly to Ala mutation site (Bella et al. 1994). These variations in triple helical structures, either locally or globally, might be able to be detected by RDCs. A program PALES was used to predict the RDCs of the two peptides indicating that the RDCs were sensitive to these differences (Zweckstetter & Bax 2000).

Experimental RDCs of the two peptides were also obtained using neutral compressed polyacrylamide gels (Table 2.1). For both peptides, Gly24 in the POG ends showed similar RDC values of ~-3. For peptide T3-785, residues in the

central region showed pretty variable RDC values. For Leu16 and Gly18, they even showed positive values in one chain whereas negative values in the other two chains, indicating that the residues at different positions or in different chains in the imino acid poor region have dissimilar orientations to the alignment tensor (Table 2.1). For peptide (POG)₁₀Gly→Ala, Ala15 at the mutation site showed different RDC values from Gly24, indicating the orientations of amide protons affected by mutations (Table 2.1).

Here, the RDC measurements were demonstrated to be able to probe the variations in triple helix structure. In the future, these RDCs will be incorporated with other NMR restrains for structure determination. The RDC measurements will also be performed in other media in order to non-ambiguously define the orientations in the alignment tensor. RDCs of other inter-nuclear vectors, such as $^1D_{HA-CA}$, $^1D_{C-CA}$ and $^2D_{C'-HN}$, will also be measured to obtain more restraints for NMR structure determination (Prestegard et al. 2000; Bax et al. 2001).

Table 2.1 X-ray structure and RDCs

T3-785 (POG) ₃ -ITG-ARG- <u>LAG</u> -(POG) ₄				(POG) ₁₀ Gly→Ala (POG) ₄ -POA- <u>(POG)</u> ₅			
							
RDC	1	2	3	RDC	1	2	3
G15	n/a	n/a	-3.6	A15	-1.5	-1.4	-2.1
L16	1.7	-2.8	-0.4	G24	-2.8		
A17	-2.1	-0.9	-1.0				
G18	-4.1	0.8	-3.2				
G24	-3.1						

Upper panel shows the ribbon models of two peptides. ¹⁵N labeled residues are underlined in the sequence. Principle axis is colored in green. Lower panel shows the experimental RDC values for labeled residues in all three chains within the peptides.

2.3.6 Structural insights from ^{15}N relaxation data

2.3.6.1 Derivation of diffusion tensor from ^{15}N relaxation data

The relative ratio of the principal components of the inertia tensor calculated from the x-ray structure of T3-785 is 1:1:0.03, indicating that the peptide can be modeled as an axially symmetric rotor (Tjandra et al. 1995). Using a simple prolate ellipsoid model, the ratio of the principal values of the diffusion tensor D_{\parallel}/D_{\perp} is calculated as 11.9, indicating that the peptide has an extremely large anisotropy (Copie et al. 1998). The ratio of the ^{15}N transverse and longitudinal relaxation rates, R_1 and R_2 , of a protein with axially symmetric diffusion, in the absence of large amplitude motion and conformational exchange, depends on the θ angle between the N-H bond vector and the unique axis of the diffusion tensor, and on the diffusion tensor constants D_{\parallel} and D_{\perp} . These can be fit to the observed R_2/R_1 values using the program $R_2R_1_diffusion$, assuming values for the θ angles from the crystal structure (Tjandra et al. 1995).

Relaxation rates R_1 and R_2 were obtained at 500MHz for labeled residues in the central region of T3-785 including Gly15, Leu16, Ala17, and Gly18 and at the C-terminal end, Gly24 (Table 2.2). The R_2 values were similar for all the labeled residues, while the R_1 values showed small differences with Gly15 and Gly18 having the largest R_1 values and Ala17 the smallest values. The differences in R_1 resulted in a range of R_2/R_1 ratios from 5.6 to 6.64. Relaxation rates R_1 and R_2 were also obtained at 600MHz and 800MHz (Table 2.3, 2.4). Similar trends for the R_1 values were observed: Gly15 and Gly18 have the largest values and Ala17 the

smallest values, though the differences between Leu16 and Ala17 became smaller. It was also observed that Ala17 had a little larger R_2 values than Leu16. For all the three fields, Ala17 had the largest R_2/R_1 ratios, while Gly15 and Gly18 had the smallest values. Relaxation data at multiple fields gave us more information to calculate the diffusion tensor and characterize the fast timescale motion using Model-free formalism.

Residue	R ₁	Error	R ₂	Error	R ₂ /R ₁	Error	NOE	Error
1G15	1.99	0.03	11.4	0.2	5.7	0.11	0.63	0.01
2G15	2.02	0.02	11.8	0.1	5.87	0.12	0.68	0.02
3G15	1.99	0.02	11.8	0.2	5.95	0.12	0.64	0.02
1L16	1.9	0.02	11.5	0.2	6.06	0.12	0.68	0.03
2L16	1.96	0.03	11.4	0.2	5.78	0.12	0.72	0.06
3L16	1.93	0.02	11.5	0.2	5.95	0.12	0.72	0.01
1A17	1.82	0.03	11.8	0.3	6.49	0.13	0.60	0.05
2A17	1.79	0.02	11.7	0.2	6.5	0.13	0.65	0.04
3A17	1.81	0.02	12	0.2	6.64	0.13	0.69	0.02
1G18	2.02	0.03	11.6	0.3	5.72	0.11	0.67	0.02
2G18	2.02	0.03	11.3	0.2	5.6	0.11	0.71	0.01
3G18	2.04	0.03	11.9	0.1	5.86	0.12	0.72	0.02
G24	2.04	0.03	11.5	0.2	5.68	0.11	0.60	0.01

Table 2.2 The ¹⁵N longitudinal and transverse relaxation rates R₁ and R₂ and heteronuclear NOEs for labeled residues in T3-785 at 20°C at 500 MHz.

Residue	R₁	Error	R₂	Error	R₂/R₁	Error	NOE	Error
1G15	1.64	0.01	13.3	0.5	8.09	0.16	0.70	0.01
2G15	1.66	0.01	12.9	0.2	7.76	0.16	0.68	0.01
3G15	1.66	0.01	12.9	0.7	7.79	0.16	0.71	0.02
1L16	1.53	0.01	12.4	0.2	8.13	0.16	0.72	0.01
2L16	1.55	0.01	12.2	0.2	7.89	0.16	0.74	0.02
3L16	1.54	0.01	12.3	0.3	8	0.16	0.73	0.01
1A17	1.50	0.01	12.5	0.4	8.35	0.17	0.70	0.00
2A17	1.49	0.01	12.6	0.3	8.48	0.17	0.72	0.01
3A17	1.48	0.01	13.3	0.2	8.96	0.18	0.72	0.01
1G18	1.66	0.01	12.3	0.3	7.4	0.15	0.75	0.02
2G18	1.65	0.01	12.5	0.1	7.58	0.15	0.72	0.03
3G18	1.66	0.01	12.9	0.2	7.74	0.16	0.76	0.01
G24	1.68	0.02	12.2	0.2	7.24	0.15	0.67	0.01

Table 2.3 The ¹⁵N longitudinal and transverse relaxation rates R₁ and R₂ and heteronuclear NOEs for labeled residues in T3-785 at 20°C at 600 MHz.

Residue	R ₁	Error	R ₂	Error	R ₂ /R ₁	Error	NOE	Error
1G15	1.32	0.01	14.6	0.1	11.03	0.11	0.74	0.03
2G15	1.30	0.01	15.2	0.1	11.64	0.12	0.75	0.02
3G15	1.30	0.01	15.1	0.0	11.66	0.12	0.73	0.03
1L16	1.17	0.01	14.3	0.1	12.15	0.12	0.76	0.03
2L16	1.18	0.01	15.0	0.2	12.71	0.13	0.78	0.03
3L16	1.17	0.01	14.6	0.2	12.55	0.13	0.80	0.03
1A17	1.16	0.01	15.2	0.1	13.07	0.13	0.78	0.03
2A17	1.15	0.01	14.9	0.1	12.91	0.13	0.74	0.03
3A17	1.15	0.01	15.3	0.2	13.33	0.13	0.76	0.03
1G18	1.28	0.01	15.5	0.1	12.17	0.12	0.84	0.03
2G18	1.28	0.02	15.3	0.1	11.92	0.12	0.79	0.03
3G18	1.30	0.02	15.3	0.1	11.77	0.12	0.81	0.02
G24	1.33	0.04	15.4	0.0	11.61	0.12	0.75	0.01

Table 2.4 The ¹⁵N longitudinal and transverse relaxation rates R₁ and R₂ and heteronuclear NOEs for labeled residues in T3-785 at 20°C at 800 MHz.

The derivation of the diffusion tensor from R_2/R_1 requires using residues for which there is the absence of conformational exchange and large amplitude internal motion. All labeled residues showed no evidence of conformational exchange as determined from $R_2^{\text{Hahn-echo}}$ experiments and no evidence of large amplitude motions as seen from NOE values that are uniformly greater than 0.6 (Table 2.2, 2.3, 2.4). All the labeled residues could in principle be included in the derivation of the diffusion tensor, but it was not possible to use the relaxation data for all four residues and obtain meaningful results at any magnetic field 500MHz, 600MHz or 800MHz. For example, when the relaxation data for four labeled residues Gly15-Leu16-Ala17-Gly18 or for only two glycine residues Gly15 and Gly18 at 500 MHz was used, the anisotropic model showed no improvement in the fit relative to the isotropic model, as indicated by low F-values (1.8 and 2.5, respectively) and by the fact that the unique axis of the diffusion tensor was not parallel to the long symmetric axis of the peptide (Figure 2.8A). However, when using the six data points for Leu16-Ala17 only, a high F-value of 34.1 indicated that the anisotropic model was significantly better than the isotropic model and the unique axis of the diffusion tensor was aligned properly with the long axis of the peptide (Figure 2.8A). The ratio of the principal values of the diffusion tensor D_{\parallel}/D_{\perp} is 13.1, comparable to the calculated value of 11.9. The overall correlation time τ_c is 6.92ns. If the peptide is approximated from the x-ray structure as a cylinder with length ~ 93 Å and width ~ 15 Å, the calculated D_{\parallel}/D_{\perp} and τ_c are 12.3 and 6.98ns (Fan et al. 1993; Kramer et al. 1999). Consequently, the diffusion tensor

obtained by fitting the relaxation data of Leu16-Ala17 is consistent with theoretical calculations.

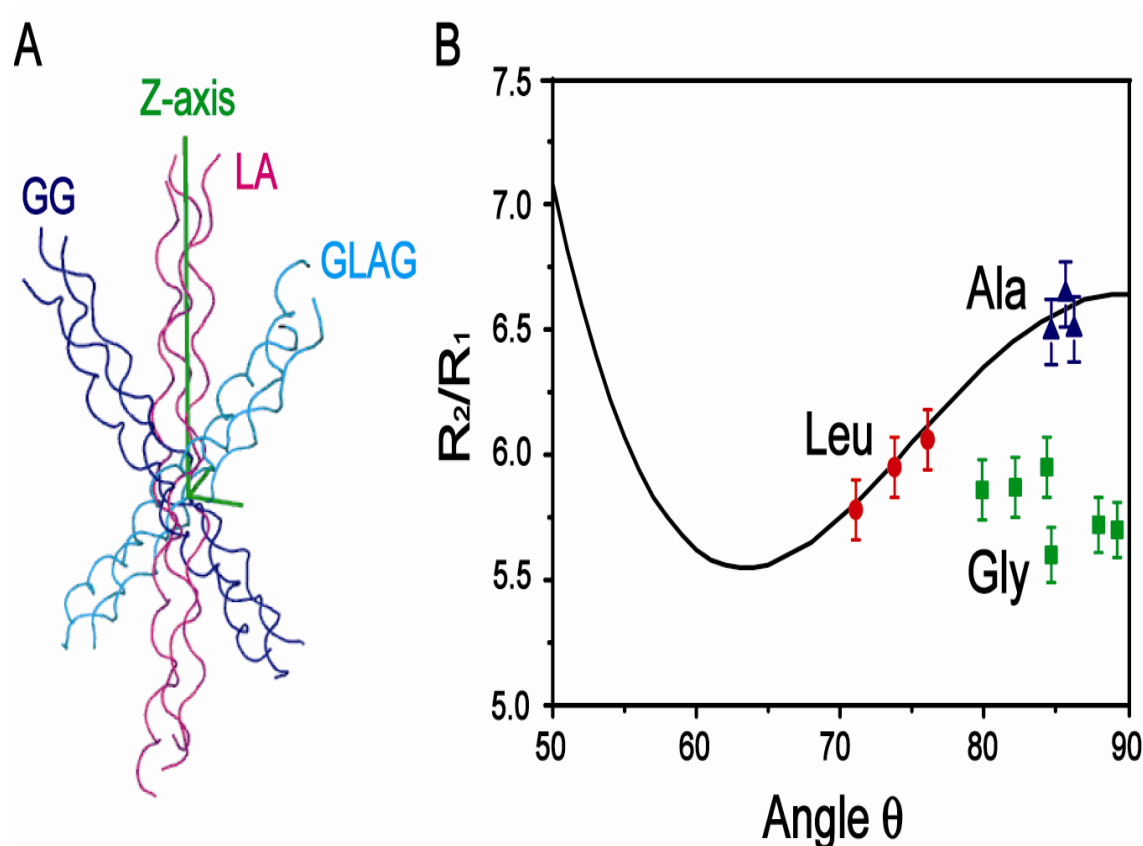


Figure 2.8 (A) The relative position of triple helical peptide in the diffusion tensor frame obtained by fitting the relaxation data of Gly15 and Gly18 only, Leu16 and Ala17 only, or all the labeled residues GLAG. (B) Plots of R_2/R_1 vs the θ angle. The line shows the back-calculated R_2/R_1 based on the diffusion tensor and the colored dots (Gly in green, Leu in red, Ala in blue) show the experimental values; The figure assumes that $\theta' = \theta$.

2.3.6.2 Deviation of experimental and theoretical ^{15}N relaxation data

R_2/R_1 values were back calculated for all the labeled residues Gly15-Leu16-Ala17-Gly18 given the diffusion tensor obtained from fitting of Leu16-Ala17 data (Figure 2.8B). The back-calculated R_2/R_1 values for Gly15 and Gly18 deviate from the experimental values and are too large by approximately 12%. Possible reasons for these deviations could result either from the non-collinearity of the principal axis of the shielding tensor with the NH bond or from the uncertainty in the positions of the H-atoms relative to the backbone (Boyd & Redfield 1998). Studies have indicated that the principal axis of the ^{15}N shielding tensor can be inclined away from the NH bond by an angle $\alpha \sim 20^\circ$ (Boyd & Redfield 1998). The dipolar terms of the spectral density function depend on the θ angle between the NH vector and the principal axis of the diffusion tensor, while the CSA terms of the spectral density function depend on the angle θ' between the principal axis of the ^{15}N shielding tensor and the principal axis of the diffusion tensor. The θ angles for Gly15 and Gly18 are about 85° and therefore the θ' angle for the two Gly could be as small as 65° . Re-calculating the R_2/R_1 values of Gly15 and Gly18 assuming an angle of $\theta=85^\circ$ and $\theta'=65^\circ$ results in a value of 6.38, still far from the experimental R_2/R_1 value of ~ 5.78 , suggesting that non-collinearity of the principal axis of the shielding tensor with the NH bond does not account for the inability to fit the Gly residues. As the CSA term contributes only $\sim 18\%$ to the R_1 and R_2 values, the adjusted θ' value would not change the values R_2/R_1 very significantly.

The divergence of the experimental data from the theoretical data can be improved by altering the positions of the backbone amide protons by adjusting the angle θ to $\sim 71^\circ$ (Figure 2.8B). The X-ray structure does not contain information about protons, as their electron density is too weak to permit accurate determination. The amide protons are added to the X-ray structure with the program REDUCE with the assumption that the amide proton is positioned in the C'-N-C α plane, while the NH vector is a little inclined away from the line that bisects the C'-N-C α angle (~ 4 degrees) toward the N-C α bond (Word et al. 1999). In order to make the calculated R_2/R_1 values consistent with the experimental values, the angle θ for the Gly15 and Gly18 is adjusted by placing the amide proton out of the C'-N-C α plane by 17.0 ± 5.3 degrees. There is precedence for this as a 1\AA structure of BPTI obtained by neutron diffraction and X-ray data have shown that some amide protons are out-of-plane by $0.4 \pm 4.8^\circ$ (Wlodawer et al. 1984). More recently, RDC studies of the third IgG-binding domain of protein G (GB3) indicated that most NH vectors were located in the C'-N-C α plane, while 38 of them were out of the plane with a maximum out-of-plane angle of 11.5 for K13 (Ulmer et al. 2003).

2.3.6.3 Hydrogen bonding leads to reorientation of Gly amide protons

It has been suggested that deviations of the amide proton from its standard position may arise from secondary structure differences or from H-bonding (Ulmer et al. 2003). In the triple helical conformation, Gly is closely packed in the center of the triple helix and the amide proton of Gly forms H-bonds with the carbonyl of the

X-residue in the adjacent chain. The amide protons of the X and Y residues are pointing outward and are not involved in direct H-bonding with adjacent chains. After modification of the positions of the glycine amide protons according to the new theta angle, the hydrogen bond angles were re-calculated. The angles have improved relative to the ones without modification; on average, the N-H-O angle is closer to 180° and the H-O distances are shorter (Table 2.6). No clashes were caused by the repositioning of the amide protons as checked by the program MolProbity (Davis et al. 2007). With the modified theta angle for the Gly NH, the diffusion tensor can be obtained by using 500, 600 or 800MHz NMR relaxation data simultaneously for all the labeled residues without the need to consider the non-collinearity of the CSA.

We suggest that in the triple helix the H-bonding rather than uniform PPII secondary structure may be responsible for the deviation of the Gly amide protons from their standard position. This is supported by the fact that only the H-bonded Gly residues require a modification of their theta angles while the X and Y residues remain in their standard positions. Additionally, the re-positioning of the Gly NH vectors results in improved H-bond characteristics. These long-range orientational restraints obtained from the ^{15}N relaxation data can be used to complement the short range NOEs and J-coupling values in order to provide additional restraints for structure determination of triple helical model peptides of collagen.

H-bond	Before modification		After modification	
	Angle N-H-O	Dist H-O	Angle N-H-O	Dist H-O
1G15 NH -2A13 C=O	158	1.83	168	1.8
1G18 NH -2L16 C=O	155	1.89	175	1.84
2G15 NH -3A13 C=O	145	2.05	175	1.93
2G18 NH -3L16 C=O	154	2.01	161	1.99
3G15 NH -1L16 C=O	163	1.91	169	1.9
3G18 NH -1P19 C=O	146	2.15	170	2.05

Table 2.5 Hydrogen bond information of the labeled Glycines before and after the modification of the amide protons.

2.3.7 Dynamics of T3-785 at fast picosecond-nanosecond timescale

Dynamic parameters were derived from the R_1 , R_2 and NOE data at multiple fields using Model-free formalism (Figure 2.9) (Mandel et al. 1995). In the formalism, there are two parameters τ_c and D_{\parallel}/D_{\perp} to characterize the overall motion for the whole molecule and three parameters S^2 , τ_e and R_{ex} for internal motion. Specifically, four models, which contain some of S^2 , τ_e and R_{ex} , are built to characterize the internal motion: (1) S^2 , (2) S^2 , τ_e , (3), S^2 , R_{ex} , (4) S^2 , τ_e , R_{ex} . Model 1 is simplest assuming internal motion is very fast ($\tau_e \rightarrow 0$) and no chemical exchange contribution ($R_{ex}=0$); Model 4 is most complicated, in which both τ_e and R_{ex} cannot be negligible. In model-free analysis, τ_c and D_{\parallel}/D_{\perp} are first estimated from R2R1_diffusion program as shown in section 2.3.6.1. Then with fixed values of τ_c and D_{\parallel}/D_{\perp} , the best model for a residue is selected by fitting the experimental

data to the different models separately. After model selection is completed, all parameters (τ_c , D_{\parallel}/D_{\perp} , S^2 , τ_e , R_{ex}) are optimized simultaneously.

The first model, which includes only generalized order parameter, was found to be the best model for all labeled residues, indicating really fast internal motion and no conformational exchange. The generalized order parameter S^2 was obtained for labeled residues: Gly15 and Gly18 had the largest values of ~ 0.9 , while Ala17 had the smallest values of ~ 0.73 (Figure 2.9). S^2 values in theory can vary from 0 to 1, while 0 means absolutely free motion and 1 indicates totally restricted motion. All the residues here had high S^2 values, indicating that all the residues had no significant internal motion and Gly15/Gly18 were most rigid.

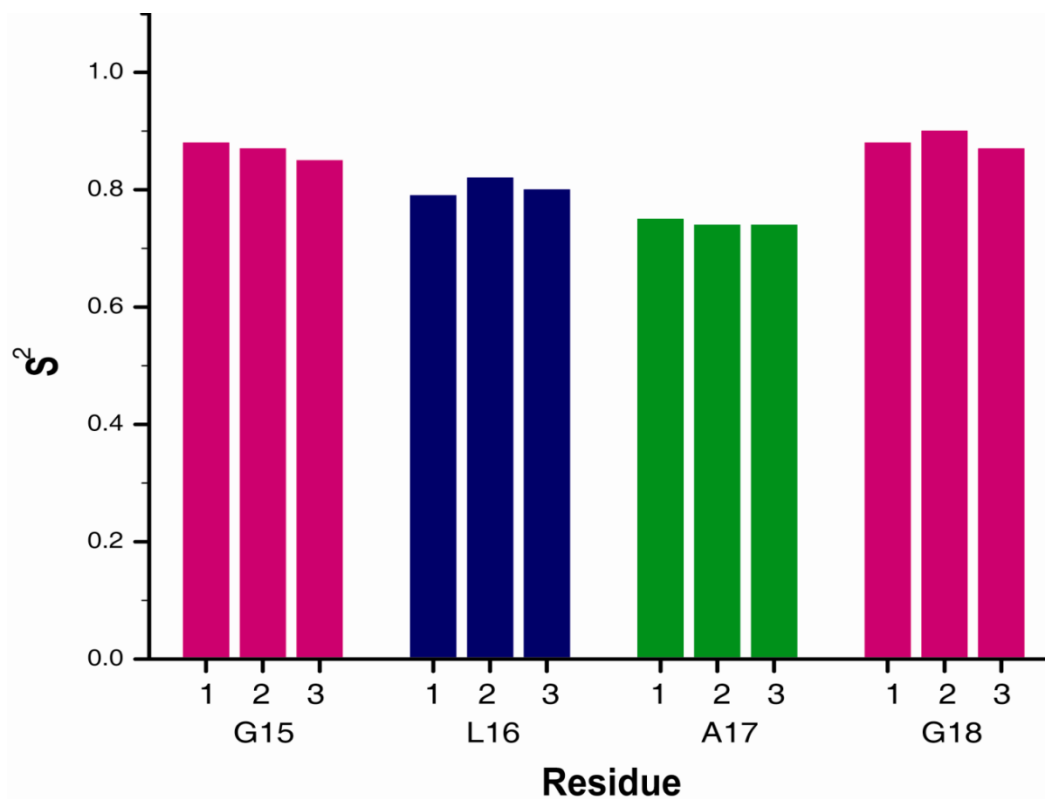


Figure 2.9 S^2 of labeled residues in T3-785.

2.4 Conclusions

Here, a strategy combining molecular modeling and NMR spectroscopy have been developed to obtain model structure of triple helical peptides. Residual dipolar coupling, a powerful tool to obtain long-range order restraints, has been shown to be sensitive to structural variations in triple helix for the first time. Particularly, we demonstrate that the dependence of ^{15}N relaxation measurements on rotational diffusion anisotropy can be used to provide detailed information about the orientation of the N-H bonds in the very anisotropic triple helical peptide model of collagen. The analysis of the divergence of experimental R_2/R_1 values from the theoretical values shows that the Gly amide protons in the triple helical peptide

need to be repositioned relative to the protein backbone in order to obtain experimentally consistent theta angles. After placing the Gly amide protons out of the C'-N-C α plane, the hydrogen bond angles are re-calculated and shown to be both closer to 180° and shorter. The orientational restraints obtained from the ^{15}N relaxation data enable us to understand the details of the H-bonding of the Gly residues, which is crucial in defining collagen structure and recognition.

Chapter 3

NMR studies on collagen peptides modeling

Gly mutations involved in Osteogenesis

Imperfecta

3.1 Introduction

Collagen, the most abundant protein in mammals, provides mechanical strength and structural integrity to the human body, supporting and protecting the softer tissues and internal organs (Brodsky & Persikov 2005; Shoulders & Raines 2009). Type I collagen, representing over 90% of all collagens, is the major structural component of skin, bone, tendon and ligament (Di Lullo et al. 2002). As a heterotrimer of two $\alpha 1$ and one $\alpha 2$ chains, Type I collagen contains a characteristic triple helix motif consisting of three extended polyproline II-like chains supercoiled about a common axis with one-residue staggering. The close packing of the three chains can only accommodate Gly as every third residue, generating the repetitive (Gly-X-Y)_n sequence pattern. The small size of Gly allows it to be deeply buried near the center of the triple helix, forming interchain Gly N-H...C=O (X) hydrogen bonds, which are crucial in defining collagen structure (Ramachandran & Kartha 1955; Rich & Crick 1955). The X and Y can be almost any amino acid, but they are frequently Pro and Hydroxyproline (Hyp or O), making Gly-Pro-Hyp the most common and most stabilizing triplet in collagen (Persikov et al. 2000).

Type I collagen maintains a perfect Gly-X-Y repeating pattern throughout its 1014 residue long triple helix domain (van der Rest et al. 1991; Kielty & Grant 2002). Even a single Gly mutation results in Osteogenesis imperfecta (OI), or brittle bone disease, which is distinguished by fragile bones and easy susceptibility to fracture (Myllyharju & Kivirikko 2001; Byers & Cole 2002; Marini et al. 2007).

More than 600 cases of OI were observed resulting from missense mutations that replace a Gly by another bulkier amino acid in Type I collagen (Marini et al. 2007). The phenotypes of the disease vary in a wide range from mild cases with multiple fractures to perinatal lethal cases (Byers & Cole 2002). Considerable efforts have been made to understand the mechanism of how different Gly substitutions correlate to different severities (Baum & Brodsky 1999; Radmer & Klein 2004; Bodian et al. 2008; Li et al. 2009).

One determining factor of the lethality of OI is the identity of amino acid that substitutes Gly (Kuivaniemi et al. 1997; Byers & Cole 2002; Marini et al. 2007). Among all the seven bulkier residues (Ser, Ala, Cys, Val, Arg, Asp, Glu) replacing Gly in the $\alpha 1$ chain of type I collagen, it is noted that the frequency of lethal outcomes is A < S, C < R, E, D, V, where the Ala substitution is least likely to cause lethal phenotype. Model peptide studies on a host-guest (Gly-Pro-Hyp)₈ peptide system suggest that the order of destabilization of triple helix for different Gly substitutions is A, S < C < R < V < E, D, which shows an obvious correlation with the lethality of OI mutations (Beck et al. 2000; Persikov et al. 2004). However, recent studies on 41 different Gly substitutions in native collagen from 47 OI patients didn't demonstrate any correlation between T_m and the residue replacing Gly, indicating that the effect of the mutation in collagen could be very local (Makareeva et al. 2008).

Another important factor determining the degree of OI severity is the nearest neighboring sequence surrounding the Gly mutation (Yang et al. 1997; Radmer & Klein 2004). It has been suggested that rigidity or stability of the local sequences correlates with severity, with mutations in more stabilized triplets leading to more serious disease (Radmer & Klein 2004). Statistical and computational studies have shown strong correlations between the presence of a nearby Pro and lethal OI, as well as the presence of a nearby Ala and nonlethal OI (Radmer & Klein 2004). The observation that a Pro near an OI mutation induces more structural disruption than an Ala supports the hypothesis that the flexibility of the nearby residues around Gly mutation would be important in permitting the re-formation of the triple helix (Radmer & Klein 2004). In contrast, recent studies indicate that a lower-stability triplet C-terminal to a Gly to Ser/Cys mutation is more inclined to cause more severe phenotype (Bodian et al. 2008). Other factors, including the location of the mutation on the chain and the proximity to interaction sites or salt bridges, are also proposed to be related to OI phenotypes (Bodian et al. 2008; Xu et al. 2008). However, the molecular mechanism of how those factors influence various clinical phenotypes of OI remains poorly understood.

Here, NMR spectroscopy is used to investigate the structural consequences of the identity of residues replacing Gly and the local sequence environment in triple helical peptides modeling Gly mutations involved in OI, in order to elucidate the structural bases of various OI phenotypes. The disruptive effects of Gly to Ala mutation on triple helical conformation have been examined, indicating a

dependence on N-terminal sequences (Bella et al. 1994; Hyde et al. 2006). The X-ray crystal structure of the Gly→Ala peptide shows that Gly to Ala mutations can be tolerated in a highly stabilized (POG)₁₀ sequence environment with localized perturbation (Bella et al. 1994), while studies on a peptide modeling a Gly to Ala substitution at site 901 in the $\alpha 1$ chain of type I collagen in a physiological sequence context indicate that it could not fold through the mutation site (Hyde et al. 2006). However, replacing the N-terminal sequence with GPO(GAO)₃ make the peptide fully folded (Hyde et al. 2006).

The ability to construct a peptide which successfully folds through an Ala mutation is very encouraging, as it best mimics the real mutation in native collagen. Here, we extend our studies to the substitution of Gly by larger residues Ser, Arg and Asp. The equilibrium states of a series of peptide models have been characterized to determine how the different substituting residues and sequence environment impact on the ability to form a standard triple helix. We demonstrate that bulkier residues are more disruptive and multiple strategies including strong renucleation and local flexibility need to be developed to incorporate them into triple helix. This project was performed in collaboration with Dr. Madhan Balaraman from Prof. Barbara Brodsky's lab, who did CD measurements, and Dr. Yingjie Li, who did early NMR studies.

3.2 Materials and Methods

3.2.1 Sample preparation

The T1-898 peptide sets were synthesized by the Tufts University Core Facility (Boston, MA) with ^{15}N labeled amino acids at selective positions for NMR characterization (Table 3.1). Peptide T1-898 was selectively ^{15}N -labeled at positions G7, A18 and G28, and peptide T1-898[G16A] was labeled at positions G7, A16 and G28. Peptides T1-898[G16S] and GPOT1-898[G16S] were labeled at positions (G7, V15, A18 and G28) and (G7, V15, A18 and G22), respectively. Peptides T1-898[G16R], GPOT1-898[G16R] and GAAT1-898[G16R] were labeled at positions (G7, G13 and G28), (G7, V15, A18 and G22), and (G7, A15 and G22), respectively (Table 3.1). Samples for all peptides were prepared in 10% D_2O /90% H_2O at pH ~2.5 with concentrations of ~6 mM.

3.2.2 NMR spectroscopy

NMR experiments were performed on a Varian Inova 500MHz spectrometer or a Varian Inova 600MHz spectrometer equipped with a cryoprobe. ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) (Kay et al. 1992) was performed at 0°C (except 5°C for peptide GPOT1-898[G16S]). 3D ^{15}N edited NOESY-HSQC experiments (Fesik & Zuiderweg 1988; Marion et al. 1989; Messerle et al. 1989) were performed to assign the labeled residues and the R_2 measurements (Farrow et al. 1994) were performed to distinguish trimer resonances from monomer resonances. All data were processed using the Felix 2004 software package (MSI,

San Diego, CA) or NMRPipe (Delaglio et al. 1995), and then were analyzed by Felix or NMRView (Johnson & Blevins 1994).

3.2.3 NMR diffusion experiments

Convection compensated LED followed by HSQC (CCLED-HSQC) experiments (Buevich & Baum 2002; Li et al. 2005) were used to measure the residue-specific diffusion coefficients of the selectively labeled peptides at 0°C (except 5°C for peptide GPOT1-898[G16S]) and 40°C (Table 3.1). Peptide GPOT1-898[G16S] has weak monomer resonances at low temperature and satisfied data could be achieved only at 5°C. Temperatures were calibrated with methanol at low temperature and with ethylene at high temperature. The gradient strength was calibrated on a standard doped 1% H₂O in 99% D₂O sample using the value of $1.90 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ for the diffusion coefficient of HDO at 25°C (Mills 1973; Holz & Weingaertner 1991).

The translational diffusion coefficients (D) were derived from non-linear regression of resonance intensity (I_t) to pulsed field gradient strength (G_t) as defined by the following equation: $I_t = I_0 \cdot \exp[-2D\gamma^2 G_t^2 \delta^2 (\Delta + 2\delta/3 + 3\tau/4)]$, where γ is ¹H gyromagnetic ratio, and Δ , δ , and τ are 280ms, 2ms and 0.4ms, respectively (Li et al. 2005). The diffusion coefficients at temperature T (0°C or 5°C) were then normalized to values at 40°C by using the Stokes-Einstein equation, $D_{\text{normalized}, T} = D_T \cdot 313.15 \cdot \eta_T / (T \cdot \eta_{40^\circ\text{C}})$, where D_T is the directly measured diffusion coefficient at T, η_T is the viscosity calculated for the solution of 90% H₂O and 10%

D2O at a specific temperature (Cho et al. 1999) (Table 3.2). Errors in the diffusion measurements were fitting errors from the variance-covariance matrix. Then a ratio of $D_{\text{normalized},T}$ over $D_{40^\circ\text{C}}$ was calculated as the normalized diffusion coefficient for each residue, denoted as $D_{M,T}$ ($D_{M0^\circ\text{C}}$ for all peptides except $D_{M5^\circ\text{C}}$ for GPOT1-898[G16S]). The error of $D_{M,T}$ was calculated as follows:

$$(D_{M,T})_{\text{error}} = \left(\frac{D_{\text{normalized},T}}{D_{40^\circ\text{C}}} \right)_{\text{error}} = \left(\frac{D_{\text{normalized},T}}{D_{40^\circ\text{C}}} \right) * \sqrt{\left(\frac{D_{\text{normalized},T,\text{error}}}{D_{\text{normalized},T}} \right)^2 + \left(\frac{D_{40^\circ\text{C},\text{error}}}{D_{40^\circ\text{C}}} \right)^2} \quad (1)$$

The diffusion coefficients focus on species with the same monomer chemical shifts, which represent only pure monomers at 40°C (a temperature higher than the melting temperature) but could represent a mixture of pure monomers and partially folded species at a lower temperature 0°C or 5°C . Then $D_{\text{normalized},0^\circ\text{C}}$ or $D_{\text{normalized},5^\circ\text{C}}$ is less than or equal to $D_{40^\circ\text{C}}$, since partially folded species has a smaller diffusion coefficient than pure monomer due to its larger size. As a result, the value for normalized diffusion coefficients $D_{M0^\circ\text{C}}$ or $D_{M5^\circ\text{C}}$ is at most equal to 1, with a value of 1 indicating the non-existence of partially folded species.

The equation defining resonance intensity (I_t) to gradient strength (G_t) were modified as follows:

$$I_t = P_m * \exp[-2D_m \gamma^2 G_t^2 \delta^2 (\Delta + 2\delta/3 + 3\tau/4)] + P_{pf} * \exp[-2D_{pf} \gamma^2 G_t^2 \delta^2 (\Delta + 2\delta/3 + 3\tau/4)],$$

assuming relative slow exchange rate between partially folded species and pure monomer on the diffusion time scale (Li et al. 2005). $D_{\text{normalized},0^\circ\text{C}}$ of the monomer and trimer resonances of G28 in T1-898[G16A] (or $D_{\text{normalized},5^\circ\text{C}}$ for GPOT1-898[G16S]) were approximated as D_m and D_{pf} (standard diffusion

coefficients for monomer and partially folded species), respectively. The ratio of P_{pf} and P_m represented relative populations of partially folded species to pure monomer for each labeled residue, which were derived from two-parameter fitting of resonance intensity (I_t) to gradient strength (G_t) using the above equation.

Peptide name	Peptide sequence ^a	T _m (°C) ^b
T1-898	Ac-GPO-GAO- <u>G</u> AO-GAO- <u>G</u> PV- G PA-GAR- GPO-GPO- <u>G</u> PQ-GPO-GY-CONH ₂	35 ^c
Gly→Ala T1-898[G16A]	Ac-GPO-GAO- <u>G</u> AO-GAO-GPV- A PA-GAR- GPO-GPO- <u>G</u> PQ-GPO-GY-CONH ₂	15.5 ^c
Gly→Ser T1-898[G16S]	Ac-GPO-GAO- <u>G</u> AO-GAO-GPV- S PA-GAR- GPO-GPO- <u>G</u> PQ-GPO-GY-CONH ₂	14.5 ^d
GPOT1-898[G16S]	Ac-GPO-GPO- <u>G</u> PQ-GPO-GPV- S PA-GAR- GPO-GPO- <u>G</u> PQ-GPO-GY-CONH ₂	28 ^e
Gly→Arg		
T1-898[G16R]	Ac-GPO-GAO- <u>G</u> AO-GAO- <u>G</u> PV- R PA-GAR- GPO-GPO- <u>G</u> PQ-GPO-GY-CONH ₂	7 ^d
GPOT1-898[G16R]	Ac-GPO-GPO- <u>G</u> PQ-GPO-GPV- R PA-GAR- <u>G</u> PQ-GPO-GPO-GPO-GY-CONH ₂	5 ^e
GAAT1-898[G16R]	Ac-GPO-GPO- <u>G</u> PQ-GAO-GAA- R PA-GAR- <u>G</u> PQ-GPO-GPO-GPO-GY-CONH ₂	13 ^e
Gly→Asp		
T1-898[G16D]	Ac-GPO-GAO- <u>G</u> AO-GAO- <u>G</u> PV- D PA-GAR- GPO-GPO- <u>G</u> PQ-GPO-GY-CONH ₂	<5 ^d

Table 3.1 Sequences and thermal stability of model peptides. a. ¹⁵N labeled residues are underlined and bolded. Glycine substitutions are boxed; b. T_m determined from CD melt experiments; c. Hyde et al. 2006; d. Personal communications with Michael Bryan; e. Personal communications with Madhan Balaraman.

Peptide	Residue	$D_{0^{\circ}\text{C}}$	Error	$D_{\text{normalized}, 0^{\circ}\text{C}}$	Error	$D_{40^{\circ}\text{C}}$	Error
T1-898	G7	6.12	0.78	19.52	2.49	20.79	0.22
	A18	7.14	0.66	22.77	2.10	20.04	0.42
	G28	6.51	0.13	20.76	0.41	21.73	0.63
T1-898[G16A]	G7	5.69	0.34	18.14	1.08	20.54	0.23
	A16	4.79	0.31	15.27	0.99	20.12	0.16
	G28	6.35	0.38	20.25	1.21	20.57	0.09
T1-898[G16S]	G7	4.48	0.35	14.29	1.12	18.54	0.20
	V15	4.77	0.19	15.21	0.61	20.33	0.19
	A18	4.54	0.18	14.48	0.57	18.25	0.27
	G28	6.12	0.24	19.52	0.77	19.51	0.20
GPOT1-898[G16S]	G7	6.35	0.64	16.74	1.67	16.60	0.21
	V15	6.11	0.31	16.10	0.81	16.40	0.17
	A18	6.58	0.34	17.35	0.90	15.60	0.18
	G22	5.78	0.32	15.22	0.84	15.90	0.12
T1-898[G16R]	G7	3.54	0.08	11.29	0.26	19.02	0.20
	G13	3.70	0.03	11.80	0.10	19.56	0.08
	G28	3.70	0.03	11.80	0.10	19.56	0.08
T1-898[G16D]	G7	3.31	0.06	10.55	0.19	20.91	0.17
	G13	3.71	0.08	11.83	0.26	21.25	0.08
	G28	3.71	0.08	11.83	0.26	21.25	0.08

Table 3.2 Monomer diffusion coefficients of model peptides. Diffusion coefficients at 0°C ($D_{0^{\circ}\text{C}}$), diffusion coefficients at 0°C normalized to values at 40°C ($D_{\text{normalized}, 0^{\circ}\text{C}}$), and diffusion coefficients at 40°C ($D_{40^{\circ}\text{C}}$) are shown in the units of $10^{-7}\text{m}^2\text{s}^{-1}$. The experiments are performed at 500MHz NMR for all peptides except 600MHz for GPOT1-898[G16S]. $D_{0^{\circ}\text{C}}$ for GPOT1-898[G16S] is obtained at 5°C instead of 0°C , as the peptide has weak monomer peaks at low temperature and satisfied data could be achieved only at 5°C . Monomer resonances of G13 and G28 in peptides T1-898[G16R] and T1-898[G16D] are overlapped.

3.3 Results

3.3.1 Design of model peptides

Studies on peptides modeling a Gly to Ala substitution at site 901 in the $\alpha 1$ chain of type I collagen indicated that replacing the N-terminal physiological sequence GDRGETGPAGPA with a stronger renucleation sequence GPO(GAO)₃ makes the originally partially folded peptide fully folded (Hyde et al. 2006). Here, we utilize the NMR diffusion experiments to further explore the structural consequences of the Gly→Ala substitution (denoted as T1-898[G16A]) and extend our studies to the substitution of Gly by bulkier residues Ser, Arg and Asp (denoted as T1-898[G16S], T1-898[G16R] and T1-898[G16D]) (Table 3.1). An even stronger renucleating sequence (GPO)₄ was used to replace GPO(GAO)₃ for Gly→Ser and Gly→Arg substitutions to attempt to push the partially folded species to form a fully folded triple helix (denoted as GPOT1-898[G16S], GPOT1-898[G16R]) (Table 3.1). For Gly→Arg substitutions, a further replacement of the relatively rigid triplet GPV immediately N-terminal to the Gly→Arg mutation site by a more flexible triplet GAA in GPOT1-898[G16R] was made to give more room to accommodate the very bulky Arg into triple helix (denoted as GPOT1-898[G16R]) (Table 3.1).

3.3.2 NMR characterization of T1-898, T1-898[G16A], and T1-898[G16S]

The OI phenotype resulting from a Gly mutation in type I collagen has been shown to correlate with the identity of the amino acid replacing Gly. Here NMR studies were performed on the T1-898 peptide set (Table 3.1) to gain insights into

the conformational changes that arise from the different substituting residues (Ala, Ser, Arg and Asp). As previously reported, the HSQC spectrum of the peptide T1-898 contains the typical features for a triple helical conformation with all ^{15}N labeled residues (G7, A18, and G28) showing one or more trimer resonances (Figure 3.1A) (Hyde et al. 2006). The distinct monomer and trimer resonances are observed for all residues supporting a monomer:trimer equilibrium in the solution. Residue specific translational diffusion measurements were performed on the monomer peaks of T1-898 peptide to characterize equilibrium species, following the methodology developed by Li et al (Li et al. 2005) (Table 3.2). For easy comparison, the monomer diffusion coefficients at 0°C are normalized to those at 40°C and denoted as $D_{M0^\circ\text{C}}$. At 0°C , monomers (more specifically, species with monomer chemical shifts), could exist in the states of pure unfolded monomer and partially folded trimer, while the monomers represent only pure monomers at 40°C . As partially folded species has a smaller diffusion coefficient than pure monomer due to its larger size, $D_{\text{normalized},0^\circ\text{C}}$ will be less than $D_{40^\circ\text{C}}$ if there is partially folded species at 0°C or equal to $D_{40^\circ\text{C}}$ if there is not. As a result, the value for normalized diffusion coefficients $D_{M0^\circ\text{C}}$ is at most equal to 1, with a value of 1 indicating the non-existence of partially folded species. For all labeled residues of T1-898 peptide, normalized diffusion coefficients at 0°C are close to 1, which is consistent with sampling a similar pure monomer state (Figure 3.1E). The HSQC and diffusion experiments support a simple monomer (M) to fully folded trimer (FT) equilibrium model.

As previously reported (Hyde et al. 2006), the replacement of the central Gly by Ala results in a peptide still containing the features of a fully folded triple-helix, with distinct trimer resonances for all labeled residues at positions G7, A16, and G28 in the HSQC spectrum (Fig. 3.1B). Normalized diffusion coefficients at 0°C ($D_{M0^{\circ}C}$) were measured for all the labeled residues in peptide T1-898[G16A] (Figure 3.1F). $D_{M0^{\circ}C}$ of G7 and G28 are close to 1, indicating that the two terminal residues are not in partially folded states, suggesting that the two ends of the peptide are folded. The much lower $D_{M0^{\circ}C}$ of A16 than G7 and G28 indicates a triple helical conformation with a looser middle region, namely partially folded trimer with a break in the middle (PFT-B). Hence for the peptide T1-898[G16A], along with the fully folded trimer, other two species can exist and the ratio of the species namely M : PFT-B is further quantified to be 1 : 1.2, indicating about half of the species of monomer chemical shift is truly partially folded trimer.

A peptide with a central Gly replaced by Ser (T1-898[G16S]) models a mild case of OI resulting from a Gly to Ser mutation at position 901 in the $\alpha 1$ chain of type I collagen (Hyde et al. 2006). The HSQC spectrum (Figure 3.1C) of T1-898[G16S] shows distinct trimer resonances for all labeled residues G7, V15, A18, and G28, indicating that the Gly→Ser substitutions can be incorporated in a triple helical conformation. The residue specific monomer diffusion coefficients of this peptide are shown in Figure 3.1G. Partially folded species with folded C-terminus but unfolded N-terminal and mutation site (PFT-C) can be deduced from the observation that $D_{M0^{\circ}C}$ for G7, V15 and A18 are similar and smaller than 1

while $D_{M0^{\circ}\text{C}}$ for G28 is around 1. Therefore, peptide T1-898[G16S] contains partially folded trimer state as well as monomer and trimer, and the ratio of M : PFT-C is calculated as 1 : 1.8. One argument for the difference in the conformation of the two ends is that C-terminus has a stronger nucleation sequence $(\text{GPO})_4$ than $\text{GPO}(\text{GAO})_3$ at N-terminus.

3.3.3 Effects of renucleation sequence on the incorporation of Gly to Ser substitution in triple helix

Peptide GPOT1-898[G16S] was designed with a more stabilizing $(\text{GPO})_4$ to replace $\text{GPO}(\text{GAO})_3$ at the N-terminus, in order to improve the folding of the peptide with Gly to Ser substitution (Table 3.1). As expected, the thermal stability of the peptide GPOT1-898[G16S] increased significantly to 28 °C compared to the T_m of 14.5 °C for T1-898[G16S] (Table 3.1) (Michael Bryan, personal communications). The HSQC spectrum of GPOT1-898[G16S] shows nice trimer resonances for all labeled residues at G7, V15, A18 and G22, indicating a fully folded trimer (Figure 3.1D). The monomer diffusion experiments for GPOT1-898[G16S] were carried out at 5°C and 40°C and the results are shown in Figure 3.1H. The $D_{M5^{\circ}\text{C}}$ of G7, V15, A18 and G22 are all close to 1, suggesting that the peptide has monomer in equilibrium with fully folded trimer conformation.

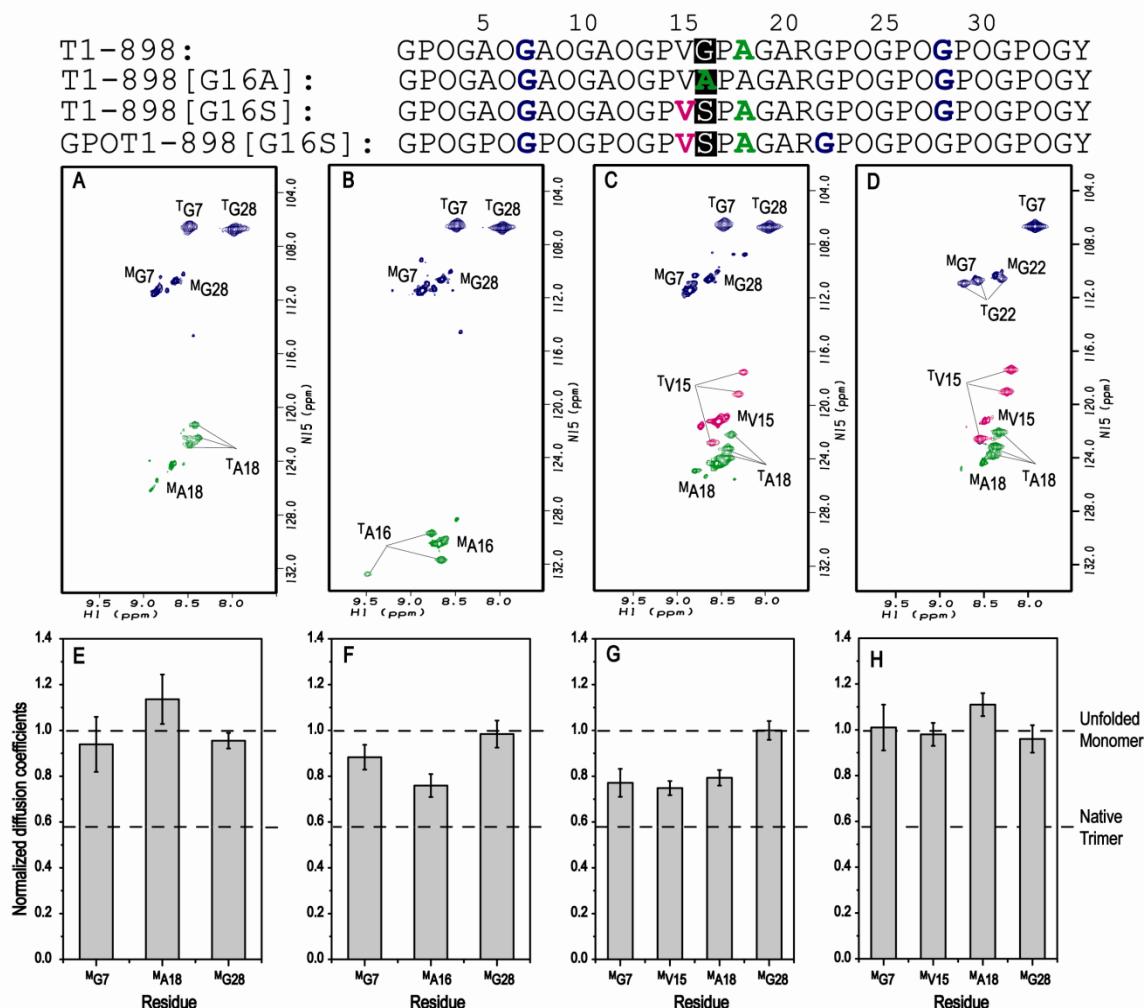


Figure 3.1 Effect of Gly to Ala/Ser substitution on equilibrium conformations of peptide T1-898. ^1H - ^{15}N HSQC spectra of T1-898 (A), T1-898[G16A] (B), T1-898[G16S] (C) and GPOT1-898[G16S] (D): the peaks corresponding to the monomer and trimer states are denoted with a superscript M or T, respectively. Histogram of the residue-specific monomer diffusion coefficients of ^{15}N -labeled residues for peptide T1-898 (E), T1-898[G16A] (F), T1-898[G16S] (G) and GPOT1-898[G16S] (H): monomer diffusion coefficients at 0°C (except 5°C for peptide GPOT1-898[G16S]) are normalized to those at 40°C. Peptide GPOT1-898[G16S] has weak monomer peaks at low temperature and satisfied data could be achieved only at 5°C. The black dashed line represent the normalized diffusion coefficients for unfolded monomer (theoretically with value of 1) and native trimer (with a value of 0.58 for G28 in peptide T1-898[G16A]), respectively.

3.3.4 NMR characterization of T1-898[G16D]

The peptide T1-898[G16D] modeling a Gly to Asp mutation shows very poor thermal stability compared to the control T1-898 and other mutant peptides (Table 3.1). Peptide T1-898[G16D] could not fold completely into a trimer since the trimer resonance of G13 is not observed in the HSQC spectrum (Figure 3.2A). The trimer resonances for G7 and G28 are present and have the same chemical shifts as in other T1-898 series peptides (Figure 3.1). It suggests that peptide T1-898[G16D] did not go through the substitution site to form a fully folded trimer, even though the two ends could be in a triple helical conformation. The peptide might be partially folded at only one end (C or N terminus) or it might be able to form a conformation with two folded ends but an unfolded or distorted middle region. The diffusion measurements on this peptide indicate predominately populations of partially folded trimers at the C-terminus (PFT-C) or at the N-terminus (PFT-N), as $D_{M0^{\circ}C}$ of all residues are very low and close to the trimer diffusion coefficients (Figure 3.2B).

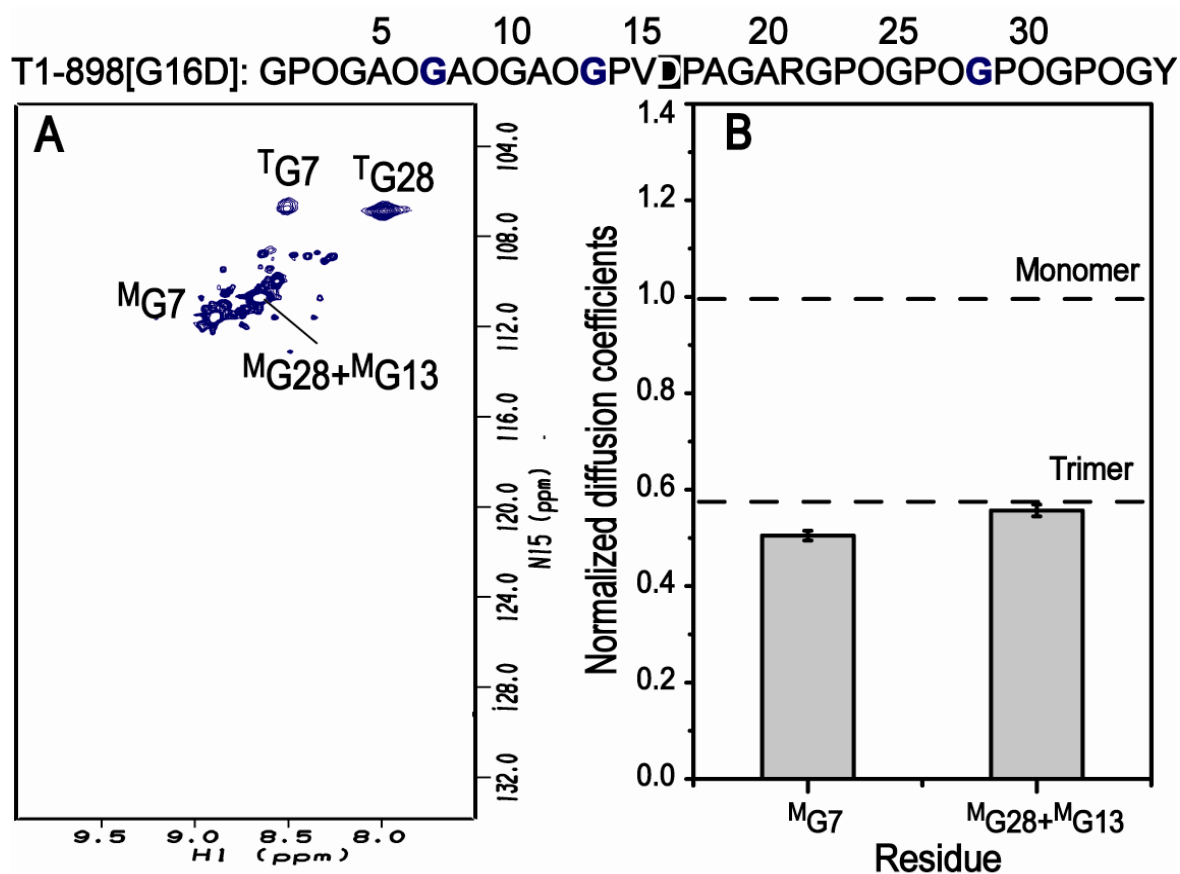


Figure 3.2 Effect of Gly to Asp substitution. ^1H - ^{15}N HSQC spectra of T1-898[G16D] (A): the peaks corresponding to the monomer and trimer states are denoted with a superscript M or T, respectively. The large number of monomer peaks arises due to cis-trans isomerization. Histogram of the residue-specific monomer diffusion coefficients of ^{15}N -labeled residues for peptide T1-898[G16D] (B): monomer diffusion coefficients at 0°C are normalized to those at 40°C . The black dashed line represent the normalized diffusion coefficients for unfolded monomer (theoretically with value of 1) and native trimer (with a value of 0.58 for G28 in peptide T1-898[G16A]), respectively.

3.3.5 Effect of renucleation sequence and nearest neighboring sequence in Gly to Arg substitution

The peptide T1-898[G16R] shows a similar HSQC spectrum as T1-898[G16D], with the absence of trimer resonance of G13, indicating that the peptide is only partially folded (Figure 3.3A). The normalized diffusion coefficients $D_{M0^{\circ}C}$ of peptide T1-898[G16R] are also similar to those of T1-898[G16D], indicating predominately populations of partially folded trimers at the C-terminus (PFT-C) or at the N-terminus (PFT-N).

To fold through the Gly to Arg mutation site, a similar design with incorporation of (GPO)₄ in GPOT1-898[G16S] was adopted for the new peptide GPOT1-898[G16R] (Table 3.1). Surprisingly the stability of GPOT1-898[G16R] was found to be lower than T1-898[G16R] (Table 3.1), indicating that the rigid GPO triplets destabilize rather than stabilize the triple helix formation for Gly to Arg mutation. The HSQC spectrum of GPOT1-898[G16R] shows a weak trimer peak for residue G7 and a trimer peak for A18, which is almost overlapped with the monomer peak (Figure 3.3B). The trimer resonances for G22 might be overlapped with the monomer resonances, as shown in GPOT1-898[G16S], which contains the same surrounding sequences for G22 as GPOT1-898[G16S] (Figure 3.1D). Residue V15 next to the Gly substitution site does not show any trimer peak, suggesting that the peptide couldn't fold through the mutation site and only partially folded species could exist.

Flexible local sequence environments might help incorporate Gly substitutions as suggested by previous studies, therefore, more flexible residues are introduced to the N-terminal side of the Gly to Arg mutation (Radmer & Klein 2004). Peptide GAAT1-898[Gly16R] was designed by modifying the triplet next to mutation site from GPV to arguably most flexible GAA, and N-terminal renucleation region from (GPO)₄ to (GPO)₃GAO. The thermal stability of the new peptide is found to increase significantly compared with the other two Gly to Arg peptides (Table 3.1). The HSQC spectrum of the peptide reveals a single strong trimer resonance for Gly7 and three well-dispersed trimer resonances for A15 (Figure 3.3C). Compared with GPOT1-898[Gly16R], the trimer resonance for Gly7 in GAAT1-898[Gly16R] becomes much stronger; particularly, the A18 near the mutation site in GPOT1-898[Gly16R] shows only a poorly-defined trimer resonance, while the A15 just beside the mutation site shows clearly three nice trimer resonances. Similar to GPOT1-898[Gly16R], the trimer resonances for G22 in GAAT1-898[Gly16R] are not observed probably because of overlapping. To solve the overlapping problem, the peptide is lyophilized and re-dissolved in 100% D₂O. As the NH of Gly in monomer exchanges much faster than those in trimer, after several hours in D₂O, the intensities of monomer resonances would be much more decreased compared with trimer resonances, giving an opportunity to observe the trimer resonances of G22. The HSQC spectrum of the peptide after ~ 20h in D₂O is taken and it shows three trimer resonances of Gly22. The evidence of trimer resonances of G7, A15 and G22 suggests that the peptide GAAT1-898[Gly16R] formed a fully folded trimer conformation. However, the

existence of partially folded trimer cannot be excluded, since the monomer diffusion measurements of G7 and G22 are interfered by overlapped trimer resonances and therefore cannot give valuable information of partially folded species.

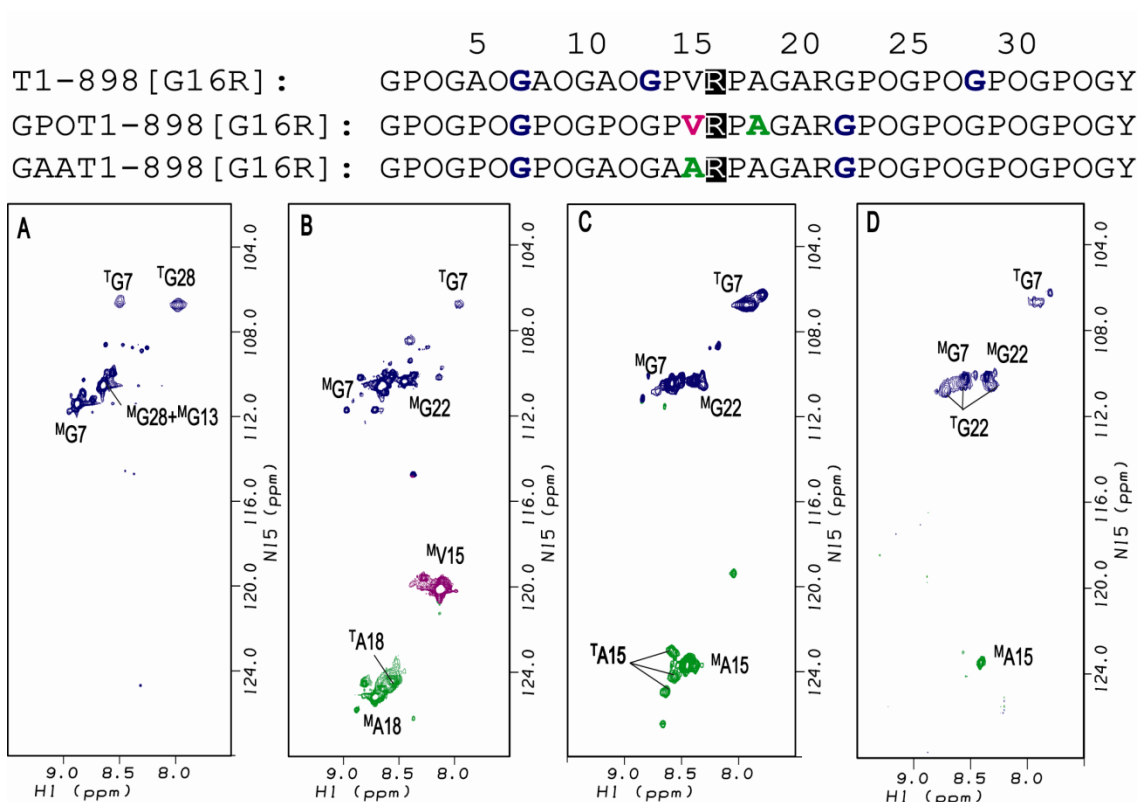


Figure 3.3 Effect of Gly to Arg substitution in different environments. ^1H - ^{15}N HSQC spectra of T1-898[G16R] (A), GPOT1-898[G16R] (B), GAAT1-898[G16R] (C) and GAAT1-898[G16R] after ~20h in 100% D_2O (D): the peaks corresponding to the monomer and trimer states are denoted with a superscript M or T, respectively. After ~20 h in D_2O , the monomer peaks of Gly in GAAT1-898[G16R] have much more decreased intensities compared with the trimer peaks of Gly, as the NH of Gly in monomer exchanges much faster than those in trimer.

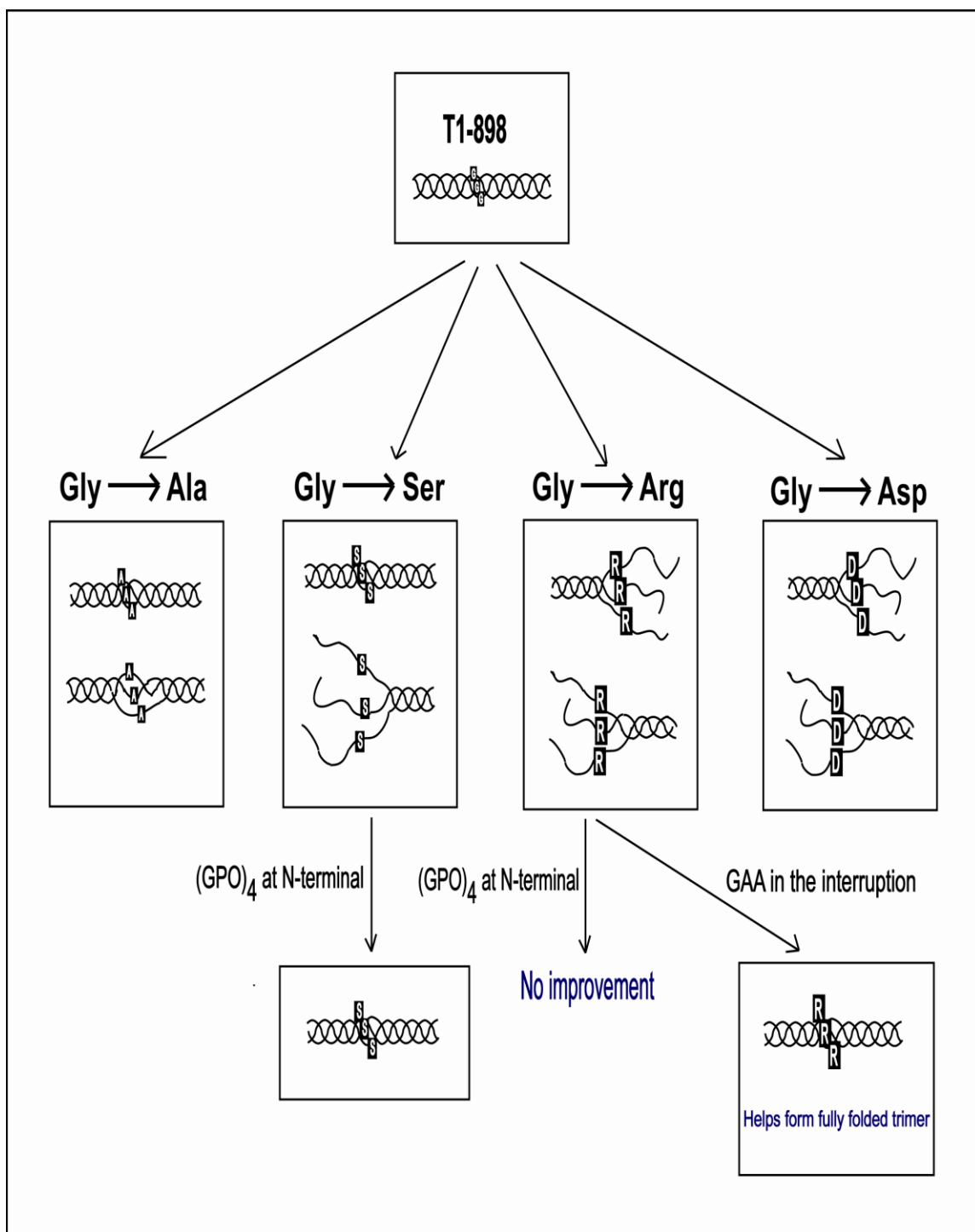


Figure 3.4 Effect of different Gly substitutions on the equilibrium conformations.

3.4 Discussion

Missense Gly substitutions by bulkier residues (Ser, Ala, Cys, Val, Arg, Asp, Glu) lead to various OI phenotypes (Kuivaniemi et al. 1997; Byers & Cole 2002; Marini et al. 2007). Studies have shown that the identity of the amino acid replacing Gly is correlated with the lethality of OI (Byers & Cole 2002; Bodian et al. 2008). However, the molecular mechanism for this correlation is still poorly understood. It is known that Gly mutations in collagen delay helix folding (Engel & Prockop 1991; Raghunath et al. 1994). The Gly to Ala mutations have also been shown to disrupt the triple helix structure to some extent in collagen model peptides (Bella et al. 1994; Hyde et al. 2006). Here, a series of triple helical peptides modeling Gly mutations involved in OI are designed to investigate the structural consequences of the identity of residues replacing Gly and the local sequence environment, in order to further our understanding of the structural bases of various OI phenotypes.

Peptide set T1-898, T1-898[G16A], T1-898[G16S], T1-898[G16D] and T1-898[G16R] models Gly substitution by Ala, Ser, Arg and Asp in the same sequence environment and it displays an inverse relationship between the size of the amino acid replacing Gly and the stability of the peptide (Table 3.1). NMR HSQC and diffusion studies on this set of peptides show that Gly to Ala substitutions can be incorporated into the triple helix via a partially folded trimer with a break in the middle (Figure 3.4). The Ser substitution of Gly in T1-898 could also get incorporated into the triple helix but a relatively larger proportion of

partially folded species with folded C-terminal but unfolded N-terminal was found (Figure 3.4). It seems that substitution of Gly by small residues Ala/Ser can be tolerated in a trimer conformation at the expense of generating some partially folded species. However, introduction of big and charged residues Arg/Asp within this context stops the trimer formation and only allows partially folded conformations (Figure 3.4). In all, the level of structural disruption correlates well with the lethality of OI, while Arg/Asp causes larger disruption and is more likely to result in lethal OI. It is proposed that Gly substituting residues Arg/Asp/Glu/Val may have a different mechanism of OI compared with Ser/Cys (Bodian et al. 2008), and studies here show big and charged residues Arg/Asp are much more disruptive than small residues Ala/Ser, suggesting the possible existence of two mechanisms.

To better accommodate Gly→Ser and Gly→Arg mutations into the triple helix, a stronger renucleating sequence (GPO)₄ was used to replace the N-terminal sequence GPO(GAO)₃, making new peptides GPOT1-898[G16S] and GPOT1-898[G16S]. The GPO stretch forces all the partially folded species into fully folded trimer conformation for peptide GPOT1-898[G16S], however, it destabilizes the triple helix formation for GPOT1-898[G16S] (Figure3.4). After further modifying the triplet next to the mutation site from GPV to GAA as well as N-terminal renucleation sequence from (GPO)₄ to (GPO)₃GAO to make a more flexible local environment, fully folded trimer is formed for Gly→Arg mutations (Figure3.4).

Two factors seem to be important for tolerating a Gly→X substitution in a triple helix: strong renucleation and local flexibility. Studies on peptide T1-892 showed that even a Gly→Ala substitution will stop the trimer formation at the substitution site (Hyde et al. 2006), while a strong renucleation sequence helps the formation of a fully folded trimer. Since Ser is a small residue, substituting Gly by Ser did get incorporated into a triple helix. Both (GPO)₄ and GPO(GAO)₃ are strong renucleation sequences and they both have the ability to allow the peptides to fold through the substitution site. (GPO)₄ is an even stronger renucleation sequence than GPO(GAO)₃, therefore, it is able to convert the remaining partially folded species to fully folded trimer in the case of peptide GPOT1-898[G16S]. However, substituting Gly by Arg may cause serious problems for incorporation into a trimer conformation, as Arg being a large amino acid can cause more steric distortions than Ala and Ser. A more flexible environment around the mutation site will give the residue more room to adjust and refold into the triple helix. For T1-898[G16R] and GPOT1-898[G16R], the presence of bulky Val15 next to the Arg provides a rigid rather than flexible environment. Incorporation of a more flexible environment immediately N-terminal to the Gly mutation site in peptide GAAT1-898[G16R] helps accommodate the Gly→Arg mutation into a fully folded trimer. To conclude, strong renucleation and local flexibility need to be balanced for accommodating bulky residues in the triple helix.

OI collagens are capable to fold through any type of Gly substitutions. This could be due to the larger size of the collagen molecule (338 triplets) compared

with the peptides (31 triplets), or could reflect the heterotrimeric nature of mutant type I OI collagens compared with the homotrimeric peptide models. Here, we have successfully developed multiple strategies to accommodate different residues replacing Gly, particularly bulky Arg, into the triple helix for peptides, which provide better mimics for OI mutations than unfolded or partially folded peptides.

It is interesting to note that the ability to incorporate the residue replacing Gly within peptide models correlates with the clinical severity of OI cases: Gly→Ala replacements are largely non-lethal, Gly→Ser replacements are mixtures of lethal and non-lethal, while all Gly→Arg replacements at positions greater than residue 178 are lethal. Computational analysis also indicates a relationship between the degree of conformational perturbation in terms of Hydrogen bonding and the identity of the residues near the mutation site (Radmer & Klein 2004). In contrast, Makareeva et al found no relation between the decrease in thermal stability of different OI collagens and the residue replacing Gly (Makareeva et al. 2008). Our data suggest that how far the structural perturbation extends along the chain in the mutant collagen depends on the residue replacing Gly, and this could affect the interactions of mutant molecules within fibrils with proteoglycans. The larger substituting residues lead to an increased population of partially folded trimers, with larger unfolded regions for Arg>Ser>Ala, and this could impact recognition and degradation triggering the unfolded protein response in the e.r.

Chapter 4

NMR Studies on heterotrimer peptides modeling a natural interruption in type IV collagen

4.1 Introduction

Collagen, the predominant protein in the human body, contains a characteristic triple-helical conformation, consisting of three left-handed polyproline II-like helices supercoiled around a common axis (Brodsky & Persikov 2005; Shoulders & Raines 2009). The close packing of the three chains can only accommodate Gly as every third residue, generating the repetitive (Gly-X-Y)_n sequence pattern. The residues at the X and Y positions are frequently Pro and Hydroxyproline (Hyp), respectively, serving to stabilize the triple helix formation (Persikov et al. 2000). Twenty eight distinct types of collagens, including both homotrimers with three identical chains and heterotrimers composed of chains with different amino acid sequences, have been identified to date (Kielty & Grant 2002; Shoulders & Raines 2009). The strict requirement of Gly as every third residue has been perfectly maintained throughout the triple helical domain of fibrillar collagens (Types I, II, III, V, XI) and even a single mutation leads to diseases (Myllyharju & Kivirikko 2001; Byers & Cole 2002; Di Lullo et al. 2002; Marini et al. 2007). In contrast, interruptions in the repetitive Gly-X-Y sequences have been observed frequently in the triple helical domain of all non-fibrillar collagens (Kielty & Grant 2002; Myllyharju & Kivirikko 2004; Thiagarajan et al. 2008). These natural breaks are suggested to play important roles in molecular structure, self-association or binding (Brazel et al. 1987; Miles et al. 1995; Mohs et al. 2006). However, one extra Gly to X mutation other than the natural interruptions, results in pathological conditions (Hudson et al. 2003; Kalluri 2003).

Type IV collagen, a critical non-fibrillar collagen in basement membranes, forms a meshwork-like suprastructure, which serves as a filtration barrier for blood vessels and exerts regulation of angiogenesis and tumor growth (Yurchenco & Ruben 1988; Schittny & Yurchenco 1989; Hudson et al. 1993; Kalluri 2003). Type IV collagen consists of six distinct α -chains, namely $\alpha 1(\text{IV})$, $\alpha 2(\text{IV})$, $\alpha 3(\text{IV})$, $\alpha 4(\text{IV})$, $\alpha 5(\text{IV})$ and $\alpha 6(\text{IV})$, which can selectively form three types of heterotrimers, $[\alpha 1(\text{IV})]_2\alpha 2(\text{IV})$, $\alpha 3(\text{IV})\alpha 4(\text{IV})\alpha 5(\text{IV})$, and $[\alpha 5(\text{IV})]_2\alpha 6(\text{IV})$, whose relative populations depend on the types of membranes (Borza et al. 2001). All isoforms of type IV collagen contain three domains: a small N-terminal 7S domain, a middle triple helical domain and a C-terminal globular non-collagenous (NC1) domain (Figure 4.1) (Hudson et al. 2003). The NC1 domain initiates the trimeric self-assembly of monomer α -chains and controls the selection of individual chains (Kalluri 2003). Unlike fibrillar collagens, type IV collagen contains at least 20 interruptions in the Gly-X-Y pattern in each chain along its ~1400 residue long triple helix (Brazel et al. 1987; Zhou et al. 1992). The interruptions can be categorized by the number of residues in the break, while interruptions in the form of X-Y-Gly-(AA)_n-Gly-X-Y are called GnG type interruption. It is observed that G1G and G4G are the most common types of interruptions (Thiagarajan et al. 2008). These interruptions often occur simultaneously in all three chains of Type IV collagen, while there might be different lengths of interruptions at a given site in each chain with a typical example of G1G interruptions in two $\alpha 1(\text{IV})$ chains overlapped with a G4G interruption in the $\alpha 2(\text{IV})$ chain (Mohs et al. 2006; Thiagarajan et al. 2008).

The function and structural consequences of the interruptions have been extensively investigated (Hofmann et al. 1984; Kilchherr et al. 1985; Brazel et al. 1987; Li et al. 2007). A number of flexible regions around the interruption sites have been identified in type IV collagen by rotary shadowing techniques (Hofmann et al. 1984). X-ray structure studies of a peptide (POG)₄PG(POG)₅ containing a GPG interruption indicated loss of helix registration as well as disruption of hydrogen bonding (Bella et al. 2006). Besides the structural effects, some interruptions are known to play a role in binding to integrins in tumor cell adhesion or serving as specific cleavage sites of matrix metalloproteinase (Furcht et al. 1994; Miles et al. 1994). Structural perturbations such as flexible sites may make the interruptions sites unique for recognition by other biomolecules.

Triple-helical peptides can serve as valuable models for collagen (Brodsky et al. 2008). It is known that peptides with Gly as every third residue and a high content of imino acids spontaneously self-assemble into homotrimers with a triple-helical structure. Such stable homotrimer peptides have been well characterized in terms of stability, folding and dynamics, and molecular structures have been obtained by NMR and X-ray crystallography (Bella et al. 1994; Baum & Brodsky 1999; Brodsky & Persikov 2005; Li et al. 2007). These homotrimer peptides serve as good models for collagen molecules with three identical chains, such as type II collagen in cartilage. It has proved more difficult to obtain and characterize hybrid triple-helical peptides composed of chains with different amino

acid sequences (heterotrimers), which can serve as models for heterotrimeric collagens such as type IV collagen.

Heterotrimer peptide design strategies have included covalent linkage to force the selection of three chains and their alignment (Fields & Prockop 1996; Fiori et al. 2002; Raynal et al. 2006), while recent studies used electrostatic interactions within the (Gly-X-Y)_n sequences to direct desired self-assembly (Gauba & Hartgerink 2007; Gauba & Hartgerink 2008). Thus far, techniques for biophysical characterization of heterotrimers have probed average properties of the triple-helix. For instance, Gauba and Hartgerink used circular dichroism (CD) spectroscopy to follow the thermal stabilities of single peptide vs. mixed peptide solutions, and differences in stability are interpreted in terms of homotrimer and heterotrimer molecules (Gauba & Hartgerink 2007; Gauba & Hartgerink 2008). In contrast, NMR has the capacity to follow the properties of individual residues and individual chains within trimers (Buevich & Baum 2001). Complexes where one subunit is labeled and others are not have been studied by NMR to define the features of the labeled complex and its interaction (Clore & Gronenborn 1998). Here, a similar strategy is applied to the triple-helix, using NMR to monitor labeled residues within one peptide chain to follow specific structural and chemical properties within a heterotrimer versus a homotrimer context.

NMR characterization of triple helical peptides has provided useful insights into structural features of the interruption sites (Li et al. 2007; Thiagarajan et al.

2008). Studies have shown that both G1G and G4G interruptions cause localized perturbations in conformation, while the locations of the hydrophobic residue in the two interruptions are very different (Li et al. 2007; Thiagarajan et al. 2008). These studies were done in a homotrimer environment, whereas the real interruptions in a number of collagens such as type IV occur in heterotrimeric surroundings. The structural effects of interruptions may be interleaved with homotrimer vs heterotrimer context, therefore, heterotrimer triple helical peptides are designed to model interruptions and NMR will be utilized to characterize the conformation, dynamics and folding. This project was performed in collaboration with Prof. Barbara Brodsky and Dr. Madhan Balaraman from her laboratory.

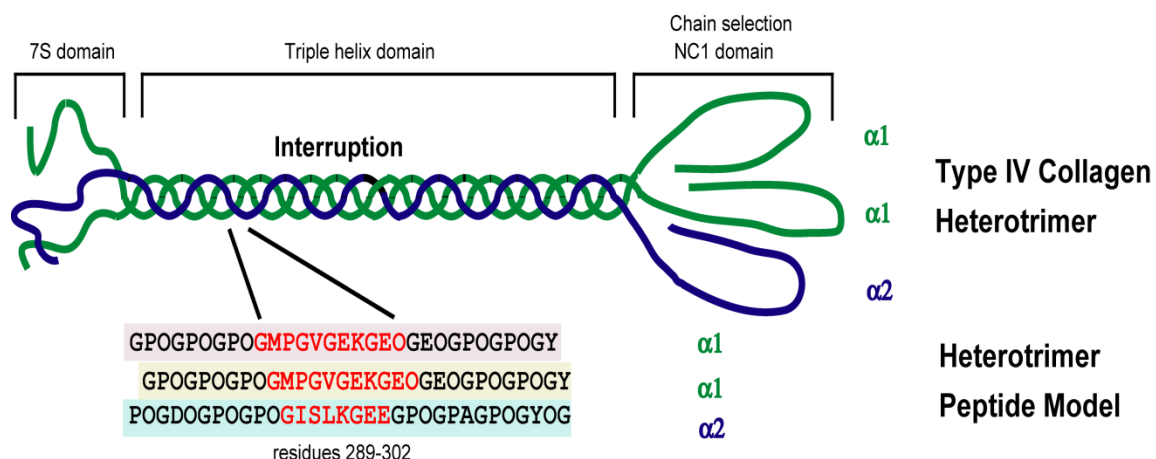


Figure 4.1 Heterotrimer peptide models of interruptions in type IV collagen. The major isoform of type IV collagen $\alpha1(\text{IV})_2\alpha2(\text{IV})$ consists of two chains with identical amino acid sequences (green) and one chain with a different sequence (blue). All type IV collagen contain three domains: a small N-terminal 7S domain, a middle triple helical domain and a C-terminal globular non-collagenous (NC1) domain. The NC1 domain mediates the chain selections of trimeric self-assembly of monomer molecules. Natural interruptions in the repeating Gly-X-Y sequence pattern in the triple helical domain are likely to play important roles in collagen structure and function. A heterotrimer peptide is designed to model a G1G interruption in the $\alpha1(\text{IV})$ chains overlapped with a G4G interruption in the $\alpha2(\text{IV})$ chain at sites 289-302 of type IV collagen (residues colored red).

4.2 Materials and Methods

4.2.1 Sample preparation

Peptides A, B and C were synthesized by Tufts University Core Facility (Boston, MA) (Table 4.1). Peptide A was unlabeled. Peptide B was selectively $^{15}\text{N}/^{13}\text{C}$ doubly labeled at positions G14, V15, G16 and G25. Peptide C was synthesized with $^{15}\text{N}/^{13}\text{C}$ labeled amino acids at positions I13, S14, L15 and A25. Samples of homotrimer peptides A, B and C were prepared in 10% $\text{D}_2\text{O}/90\% \text{H}_2\text{O}$ with a concentration of 6 mM. Peptides A and C were mixed in solution at a molar ratio of 2A:1C (4mM:2mM) to look for formation of heterotrimeric molecule A_2C . Similarly, heterotrimer peptides ABC and B_2C were made by mixing 1A:1B:1C (2mM:2mM:2mM) or 2B:1C (4mM:2mM), respectively. The mixtures were heated to 50°C, cooled slowly to 0°C and incubated for 30 hrs.

4.2.2 NMR spectroscopy

NMR experiments were performed on a Varian INOVA 600 MHz spectrometer with a cold probe or a Varian VNMRs 800MHz spectrometer. To accomplish sequential assignments for peptide B_2C , ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) (Kay et al. 1992) and 3D HNCA experiments (Ikura et al. 1990) were carried out at 5°C. 3D HNCA experiments consisted of $100(t_1) \times 30(t_2) \times 1024(t_3)$ complex points, and was recorded with spectral widths of 8000(F_1), 2800(F_2) and 8000(F_3) Hz. 3D ^{15}N edited NOESY-HSQC experiments (Fesik & Zuiderweg 1988; Marion et al. 1989; Messerle et al. 1989) were performed with a mixing time of 50 or 60ms at 5°C, 9°C and 10°C. Relaxation R_1 , R_2 and

heteronuclear NOE measurements (Fan et al. 1993; Palmer 1993; Farrow et al. 1994) were done at 10°C as described in section 2.2.8.1. Similar HNCA, NOESY-HSQC and relaxation experiments were done on other peptides. To obtain residue-specific melting curve and calculate amide proton temperature gradients, ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) was performed at 0-29 degrees for C, A₂C and B₂C, 3-19 degrees for B and 3-25 degrees for ABC. All data were processed using the FELIX 2004 software package (MSI, San Diego, CA), and analyzed with FELIX 2004 or NMRView.

4.2.3 NMR folding experiments

The NMR folding experiments were performed as previously described (Liu et al. 1996; Buevich et al. 2004). The sample was denatured by heating to 50°C for 20 min, cooled to 10°C by immersion in an ice/water bath for 10s, and then transferred to the spectrometer which was equilibrated at 10°C. A series of HSQC spectra were acquired every 5 min immediately after the sample was placed in the probe. 64 t_1 increments and 2 scans per increment were used. The kinetics of folding was monitored by measuring cross-peak intensities as a function of time. For easy comparison among different residues, the intensities of the monomer peaks were normalized following this equation: $I(t) = (I_t - I_f) / (I_0 - I_f)$, where I_t , I_0 , and I_f are the intensities at time t , at time 0 and at final time point when monomer:trimer equilibrium has been reached. The rate constant k is then obtained by fitting of normalized intensity I against time t for each residue using the following equation: $I = a + b \exp(-kt)$.

4.3 Results and discussion

4.3.1 Design of model peptides

The design strategy presented here is mainly based on mixing a peptide chain (C) which has a low propensity to self-associate into a triple-helix together with a peptide chain (A or B) of different sequence and high triple-helix propensity. The peptides model a natural interruption in the Gly-X-Y repeating sequence of both the $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ chains (Kuhn 1995; Khoshnoodi et al. 2008). Peptides A and B, include residues 289-302 (Table 4.1) from the sequence of the $\alpha 1(\text{IV})$ chain of type IV collagen with a G1G interruption. Peptide B has the same sequence as A except differences in 4 residues: one extra residue O at the N-terminal for B only, residue O in the first and fourth triplets of A mutated to residue K and P in peptide B, respectively, and the residue K in the eighth triplet of A mutated to residue E in peptide B. More importantly, peptide A is unlabeled, in order to simplify the NMR spectrum of heterotrimer A_2C and allow easier comparison with the spectrum of homotrimer C; peptide B, however, contains selectively $^{15}\text{N}/^{13}\text{C}$ labeled residues GVG at the interruption site and a Gly25 at the C-terminal, in order to obtain detailed structural and dynamic features of the GVG interruption in heterotrimer B_2C (Figure 4.1). Peptide C includes residues 288-295 from the corresponding sequence in the $\alpha 2(\text{IV})$ chain at the same site with a corresponding G4G interruption with selectively $^{15}\text{N}/^{13}\text{C}$ labeled residues ISL at the interruption site and an Ala25 at the C-terminal (Table 4.1). All collagen sequences are flanked by stabilizing triplets and a terminal Tyr for concentration determination. In order to push the heterotrimer formation, extra residues O or P were used to make blunt

ends to stabilize the heterotrimer ABC; Moreover, the overlapping of G1G and G4G interruptions in the heterotrimer makes it possible that there is one and only one Gly in the vertical 3-residue window across the whole triple helix; however, the homotrimers containing G1G or G4G interruptions break that Gly rule. Finally, favorable electrostatic interactions between K and D at the N-terminal may help the formation of heterotrimer B₂C, whereas unfavorable electrostatic interactions between the same charged residues K or D may destabilize homotrimer B or C.

Peptide name	Peptide sequence
A	Ac-GPOGPOGPOG <u>MOGVGEKGEOGKOG</u> POGPOGY-NH ₂
B	Ac-OGPKGPOGPOG <u>MPGVGEKGEOG</u> POGPOGY-NH ₂
C	Ac-POGDOGPOGPOG <u>ISLKGEEG</u> POGPAGPOGYOG-NH ₂

Table 4.1 Sequences of polypeptides A, B and C. Underlined residues are sequences from the $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ chains. Bold residues in B and C are $^{15}\text{N}/^{13}\text{C}$ doubly labeled, whereas no residue in A is labeled. Peptide B has the same sequences as A except differences in 4 residues: one extra residue O at the N-terminal for B only, residue O in the first and fourth triplets of A mutated to residue K and P in peptide B, respectively, and the residue K in the eighth triplet of A mutated to residue E in peptide B. The change of charged residues is aimed to allow favorable electrostatic interaction in heterotrimer ABC or B₂C.

4.3.2 NMR monitoring of chain-specific stability of heterotrimer A₂C

NMR studies were carried out on peptide C and a 2A:1C peptide mixture to evaluate if heterotrimer A₂C was formed in the mixture. 3D ¹⁵N edited NOESY-HSQC experiments were performed to obtain sequential assignments by using the strong inter-residue NOEs between the NH of residue i and the HA of residue i-1. The ¹H-¹⁵N HSQC spectra of peptide C alone (blue) and peptide C in a 2A:1C mixture (red) were compared, while unlabeled peptide A in the mixture shows no peak in the spectrum (Figure 4.2a). As seen in previous spectra of triple helical peptides, the HSQC spectrum of the homotrimer peptide C shows monomer and trimer peaks for each labeled residue (Figure 4.2a). Three trimer peaks are seen for Leu, indicating different environments for the Leu residues located in the interruption region in each of the three chains of the homotrimer. The absence of three distinct trimer peaks for Ile, Ser and Ala residues, which show only one or two trimer resonances, is likely due to overlapping trimer resonances. For the HSQC spectrum of peptide C in a 2A:1C mixture after heating and cooling, there is only 1 Leu trimer resonance, compared with 3 distinct resonances for the C homotrimer (Figure 4.2a). This is consistent with an A₂C composition where there is only 1 C chain in a heterotrimer. A single peak is seen for Ile, but its position is shifted compared with the C homotrimer, suggesting a different environment in the heterotrimer compared with the homotrimer. A single resonance is seen for the Ala trimer and no Ser resonance is seen for the mixed peptides, which is likely due to overlap with monomer resonances. The differences in the number of trimer

resonances for each labeled residue and their altered chemical shifts between peptide C and the 2A:1C mixture, suggesting heterotrimer formation in the mixture.

The temperature dependence of the trimer resonances arising from the C chain in a homotrimer peptide C solution as well as that in the mixed 2A:1C system was also compared. The melting temperatures of trimer resonances in the mixture differ markedly from those seen for the C homotrimer. All three Leu peaks in homotrimer C show a thermal transition at the same temperature ($T_m=7^\circ\text{C}$), however, the stability of the single Leu trimer resonance in the mixture ($T_m\sim 17^\circ\text{C}$) is markedly higher (Figure 4.2b). The enhancements in the thermal stabilities of chain C due to hybrid chain association with A were seen more clearly by comparing the HSQC spectra of homotrimer C and the 2A:1C mixture at different temperatures (Figure 4.3). At 5°C , the trimer resonances are seen for both homotrimer C and the mixture. Raising the temperature to 11°C leads to loss of almost all trimer peaks in homotrimer C, indicating the complete melting of the C homotrimer triple helix by this temperature (Figure 4.3). However, all the trimer resonances of C in the mixed system remain strong at 11°C and they do not disappear until the temperature is higher than 20°C (Figure 4.2b, 4.3). The high stability of the residues in peptide C must arise as a result of interactions involved in heterotrimer association with A.

The formation and characterization of hybrid triple-helical peptide systems is important for modeling heterotrimeric collagens and for modeling diseases where

there are one or two mutations in a collagen triple-helix. Here, NMR has been successfully used to monitor the chemical shifts and stability of defined residues within one chain in a homotrimer compared with a heterotrimer setting. The NMR results show that in a 2:1 ratio mixture of A and C chains, peptide chain C is present in mixed trimers containing both A and C chains, and supports the presence of a single A_2C molecular species. If both A_2C and AC_2 species were present, multiple resonances with a more heterogeneous spectrum would be expected. It is not possible to eliminate the possibility of some A homotrimers, since these chains are not labeled. The simple approach presented here can be done on any hybrid system and provides evidence of the incorporation of a chain into a heterotrimer triple-helix as well as residue specific information about that chain.

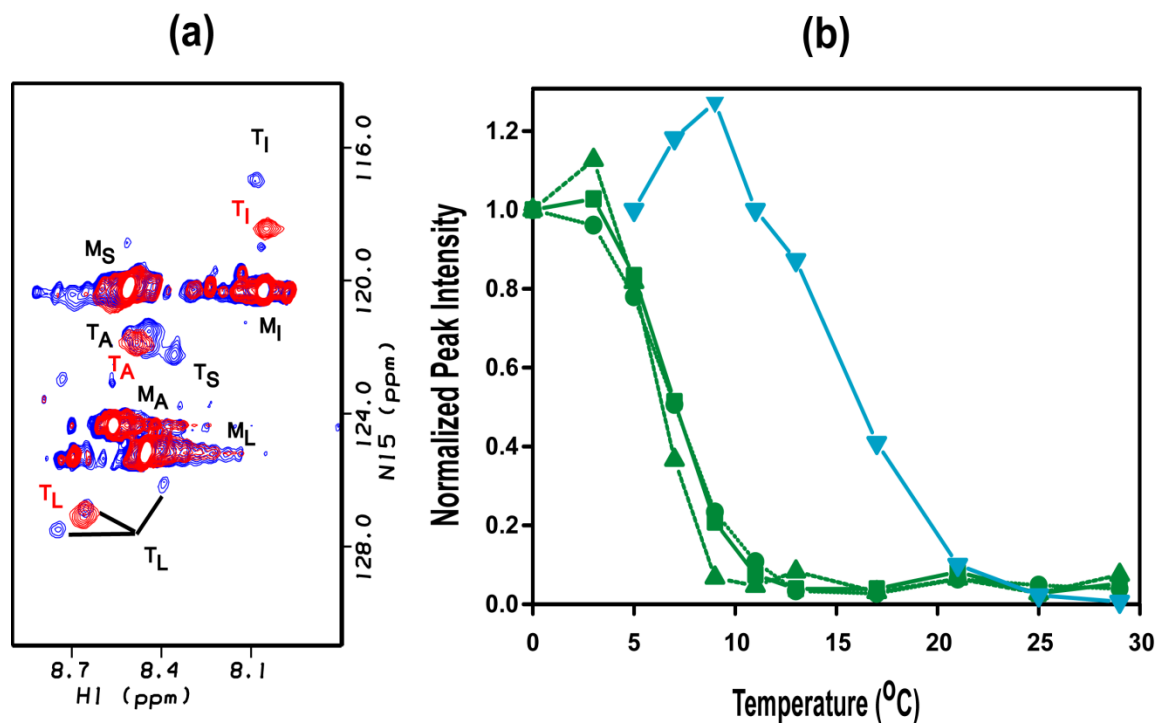


Figure 4.2 (a) Overlapped ^1H - ^{15}N HSQC spectra of peptide C (blue) and a 2A:1C peptide mixture (red). The peaks corresponding to monomer and trimer state are denoted with superscript M or T. The large number of minor monomer peaks are due to cis/trans isomerization in the unfolded state in this proline rich chain. Other small peaks between the monomer and trimer peaks may be due to noise or impurities. (b) Melting curves of trimer resonances of Leu in peptide C (green) and in a 2A:1C peptide mixture (cyan). Three Leu in homotrimer peptide C have a melting temperature of 7 $^{\circ}\text{C}$, while the single Leu in the 2A:1C peptide mixture shows a melting temperature of 17 $^{\circ}\text{C}$. NMR melting curves are not obtained for peptide A, which contains no $^{15}\text{N}/^{13}\text{C}$ labeled residues.

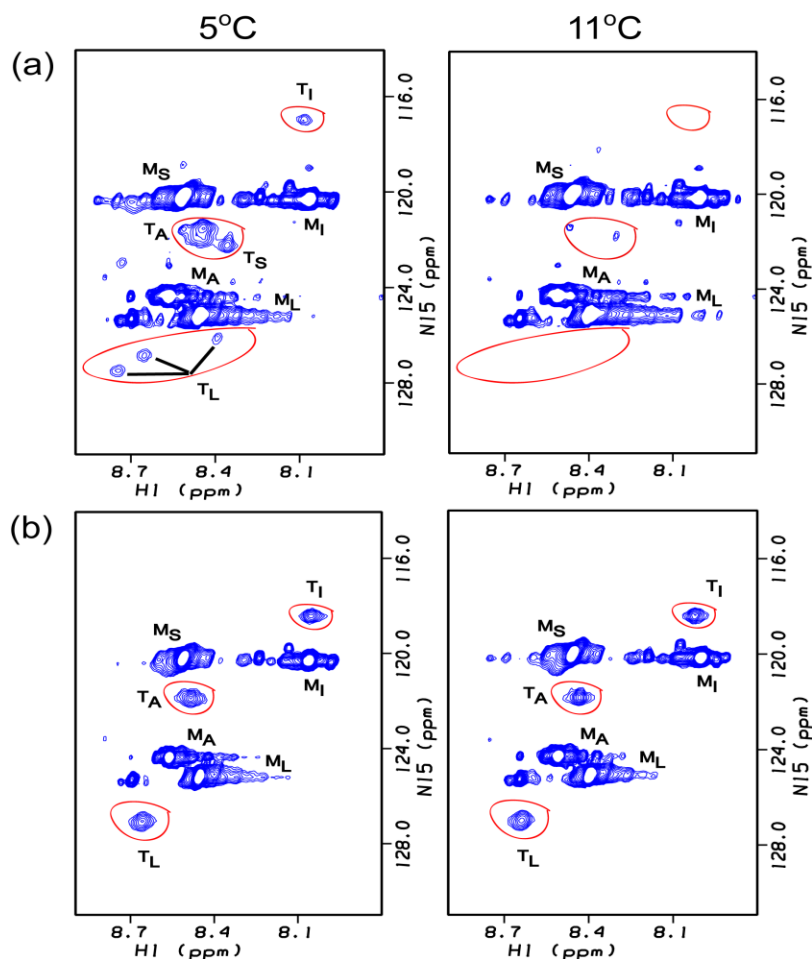


Figure 4.3 Comparison of trimer resonances of labeled Ile, Ser, Leu and Ala residues in peptide C (a) and in a 2A:1C peptide mixture (b) at different temperatures. Trimer peaks for all residues of peptide C in a homotrimer (a) are present at 5°C (left) and absent at 11°C (right), while trimer peaks are present at both temperatures when peptide C is present in a 2A:1C peptide mixture. The peaks corresponding to monomer and trimer state are denoted with superscript M or T.

4.3.3 NMR shows heterotrimer formation of ABC

NMR studies were performed on peptide B and a 1A:1B:1C peptide mixture to evaluate if heterotrimer ABC was formed in the mixture. The ^1H - ^{15}N HSQC spectra of peptide B alone and a 1A:1B:1C peptide mixture were obtained, and the sequential assignments were based on 3D HNCA experiments (Figure 4.4). The HSQC spectrum of homotrimer peptide B shows three trimer peaks for labeled residue V15, G16 and G25, indicating distinct chemical environments in three different chains. A trimer peak for G14 ($^2\text{G14}$) was missing possibly due to overlapping with other resonances. For the HSQC spectrum of the 1A:1B:1C mixture system, only 1 trimer resonance is observed for all labeled residues I13, S14, L15 and A25 in chain C, indicating only 1 C chain is included in a heterotrimer. There is also only 1 trimer resonance for labeled residues G14, G16 and G25 in chain B, indicating the presence of 1 B chain in a heterotrimer and the absence of homotrimer B. A trimer peak for V15 was missing possibly due to overlapping. The observance of one and only one trimer resonance for all labeled residues in chain B and C suggests the formation of heterotrimer ABC and excludes the existence of homotrimer B, even though it is stable.

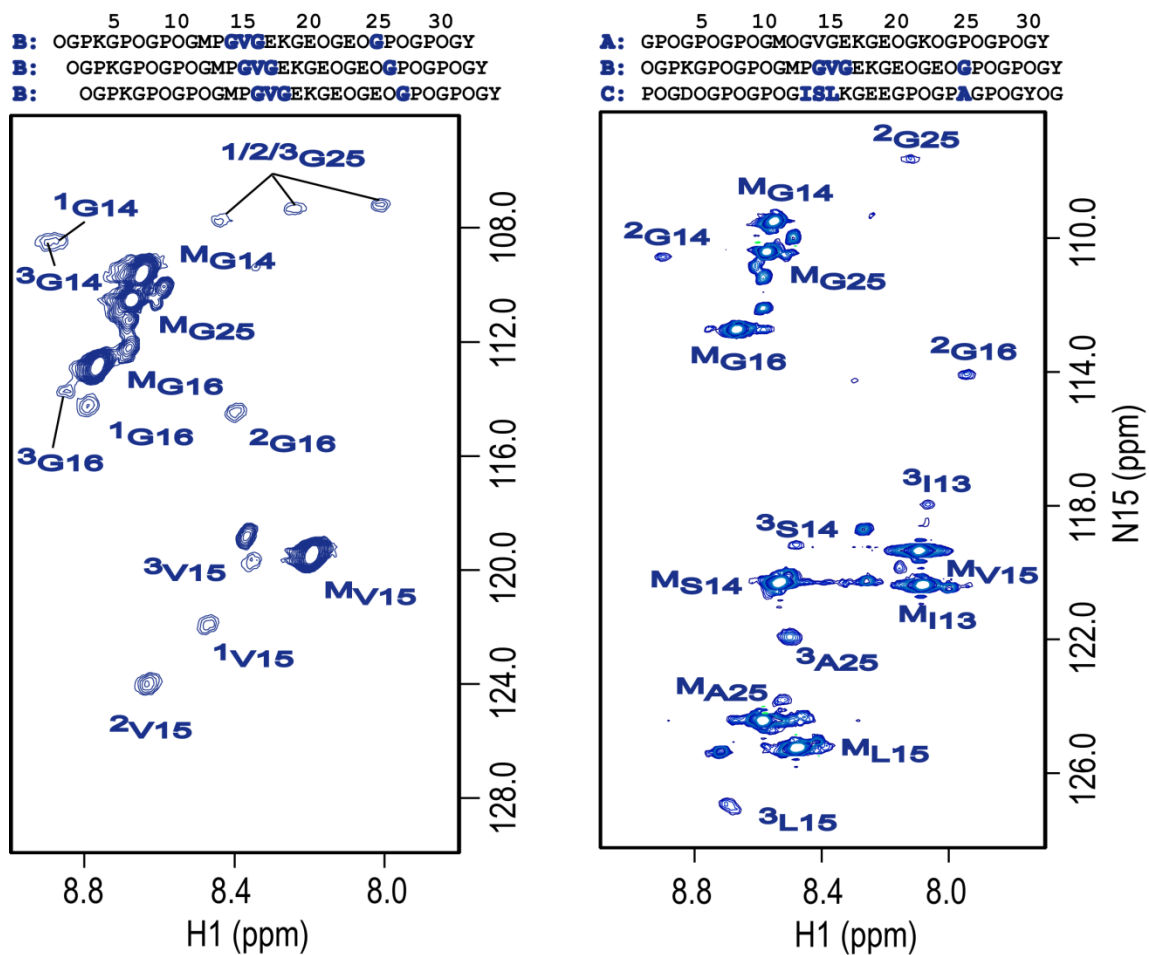


Figure 4.4 ^1H - ^{15}N HSQC spectra of homotrimer B at 3°C and heterotrimer ABC at 0°C. The peaks corresponding to monomer and trimer state are denoted with superscript M or numbers 1, 2 and 3, which not necessarily corresponds to the leading, middle and lagging chains, due to lack of inter-chain NOEs for chain assignments.

4.3.4 NMR conformation, dynamics and folding studies of heterotrimer B₂C

4.3.4.1 Sequential and chain assignments of heterotrimer B₂C

The above results have demonstrated that heterotrimer peptides can be successfully made using our strategies and NMR can be used to investigate chain-specific properties. To explore the detailed conformational and dynamic features at the interruption sites in a heterotrimer environment, heterotrimer B₂C composed of all selectively labeled chains was obtained in a 2B:1C peptide mixture and investigated by NMR. The HSQC spectrum of heterotrimer B₂C is well resolved (Figure 4.5a). All labeled residues in chains B have two trimer resonances, while the labeled residues in chain C all show a single trimer resonance, which is consistent with the composition of the heterotrimer as 2 B chains and 1 C chain. One monomer resonance is also observed for each residue, supporting a trimer state in equilibrium with monomer. No other trimer resonance is found for the labeled residues, indicating that no homotrimer is formed. Residues with the same superscripted number (eg. ¹G14, ¹V15) in the HSQC spectrum belong to the same chain. These chain-specific sequential assignments were based on the 3D HNCA experiments (Figure 4.6). By matching the alpha carbon chemical shifts, three groups of sequentially connected residues were identified. G25 contains two trimer resonances, which cannot be unambiguously assigned to chain 1 or 2, as G25 is separated from other labeled residues in chain B.

The superscripted number 1, 2 and 3 in the HSQC spectrum indicates the leading, middle and lagging chains, respectively. These chain assignments were

based on inter-chain NOEs, which were summarized in the experimental contact map (Figure 4.5b) (Li et al. 2007). The colored circles were utilized to represent experimentally observed NOEs ((○) HN-HN; (○) HN-HA; (○) HN-side chain protons). The shaded squares in the background denote intra-chain NOEs. Only a few inter-chain NOEs including $^1\text{G14 NH} - ^3\text{I13 } \alpha\text{H}$, $^1\text{V15NH} - ^3\text{I13 } \alpha\text{H}$, $^1\text{G16 NH} - ^3\text{S14 } \alpha\text{H}$, $^2\text{G14 NH} - ^1\text{V15 } \gamma\text{H}$ and $^3\text{I13 NH} - ^{1/2}\text{P13 } \alpha\text{H}$ could be observed. The inter-chain NOEs between $^1\text{G14 NH} - ^3\text{I13 } \alpha\text{H}$ and $^2\text{G14 NH} - ^1\text{V15 } \gamma\text{H}$ indicated that the 1st chain was the leading chain, while the 3rd chain was the lagging chain. All other inter-chain NOEs supported these chain assignments.

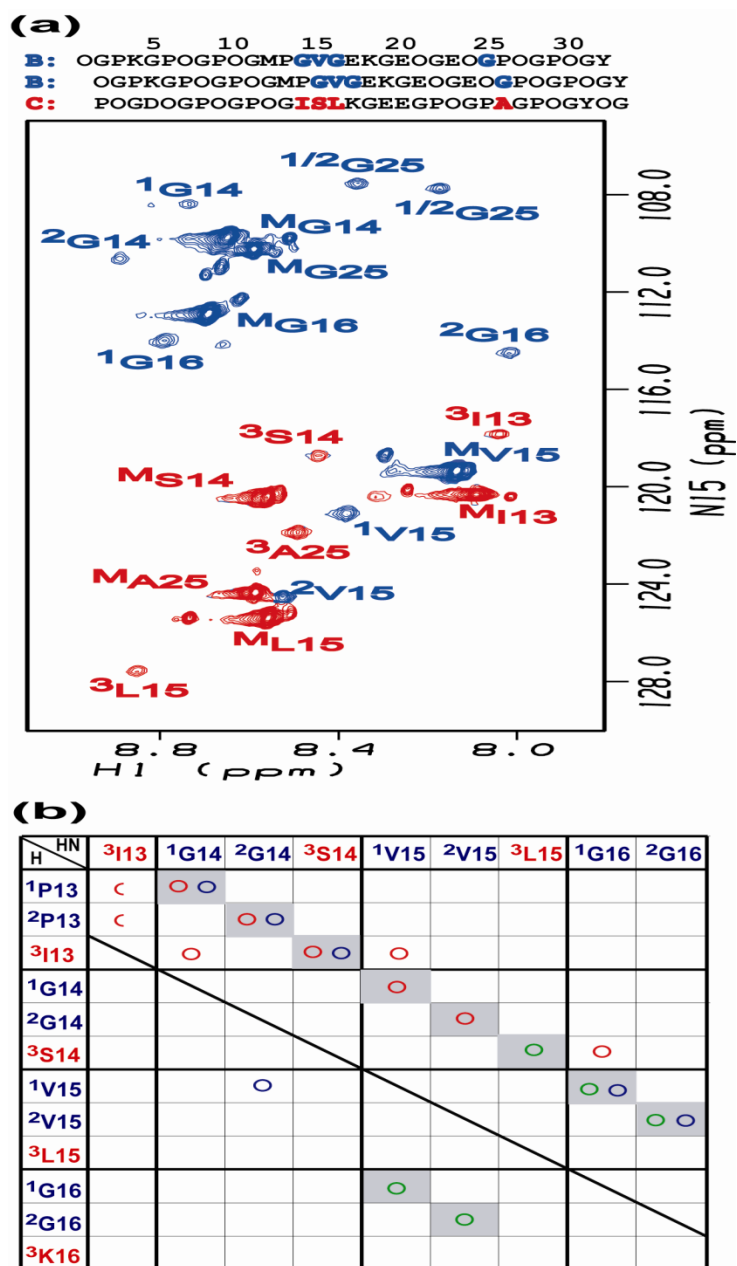


Figure 4.5 (a) ^1H - ^{15}N HSQC spectrum of heterotrimer peptide B₂C. Labeled residues are colored blue for chain B and red for chain C. The peaks corresponding to the monomer and trimer states are denoted with a superscript M or T, respectively. (b) Contact map generated from NH-H experimental NOEs of peptide B₂C. Experimental NOEs are represented by circles ((o) HN-HN; (o) HN-HA; (o) HN-side chain protons).

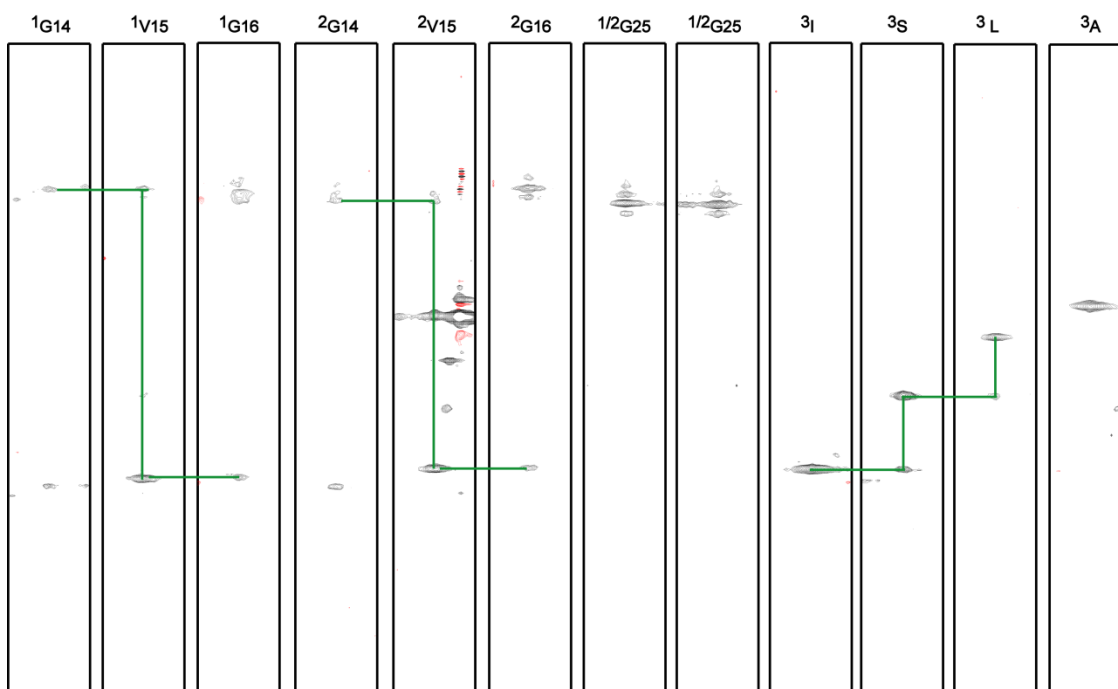


Figure 4.6 Strips from the HNCA spectra of heterotrimer B2C. Sequential connections in individual chains are traced out as shown in green lines. The numbers 1, 2 and 3 corresponds to the leading, middle and lagging chains, respectively, whose assignments are based on inter-chain NOEs. Two trimer resonances of G25 cannot be unambiguously assigned to chain 1 or 2.

4.3.4.2 Hydrogen bonding of heterotrimer versus homotrimer

Amide proton temperature gradients were measured for homotrimer B and C and heterotrimer B₂C (Figure 4.7). Amide protons showing the temperature gradients with a more positive value than -4.6 have been suggested to be involved in hydrogen bonding (Baxter & Williamson 1997; Cierpicki & Otlewski 2001). For homotrimer B, Gly25 in all three chains of triple helix have amide proton temperature gradients ~ -2 , indicating the formation of hydrogen bonds, while the residues Gly14-Val15-Gly16 at the interruption sites of triple helix all show

temperature gradients more negative than -6, indicating that the hydrogen bonds of Gly, which are preserved for continuous Gly-X-Y region, have been broken up (Figure 4.7). For homotrimer C, all labeled residues including Ile13, Ser14, and Leu15 at the interruption sites and Ala25 at the Y position in the Gly-X-Y repeating triplets show temperature gradients more negative than -6, indicating the absence of hydrogen bonds.

However, in heterotrimer B₂C, Leu15, which could be considered to replace a Gly, has temperature gradient with a value of -2.8, indicating the formation of hydrogen bonds. Besides Leu15 in the third chain, Gly14 in the first and second chains have temperature gradients with values of -4.4 and -4.6, indicating the possible existence of hydrogen bonds. Compared with Gly14 in heterotrimer B₂C, the Gly16 has temperature gradients more negative than -6, indicating the loss of hydrogen bond. Gly25 in heterotrimer B₂C forms hydrogen bonds, as indicated by its temperature gradients with values more positive than -3. In all, there is no hydrogen bond formation at the interruption sites in the homotrimer environment, but there are some in the heterotrimer context, indicating the interruption sites are better accommodated in heterotrimer, at least in terms of hydrogen bonding, which is crucial to stabilize the triple helical structure. The asymmetric effects of interruptions on Gly are also observed in heterotrimer, as Gly14 at the N-terminal of the interruption sites shows hydrogen bonds, but Gly16 at the C-terminal does not.

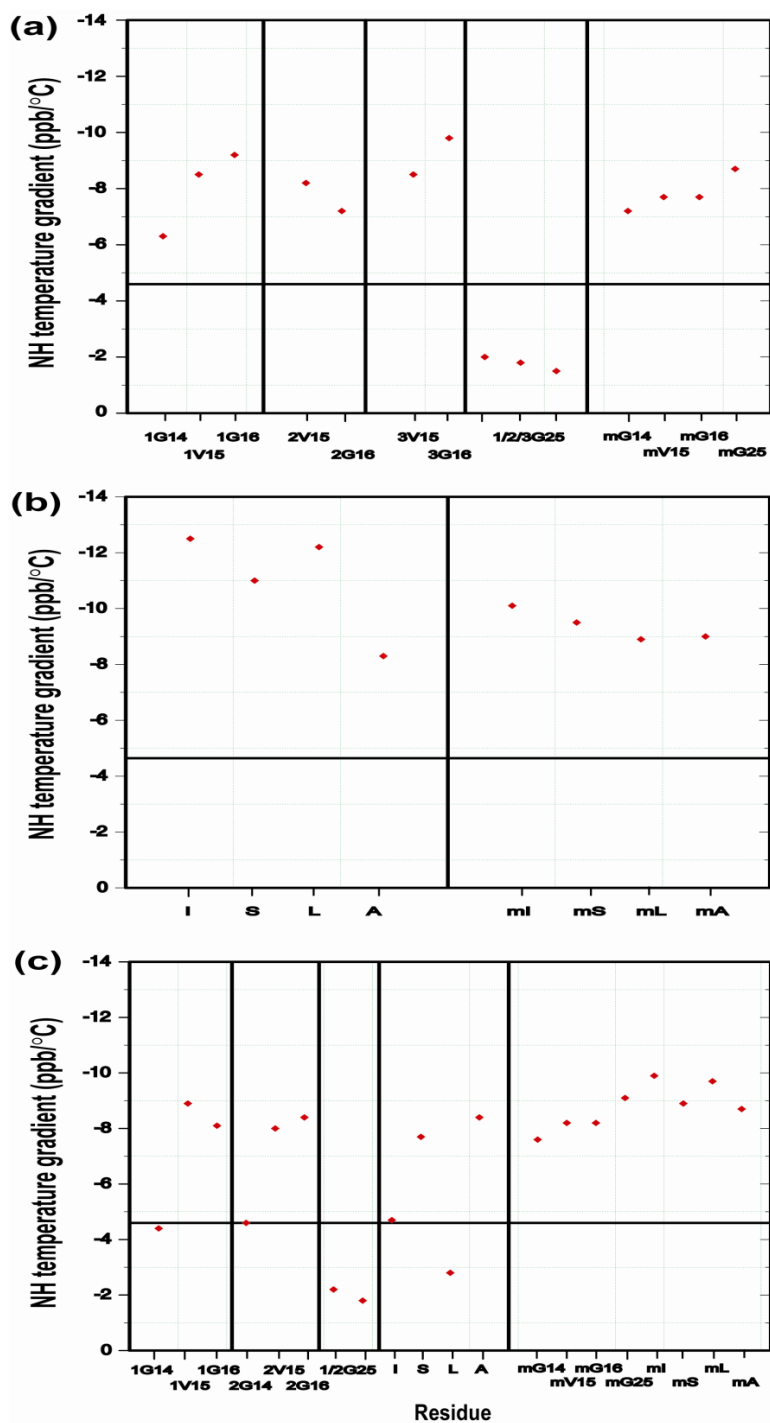


Figure 4.7 Amide proton temperature gradients of (a) homotrimer B and (b) homotrimer C and (c) heterotrimer B₂C. The black horizontal line corresponds to a value of -4.6, a cutting-off for hydrogen bonding, with less negative values indicative of hydrogen bonding.

4.3.4.3 Model structure of heterotrimer B₂C

A model structure was built by molecular modeling strategy using Molecular Operating Environment (MOE) (Chemical Computing Group Inc., Montreal, Canada) as described in section 2.3.4. The NMR structure of peptide (POG)₄-GF-(POG)₄ (Thiagarajan et al. 2008) was used as the starting structure with the sequences mutated to the appropriate sequences for B₂C. Then energy minimization was performed on the 13 central residues around the interruption sites (GVG plus 5 more on each side) with distance restraints from inter-chain NOEs. The model structure shows that the triplet C-terminal to the interruption sites GVG is more loosely bundled than the triplet N-terminal to the site (Figure 4.8). Gly14 in chains 1 and 2 are more tightly packed in the center of triple helix than Gly16. For Leu15 in chain 3, its amide proton is closely buried inside the triple helix, while the side chain is located toward outside. These results are consistent with the hydrogen bonding observations that Gly14 in chains 1 and 2 and Leu15 in chain 3 form hydrogen bonds, whereas Gly16 in chains 1 and 2 does not.

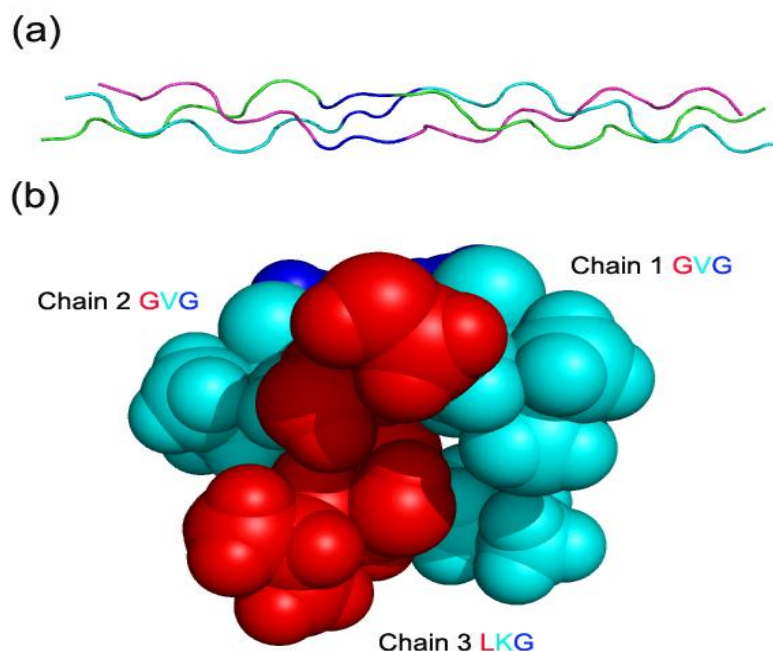


Figure 4.8 Model structure of heterotrimer B₂C. (a) Cartoon diagram of the model structure of heterotrimer B₂C. The labeled residues at the interruption sites (GVG in chains 1/2 and LKG in chain 3) are colored blue. (b) Sphere model of a segment (GVG in chains 1/2 and LKG in chain 3) of heterotrimer B₂C in the cross section view from N-terminal to C-terminal. Gly14 in chains 1 and 2 are more tightly packed in the center of the triple helix than Gly16. The amide proton of Leu15 in chain 3 is closely buried inside the triple helix, while its side chain is located toward outside.

4.3.4.4 Dynamics of heterotrimer B₂C at fast picosecond timescale

Heteronuclear ^1H - ^{15}N NOE measurements provide useful information on the fast, picosecond timescale motion such as bond vibration and sidechain rotation, while smaller NOE values indicate larger mobility on this time scale (Xu et al. 2003). The flexibility of heterotrimer B₂C was estimated by measuring the ^1H - ^{15}N NOE (Figure 4.9). Among the labeled residues Gly14-Val15-Gly16, G14 is the most rigid, while G16 is most flexible, in both the 1st and 2nd chains. Gly25 in the C-terminal region in both chains are very rigid. For the 3rd chain, the labeled residues Ile13-Ser14-Leu15 at the interruptions sites as well as Ala25 in the non-disrupted C-terminal region all have NOE values larger than 0.6, indicating all the residues in the 3rd chain are rigid. In all, most residues at the interruption sites are rigid except two Gly16 at the C-terminal of the sites.

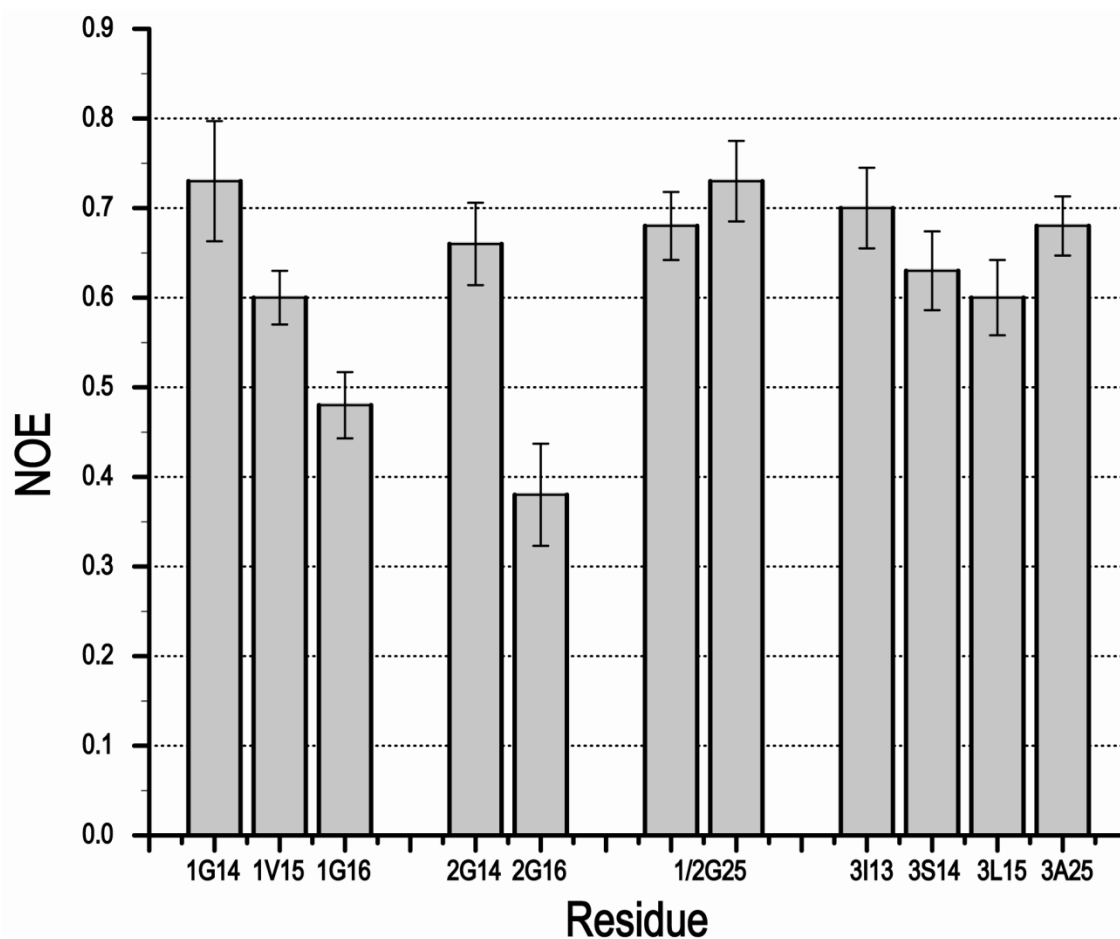


Figure 4.9 Heteronuclear ^1H - ^{15}N NOE values of heterotrimer peptide B2C. Numbers 1, 2 and 3 indicate the leading, middle and lagging chains, respectively. Gly25 in chains 1 and 2 cannot be distinguished from each other. Decreasing NOE values indicate increasing mobility on the fast, pico-second time scale. G16 in chains 1 and 2 have the smallest NOE values, suggesting that they are most flexible among all labeled residues.

4.3.4.5 Folding properties of heterotrimer B₂C

Folding of individual residues in heterotrimer B₂C was investigated by monitoring the decrease in intensities of monomer resonances (Figure 4.10) (Hyde et al. 2006). All the labeled residues including Gly14-Val15-Gly16 at the interruption sites and Gly25 in the C-terminal region in chains B show the same pattern of folding kinetics. Chain C also shares the same pattern in folding for labeled residues. Two steps of nucleation and elongation have been proposed for collagen folding, and studies on homotrimer peptides have shown that monomer resonances of residues at N- or C- terminals which are responsible for nucleation disappears faster than central residues (Liu et al. 1996). For heterotrimer B₂C, however, there is no difference in the folding of terminal or central residues, suggesting that nucleation rather than elongation is the rate-limiting step. The first monomer resonance in chain B shows lower intensities than that in chain C. The calculated rate constant is also higher for chain B ($\sim 0.035 \text{ min}^{-1}$) than chain C ($\sim 0.03 \text{ min}^{-1}$), indicating that chain B folds faster than chain C. Folding kinetics cannot be followed by chains in homotrimers. The heterotrimer here enables us to evaluate the chain-specific folding kinetics for the first time.

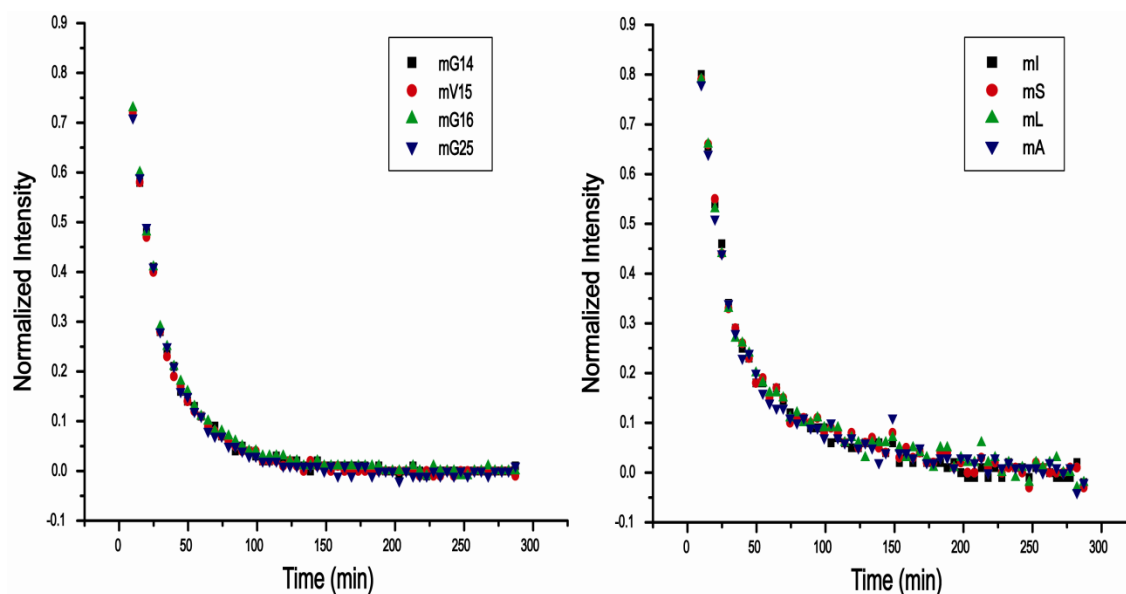


Figure 4.10 Residue-specific folding of chain B (Left) and chain C (right) in heterotrimer peptide B₂C at 10°C. The intensities of monomer resonances were monitored for ~300 min.

Chapter 5

**NMR studies on triple helical peptides
modeling the collagenase cleavage sites**

5.1 Introduction

Collagens are the major structural proteins of connective tissues such as skin, bone, tendon and ligament (Miller & Gay 1987; Brodsky & Persikov 2005). There are at least 28 types of collagens in vertebrates (Kielty & Grant 2002; Shoulders & Raines 2009). Types I, II and III collagens, also called interstitial collagens, are most abundant (van der Rest et al. 1991). Interstitial collagens contain a characteristic triple-helical conformation, which consists of three polyproline II-like helices supercoiled around a common axis (Rich & Crick 1961; Ramachandran 1967; Zijlstra et al. 2004). The close packing of the three chains can only accommodate Gly as every third residue, generating the repetitive (Gly-X-Y)_n sequence pattern. The residues at the X and Y positions can be almost any amino acid, but they are frequently Pro and Hydroxyproline (Hyp or O), respectively.

The degradation of collagen is an integral part in many biological processes such as wound healing, cell migration, tissue remodeling and organ morphogenesis (Baragi et al. 1997; Cawston et al. 1999; Brinckerhoff & Matrisian 2002; Visse & Nagase 2003). However, accelerated breakdown of collagen may result in many diseases such as arthritis, tumor cell invasion, glomerulonephritis and metastasis (Riley et al. 1995; Woessner 1998; Sternlicht & Werb 2001). The uniform triple helix makes collagen be degraded by only a few proteinases including a group of matrix metalloproteinases (MMPs) (Aimes & Quigley 1995; Sternlicht & Werb 2001). These MMPs (MMP-1, MMP-2, MMP-8, MMP-13,

MMP-14, MMP-18) can bind and cleave collagens at a unique locus approximately three-fourths away from the N-terminus of the collagens (Fields & Van Wart 1992). The cleavage site is after one Gly residue in the sequence of Gly-[Ile/Leu]-[Ala/leu]. Interestingly, there are dozens of other sites in the collagens that contain the cleavage scissile bond Gly-Ile/Leu but are not hydrolyzed (Fields 1991).

Early hypotheses concerning recognition of collagenase cleavage sites suggested that the distribution of 'weak helix' triplets, charged residues as well as the side-chain volume concentration could provide recognition information for MMPs (Gross et al. 1980). It has been proposed that the cleavage site region is locally unstable due to deficiency in imino acids and therefore the scissile bond could be exposed to the enzyme attack (Miller et al. 1976b; Highberger et al. 1979; Gross et al. 1980). Types I and III collagens can be cleaved by trypsin and α -chymotrypsin around the collagenase cleavage site, suggesting relaxation of the helix in this region (Miller et al. 1976; Ryhanen et al. 1983). Fields did a systematic sequence analysis and proposed that all the information necessary for efficient collagenolysis is contained in a 25-residue stretch, which is distinguished by a 'tight' triple helix with high imino acid content prior to the cleavage site and a 'loose' triple helix with low imino acid content following the cleavage site (Fields 1991). The correct distribution of imino acids has been shown to be important for cleavage, as mutation of Ile₇₇₆ or Gln₇₇₄ and Ala₇₇₇ to Pro residues eliminated MMP-1 and MMP-8 activity (Wu H 1990; Hasty KA 1993).

Besides the amino acid sequence, the unique triple helical conformation of collagen also plays important roles in the high specificity of collagen's recognition by MMPs, as indicated by huge differences in native collagen and denatured collagen (Welgus et al. 1981; Welgus et al. 1982; Mallya et al. 1990; Netzel-Arnett et al. 1991). Native collagen can only be degraded by a few proteinases, while denatured collagen is susceptible to a variety of proteinases. Collagen is cleaved at a single site, but denatured collagen is cleaved at multiple sites containing the sequence Gly-[Ile or Leu]-Y (Welgus et al. 1982). Also, native collagen is a much better substrate than denatured collagen for MMPs (Welgus et al. 1982). It is thus believed that the variation in sequence would lead to a change in the triple helical conformation and/or dynamics of collagen at the collagenase cleavage site, which can be identified by MMPs.

It was soon noted that the active site cleft in the catalytic domain of the enzyme was too narrow to accommodate a triple helix when a number of X-ray and NMR structures of MMPs were solved (Lovejoy et al. 1994; Li et al. 1995; Moy et al. 1998; Bode et al. 1999; Moy et al. 2000). It was proposed that the triple helix needs to unwind locally so that a single chain can enter into the active site of MMPs or the active site of MMPs undergo large conformational changes (Stultz 2002; Chung et al. 2004). Nagase's group showed that mutation of the catalytically essential residue Glu200 of MMP-1 to Ala resulted in a catalytically inactive enzyme, but in its presence noncollagenolytic proteinases digested collagen into typical $\frac{1}{4}$ and $\frac{3}{4}$ fragments, indicating that the MMP-1(E200A) mutant unwinds the

triple-helical collagen, which can then be cleaved by other proteinases (Chung et al. 2004). These studies indicate that the dynamics at the collagenase cleavage sites is critical for collagen recognition by MMPs.

A number of triple helical peptides were designed to model the cleavage site and they were hydrolyzed by MMPs in a similar fashion to native collagens, indicating that model peptides contain necessary information for recognition (Ottl et al. 1996; Lauer-Fields et al. 2000; Lauer-Fields & Fields 2002; Lauer-Fields et al. 2003; Minond et al. 2004). Here, a series of peptides, which model the natural cleavage site and a potential but noncleavable site just 2-triplets away from the natural site in type III collagen, are designed to investigate the relationship of the conformation and dynamics at the collagenase cleavage site to the collagenase activities of the peptides. Our NMR studies show that that a single Ile at the natural cleavage site has distinct chemical shift, unusual J-coupling values, increased dynamics and decreased local stability, suggesting that the Ile may be released from the restricted triple helical conformation and recognized by MMPs. The studies also present evidence that the neighboring imino acids affect the local conformation and dynamics at the cleavage site and thus the specificities of the recognition by MMPs. This project was performed in collaboration with Dr. Janelle Lauer-Fields from Prof. Gregg Fields' lab, who did enzymatic kinetics measurements.

5.2 Materials and Methods

5.2.1 Sample preparation

The peptide sets were synthesized by the Tufts University Core Facility (Boston, MA) with ^{15}N and/or ^{13}C labeled amino acids at selective positions for NMR characterization (Table 5.1). Peptides T3-778, T3-778IT-PO and T3-778P-A were selectively ^{15}N labeled at positions of Gly16, Ile17, Gly22, Ala23 and Gly31. Peptide T3-778I-L was selectively ^{15}N labeled at positions Gly16, Leu17, Gly22, Ala23 and Gly31, while peptide T3-785 was selectively $^{15}\text{N}/^{13}\text{C}$ doubly at positions Gly15, Leu16, Ala17, Gly18 and Gly24 (Table 5.1). The NMR samples for all peptides were prepared in 10% $\text{D}_2\text{O}/90\% \text{H}_2\text{O}$ at pH ~ 3.1 with concentrations of ~ 6 mM.

5.2.2 NMR spectroscopy

NMR experiments were performed on a Varian Inova 500MHz spectrometer or a Varian INOVA 600 MHz spectrometer with a cold probe. ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) (Kay et al. 1992) and 3D ^{15}N edited NOESY-HSQC experiments (Fesik & Zuiderweg 1988; Marion et al. 1989; Messerle et al. 1989) with a mixing time of 50ms were carried out for assignments of NMR resonances at 20°C . 3D HNHA experiment was performed with an H-H coupling period of 25ms at 20°C to measure $^3J_{\text{HNH}\alpha}$ coupling constants, which provided information about phi angles (Vuister & Bax 1993). Relaxation R_1 , R_2 and heteronuclear NOE measurements (Fan et al. 1993; Palmer 1993; Farrow et al. 1994) were done at 20°C as described in section 2.2.8.1. All data were processed

using the FELIX 2004 software package (MSI, San Diego, CA) or NMRPipe (Delaglio et al. 1995), and analyzed with FELIX 2004 or NMRView (Johnson & Blevins 1994).

5.2.3 Hydrogen exchange experiments

5.2.3.1 Experimental procedures

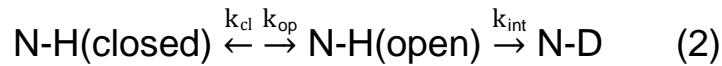
Hydrogen exchange experiments were performed at 10°C and pD 2.3-2.8 for the peptide sets (2.74 for T3-778, 2.34 for T3-778I-L, 2.56 for T3-778P-A, 2.53 for T3-778IT-PO and 2.41 for T3-785), where the pD is corrected for the glass electrode solvent isotope artifact (Glasoe & Long 1960). All samples were equilibrated in H₂O at 10°C for a minimum of 48 hours to assure the monomer:trimer equilibrium to be reached. The samples were then lyophilized, re-dissolved in 100% D₂O and quickly transferred to spectrometer which was equilibrated at 10°C. A series of HSQC spectra were acquired every 5 min (except every 10 min for T3-778) immediately after the sample was placed in the probe to allow the measurements of fast-exchanging amide protons, which is replaced by deuterium within several hours. Then the HSQC spectra were acquired every 40 min (64 t_1 increments and 2 scans per increment) to monitor the slow-exchanging amide protons. The hydrogen exchange rates k_{ex} were determined by a non-linear least squares fit using the equation $I(t)=(I_0-I_\infty)\exp(-k_{ex}t)$, where I_t , I_0 , and I_∞ are the intensities at time t , at time 0 and at final time point when the exchange is completed. The protection factor P is defined as $P=k_{int}/k_{ex}$, where k_{int} is the intrinsic exchange rate of amide protons in the monomer state at a specific pH and

temperature (Bai et al. 1993). The error of k_{int} was estimated by assuming an experimental uncertainty of $\pm 0.1^\circ\text{C}$ and ± 0.02 pH units. The error of the protection factor P is calculated as follows:

$$P_{\text{error}} = \left(\frac{k_{int}}{k_{ex}}\right) \text{error} = \left(\frac{k_{int}}{k_{ex}}\right) * \sqrt{\left(\frac{k_{int, \text{error}}}{k_{int}}\right)^2 + \left(\frac{k_{ex, \text{error}}}{k_{ex}}\right)^2} \quad (1)$$

5.2.3.2 Local stabilities derived from hydrogen exchange measurements

The hydrogen exchange process under native conditions is proposed to follow a two step model (Hvidt & Nielsen 1966; Englander & Krishna 2001; Huyghues-Despointes et al. 2001):



where k_{cl} and k_{op} are the rates constants for structural closing and opening, respectively, and k_{int} is the intrinsic rate constant at which the amide protons in the open form, or random coil state, exchanges with deuterium. There is an equilibrium between the open state and closed state for amide protons, while the exchange of the open form N-H to N-D is irreversible. Therefore, the measurable rate constant k_{ex} can be derived as follows:

$$k_{ex} = k_{op}k_{int} / (k_{op} + k_{cl} + k_{int}) \quad (3)$$

When $k_{cl} \gg k_{op}$ and $k_{cl} \gg k_{int}$ (the EX2 exchange conditions), k_{ex} can be defined by a simpler equation:

$$k_{ex} = k_{op}k_{int}/k_{cl} = (k_{op}/k_{cl}) k_{int} = K_{op} k_{int} \quad (4)$$

where K_{op} is the equilibrium constant for structural opening and can be related to the free energy for NH unfolding:

$$\Delta G_{HX} = -RT \ln K_{op} = -RT \ln (k_{ex}/k_{int}) = -RT \ln P \quad (5)$$

where R is the gas constant, T is the temperature and P is the protection factor defined as above. Therefore, ΔG_{HX} , which can represent the local conformational stabilities for each residue, can be calculated from the values of protection factor P derived from hydrogen exchange measurements.

5.3 Results

5.3.1 5-triplet model of the collagenase cleavage sites

The amino acid sequences around the natural collagenase cleavage sites in human types I-III collagen are examined to evaluate if they share any unique features (Figure 5.1a). The triplet containing the cleavage scissile bond is colored in red. The four chains of interstitial collagens show three conserved residues besides Gly: one Hydroxyproline and one Proline in the two triplets immediately preceding the cleavage sites and one Leu in the third triplet after the cleavage sites (Figure 5.1a). The $\alpha 1$ chains of type I and type II collagens share the most common sequences. They have almost the same 9 residues around the collagenase cleavage sites (GPQ**GI**AGQR for Type I and GPQ**GL**AGQR for Type II). The $\alpha 1$ chains of type I and type III collagens contain the same cleavage triplet GIA and the two triplets immediately following the cleavage sites don't have any imino acids. However, the $\alpha 2$ chain of type I collagen contain cleavage triplet GLL, which has a large residue Leu (rather than a small Ala) following the scissile bond as well as a Hydroxyproline in the following triplet.

The cleavage triplet GIA in the $\alpha 1$ chains of type I and type III collagens is 100% conserved across different species (Figure 5.2). The cleavage triplet GLA in the $\alpha 1$ chains of type II collagen has Leu 100% conserved, but 40% of the time Ala is substituted by another small residue Ser. However, the cleavage triplet GLL in the $\alpha 2$ chain of type I collagen has two occurrences of the scissile bond GL converted to GI or GV and it has the conserved Hyp in the preceding two triplets substituted by Ala 3 out of 7 times (Figure 5.2). The triplet GLL occurs only once in the types I-III collagens, while GIA and GLA each occurs 9 times. Compared with the $\alpha 1$ chains, the $\alpha 2$ chain of type I collagen also has the least imino acids (Figure 5.3). The less conserved GLL, and the presence of Hydroproline following the cleavage triplet despite the least content of imino acids in the $\alpha 2$ chain, suggest that the cleavage triplet GLL in the $\alpha 2$ chain of type I collagen may display a different pattern for collagenase recognition, compared with other cleavage triplets GIA/GLA in the $\alpha 1$ chains. A 5-triplet model of the collagenase cleavage sites could be built for the $\alpha 1$ chain of types I-III collagen (Figure 5.1b). The cleavage triplet GIA/GLA requires the presence of two preceding imino acid containing triplets as well as two following non-imino acid triplets.

(a)

	763	768	773	778	783	788
a1 (I)	GPA-GAO-GTO	GPQ	GIA	GQR-GVV	GLO	GQR-GER
a2 (I)	GTA-GPO-GTO	GPQ	GLL	GAO	GIL-GLO	GSR-GER
a1 (II)	GPS-GAE-GPO	GPQ	GLA	GQR-GIV	GLO	GQR-GER
a1 (III)	GAQ-GPO-GAO	GPL	GIA	GIT-GAR	GLA	GPO-GMO

(b)

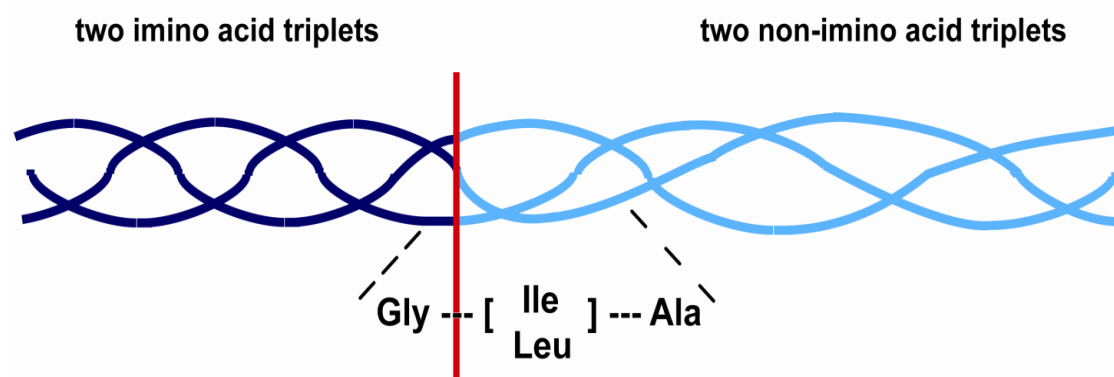


Figure 5.1 (a) Natural collagenase cleavage sites in human types I-III collagens. The triplet containing the cleavage scissile bond is colored in red. Conserved residues, including two imino acids in the two triplets immediately preceding the cleavage sites, are highlighted in the box. The two triplets immediately following the cleavage sites don't contain any imino acids, except one Hyp (colored in green) in the $\alpha 2$ chain of type I collagen. (b) 5-triplet model of the collagenase cleavage sites in the $\alpha 1$ chain of types I-III collagen. The cleavage triplet GIA/GLA requires the presence of two preceding imino acid triplets as well as two following non-imino acid triplets. The cleavage triplet GLL in the $\alpha 2$ chain of types I collagen shows a dissimilar pattern, which is not included in this model.

Collagen I (Alpha-1)

CO1A1_RAT	GPAGSPGTPGPQGIAGQRGVVGLPGQR
CO1A1_MOUSE	GPAGSPGTPGPQGIAGQRGVVGLPGQR
CO1A1_HUMAN	GPAGAPGTPGPQGIAGQRGVVGLPGQR
CO1A1_BOVIN	GPAGAPGTPGPQGIAGQRGVVGLPGQR
CO1A1_Canis	GPAGAPGTPGPQGIAGQRGVVGLPGQR
CO1A1_CHICK	GPIGAPGTPGPQGIAGQRGVVGLPGQR
CO1A1_CYNPY	GPAGAPGIPGPQGIAGQRGVVGLPGQR

Collagen I (Alpha-2)

CO1A2_HUMAN	GTAGPPGTPGPQGLLGAPGILGLPGSR
CO1A2_CANFA	GTAGPPGTPGPQGLLGAPGILGLPGSR
CO1A2_BOVIN	GTAGPPGTPGPQGLLGAPGFLGLPGSR
CO1A2_MOUSE	GTAGAPGTAGPQGLLGAPGILGLPGSR
CO1A2_RAT	GTTGPPGTAGPQGLLGAPGILGLPGSR
CO1A2_CHICK	GAAGPPGTPGPQGILGAPGILGLPGSR
CO1A2_RANCA	GPAGPPGAAGPSGVLGARGILGLPGTR

Collagen II (Alpha-1)

CO2A1_RAT	GPSGSDGPPGPQGLAGQRGIVGLPGQR
CO2A1_MOUSE	GPSGLDGPPGPQGLAGQRGIVGLPGQR
CO2A1_HUMAN	GPSGAEGPPGPQGLAGQRGIVGLPGQR
CO2A1_XENTR	GPSGPDGPPGPQGLSGQRGIVGLPGQR
CO2A1_XENLA	GPSGPDGPPGPQGLSGQRGIVGLPGQR

Collagen III (Alpha-1)

CO3A1_MOUSE	GAQGPPGSPGPLGIAGLTGARGLAGPP
CO3A1_RAT	GAQGPPGSPGPLGIAGLTGARGLAGPP
CO3A1_HUMAN	GAQGPPGAPGPLGIAGITGARGLAGPP
CO3A1_BOVIN	GPQGPPGAPGPLGIAGLTGARGLAGPP

Figure 5.2 Alignments of the sequences at the collagenase cleavage sites in types I-III collagens across different species. Cleavage triplet is colored in red. Unconserved residues are colored in blue.

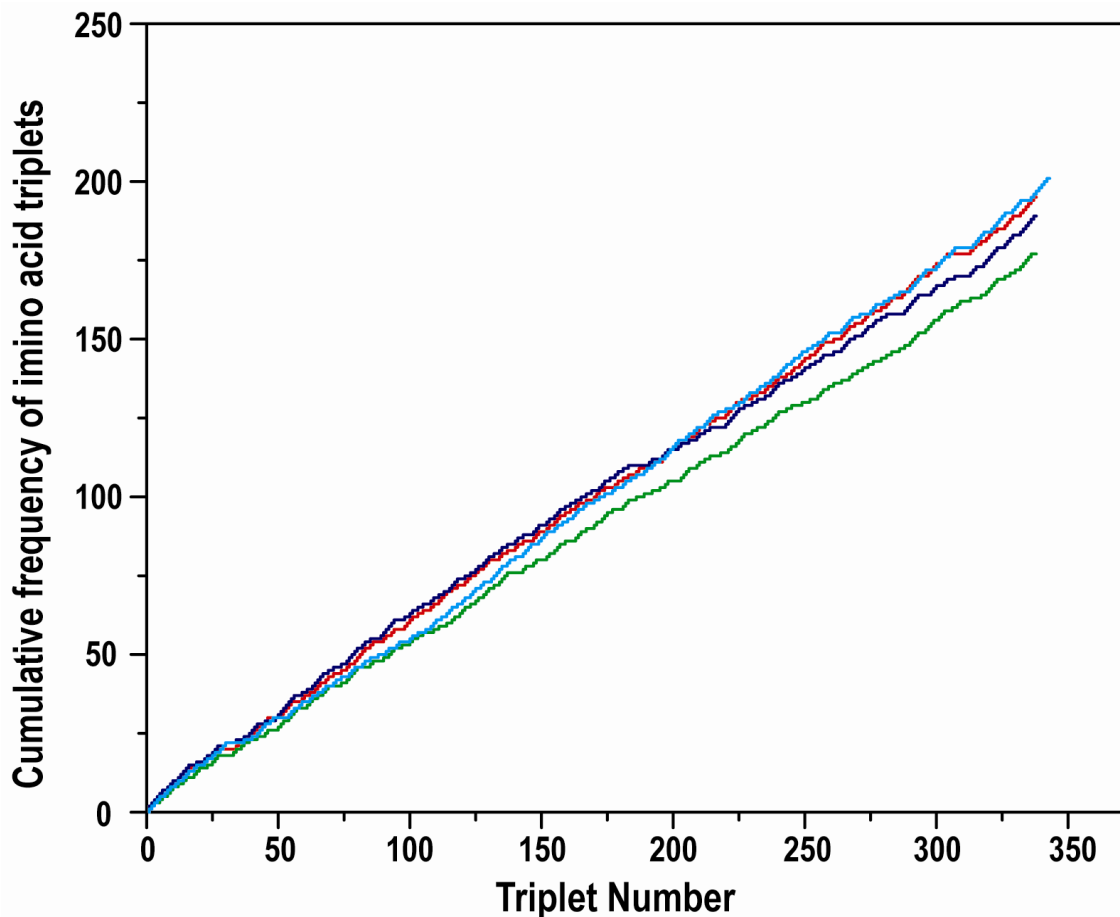


Figure 5.3 Cumulative frequencies of imino acid triplets in CO1A1 (red), CO1A2 (green), CO2A1 (blue) and CO3A1 (cyan). CO1A1, CO1A2, CO2A1 and CO3A1 have totally 195, 177, 189 and 201 imino acid triplets, respectively.

5.3.2 Distribution analysis of imino acid triplets and non-imino acid triplets

The 5-triplet model highlights the importance of imino acids, as the model can be restated as the cleavage sites characteristic of two imino acid containing triplets followed by three non-imino acid triplets. The distribution of imino acid triplets and non-imino acid triplets in interstitial collagens was thus analyzed, to see if any other sites besides the natural collagenase cleavage sites display the

unique features claimed in the model. The cumulative frequency plot of the imino acid containing triplets was first done, and its almost linear shape indicated that the imino acid triplets are pretty evenly distributed across collagen and there is no obvious region extremely rich or poor in imino acids (Figure 5.3).

Further analysis was done to identify any pattern of the distribution of imino acid triplets and non-imino acid triplets (Figure 5.4). Collagen sequences were categorized into stretches of m non-imino acid triplets preceded by imino acid triplets ($m+n$ triplet pattern). Type III collagen contain 87 such stretches (excluding 8 terminal triplets): 30 stretches contain only one or two non-imino acid triplets preceded by one imino acid triplet while only 7 stretches contain at least 3 non-imino acid triplets preceded by at least two imino acid triplets (Figure 5.4). Similar results were obtained for other collagens: the $\alpha 1$ chains of types I and type II collagen only have 9 and 5 stretches containing at least 3 non-imino acid triplets preceded by at least two imino acid triplets, including one stretch containing the natural cleavage sites. The frequency analysis of the $m+n$ triplet pattern indicated that imino acid distribution could limit the potential cleavage sites to only a few candidates. If charge is also considered, as no charged residue is required in the first non-imino acid triplet, only two stretches can be found in the $\alpha 1$ chains of types I-III collagens, while one of them is the natural cleavage site.

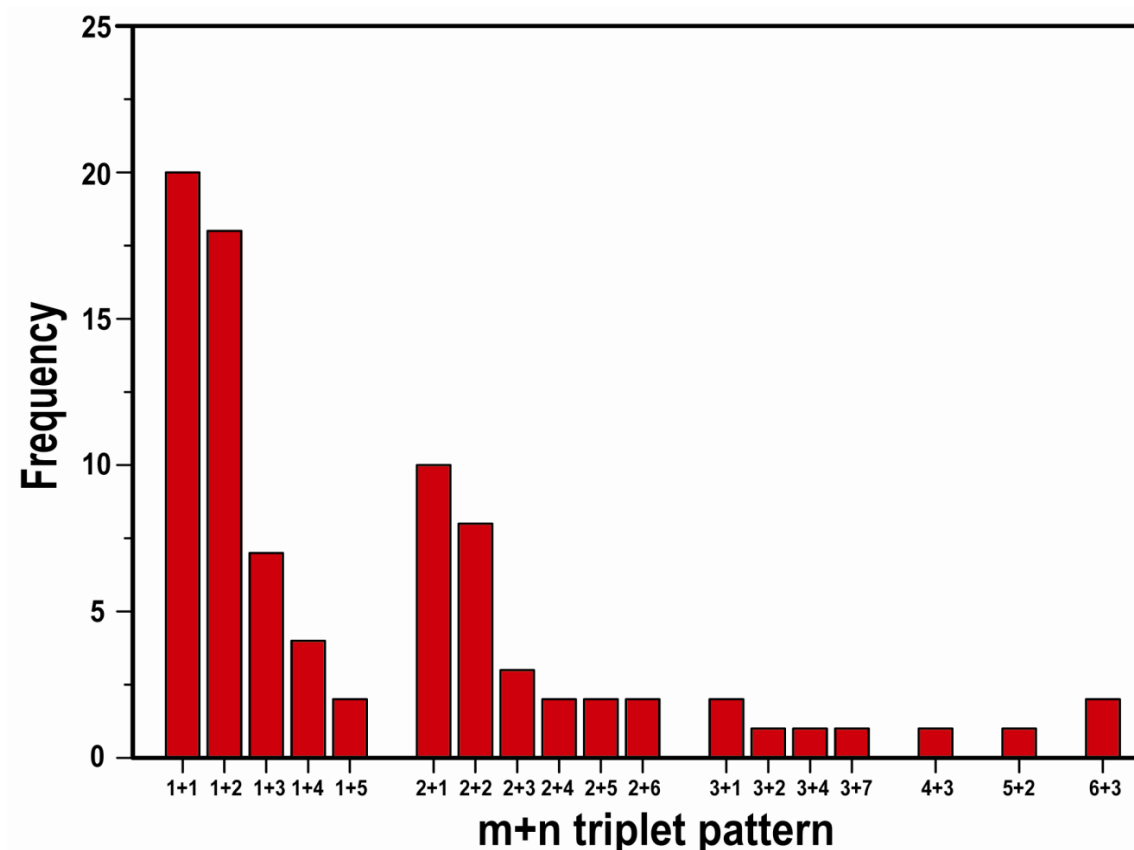


Figure 5.4 Frequency of m+n triplet pattern in human type III collagen. Collagen sequences can be categorized into stretches of m non-imino acid triplets preceded by imino acid triplets (m+n triplet pattern). Type III collagen contain 87 such stretches (excluding 8 terminal triplets), while only 9 stretches contain at least 3 non-imino acid triplets and could be potential collagenase cleavage sites.

5.3.3 Analysis of the Ile/ Leu in the X position

There are 30 Ile and 66 Leu in the $\alpha 1$ chains of types I-III collagens. 67% of the Ile and 94% of the Leu are located in the X position, forming the Gly-Ile or Gly-Leu bonds, which could be potentially cleaved by MMPs. The Ile/Leu in the X-position were categorized according to the presence or absence of surrounding

imino acids, to see if the Ile/Leu at the cleavage site could distinguish itself from those at non-cleavage sites (Figure 5.5). Four types of Ile/Leu in the X-position could be identified: Ile/Leu immediately followed by a Hyp in the Y position (red); Ile/Leu followed by two triplets at least one of which contained imino-acids (green); Ile/Leu preceded by two triplets at least one of which contained no imino-acids (blue); Ile/Leu with two preceding imino-acid triplets and no imino acids in the Y position or in the following two triplets (cyan) (Figure 5.5). Most Ile/Leu have a Hydroxyproline in the Y position or some imino acids in the following two triplets. Only one Ile in the $\alpha 1$ chains of type I and type III collagens and one Leu in type II collagen follow the last pattern (cyan) and they are the real collagenase cleavage sites. In other words, the 5-triplet model of GIA/GLA with two preceding imino acid triplets and two following non-imino acid triplets could correctly predict the unique collagenase cleavage sites in all the $\alpha 1$ chains of types I-III collagens.

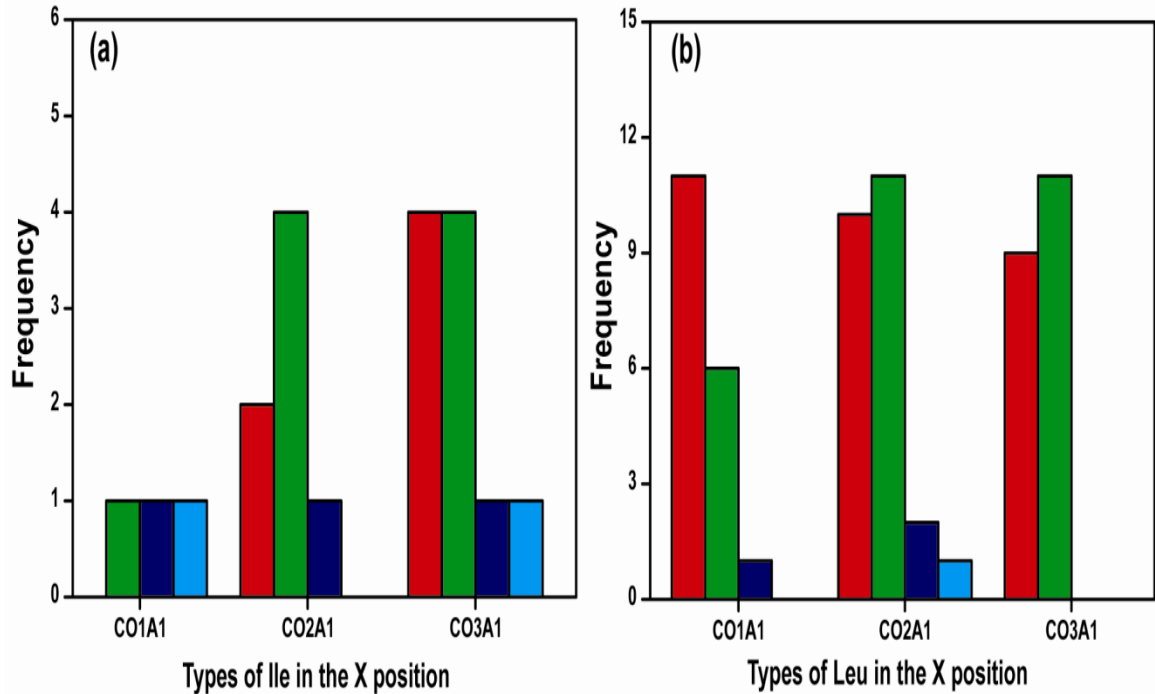


Figure 5.5 Analysis of Ile/Leu in the X-position of the $\alpha 1$ chain of types I-III collagen. Ile/Leu in the X-position can be classified into four categories by the distribution of imino acids: Ile/Leu immediately followed by a Hyp in the Y position (red); Ile/Leu followed by two triplets at least one of which containing imino-acids (green); Ile/Leu preceded by two triplets at least one of which containing no imino-acids (blue); other Ile/Leu with two preceding imino-acid triplets and no imino acids in the Y position or in the following two triplets (cyan). There are only one Ile in the $\alpha 1$ chains of type I and type III collagens and one Leu in type II collagen following the last pattern (cyan) and they are the real collagenase cleavage sites.

5.3.4 Peptide Design and Activities of the model peptides

As a unique 5-triplet pattern has been identified for the collagenase cleavage sites, we would like to design model peptides to understand what distinguishes this pattern from others in terms of conformation and dynamics. It is noted that in the collagenase cleavage site region in type III collagen (772-801: GPOGAOGPL**GIA**GITGAR**GLA**GPOGMOGPR), only two triplets away from the natural cleavage triplet GIA, there is another potential cleavage triplet GLA that cannot be cleaved. Therefore, two peptides T3-778 and T3-785 were designed to model the natural cleavage site and potential but non-cleavable site (Table 5.1). Peptide T3-778 incorporates 4 triplets GPLGIAGITGAR from the cleavage site, while peptide T3-785 includes 4 triplets GITGARGLAGPO from the non-cleavable site. The two peptides share the same two triplets GITGAR and they have similarities in the other triplet GPL vs GPO. However, the two peptides show significant differences in the relative position of the imino acid triplet to the cleavage triplet GIA/GLA: GPL is located at the N-terminal to GIA while GPO is at the C-terminal to GLA.

For both peptides, GPO triplets are added at the two ends to enhance the formation of the triple helix. According to our 5-triplet model, incorporating 5 triplets in our model peptides could be a better choice for MMP recognition. However, triple helical peptides behave like big molecules and a longer peptide would decrease the NMR resolution (Li et al. 1993). Therefore, 4 triplets were included in our model peptides. It has to be pointed out, though, that the fifth triplet not

included was GAO for T3-778 and GMO for T3-778, which could be considered to be replaced by a similar imino acid triplet GPO already present in our peptides.

As stated in the 5-triplet model, for the site to be cleaved in the $\alpha 1$ chains of types I-III collagens, it requires a cleavage triplet GIA/GLA with two preceding imino acid triplets and two following non-imino acid triplets (Figure 5.1b). To investigate if there is any difference between GIA and GLA, a peptide T3-778I-L with the Ile mutated to Leu was designed (Table 5.1). To investigate the effect of the preceding imino acid triplets, a peptide T3-778P-A with the conserved Pro at the 3 subsite mutated to Ala was designed. To investigate the effect of the following non-imino acid triplets, a peptide T3-778IT-PO with the Ile-Thr at the 4' and 5' subsites mutated to Pro-Hyp was designed. The sequence GPLGIAGPOGAR modeled by the peptide T3-778IT-PO, in fact, can also be found at a site just 20 residues away from the natural cleavage site in type I collagen. The series of T3-778 peptides contain five selectively ^{15}N labeled residues: G16, I17, G22, A23 and G31 (L17 instead of I17 for peptide T3-778I-L). The peptide T3-785 contains five selectively $^{15}\text{N}/^{13}\text{C}$ doubly labeled residues: G15, L16, A17, G18 and G24. All the peptides have a residue Tyr added at the C-terminal for concentration determination (Table 5.1).

The kinetics for all the peptides has been measured in Prof. Gregg Fields' lab (Table 5.1) (Personal communication, Janelle Lauer-Fields). The $k_{\text{cat}}/k_{\text{M}}$ values for peptides T3-778 and T3-778I-L are the largest, indicating that both peptides

modeling the natural cleavage sites are most active and GLA does not make a big difference from GIA. Compared with T3-778, the k_{cat}/k_M values for peptides T3-778P-A and T3-778IT-PO were decreased by 63% and 34%, indicating both the P-A and IT-PO mutations leading to decreased activities. The peptide T3-785, which models the potential but non-cleavable site, was totally inactive.

Table 5.1 Triple helical peptides modeling the collagenase cleavage sites and their activities

Peptide	Sequence*	k_{cat}/k_M ($M^{-1}sec^{-1}$)
T3-778	(GPO) ₄ -GPL- <u>G</u> IA-GIT- <u>G</u> AR-(<u>G</u> PO) ₅ GY	1119 ^a , 2,341 ^b
T3-778I-L	(GPO) ₄ -GPL- <u>G</u> <u>L</u> A-GIT- <u>G</u> AR-(<u>G</u> PO) ₅ GY	1533 ^a
T3-778P-A	(GPO) ₄ -G <u>A</u> L- <u>G</u> IA-GIT- <u>G</u> AR-(<u>G</u> PO) ₅ GY	414 ^a
T3-778IT-PO	(GPO) ₄ -GPL- <u>G</u> IA-G <u>P</u> O- <u>G</u> AR-(<u>G</u> PO) ₅ GY	1550 ^b
T3-785	PO(GPO) ₂ -GIT-GAR- <u>G</u> LA- <u>G</u> PO-(<u>G</u> PO) ₃ GY	0 ^a

* Residues that are ¹⁵N labeled are underlined; each peptide has five labeled residues including a Gly at the central position in its (GPO)₅/(GPO)₃ C-terminal end; mutation residues in peptides T3-778I-L, T3-778P-A and T3-778IT-TO are colored in red. a. The kinetics measurements are done at 20°C (Personal communication, Janelle Lauer-Fields). b. The kinetics measurements are done at 37°C (Personal communication, Janelle Lauer-Fields).

5.3.5 Reverse activities of monomer peptides

Peptides T3-778 and T3-778IT-PO models the natural cleavage site and an IT-PO mutation around the cleavage site in a triple helical environment, as GPO triplets are added to force the trimer formation. Two monomer peptides M3-778 and M3-778IT-PO, which model the same sequences as T3-778 and T3-778IT-PO but don't include the GPO triplets, were designed to investigate the activities of those sequences in the monomer state (Table 5.2). Peptides T3-778 and M3-778 share the same sequences while T3-778IT-PO and M3-778IT-PO share the same sequences. Therefore, it is interesting to compare their relative activities in the trimer state versus the monomer state.

The kinetics measurement shows that M3-778IT-PO has much larger $k_{\text{cat}}/k_{\text{M}}$ value than M3-778, indicating that M3-778IT-PO is far more active than M3-778 (Table 5.2) (Personal communication, Janelle Lauer-Fields). However, T3-778 is more active than T3-778IT-PO (Table 5.1). As single chain activity is reversed in the triple helix, it suggests that sequence alone is not the critical component but that conformation and dynamics of the trimer are critical in modulating activity.

Table 5.2 Monomer peptides incorporating the sequences at the collagenase cleavage sites and their activities

Peptide	Sequence	$k_{\text{cat}}/k_{\text{M}}$ ($\text{M}^{-1}\text{sec}^{-1}$)*
M3-778	Mca-GPL-GIA-GIT-GA-Lys(Dnp)	5,637.2 \pm 973.1
M3-778IT-PO	Mca-GPL-GIA-GPO-GA-Lys(Dnp)	61,717.5 \pm 4,372.5

*Personal communication, Janelle Lauer-Fields

5.3.6 NMR HSQC spectra of the model peptides

NMR HSQC spectra of the triple helical model peptides (Table 5.1) show that all labeled residues have trimer as well as monomer peaks, indicating that the peptides contain typical triple helix conformation in equilibrium with monomer conformation (Figure 5.6, 5.7). For peptide T3-778, residues G16, I17, G22 and A23 have three well-dispersed trimer peaks as a result of the three non-equivalent chains, while residue G31 shows only a single resonance due to the repeating GPO environment. One unique feature of the peptide is that the Ile in the first chain ($^1\text{T}^1\text{I17}$) of the triple helix shows distinct chemical shifts from the Ile in the other two chains (Figure 5.6). However, for peptide T3-785, the chemical shifts of the Leu in the first chain ($^1\text{T}^1\text{L16}$) become closer to the Leu in the other two chains (Figure 5.6).

The mutated T3-778 series peptides show similar HSQC spectra as T3-778 (Figure 5.7). All the Ile/Leu in the first chain shows distinct chemical shifts from the Ile/Leu in the other two chains (Figure 5.7). Particularly, peptide T3-778I-L shows

almost the same spectra as T3-778; the P-A mutation mainly changes the chemical shifts of Gly16, while the IT-PO mutation largely affects Gly22 and Ile17 (particularly the Ile in the third chain (^3T I17)) (Figure 5.7). It indicates that the P-A and IT-PO mutations provide a different chemical environment for G16 or I17 at the collagenase cleavage site, which could play a significant role in the specificity of collagen recognition by MMP.

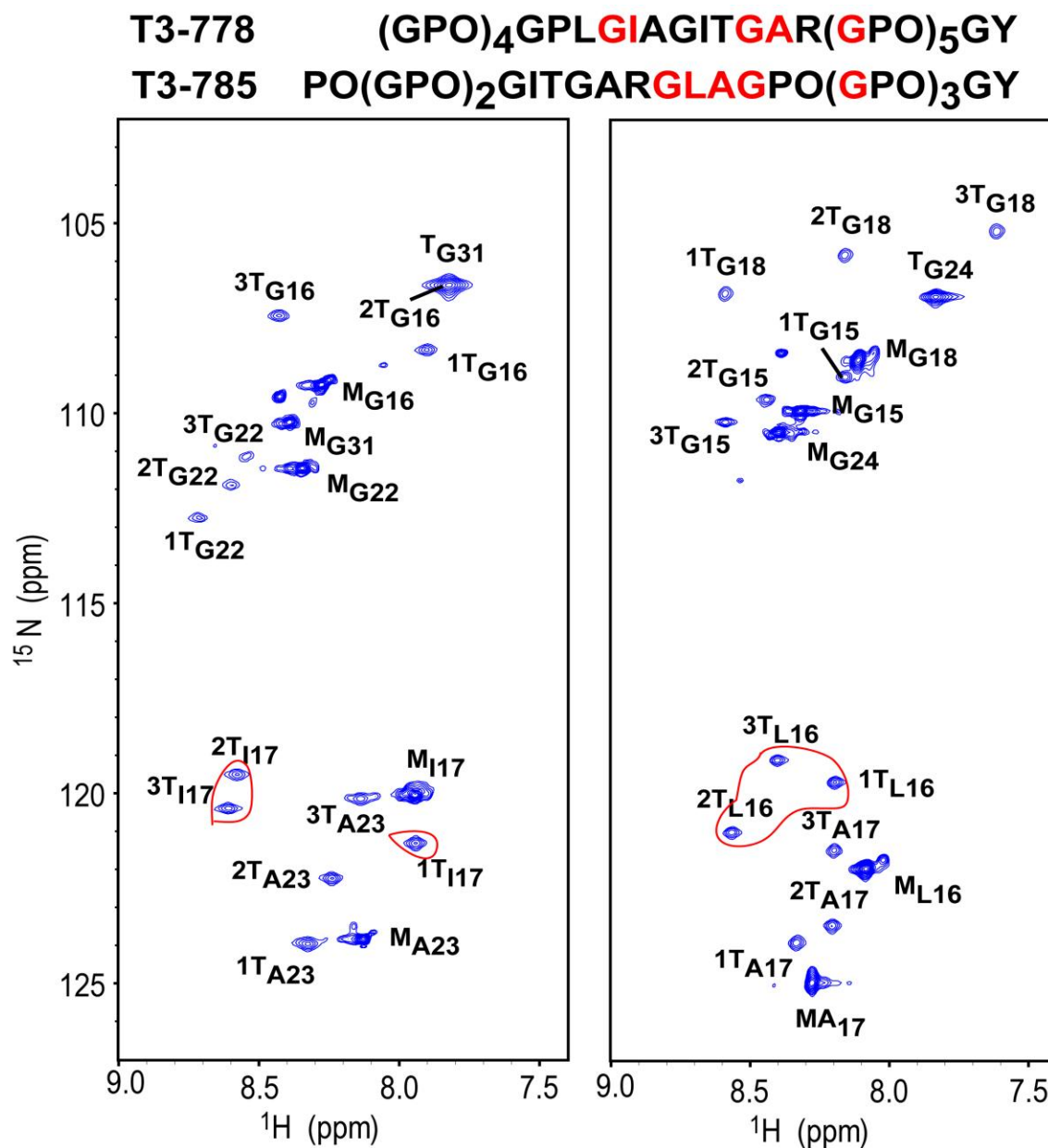


Figure 5.6 HSQC spectra of active peptide T3-778 (left) and inactive peptide T3-785 (right). Labeled residues are colored red in the above sequences. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T. The superscripted number 1, 2 and 3 correspond to the leading, middle and lagging chains, respectively.

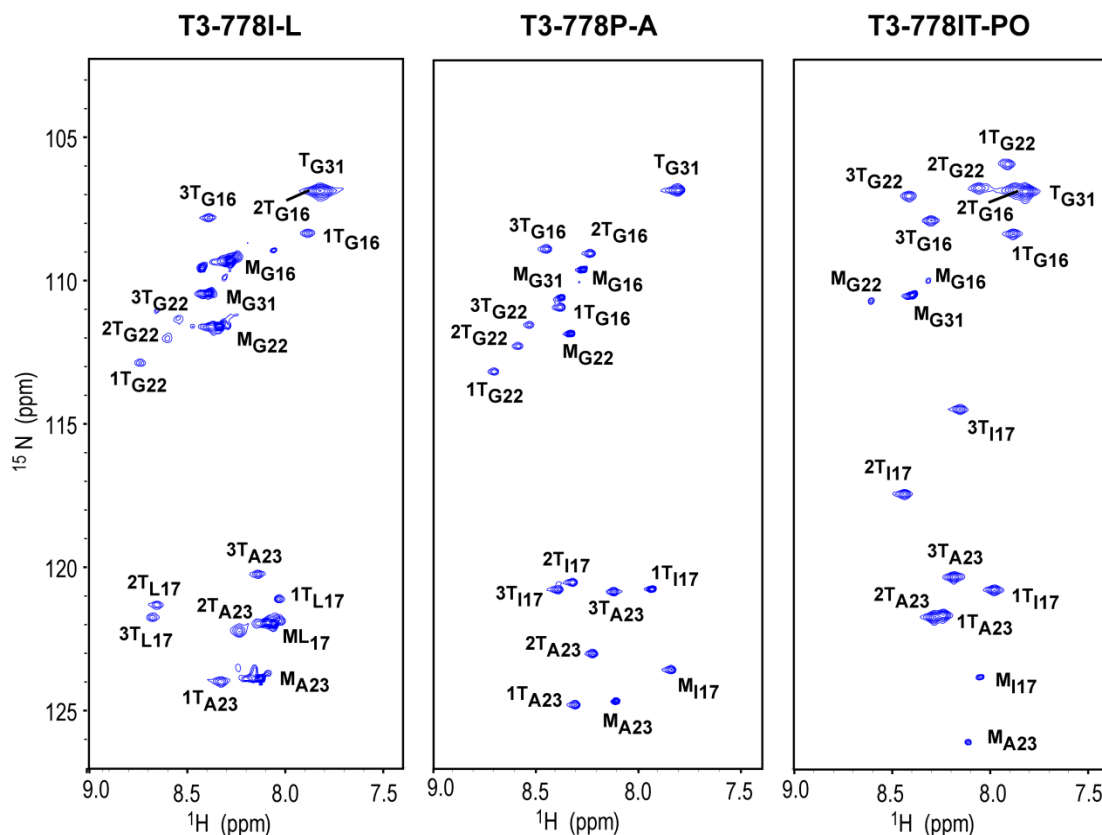


Figure 5.7 HSQC spectra of peptide T3-778I-L, T3-778P-A and T3-778IT-PO. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T. The superscripted number 1, 2 and 3 correspond to the leading, middle and lagging chains, respectively.

5.3.7 Conformation of the model peptides

J-coupling constants provide a good measure of backbone conformation, while residues in the triple helical conformation containing phi angles from -55 to -75 degrees have corresponding J coupling values from 4 to 6 Hz. All residues in the triple helix of the model peptides have J-coupling values between 4-6 Hz except a single Ile in T3-778 ($^1\text{T}_{117}$ with a value of 6.74 Hz), which has also shown

a distinct chemical shift (Figure 5.8). Particularly, T3-785 has very uniform J-coupling values of 5 Hz for the Leu16 in three different chains. It is also noted that the Ile17 in peptides T3-778P-A and T3-778IT-PO show a similar pattern in J coupling values: $^1\text{T}17$ and $^3\text{T}17$ have similarly larger values than $^2\text{T}17$, suggesting that the P-A and IT-PO mutations have similar effects on the conformation of Ile at the cleavage site.

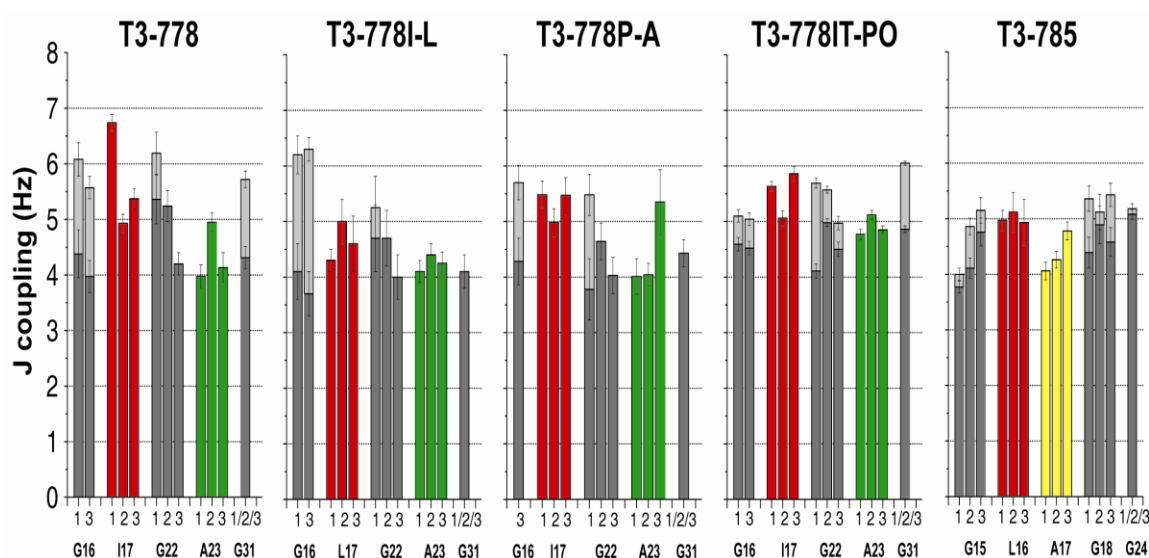


Figure 5.8 J-coupling values of model peptides. Gly are colored in gray, Ile/Leu in the X position are colored in red, while Ala are colored green in the X position and yellow in the Y position. All labeled residues in all peptides except $^1\text{I}17$ in T3-778 have J-coupling values between 4-6 Hz, which are typical for triple helical conformation.

5.3.8 Dynamics of the model peptides on the slow second time scale

Hydrogen exchange experiments were performed on all the model peptides to explore the protection of amide protons. The high protection factors of all the Gly residues indicate that the Gly are well protected due to the existence of hydrogen bonding (Table 5.3). Residue-specific local stabilities calculated from the protection factors are shown in Figure 5.9. All the Gly of all the peptides show similarly high stabilities; all the Ala in the X position show uniformly decreased stabilities. However, the Ile/Leu show significantly decreased and quite variable stabilities in the T3-778 series peptides, but not in T3-785. The Leu in T3-785 are uniformly very stable, indicating the existence of indirect hydrogen bonding. However, other Ile/Leu are very unstable, indicating the loss of hydrogen bonding and change of hydration structure.

For all the T3-778 series peptides, the Ile/Leu in the first chain is less stable than the Ile/Leu in the third chain, while the Ile/Leu in the third chain is less stable than the Ile/Leu in the second chain. However, the difference in stabilities of Ile/leu in different chains varies greatly from peptide to peptide. For T3-778 and T3-778I-L, the Ile/Leu in the first chain is much less stable than the Ile/Leu in the other two chains, which have similar stabilities. For T3-778P-A and T3-778IT-PO, the Ile in the first chain has similar stabilities as the Ile in the third chain. However, the P-A mutation makes the Ile in the first chain more stable, while the IT-PO mutation makes the Ile in the third chain less stable.

Table 5.3 Hydrogen Exchange data of model peptides.

Residues	T3-778	T3-778P-A	T3-778IT-PO	T3-778I-L	Residues	T3-785
1G16	1598+/- 824	584+/-70.1	2709+/-750	2254+/-345	1G15	525.2+/-39.6
3G16	753+/-144	647+/-88.7	857+/-110	563+/-36.5	3G15	541+/-56.2
1I17	1.7+/-0.2	3.2+/-0.2	1.1+/-0.2	1.2+/-0.5*	1L16	98.3+/-5.9
2I17	10.9+/-0.9	6.1+/-0.4	10.5+/-1.2	5+/-0.5*	2L16	213.4+/-13.9
3I17	6.2+/-0.6	3.8+/-0.2	1.5+/-0.3	3.2+/-0.3*	3L16	236.5+/-15.9
1G22	658+/-161	2072+/-1631	388+/-27.9	1170+/-137	1A17	27+/-2
2G22	n/a	856+/-425	847+/-109	4079+/-2962	2A17	14+/-1.4
3G22	622+/-293	636+/-236	1026+/-204	n/a	3A17	14.7+/-1.7
1A23	9.5+/-1	8.9+/-0.7	17.9+/-2.5	12.8+/-1.4	1G18	649+/-65.1
2A23	10+/-1.3	11.8+/-1.0	17.9+/-2.5	7.8+/-0.7	2G18	868+/-100.3
3A23	18.2+/-2.7	9.7+/- 0.6	14.1+/-2.1	9+/-0.9	3G18	612+/-41.8
G31	651+/-57.7	303+/-17.7	1185+/-80	1068+/-67	G24	325+/-16.2

* ¹⁵N labeled residues are colored in different colors (gray for Gly and Ala in the Y position, red for Ile/Leu and blue for Ala in the X position); Hydrogen exchange data for the labeled residues in T3-778 series peptides (Left) and in inactive control peptide T3-785 (Right); All Gly show high protection factors. Leu16 in T3-785 also show high protection factors, whereas Ile17 in T3-778 series peptides all show low protection factors, particularly the Ile17 in the first chain. Compared with Ile17, Ala23 in the X position shows relatively higher protection factors.

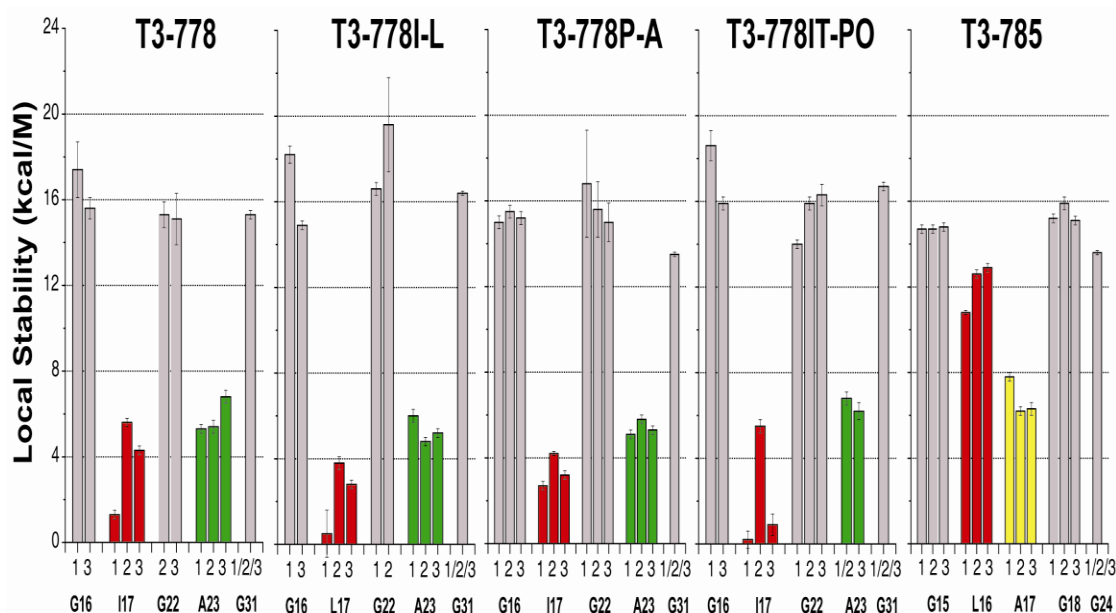


Figure 5.9 Local stabilities of model peptides calculated from protection factors.

Gly are colored in gray, Ile/Leu in the X position are colored in red, while Ala are colored green in the X position and yellow in the Y position. All the Gly of all the peptides show similarly high stabilities (~16kcal/mol); all the Ala in the X position show uniformly decreased stabilities (~6kcal/mol). However, the Ile/Leu show significantly decreased and quite variable stabilities in T3-778 series peptides, but not in T3-785.

5.3.9 Dynamics of the model peptides on the fast pico-second time scale

^{15}N relaxation R_1 , R_2 and heteronuclear NOE measurements provide useful information on the fast, picosecond timescale motion such as bond vibration and sidechain rotation (Palmer et al. 2001). The flexibility of peptide T3-778 was estimated by measuring the ^1H - ^{15}N NOE (Figure 4.9). All labeled residues have uniformly large NOE values around 0.6-0.7, indicating very small amplitude internal mobility on the fast, pico-second time scale. Other peptides display similar

patterns of the ^1H - ^{15}N NOE values for labeled residues, indicating that all the peptides are pretty rigid on this fast timescale. No difference in the flexibility is observed for the cleavage sites and other sites for all the model peptides, indicating that fast-time scale motions are probably not playing a critical role in collagen recognition by MMPs.

Little variations in the R_1 and R_2 values were also observed within each model peptide, indicating similar flexibilities for all residues. The R_2/R_1 values are pretty uniform for the residues in one peptide; however, different peptides show significantly different R_2/R_1 values (Figure 5.11). Peptide T3-778 shows the smallest R_2/R_1 value of ~ 11 , while T3-778I-L shows the largest value of ~ 16 . The R_2/R_1 values depend on the shape of a molecule, or the length and width of triple helical peptides if assuming a cylinder model (Fan et al. 1993). The four T3-778 series peptides have the same lengths; therefore, the differences in R_2/R_1 values could result from altered widths of hydration layer or aggregation. The R_2 values are very sensitive to aggregation; therefore, the R_2 measurements were performed for T3-778I-L, which shows the largest R_2/R_1 values (Figure 5.12). As peptide T3-778I-L becomes more diluted, the R_2 values of all labeled residues are reduced, indicating the presence of aggregation. At a concentration of 1.5mM, peptide T3-778I-L shows R_2 values of ~ 18 , comparable to the R_2 values of T3-778. Reduced R_2 values were also obtained for more diluted T3-778IT-PO, indicating that aggregation rather than altered hydration layer leads to variable R_2/R_1 values.

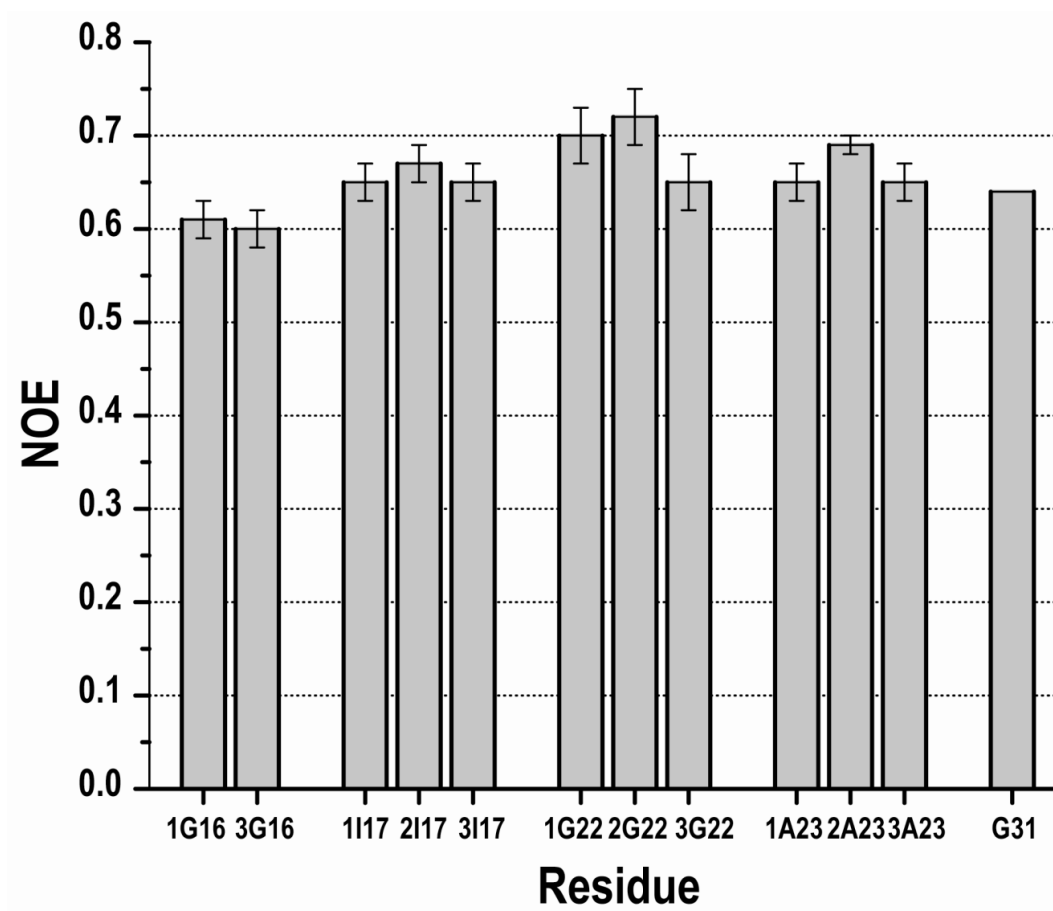


Figure 5.10 Heteronuclear ^1H - ^{15}N NOE values of heterotrimer peptide T3-778. Numbers 1, 2 and 3 indicate the leading, middle and lagging chains, respectively. All labeled residues have uniformly large NOE values, indicating small amplitude internal mobility on the fast, pico-second time scale.

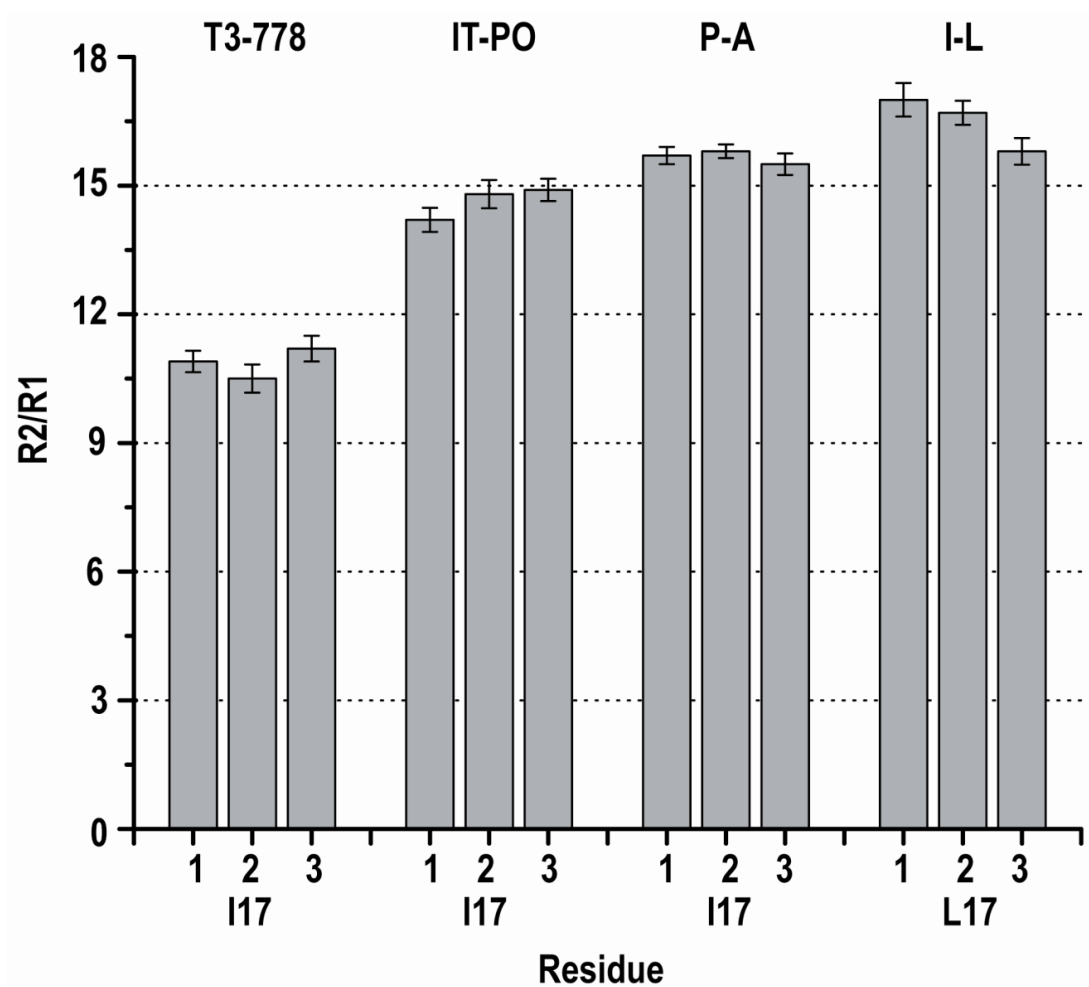


Figure 5.11 R_2/R_1 values of labeled Ile17/Leu17 at the cleavage sites in T3-778 series peptides. Numbers 1, 2 and 3 indicate the leading, middle and lagging chains, respectively. The residues have uniform R_2/R_1 values in the same peptide, while different peptides have various R_2/R_1 values, with the smallest value for T3-778.

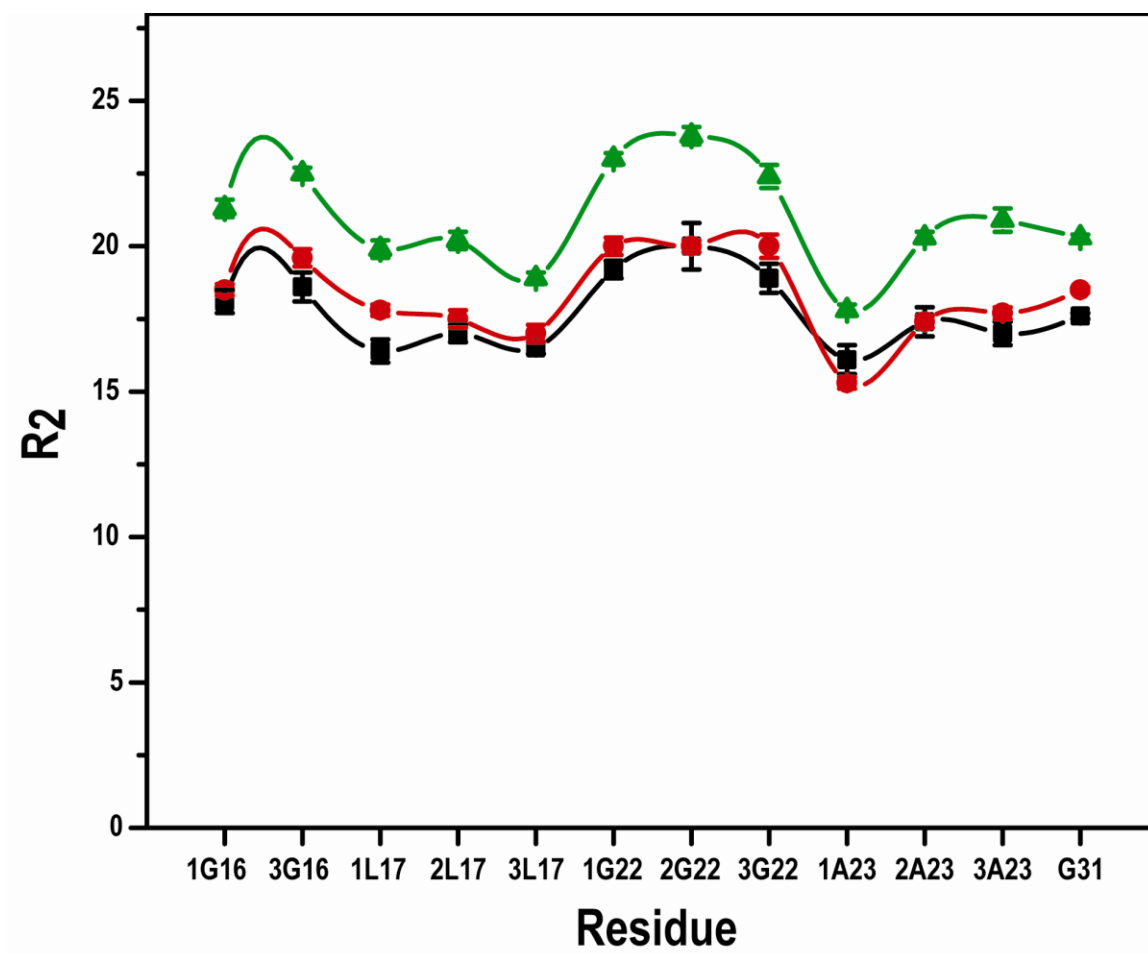


Figure 5.12 R_2 values of labeled residues in T3-778I-L at different concentrations (green for 6mM, red for 3mM and black for 1.5mM). Numbers 1, 2 and 3 indicate the leading, middle and lagging chains, respectively. As peptide T3-778I-L becomes more diluted, the R_2 values of all labeled residues are reduced.

5.4 Discussion

A key question to understand the specific recognition of collagen by MMPs is to identify the crucial residues important for recognition in collagen and investigate what special features those residues display. Fields proposed a model that 25 residues in the cleavage site region (P12-P13') dictate collagenase behavior (Fields 1991). Mutation studies indicated that at least 12 residues (P4-P8') affect the collagenolytic specificity (Fields et al. 1987; Nagase & Fields 1996). Triple helical model peptides including five triplets (residues P6-P9') from the collagenase cleavage sites show that those peptides could be cleaved not only specifically, but also selectively by different MMPs (Lauer-Fields et al. 2000; Lauer-Fields et al. 2002). Our sequence analysis, both locally on the natural cleavage sites and globally across the whole collagen, identified a five triplet model (residues P6-P9') for the $\alpha 1$ chains of types I-III collagens. The model suggests the cleavage sites require a cleavage triplet GIA/GLA with two preceding imino acid containing triplets and two following non-imino acid triplets. Among ~80 potential cleavage Gly-Ile/Leu sites in all the $\alpha 1$ chains of types I-III collagens, this model could 100% correctly predict the real collagenase cleavage sites. Sequence analysis also showed that the $\alpha 2$ chain in type I collagen is different from the $\alpha 1$ chains in types I-III collagen in terms of imino acid distribution as well as the conservation of cleavage triplet GLL, suggesting that the $\alpha 2$ chain may have a different pattern of recognition compared with the $\alpha 1$ chains.

The differences in activities of native collagen and denatured collagen suggests that collagenolytic activities must be related to the unique triple helix of collagen (Welgus et al. 1982), and the variations in sequence may cause small variations in structure and flexibility of collagen (Fan et al. 1993; Kramer et al. 1999), which are probably crucial to understand the specificity of collagen recognition by MMPs. We designed both triple helical peptides (T3-778 and T3-778IT-PO) and monomer peptides (M3-778 and M3-778IT-PO), which shared the same natural and mutated sequences. The sequences show totally different relative activities: T-3778>T3-778IT-PO, while M3-778<M3-778IT-PO. The reverse order of single chain activity versus triple helix activity suggests that conformation and dynamics are the critical components in modulating activity for triple helical peptides.

However, little information is available about the conformation and dynamics at the collagenase cleavage sites. The X-ray structure of a collagen-like peptide suggested that the imino acid rich region has different helical symmetry pattern with the imino poor region (Kramer et al. 1999). NMR studies of an unlabeled heteromeric triple helical peptide containing collagen I cleavage site indicates that the cleavage site region has less ordered structure than the ends with G-P-O repeating triplets (Fiori et al. 2002). Here, NMR studies of selectively labeled peptides present us opportunities to obtain residue-specific information about conformation and dynamics and evaluate if isoleucine in the cleavage triplet has special features compared with other residues.

Two triple helical peptides T3-778 and T3-785, one modeling the natural cleavage site and the other one modeling a potential but noncleavable site just 2-triplets away in type III collagen, show huge differences in activities as well as conformation and dynamics. T3-778 is active, while T3-785 is totally inactive. One Ile in the first chain of T3-778 shows distinct chemical shift and a larger J-coupling value than others. Considering that residues in triple helical conformation show small variations in J-coupling values, this higher value for the single Ile may be indicative of a distorted conformation. All the Ile in T3-778 are unstable, but this particular Ile is extremely unstable. However, all the Leu in T3-785 are very stable. These data may indicate that the hydrogen bonding of Ile in T3-778 has been changed. It is likely that the indirect hydrogen bonding is no longer existent, which makes the Isoleucines, particularly the Ile in the first chain, more exposed to solvents and enzyme attack. In all, in the active peptide only, the Ile residue in one chain is very different from the other 2 chains: distinct chemical shift, a larger J-coupling value and extremely unprotected.

The differences in the two peptides T3-778 and T3-785 may be caused by the different residue identity Ile vs Leu. Therefore, an I-L mutated peptide T3-778I-L is compared with T3-778. T3-778I-L is comparably active as T3-778 and the Leu in T3-778I-L shows a similar pattern in distinct chemical shift and local stabilities as the Ile in T3-778, indicating the difference in T3-778 and T3-785 may not result from the identity difference of Ile and Leu. It is also suggested that different stabilities of peptides may cause different collagenolytic activities, and a

less stable peptide would show a higher activity (Minond et al. 2004). However, T3-785 here is less stable than T3-778 and it is inactive, suggesting the activities of the peptides do not come from stabilities, but from the difference in conformation and dynamics we have observed.

Two mechanisms have been proposed for collagen to be cleaved by MMPs. The first mechanism proposed that collagen is intact and the binding of MMPs unwinds collagen (Chung et al. 2004). The second mechanism suggests that collagen is passively activated and it has some unwound state which can be recognized by MMPs (Nerenberg et al. 2008). Studies of Type I collagen indicated that an inactive MMP-1(E200A) mutant binds and unwinds the triple-helical collagen first, which can then be cleaved by the catalytic domain of MMP-1 (Chung et al. 2004). However, computational studies indicated that the cleavage site could have one chain more flexible and looped out (Nerenberg et al. 2008). Our result is the first evidence that even in a homotrimer, the Gly-Ile site of one chain may be looser than the Gly-Ile sites in the other 2 chains. The site in this chain may be more susceptible to MMP cleavage. This could be consistent with the model suggesting the site is already available for cleavage vs. a need for unwinding after binding. However, our findings may be only characteristic of Type III collagen, which is more susceptible to enzyme attack than Type I collagen and which may have more local instabilities.

The activities of collagen has been shown to be regulated by neighboring residues, particularly, imino acids (Nagase & Fields 1996). Sequence analysis shows that imino acids are generally favored at N-terminal, while unfavored at C-terminal for collagenase cleavage. The activities on T3-778 and T3-785 are consistent with this observation. However, it is poorly understood how the imino acid distribution affects the conformation and dynamics at the collagenase cleavage sites. A peptide T3-778P-A was designed to study the effects of imino acids at the N-terminal and a peptide T3-778IT-PO was designed to study the effect of imino acids at the C-terminal. Compared with T3-778, both the P-A and IT-PO mutations lead to decreased activities. It is also noted that the Ile17 in peptides T3-778P-A and T3-778IT-PO show a similar pattern in J coupling values: $^1T_{I17}$ and $^3T_{I17}$ have similarly larger values than $^2T_{I17}$, suggesting that the P-A and IT-PO mutations have similar effects on the conformation of Ile at the cleavage site. For both T3-778P-A and T3-778IT-PO, the Ile in the first chain has similar local stabilities as the Ile in the third chain. In all, removing the Pro at N-terminal or adding imino acids at C-terminal makes the unique Ile in the first chain become similar to the Ile in the third chain.

To understand why T3-778 has one special Ile in the first chain but T3-778P-A and T3-778IT-PO have two similar Ile in the first and third chains, staggering between the three chains need to be considered (Figure 5.13). The staggering makes the three chains non-equivalent. For T3-778, only the Ile in the first chain is affected by the Pro at the 3 subsite in the third chain, while other Ile do

not have any close contact with Pro. For T3-778IT-PO, the Pro at the 4' subsite would affect one additional Ile in the third chain. For T3-778P-A, removing the Pro would remove its effect on the Ile in the first chain. Thus, both T3-778IT-PO and T3-778P-A have similar first and third chains. However, the P-A mutation makes the Ile in the first chain more stable, while the IT-PO mutation makes the Ile in the third chain less stable. The higher local stability of Ile in the first chain in T3-778P-A than that in T3-778IT-PO may explain why T3-778P-A has a more decreased activity than T3-778IT-PO.

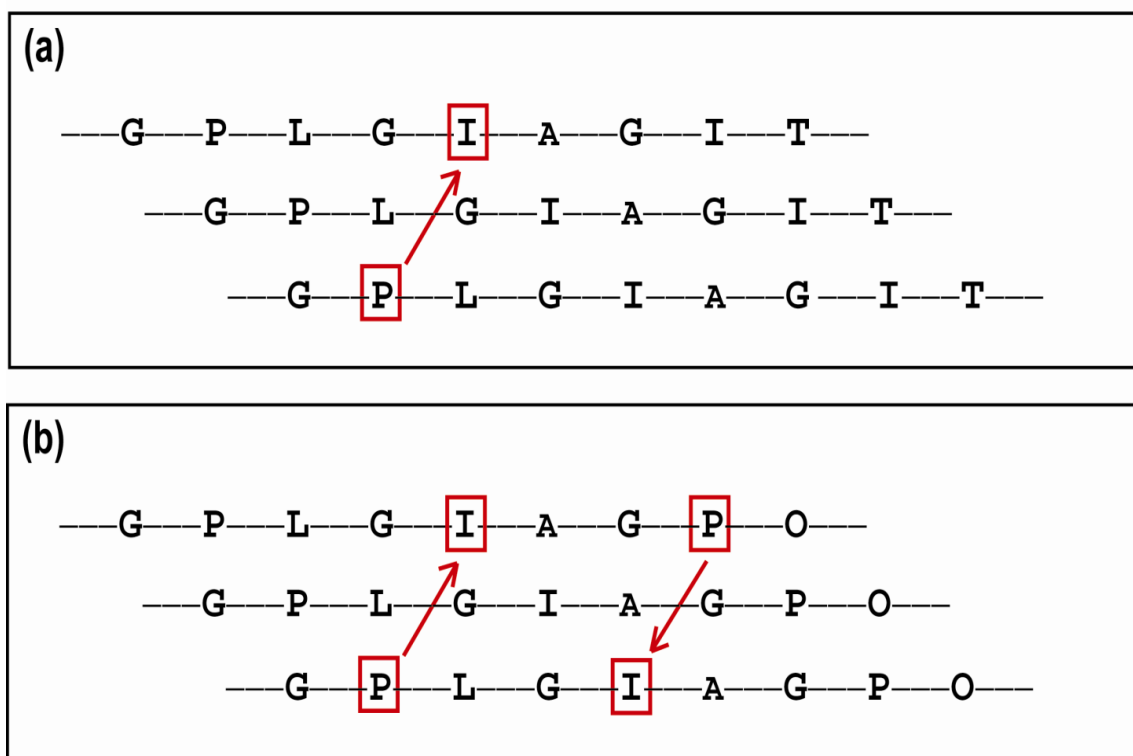


Figure 5.13 Effect of neighboring Pro on Ile at the cleavage site. (a) Pro at 4' subsite in the lagging chain affects the Ile in the leading chain; (b) Pro at 4' subsite in the leading chain affects the Ile in the lagging chain.

REFERENCES

- Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). "Primary structure effects on peptide group hydrogen exchange." *Proteins* 17, 75-86.
- Baum, J. & Brodsky, B. (1999). "Folding of peptide models of collagen and misfolding in disease." *Curr Opin Struct Biol* 9, 122-8.
- Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. (1994). "Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution." *Science* 266, 75-81.
- Bella, J., Liu, J., Kramer, R., Brodsky, B. & Berman, H. M. (2006). "Conformational effects of Gly-X-Gly interruptions in the collagen triple helix." *J Mol Biol* 362, 298-311.
- Bode, W., Fernandez-Catalan, C., Tschesche, H., Grams, F., Nagase, H. & Maskos, K. (1999). "Structural properties of matrix metalloproteinases." *Cell Mol Life Sci* 55, 639-52.
- Bodian, D. L., Madhan, B., Brodsky, B. & Klein, T. E. (2008). "Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations." *Biochemistry* 47, 5424-32.
- Boyd, J. & Redfield, C. (1998). "Defining the orientation of the ^{15}N shielding tensor using ^{15}N relaxation data for a protein in solution." *J. Am. Chem. Soc.* 120, 9692-9693.
- Brodsky, B., Thiagarajan, G., Madhan, B. & Kar, K. (2008). "Triple-helical peptides: an approach to collagen conformation, stability, and self-association." *Biopolymers* 89, 345-53.
- Buevich, A. & Baum, J. (2001). "Nuclear magnetic resonance characterization of peptide models of collagen-folding diseases." *Philos Trans R Soc Lond B Biol Sci* 356, 159-68.
- Byers, P. H. & Cole, W. G. (2002). *Osteogenesis Imperfecta. Connective tissue and its hereditary disorders*. Royce, P. M. & Steinmann, B. New York, Wiley-Liss: 385-430.
- Cho, C. H., Urquidi, J., Singh, S. & Robinson, G. W. (1999). "Thermal offset viscosities of liquid H₂O, D₂O, and T₂O." *Journal of Physical Chemistry B* 103, 1991-1994.

Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J. L., Fields, G. B., Visse, R. & Nagase, H. (2004). "Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis." *Embo J* 23, 3020-30.

Clore, G. M. & Gronenborn, A. M. (1998). "NMR structure determination of proteins and protein complexes larger than 20 kDa." *Curr Opin Chem Biol* 2, 564-70.

Cole, R. & Loria, J. P. (2002). "Evidence for flexibility in the function of ribonuclease A." *Biochemistry* 41, 6072-81.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." *J Biomol NMR* 6, 277-93.

Di Lullo, G. A., Sweeney, S. M., Korkko, J., Ala-Kokko, L. & San Antonio, J. D. (2002). "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.

Di Lullo, G. A., Sweeney, S. M., Korkko, J., Ala-Kokko, L. & San Antonio, J. D. (2002). "Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen." *Journal of Biological Chemistry* 277, 4223-4231.

Emsley, J., Knight, C. G., Farndale, R. W. & Barnes, M. J. (2004). "Structure of the integrin alpha2beta1-binding collagen peptide." *J Mol Biol* 335, 1019-28.

Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. & Liddington, R. C. (2000). "Structural basis of collagen recognition by integrin alpha2beta1." *Cell* 101, 47-56.

Engel, J. & Prockop, D. J. (1991). "The Zipper-like Folding of Collagen Triple-Helices and the Effects of Mutations that Disrupt the Zipper." *Annu. Rev. Biophys. Biophys. Chem.* 20, 137-152.

Englander, S. W. & Krishna, M. M. G. (2001). "Hydrogen exchange." *Nature Structural Biology* 8, 741-742.

Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. (1997). "Hydrogen exchange: the modern legacy of Linderstrom-Lang." *Protein Sci* 6, 1101-9.

Eyre, D. (2002). "Collagen of articular cartilage." *Arthritis Res* 4, 30-5.

Fan, P., Li, M. H., Brodsky, B. & Baum, J. (1993). "Backbone dynamics of (Pro-Hyp-Gly)₁₀ and a designed collagen-like triple-helical peptide by ¹⁵N NMR relaxation and hydrogen-exchange measurements." *Biochemistry* 32, 13299-309.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation." *Biochemistry* 33, 5984-6003.

Fesik, S. W. & Zuiderweg, E. R. (1988). "Heteronuclear three-dimensional nmr spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra." *J Magn Reson* 78, 588-593.

Fields, G. B. (1991). "A model for interstitial collagen catabolism by mammalian collagenases." *J Theor Biol* 153, 585-602.

Fields, G. B. & Van Wart, H. E. (1992). "Unique features of the tissue collagenase cleavage site in interstitial collagens." *Matrix Suppl* 1, 68-70.

Fields, G. B., Van Wart, H. E. & Birkedal-Hansen, H. (1987). "Sequence specificity of human skin fibroblast collagenase. Evidence for the role of collagen structure in determining the collagenase cleavage site." *J Biol Chem* 262, 6221-6.

Fiori, S. & Moroder, L. (2001). NMR-based structural studies of a collagenous substrate of collagenase. In: Houghton RA, Lebl M, editors. Peptides: The Wave Of The Future (Proceedings of the Seventeenth American Peptide Symposium, June 9-14, 2001, San Diego, CA). San Diego: American Peptide Symposium;.

Fiori, S., Sacca, B. & Moroder, L. (2002). "Structural properties of a collagenous heterotrimer that mimics the collagenase cleavage site of collagen type I." *J Mol Biol* 319, 1235-42.

Glasoe, P. K. & Long, F. A. (1960). "Use of glass electrodes to measure acidities in deuterium oxide." *J. Phys. Chem.* 64, 188-189.

Gross, J., Highberger, J. H., Johnson-Wint, B. & Biswas, C. (1980). Mode of action and regulation of tissue collagenases. Collagenase in normal and pathological connective tissues. Woolley, D. E. & Evanson, J. M. New York, John Wiley & Sons: 11-35.

Hasty KA, W. H., Byrne M, Goldring MB, Seyer JM, Jaenisch R, Krane SM, Mainardi CL. (1993). "Susceptibility of type I collagen containing mutated $\alpha 1(I)$ chains to cleavage by human neutrophil collagenase. ." *Matrix* 13, 181-186.

Hofmann, H., Voss, T., Kuhn, K. & Engel, J. (1984). "Localization of flexible sites in thread-like molecules from electron micrographs. Comparison of interstitial, basement membrane and intima collagens." *J Mol Biol* 172, 325-43.

Holz, M. & Weingaertner, H. (1991). "Calibration in accurate spin-echo self-diffusion measurements using proton and less-common nuclei " *J Magn Reson* 92, 115-125.

Hudson, B. G., Tryggvason, K., Sundaramoorthy, M. & Neilson, E. G. (2003). "Alport's syndrome, Goodpasture's syndrome, and type IV collagen." *N Engl J Med* 348, 2543-56.

Huyghues-Despointes, B. M., Pace, C. N., Englander, S. W. & Scholtz, J. M. (2001). "Measuring the conformational stability of a protein by hydrogen exchange." *Methods Mol Biol* 168, 69-92.

Hvidt, A. & Nielsen, S. O. (1966). "Hydrogen exchange in proteins." *Adv Protein Chem* 21, 287-386.

Hyde, T. J., Bryan, M. A., Brodsky, B. & Baum, J. (2006). "Sequence dependence of renucleation after a Gly mutation in model collagen peptides." *J Biol Chem* 281, 36937-43.

Ikura, M., Kay, L. E. & Bax, A. (1990). "A novel approach for sequential assignment of ^1H , ^{13}C , and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin." *Biochemistry* 29, 4659-67.

Johnson, B. A. & Blevins, R. A. (1994). "Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data." *J Biomol NMR* 4, 603-614.

Kalluri, R. (2003). "Basement membranes: structure, assembly and role in tumour angiogenesis." *Nat Rev Cancer* 3, 422-33.

Kay, L. E. (2005). "NMR studies of protein structure and dynamics." *J Magn Reson* 173, 193-207.

Kay, L. E. (2005). "NMR studies of protein structure and dynamics." *Journal of Magnetic Resonance* 173, 193-207.

Kay, L. E., Ikura, M. & Bax, A. (1991). "The design and optimization of complex NMR experiments. Application to a triple-resonance pulse scheme correlating $\text{H}\alpha$, NH , and ^{15}N chemical shifts in ^{15}N --- ^{13}C -labeled proteins." *J Magn Reson* 91, 84-92.

Kay, L. E., Keifer, P. & Saarinen, T. (1992). "Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity." *J. Am. Chem. Soc.* 114, 10663-10665.

Konitsiotis, A. D., Raynal, N., Bihan, D., Hohenester, E., Farndale, R. W. & Leitingner, B. (2008). "Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen." *J Biol Chem* 283, 6861-8.

Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B. & Berman, H. M. (1999). "Sequence dependent conformational variations of collagen triple-helical structure." *Nat Struct Biol* 6, 454-7.

Kramer, R. Z., Venugopal, M. G., Bella, J., Mayville, P., Brodsky, B. & Berman, H. M. (2000). "Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair." *J Mol Biol* 301, 1191-205.

Kuivaniemi, H., Tromp, G. & Prockop, D. J. (1997). "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels." *Hum Mutat* 9, 300-15.

Lauer-Fields, J. L. & Fields, G. B. (2002). "Triple-helical peptide analysis of collagenolytic protease activity." *Biol Chem* 383, 1095-105.

Lauer-Fields, J. L., Juska, D. & Fields, G. B. (2002). "Matrix metalloproteinases and collagen catabolism." *Biopolymers* 66, 19-32.

Lauer-Fields, J. L., Sritharan, T., Stack, M. S., Nagase, H. & Fields, G. B. (2003). "Selective hydrolysis of triple-helical substrates by matrix metalloproteinase-2 and -9." *J Biol Chem* 278, 18140-5.

Lauer-Fields, J. L., Tuzinski, K. A., Shimokawa, K., Nagase, H. & Fields, G. B. (2000). "Hydrolysis of triple-helical collagen peptide models by matrix metalloproteinases." *J Biol Chem* 275, 13282-90.

Li, J., Brick, P., O'Hare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E. & et al. (1995). "Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed beta-propeller." *Structure* 3, 541-9.

Li, M. H., Fan, P., Brodsky, B. & Baum, J. (1993). "Two-dimensional NMR assignments and conformation of (Pro-Hyp-Gly)₁₀ and a designed collagen triple-helical peptide." *Biochemistry* 32, 7377-87.

Li, Y., Brodsky, B. & Baum, J. (2007). "NMR shows hydrophobic interactions replace glycine packing in the triple helix at a natural break in the (Gly-X-Y)_n repeat." *J Biol Chem* 282, 22699-706.

Li, Y., Brodsky, B. & Baum, J. (2009). "NMR conformational and dynamic consequences of a gly to ser substitution in an osteogenesis imperfecta collagen model Peptide." *J Biol Chem* 284, 20660-7.

Li, Y., Kim, S., Brodsky, B. & Baum, J. (2005). "Identification of partially disordered peptide intermediates through residue-specific NMR diffusion measurements." *J Am Chem Soc* 127, 10490-1.

Lipari, G. & Szabo, A. (1982). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity." *J. Am. Chem. Soc.* 104, 4546-4559.

Liu, X., Siegel, D. L., Fan, P., Brodsky, B. & Baum, J. (1996). "Direct NMR measurement of folding kinetics of a trimeric peptide." *Biochemistry* 35, 4306-13.

Long, C. G., Braswell, E., Zhu, D., Apigo, J., Baum, J. & Brodsky, B. (1993). "Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence." *Biochemistry* 32, 11688-95.

Lovejoy, B., Hassell, A. M., Luther, M. A., Weigl, D. & Jordan, S. R. (1994). "Crystal structures of recombinant 19-kDa human fibroblast collagenase complexed to itself." *Biochemistry* 33, 8207-17.

Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). "Structure validation by C α geometry: phi,psi and C β deviation." *Proteins* 50, 437-50.

Makareeva, E., Mertz, E. L., Kuznetsova, N. V., Sutter, M. B., DeRidder, A. M., Cabral, W. A., Barnes, A. M., McBride, D. J., Marini, J. C. & Leikin, S. (2008). "Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta." *J Biol Chem* 283, 4787-98.

Mallya, S. K., Mookhtiar, K. A., Gao, Y., Brew, K., Dioszegi, M., Birkedal-Hansen, H. & Van Wart, H. E. (1990). "Characterization of 58-kilodalton human neutrophil collagenase: comparison with human fibroblast collagenase." *Biochemistry* 29, 10628-34.

Mandel, A. M., Akke, M. & Palmer, A. G., 3rd (1995). "Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme." *J Mol Biol* 246, 144-63.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989). "Three-dimensional heteronuclear NMR of ^{15}N labeled proteins." *J Am Chem Soc* 111, 1515-1517

Messerle, B. A., Wider, G., Otting, G., Weber, C. & Wüthrich, K. (1989). "Solvent suppression using a spin-lock in 2D and 3D NMR spectroscopy with H_2O solutions." *J Magn Reson* 85, 608–613.

Miller, E. J., Finch, J. E., Jr., Chung, E., Butler, W. T. & Robertson, P. B. (1976). "Specific cleavage of the native type III collagen molecule with trypsin. Similarity of the cleavage products to collagenase-produced fragments and primary structure at the cleavage site." *Arch Biochem Biophys* 173, 631-7.

Miller, E. J. & Gay, S. (1987). "The collagens: an overview and update." *Methods Enzymol* 144, 3-41.

Mills, R. (1973). "Self-diffusion in normal and heavy water in the range 1-45 degrees." *J Phys Chem* 77, 685-688.

Minond, D., Lauer-Fields, J. L., Nagase, H. & Fields, G. B. (2004). "Matrix metalloproteinase triple-helical peptidase activities are differentially regulated by substrate stability." *Biochemistry* 43, 11474-81.

Moy, F. J., Chanda, P. K., Chen, J. M., Cosmi, S., Edris, W., Levin, J. I. & Powers, R. (2000). "High-resolution solution structure of the catalytic fragment of human collagenase-3 (MMP-13) complexed with a hydroxamic acid inhibitor." *Journal of Molecular Biology* 302, 671-689.

Moy, F. J., Chanda, P. K., Cosmi, S., Pisano, M. R., Urbano, C., Wilhelm, J. & Powers, R. (1998). "High-resolution solution structure of the inhibitor-free catalytic fragment of human fibroblast collagenase determined by multidimensional NMR." *Biochemistry* 37, 1495-1504.

Myllyharju, J. & Kivirikko, K. I. (2004). "Collagens, modifying enzymes and their mutations in humans, flies and worms." *Trends Genet* 20, 33-43.

Nagase, H. & Fields, G. B. (1996). "Human matrix metalloproteinase specificity studies using collagen sequence-based synthetic peptides." *Biopolymers* 40, 399-416.

Nerenberg, P. S., Salsas-Escat, R. & Stultz, C. M. (2008). "Do collagenases unwind triple-helical collagen before peptide bond hydrolysis? Reinterpreting experimental observations with mathematical models." *Proteins* 70, 1154-61.

Netzel-Arnett, S., Fields, G. B., Birkedal-Hansen, H. & Van Wart, H. E. (1991). "Sequence specificities of human fibroblast and neutrophil collagenases." *J Biol Chem* 266, 6747-55.

Ottl, J., Battistuta, R., Pieper, M., Tschesche, H., Bode, W., Kuhn, K. & Moroder, L. (1996). "Design and synthesis of heterotrimeric collagen peptides with a built-in cystine-knot. Models for collagen catabolism by matrix-metalloproteases." *FEBS Lett* 398, 31-6.

Palmer, A. G. (2004). "NMR characterization of the dynamics of biomacromolecules." *Chemical Reviews* 104, 3623-3640.

Palmer, A. G., 3rd (1993). "Dynamic properties of proteins from NMR spectroscopy." *Curr Opin Biotechnol* 4, 385-91.

Palmer, A. G., 3rd (2001). "Nmr probes of molecular dynamics: overview and comparison with other techniques." *Annu Rev Biophys Biomol Struct* 30, 129-55.

Palmer, A. G., 3rd, Kroenke, C. D. & Loria, J. P. (2001). "Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules." *Methods Enzymol* 339, 204-38.

Persikov, A. V., Ramshaw, J. A. & Brodsky, B. (2000). "Collagen model peptides: Sequence dependence of triple-helix stability." *Biopolymers* 55, 436-50.

Persikov, A. V., Ramshaw, J. A., Kirkpatrick, A. & Brodsky, B. (2000). "Amino acid propensities for the collagen triple-helix." *Biochemistry* 39, 14960-7.

Radmer, R. J. & Klein, T. E. (2004). "Severity of osteogenesis imperfecta and structure of a collagen-like peptide modeling a lethal mutation site." *Biochemistry* 43, 5314-5323.

Raghunath, M., Bruckner, P. & Steinmann, B. (1994). "Delayed Triple Helix Formation of Mutant Collagen from Patients with Osteogenesis Imperfecta." *J. Mol. Biol.* 236, 940-949.

Ramachandran, G. N. (1967). *Structure of Collagen at the Molecular Level. Treatise on Collagen*. Ramachandran, G. N. New York, Academic Press: 103-184.

Ramachandran, G. N. & Kartha, G. (1955). "Structure of collagen." *Nature* 176, 593-5.

Renner, C., Sacca, B. & Moroder, L. (2004). "Synthetic heterotrimeric collagen peptides as mimics of cell adhesion sites of the basement membrane." *Biopolymers* 76, 34-47.

Rich, A. & Crick, F. H. (1955). "The structure of collagen." *Nature* 176, 915-6.

Rich, A. & Crick, F. H. (1961). "The molecular structure of collagen." *J Mol Biol* 3, 483-506.

Ryhanen, L., Zaragoza, E. J. & Uitto, J. (1983). "Conformational stability of type I collagen triple helix: evidence for temporary and local relaxation of the protein conformation using a proteolytic probe." *Arch Biochem Biophys* 223, 562-71.

Shoulders, M. D. & Raines, R. T. (2009). "Collagen structure and stability." *Annu Rev Biochem* 78, 929-58.

Siljander, P. R., Hamaia, S., Peachey, A. R., Slatter, D. A., Smethurst, P. A., Ouwehand, W. H., Knight, C. G. & Farndale, R. W. (2004). "Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens." *J Biol Chem* 279, 47763-72.

Skrynnikov, N. R. & Kay, L. E. (2000). "Assessment of molecular structure using frame-independent orientational restraints derived from residual dipolar couplings." *J Biomol NMR* 18, 239-52.

Stultz, C. M. (2002). "Localized unfolding of collagen explains collagenase cleavage near imino-poor sites." *J Mol Biol* 319, 997-1003.

Thiagarajan, G., Li, Y., Mohs, A., Strafacci, C., Popiel, M., Baum, J. & Brodsky, B. (2008). "Common interruptions in the repeating tripeptide sequence of non-fibrillar collagens: sequence analysis and structural studies on triple-helix peptide models." *J Mol Biol* 376, 736-48.

Tiller, G. E., Polumbo, P. A., Weis, M. A., Bogaert, R., Lachman, R. S., Cohn, D. H., Rimoian, D. L. & Eyre, D. R. (1995). "Dominant mutations in the type II collagen gene, COL2A1, produce spondyloepimetaphyseal dysplasia, Strudwick type." *Nat Genet* 11, 87-9.

Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. (1995). "Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation." *J Am Chem Soc* 117, 12562 - 12566.

Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A. & Clore, G. M. (1997). "Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy." *Nat Struct Biol* 4, 443-9.

Ulmer, T. S., Ramirez, B. E., Delaglio, F. & Bax, A. (2003). "Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy." *J Am Chem Soc* 125, 9179-91.

van der Rest, M., Aubert-Foucher, E., Dublet, B., Eichenberger, D., Font, B. & Goldschmidt, D. (1991). "Structure and function of the fibril-associated collagens." *Biochem Soc Trans* 19, 820-4.

Vuister, G. W. & Bax, A. (1993). "Quantitative J correlation: a new approach for measuring homonuclear three-bond $J(\text{H}^{\text{N}}\text{H}^{\text{A}})$ coupling constants in ^{15}N -enriched proteins." *J Am Chem Soc* 115, 7772-7777.

Wang, A. C. & Bax, A. (1996). "Determination of the backbone dihedral angles ϕ in human ubiquitin from reparametrized empirical Karplus equations." *Journal of the American Chemical Society* 118, 2483-2494.

Welgus, H. G., Jeffrey, J. J. & Eisen, A. Z. (1981). "The collagen substrate specificity of human skin fibroblast collagenase." *J Biol Chem* 256, 9511-5.

Welgus, H. G., Jeffrey, J. J., Stricklin, G. P. & Eisen, A. Z. (1982). "The gelatinolytic activity of human skin fibroblast collagenase." *J Biol Chem* 257, 11534-9.

Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984). "Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II." *J Mol Biol* 180, 301-29.

Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). "Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation." *J Mol Biol* 285, 1735-47.

Wu H, B. M., Stacey A, Goldring MB, Birkhead JR, Jaenisch R, Krane SM. (1990). "Generation of collagenase-resistant collagen by site-directed mutagenesis of murine pro $\alpha 1(\text{I})$ collagen gene. ." *Proc. Natl. Acad. Sci. USA* 87, 5888-5892.

Wuthrich, K. (2001). "The way to NMR structures of proteins." *Nat Struct Biol* 8, 923-5.

Wuthrich, K. (2003). "NMR studies of structure and function of biological macromolecules (Nobel Lecture)." *J Biomol NMR* 27, 13-39.

Zijlstra, A., Aimes, R. T., Zhu, D., Regazzoni, K., Kupriyanova, T., Seandel, M., Deryugina, E. I. & Quigley, J. P. (2004). "Collagenolysis-dependent angiogenesis mediated by matrix metalloproteinase-13 (collagenase-3)." *J Biol Chem* 279, 27633-45.

Zweckstetter, M. & Bax, A. (2000). "Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR." *Journal of the American Chemical Society* 122, 3791-3792.

Curriculum Vita

Jianxi Xiao

1999-2003	B.S. in Chemistry, Wuhan University, Wuhan, China
1999-2003	Dean's List
2004- 2008	Teaching Assistant. Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ
2004-2009	Research Assistant. Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ
2005-2006	Damle Graduate Fellowship
2008	First Prize Winner of poster contest at RUTGERS–UMDNJ 20 th Annual Molecular Biophysics Minisymposium
2008	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, 13520
2009	Ph.D. in Chemistry. Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ