APPLICATION AND DEVELOPMENT OF NMR SPECTROSCOPY TO STUDY THE CONFORMATION AND DYNAMICS OF COLLAGEN-LIKE TRIPLE HELICAL PEPTIDES

by

YINGJIE LI

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Biochemistry

written under the direction of

Professor Jean Baum

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2007

ABSTRACT OF THE DISSERTATION

Application and development of NMR spectroscopy to study the conformation and dynamics of collagen-like triple helical peptides

By YINGJIE LI

Dissertation Director:
Professor Jean Baum

This work describes the NMR conformational and dynamic characterization of collagen-like triple helical peptides. The unique features of triple helices result in limited interproton distance information and the lack of long range contacts, so the solution structure of triple helical peptides can not be solved by NMR. Alternative strategies have been developed and applied on selectively $^{13}C/^{15}N$ doubly labeled synthetic peptides, to allow the visualization of peptide solution models. This approach, including the stagger chain identification, measurement of NMR conformational parameters, molecular modeling incorporating NMR data, and dynamic and hydrogen bonding investigation, provided the detailed structural and dynamic characterization of the triple helix system.

This method is applied to two families of peptides: peptides that model natural interruptions in the non-fibrillar collagen $(Gly\text{-}X\text{-}Y)_n$ repeats and peptides that model Gly to X substitutions in fibrillar collagen disease. NMR studies are presented to define the structural and dynamic effects that small breaks have on triple-helix structure, and allow the visualization of peptide models with two most common types of breaks with different

lengths. Both breaks allow continuation of rod-like helices and maintenance of the 1-residue stagger, and have a highly localized perturbation in dihedral angle and hydrogen bonding. However, the conformation of residues within the breaks is different for the two kinds of breaks. This work has provided clues on the molecular basis of collagen binding at specific recognition sites.

The NMR methodology was also applied on peptides that model Gly to X mutations in fibrillar collagen diseases. Mutation-specific equilibrium states were identified for peptides with different Gly to X substitutions, indicating mutation specific abilities for the peptides to fold around the substitution. The first visualization of the solution conformation of a model peptide with a Gly to Ser substitution in a physiological sequence context was obtained, demonstrating the disturbed conformation and dynamic features and hydrogen bonding at the substitution site. It suggested that the substitution could alter the appearance of the exterior of the triple helix molecule as it is seen by binding partners, and represented a first step in clarifying how mutations could result in collagen disease.

Acknowledgements

Many people have helped me in the course of my Ph.D study. It is a pleasant aspect that I have the opportunity to express my gratitude to all of them.

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Jean Baum, for her support and guidance for my thesis study. Without her expert advice, kind encouragement and patience my thesis research would never have been completed. In addition to her help on my scientific research, she also provided me opportunities and supported me in my personal development.

I owe a debt of gratitude to Prof. Barbara Brodsky, a great professor directly involved in my thesis research. During our long term collaborations, I was continuously encouraged and learned a lot from her inspiration, enthusiasm and immense knowledge. I appreciate her help on my personal development as well.

I am also grateful to Dr. Seho Kim for his kind help in many aspects of NMR, from theory to experiments and applications at all times and for his patience and invaluable advice.

I would like to thank Prof. Helen Berman, Prof. Daniel Pilch and Prof. John Taylor for serving on my thesis committees and taking effort in reading whole thesis thoroughly. I also would like to thank Prof. Masayori Inouye for the collaboration on the bacterial expression of collagen peptides, and for allowing me to use his protein expression and purification facilities. Thanks to Dr. Takeshi Yoshida for his great help in the establishment of bacterial expression system.

I am also grateful to present and past members of Prof. Baum's group. Thanks to Dr. Yujia Xu, Timothy Hyde, and David Fela for giving me an introduction into the lab

and for sharing their knowledge. Thank Jianxi Xiao, Kuen-phon Wu, Lijuan Kang, Yu-Jen Chen, Chitra Narayanan, Abigail Go, and Yinan Wei. My experience at Rutgers would not have been such a pleasurable one without the presence of all the people working there.

I have years of collaborations with Prof. Brodsky's group. I thank all the members from Prof. Brodsky's group for inspiring discussions and for spreading happiness. I especially thank Dr. Angela Mohs, Dr. Geetha Thiagarajan, and Michael Bryan for co-authoring papers, which have contributed significantly to this thesis.

Finally, I would like to thank my family and friends for being an unstinting source of support and encouragement. I am forever indebted to my parents for their endless love to me.

Table of Contents

Abstract of the Dissertation……ii

Acknowledgments……..iv

Table of Contents……vi

List of Tables……xii

List of Figures……...xiii

List of Abbreviations…….

Chapter 1 Introduction……1

1.1. Introduction to collagen……2

1.2. Collagen peptides and NMR…….……....7

1.2.1. NMR structural characterization……..9

1.2.2. NMR dynamic characterization…….10

1.2.3. NMR hydrogen bonding characterization……..11

1.3. Scope of the dissertation…….13

Chapter 2 Peptide design and preparation…….14

2.1. Introduction…….15

2.2. The purification of synthetic collagen model peptides……..17

2.2.1. Sample preparation……..17

2.2.2. Effect of temperature……...18

2.2.3. Identification of two fractions at low temperature……...20

2.2.4. Effect of concentration….....26

2.2.5. Effect of peptide sequence and complementary size exclusion chromatography…….28

2.2.6. Conclusion….....30

2.3. Bacterial expression approaches…..……...31

2.3.1. Design of bacterial expression system…….31

2.3.2. Development and application of bacterial expression system on the GAAVM peptide……33

2.3.3. Conclusion……...35

Chapter 3 NMR methods applied to triple helical peptides……...37

3.1. Introduction…….38

3.2. Development of NMR methodology to characterize structural properties of triple helical peptides…….41

3.2.1. Introduction……..41

3.2.2. NMR chain assignment and the measurements of structural parameters…….….42

3.2.2.1. Materials and methods……..42

3.2.2.1.1. Sample preparation……42

3.2.2.1.2. Sequential assignments..……43

3.2.2.1.3. NOESY and TOCSY experiments…….44

3.2.2.1.4. Measurements of $^{3}J_{HNHa}$ coupling constants.....45

3.2.2.1.5. Data processing…..……....45

3.2.2.1.6. Generation of NOE contact map……..……..46

3.2.2.2. Results and discussion……..46

3.2.2.2.1. NMR chain assignments in GAAGVM and GAAVM peptides….……46

3.2.2.2.2. NOE contacts to identify chain stagger and interchain distances……51
3.2.2.2.3. Determination of J coupling constants……...56
3.2.3. Molecular modeling……59
3.2.3.1. The procedure of molecular modeling……..59
3.2.3.2. An example of the molecular modeling: GPO(G16S) peptide…….……...60
3.2.3.2.1. Starting structures and NMR restraints……..60
3.2.3.2.2. Iterative procedures of the back calculation and structure validation…….…...62
3.2.4. Conclusions……..65
3.3. Identification of partially disordered peptide intermediates through residue specific NMR diffusion measurements……..67
3.3.1. Introduction……..67
3.3.2. Materials and methods……69
3.3.2.1. Sample preparation……...69
3.3.2.2. Diffusion experiments……...69
3.3.2.3. Data analysis……71
3.3.3. Results and discussion……74
3.3.3.1. Residue specific diffusion measurements…….……74
3.3.3.2. Relative populations of the equilibrium intermediates of the T1-892c peptide……..……77
3.3.3.3. Implications on the folding……...81

3.3.3.4. Conclusions……...82

Chapter 4 Structural and dynamic consequences of natural breaks in the non-fibrillar collagen $(Gly\text{-}X\text{-}Y)_n$ repeats……..83

4.1. Introduction……84

4.2. Materials and methods…….....90

4.2.1. Sample preparation……..90

4.2.2. NMR chain assignments and structural measurements……...90

4.2.3. Hydrogen exchange and amide temperature gradients...…….91

4.2.4. Molecular modeling……...……..92

4.3. Results and discussion……..……..94

4.3.1. NMR studies on a G4G break……..94

4.3.1.1. Peptide design……...94

4.3.1.2. NMR chain assignments and chemical shifts……..95

4.3.1.3. Experimental NOE observations …..……96

4.3.1.4. Determination of J coupling constants……..99

4.3.1.5. Hydrogen exchange and amide temperature gradients...100

4.3.1.6. Molecular modeling……105

4.3.2. NMR studies on a prototypical G1G break……109

4.3.2.1. The chain assignment and conformational characterization……....109

4.3.2.2. Hydrogen bonding studies: hydrogen exchange and amide temperature gradient…….……...110

4.3.2.3. Molecular modeling……113

4.3.3. Implications for collagen……...……116

Chapter 5 Conformational and dynamic consequences of Gly to X substitutions of collagen disease in triple helical peptides……119

5.1. Introduction……...120

5.2. Materials and methods……..125

5.2.1. Translational diffusion and J coupling measurements on the T1-898 peptide series..……..125

5.2.2. Structural and dynamic characterization of the GPO(G16S) peptide……..……..127

5.2.2.1. Sample preparation…….127

5.2.2.2. NMR assignment, structural and dynamic measurements……...127

5.2.2.3. Molecular modeling……..……..129

5.3. Results and discussion……..130

5.3.1. Effect of residue identity of Gly to X substitution on the equilibrium states of T1-898 peptide series…….……130

5.3.1.1. Peptide design…….130

5.3.1.2. Translational diffusion measurements……130

5.3.1.3. J coupling measurements on T1-898(G16A) and T1-898(G16S) peptides…….134

5.3.1.4. Discussion……...135

5.3.2. Structural and dynamic consequences of a Gly to Ser substitution in the GPO(G16S) peptide……...138

5.3.2.1. Peptide design..……...138

5.3.2.2. NMR chain assignments and chemical shifts…….140

5.3.2.3. NMR conformational characterization...……143

5.3.2.4. Hydrogen exchange and amide temperature gradients...148

5.3.2.5. Molecular modeling……151

5.3.2.6. Discussion……...154

5.3.2.6.1. The conformation and hydrogen bonding of the GPO(G16S) model……...154

5.3.2.6.2. The dynamics of the GPO(G16S) peptide…...155

5.3.2.6.3. Comparison of the GPO(G16S) peptide to the Gly→Ala peptide……...……..158

5.3.2.6.4. Biological implications……160

References……164

Curriculum Vita……...177

Lists of tables

3.1 Peptide sequences of GAAGVM and GAAVM……....43

3.2 Assignment summary of peptide GAAGVM and GAAVM……..49

3.3 The relative population ratio of partially disordered intermediates (pd) to monomer (m) in T1-892c at 10°C……..81

4.1 ^{15}N R_2 and NOE of the GAAVM peptide……..……..104

5.1 Peptide sequences of the T1-898 series……...125

5.2 $^{3}J_{HNHa}$ of the Gly→Ala peptide and the corresponding multiple dihedral angle solutions from the Karplus equation and from the x-ray crystal structure, and the NH temperature gradients of Gly→Ala…….……..147

5.3 Comparison of the (GPO)G16S and Gly→Ala peptide……..…160

List of figures

1.1 Representation of the collagen triple helix conformation……3

2.1 Reversed phase chromatography of the GAAVM peptide at 22°C (A) and 4°C (B)…….19

2.2 Purification of the GAAVM peptide……..21

2.3 Purification of the GFG peptide…….22

2.4 Reversed phase chromatography of GAAVM peptide under native (black line) and heat denatured (red line) conditions……23

2.5 HSQC spectra of the ^{15}N labeled GAAVM peptide under conditions: before purification at 5°C (A); Monomer fraction from purification (B-D); Trimer fraction from purification (E-G)…..……...25

2.6 Effects of concentration of the GAAVM peptide in reversed phase chromatography…….27

2.7 Application of size exclusion chromatography on the purification of peptide GPO(G16S)……29

2.8 Design of bacterial expression system……...31

2.9 MALDI-TOF Mass spectrometry of proteinS/GAAVM mixture after treatment with thrombin…….34

2.10 ^{1}H-^{15}N HSQC spectra of ^{15}N-uniformly labeled GAAVM peptide from bacterial expression system……..36

3.1 Representation of the peptide sequence GAAGVM and the cross section view of the one residue stagger……...40

3.2 ^{1}H-^{15}N HSQC spectra of peptides GAAGVM and GAAVM at 20°C……...47

3.3 Strips from the subset 4 of (5,3)D HACACONHN (strip a) /HACA,CONHN (strip b) GFT experiment of GAAGVM peptide……...50

3.4 Strips from the HA(CA)NH (strip a)/HA(CACO)NH (strip b) spectra of GAAVM peptide……50

3.5 Comparison of experimental NOEs from the NOESY-HSQC experiment for the GAAGVM with predicted NOEs from a standard triple helical conformation….55

3.6 Stripes from HNCA-J experiment of ^{T}G24 (A) and ^{M}G24 (B) in GAAGVM peptide……58

3.7 The flowchart of molecular modeling procedure incorporating NMR restraints..60

3.8 Back calculation of the predicted NOEs from model structure of GPO(G16S) and compare it with the experimental NOEs from the NOESY-HSQC experiment…65

3.9 CCLED-HSQC pulse sequence with the convection compensated diffusion encoded CCLED part shown in box……..72

3.10 Translational diffusion spectrum and data fitting of the T1-892c peptide…….73

3.11 Histogram of the residue specific diffusion coefficients versus ^{15}N-labeled residues for T1-892n (A), T1-892c[G10A] (B), and T1-892c (C)……74

3.12 Schematic representation of the equilibrium states of T1-892n (A), T1-892c[G10A] (B) and T1-892c (C).……77

3.13 Bi- and mono-exponential fitting of T1-892c monomer diffusion decay……..80

4.1 Schematic illustration of the locations of the 22 breaks within the 1414-residue Gly-X-Y repeating region (residues 42–1456) of the triple helix in the α5 chain of human type IV collagen…….94

4.2 Comparison of experimental NOEs from the NOESY-HSQC experiment for the GAAGVM and GAAVM model peptides……98

4.3 Experimental $^{3}J_{HNHa}$ coupling constants of peptides GAAGVM (A) and GAAVM (B)……100

4.4 Histogram of H/D protection factors for the labeled residues in GAAGVM (A) and GAAVM (B) peptide……102

4.5 Amide NH Δδ/ΔT plot for GAAGVM (A, black open circles) and GAAVM (B, red filled circles) peptide……104

4.6 Model structures of peptide GAAGVM with a standard Gly-X-Y pattern and peptide GAAVM containing a type IV collagen break……...108

4.7 NMR structural, dynamic and hydrogen bonding characterization of the GFG peptide……..112

4.8 Model structures of peptide GFG……115

5.1 HSQC and translational diffusion measurements on the T1-898 peptide series.133

5.2 Schematic representation of the equilibrium states of T1-898 (A), T1-898(G16A) and T1-898(G16S) (B), T1-898(G16R) and T1-898(G16D) (C)…….133

5.3 $^{3}J_{HNHA}$ coupling constants of labeled residues in peptide T1-898 (A), T1-898(G16A) (B), and T1-898(G16S) (C)……..135

5.4 Translational diffusion measurements of the GPO(G16S) peptide…….139

5.5 The HSQC spectrum of GPO(G16S) peptide at 15°C…….142

5.6 The HSQC spectrum of the Gly→Ala peptide at 15°C……...143

5.7 Comparison of experimental NOEs of GPO(G16S) peptide with a standard triple helical conformation……145

5.8 The histogram of the $^{3}J_{HNHa}$ coupling constants are measured for the GPO(G16S) peptide……..147

5.9 Hydrogen exchange and NH temperature gradient studies on the GPO(G16S) peptide……..150

5.10 The model structure of the GPO(G16S) peptide……..…….……..152

5.11 Comparison of conformational and hydrogen bonding features of GPO(G16S) peptides to the Gly→Ala peptide…..……...157

5.12 Comparison of the cross views of Cα of residues around the substitution region in the GPO(G16S) model (A) and Gly→Ala structure (B).……163

Chapter 1

Introduction

1.1. Introduction to Collagen

Collagen, the most abundant protein in the human body, is the major constituent protein in bone, skin, tendon, ligament, cartilage and membranous tissues (Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004). Very different from typical globular proteins, collagen has a triple helix motif and a repeating Gly-X-Y sequence pattern with Gly presenting as every third residue. The collagen triple helix is not only the defining motif of all extracellular matrix collagens, but also found as a domain in various other proteins, including C1q, macrophage scavenger receptor, collagenous tail of asymmetric acetylcholinesterase, and the bacterial proteins *Streptococcus pyogenes* scl and *Bacillus anthraces* bcl1 (Boydston et al., 2005; Deprez et al., 2000; Hoppe and Reid, 1994; Kielty and Grant, 2002; Xu et al., 2002). In the triple-helical conformation, three monomer chains with left-handed poly-proline II conformation are parallel and right-handed supercoiled about a central axis with one-residue stagger, and form a long rod-shape molecule (Figure 1.1) (Bella et al., 1994; Ramachandran, 1967; Rich and Crick, 1961). The Gly residues, present as every third residue, are essential to the stability of the triple helix because Gly can be closely packed in the center of triple helix without any steric interference and conformational deformation. The X and Y positions are frequently occupied by Pro and hydroxyproline (Hyp), respectively. Both imino acids are found to be important to triple-helix stability, and Hyp confers an additional stability (Privalov, 1982). The amide protons of Gly are pointed toward the center of the molecule, while amide protons of X and Y residues are outward and exposed to solvent. Standard Rich and Crick II hydrogen bonds are formed between the amide proton of Gly and the carbonyl group of the X residue in the neighboring chain (Bella et al., 1994;

Ramachandran, 1967; Rich and Crick, 1961).

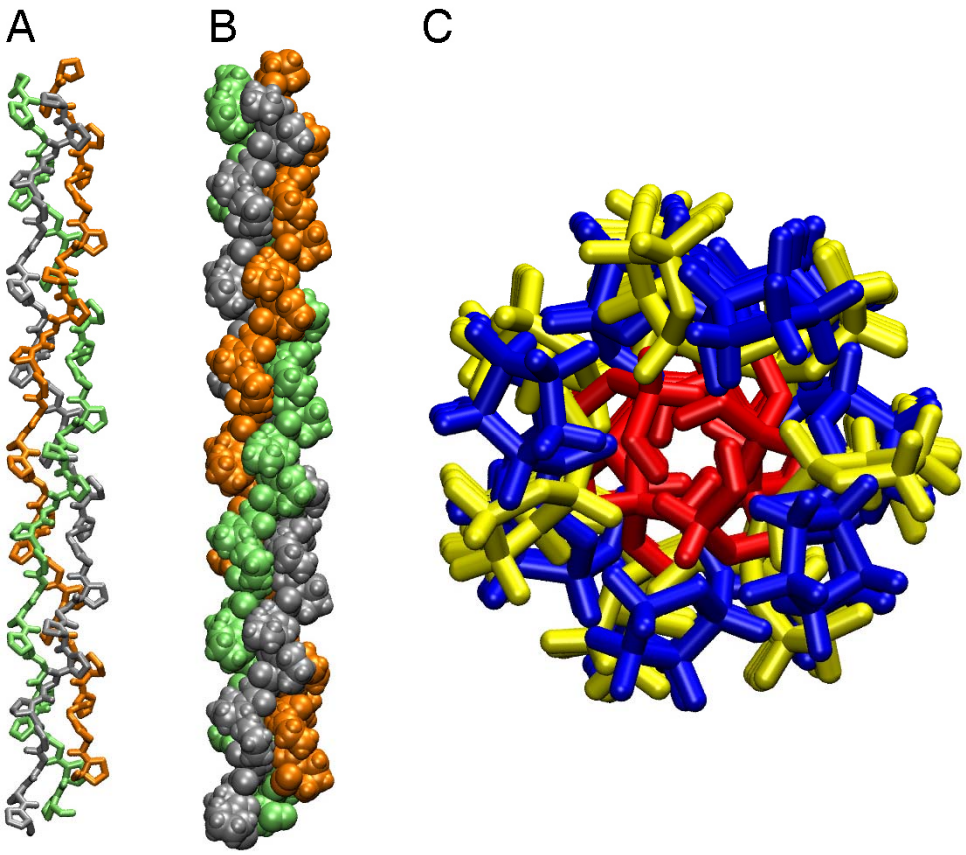

Figure 1.1 Representation of the collagen triple-helix conformation.

A, The stick model of the triple helical structure colored by three chains.

B, The spacefill model of the triple helical structure. Three polypeptide chains with left-handed polyproline II conformation supercoiled right-handed about a central axis.

C, The cross section of the triple helical structure. The Gly, X, and Y residues are colored in red, blue, and yellow, respectively. The Gly residues are closely packed in the center of the triple helix, while the residues in the X and Y positions are largely exposed to solvent. All representations are produced from the $(PPG)_{10}$ x-ray crystal structure.

As the major structural protein in the extracellular matrix in all tissues, collagen is involved a variety of biological processes and plays an important role in normal health and various diseases. Collagen binds a number of proteins such as cell surface integrins, fibronectin, proteoglycans, phospholipid vesicles and matrix metalloproteinases, and plays a critical role in many aspects of extracellular matrix interactions and development including cell migration, cell binding and cleavage involved in tumor metastasis (Di Lullo et al., 2002; Kadler, 1994; Kielty and Grant, 2002; Knight et al., 2000; Myllyharju and Kivirikko, 2004). Mutations in collagen lead to a number of extracellular matrix related diseases, while alterations in degradation, synthesis or binding of collagen play a role in various common diseases such as osteoarthritis, thrombosis, cancer and diabetes (Byers and Cole, 2002; Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004). Defects in the triple helix domain of non-collgenous proteins such as mannose binding protein results in childhood infections and increased susceptibility to infections and autoimmune disease (Summerfield et al., 1997; Super et al., 1989).

Although the collagen triple-helix can be considered as a well-characterized protein motif, there are still surprises and undefined molecular features. The $(Gly\text{-}X\text{-}Y)_n$ repeating pattern is perfectly maintained in some collagen domains but not in others (Kielty and Grant, 2002). So far 28 different types of collagens are found including fibril forming collagens and non-fibrillar collagens. The most abundant collagens are the fibrillar collagens such as types I-III collagen, which are composed of heterotrimers (type I) and homotrimers (types II and III). Fibrillar collagens form the supporting matrix of bone, tendon and other connective tissues. Non-fibrillar types include collagens found in networks, such as type IV collagen; membrane proteins, such as type XVII and type XIII;

FACIT collagens (types IX, XII, and XIV) found on the surface of periodic collagen fibrils; and type VII collagen in anchoring fibrils (Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004).

Fibrillar collagens all maintain a precise Gly-X-Y repeat throughout their ~1000-residue triple helix domain. The replacement of even one Gly by another residue as a result of a single base change results in a pathological condition (Kuivaniemi et al., 1997; Myllyharju and Kivirikko, 2001, 2004). For instance, osteogenesis imperfecta (OI), a connective tissue disease where bones are brittle and fragile, has been found to associate with single base mutations that results in the change of one Gly to one of 8 bulkier residues (C,S,R,E,V,D,A,W) in the Gly-X-Y repeating sequence of type I collagen (Byers and Cole, 2002; Kuivaniemi et al., 1997). More than 600 Gly sites along the α1 and α2 chains of type I collagen are identified and the severity of OI phenotype ranges from mild cases with skeletal fractures to perinatal lethal cases (Byers and Cole, 2002; Kuivaniemi et al., 1997; Marini et al., 2007). The identity of the residue replacing Gly is indicated as one of the factors affecting OI severity (Byers and Cole, 2002; Kuivaniemi et al., 1997). Alterations arising from Gly$\rightarrow$X mutations may occur at different levels ranging from defective folding of the triple-helix to alteration of higher order structure related to ligand binding or self recognition, yet the molecular mechanism and the correlation between substitution identity and phenotype is not clear. We have studied the effects of substitutions on the structural and dynamic properties of the triple helix, and how these alterations are correlated to the type of Gly$\rightarrow$X substitution (where X is Ala, Ser, Arg, Asp), and to the type of neighboring sequences surrounding the mutation. The studies

provided clues to elucidate the molecular basis of connective tissue diseases resulting from Gly mutations.

Contrary to fibrillar collagens that have a strict regularity of Gly as every 3rd residue throughout the triple helix domain, non-fibrillar collagens do not obey this requirement. All types of non-fibrillar collagens contain interruptions in the Gly-X-Y repeating pattern (Kielty and Grant, 2002). For instance, more than 20 breaks in the repeating $(Gly\text{-}X\text{-}Y)_n$ pattern are found in the ~1350 residues of the biologically critical type IV collagen, which forms a network-like array in basement membranes of all multicellular animals. Evidence suggests that these breaks are of functional importance, playing a role in molecular or higher order structure, or serving as recognition sites for interactions (Kielty and Grant, 2002; Miles et al., 1995). Despite the presence of these breaks in non-fibrillar collagens, the replacement of one Gly residue in the Gly-X-Y repeat can lead to a clinical phenotype (Hudson et al., 1993; Hudson et al., 2003; Lemmink et al., 1997). More than 150 Gly substitution mutations leading to Alport Syndrome with progressive kidney failure have been defined within the X-linked α5 chain of type IV collagen (Hudson et al., 1993; Hudson et al., 2003; Lemmink et al., 1997), while Gly substitution mutations in type VII collagen lead to the dystrophic form of epidermolysis bullosa, with scarring and blistering skin (Hudson et al., 1993; Hudson et al., 2003; Lemmink et al., 1997). It is of great interest to understand how fibrillar collagens can accommodate so many breaks but not Gly substitutions. The structural consequences of breaks in the triple helix are poorly understood. A kink was found at a interruption site in a non-collagenous protein (Brodsky-Doyle et al., 1976; Kilchherr et al., 1985). Alternatively, an interruption could lead to a flexible site within the triple helix (Brazel et al., 1987; Hofmann et al., 1984).

Visualization of type IV collagen chains by rotary shadowing shows some flexible sites at several sequence interruptions. In addition to resulting in structural perturbations, breaks in the $(Gly\text{-}X\text{-}Y)_n$ pattern can also serve as binding sites of biological importance. For example, the sequence of the α1 chain of type IV collagen contains a break GAKGEPGEFYFDLRLKGDKGDP, which was shown to promote the adhesion of melanoma, ovarian carcinoma, and Jurkat cells (Furcht et al., 1994; Miles et al., 1994). A kink, flexible site, or loss of helix registration would introduce a structurally distinct site that could be involved in recognition. Breaks in the form $Gly\text{-}X\text{-}Y\text{-}Gly\text{-}(AA)_n\text{-}Gly\text{-}X\text{-}Y$ are found of varying lengths in non-fibrillar collagens, where n=1, to long breaks where n=26 (Brazel et al., 1987; Zhou et al., 1992). We have looked at two most common types of breaks where n=1 and n=4 (denoted as G1G and G4G breaks), and defined the structural and dynamic perturbations and changes of hydrogen bonding of the triple helix as a result of these breaks. It has provided information about the relationship between G1G and G4G breaks and helped understand the molecular basis of structural and dynamic alterations of breaks in collagen.

1.2. Collagen peptides and NMR

Triple helical peptide models that satisfy the $(Gly\text{-}X\text{-}Y)_n$ sequence requirement and have a high content of imino acids will form stable triple helices in aqueous solution. The model peptide approach has been used to model different regions of collagen and to characterize the effect of sequence variation, folding, ligand binding, recognition, collagenolytic protease activity and other biological activities (Baum and Brodsky, 1999; Bella et al., 1994; Brodsky and Persikov, 2005; Buevich and Baum, 2001; Consonni et

al., 2000; Frank et al., 2001; Goodman et al., 1998; Heidemann and Roth, 1982; Lauer-Fields and Fields, 2002; Mayo, 1996; Okuyama et al., 2004; Renner et al., 2004; Sakakibara et al., 1973; Siljander et al., 2004). Model Peptides for biological analysis and structural studies are typically of the type $(GPO)_n$-$(GXY)_n$-$(GPO)_n$. A number of different approaches to peptide design have been developed over the years and include heterotrimeric triple helices that are covalently crosslinked and homotrimeric triple helices formed by spontaneously self-assembly of the three identical chains (Baum and Brodsky, 1999; Goodman et al., 1998; Lauer-Fields and Fields, 2002; Ottl et al., 1996). We use the homotrimeric peptide approach and the peptide models are 30-mer long, with $(GPO)_n$ repeats incorporated at either or both ends of the peptides to enhance stability, and ^{13}C and/or ^{15}N labeled for NMR studies.

NMR provides a wealth of molecular information about the structural properties and conformational dynamics of peptides. NMR on collagen model peptides allows the residue-specific characterization of molecular features of sequence variations in the triple helix. Our lab has applied NMR techniques to study collagen model peptides with a standard triple helical sequence or with sequence variations (Baum and Brodsky, 1999; Buevich and Baum, 2001; Fan et al., 1993; Hyde et al., 2006; Li et al., 1993; Liu et al., 1996). The previous studies have been primarily focused on the measurement of dynamic properties and folding rates on selectively ^{15}N labeled collagen model peptides (Buevich and Baum, 2001; Fan et al., 1993; Liu et al., 1998; Xu et al., 2003). The dynamic studies provided characterization of ^{15}N relaxation properties of the triple helix and the influence of specific amino acid sequences on the dynamics of monomer and trimer form (Bhate et al., 2002; Fan et al., 1993; Liu et al., 1998; Xu et al., 2003). The real time folding studies

helped probe individual kinetic steps in collagen triple helix folding pathways, and identified the effect of sequence variations on folding (Buevich and Baum, 2001; Buevich et al., 2000; Buevich et al., 2004; Hyde et al., 2006; Liu et al., 1996). We have extended the use of NMR spectroscopy on $^{13}C/^{15}N$ doubly labeled peptides, to obtain the first solution model of peptides containing breaks in the $(Gly\text{-}X\text{-}Y)_n$ sequence and of peptides containing a Gly to Ser mutation, to allow structural and dynamic characterization in details, and to predict the existence of hydrogen bonds. These studies helped us understand more about the consequences and molecular basis of collagen sequence variations.

1.2.1. NMR structural characterization

NMR techniques for the determination of three-dimensional biomolecules structure have been rapidly developed during the past decades (Bax et al., 2001; Bax et al., 1994; Grzesiek et al., 1993; Guntert, 1998; Kim and Szyperski, 2003; Pervushin et al., 1997; Sutcliffe and Dobson, 1991; Tjandra and Bax, 1997a; Wuthrich, 2001). NMR and x-ray crystallography provide complementary techniques for protein structure determination. X-ray crystallography needs single protein crystals, whereas NMR allows the structure determination under near-physiological conditions in solution. NMR structure determination for normal globular proteins has played a very important role in understanding structure-function relationship (Wuthrich, 2001). The triple helical peptide, however, is very different from normal globular proteins and is a challenging system for NMR studies because of its very anisotropic shape, Gly-X-Y repeating sequence, and three homotrimeric chains. The long rod like shape of the triple helix gives

rise to a high molecular tumbling rate and broad NMR resonances. An increase in the linewidth of a peak will decrease the signal to noise ratio of the peak. The NMR spectra of the triple helical peptides encounter similar sensitivity problems as proteins with very high molecular weight. The Gly-X-Y repeating sequence and three homotrimeric chains result in a high degree of chemical shift degeneracy and a difficulty in identifying three non-equivalent chains. The limited data and lack of long-range distance information resulting from the shape and the sequence of triple helical peptides makes the NMR structure determination intractable. Therefore, alternative strategies of NMR incorporated molecular modeling have been developed for triple helix peptides other than those for globular proteins. We have developed strategies for structural characterization of triple helical peptides, including the assignment of three non-equivalent chains, the measurement of NMR structural parameters related to distances and dihedral angles, and the visualization of solution structures using molecular modeling incorporating NMR data. The above strategies were applied to the collagen peptides to define the structural consequences resulting from Gly to X substitutions and breaks in the triple helix sequences, to study whether structural perturbations such as a bulge or untwisting exist at the sequence variation sites, and to help elucidate the structure–function relationships of collagen and the molecular basis of collagen diseases.

1.2.2. NMR dynamic characterization

NMR spectroscopy is a useful tool to probe the internal molecular mobility and dynamic features of proteins (Cavanagh et al., 1996; Englander and Krishna, 2001; Kay, 2005; Palmer, 2004). The powerful and versatile NMR dynamic studies can yield detailed

insights into the conformational dynamic basis of protein functions. A variety of NMR dynamics measurements cover a wide range of time scales spanning from picoseconds to hours (Englander and Krishna, 2001; Huyghues-Despointes et al., 2001; Kay, 2005; Palmer, 2004, 2001; Palmer and Massi, 2006; Palmer et al., 2005). Hydrogen exchange measurements can be used to detect local unfolding dynamics in the slow millisecond to second time scales (Englander and Krishna, 2001; Huyghues-Despointes et al., 2001). In the hydrogen exchange experiment, the exchange rates of the amide protons with solvent are measured. For a buried proton, dynamic processes occur to expose the proton and make it solvent accessible. For triple helical peptides, Gly residues are buried in the center of the molecule while X and Y residues are exposed to solvent. Local unfolding or "breathing" of the triple helix allows the Gly residues to exchange with solvent but with a much slower rate than the X and Y residues in a standard triple helical conformation (Fan et al., 1993). We have applied hydrogen exchange measurements on triple helical peptides with Gly to X substitutions and breaks, and characterized alterations of the flexibility around the sequence variation sites, to understand the dynamic fluctuations of triple helices that could be involved in the binding and interaction phenomena of collagen.

1.2.3. NMR hydrogen bonding characterization

NMR spectroscopy provides detailed characterization of hydrogen bonding properties, which play a key role in the formation and stabilization of protein structures. Many NMR parameters have been shown to correlate with hydrogen bonds, such as hydrogen exchange (Englander and Krishna, 2001; Englander et al., 1996), amide proton

temperature gradient (Andersen et al., 1997; Baxter and Williamson, 1997; Cierpicki and Otlewski, 2001), chemical shifts (Tjandra and Bax, 1997b; Wagner et al., 1983), and scalar couplings across hydrogen bonds (Cordier and Grzesiek, 2002; Grzesiek et al., 2001). In the hydrogen exchange experiments, the protection of labile amide protons from solvent are related to hydrogen bonding, solvent exposure, as well as local unfolding events (Englander and Krishna, 2001; Huyghues-Despointes et al., 2001). Amide proton temperature gradients have been used as a complementary method to hydrogen exchange to investigate hydrogen bonding properties in proteins (Andersen et al., 1997; Baxter and Williamson, 1997; Cierpicki and Otlewski, 2001; Merutka et al., 1995). Amide temperature gradients correlate with hydrogen bonds and solvent exposure, although they are less influenced by exposure (Baxter and Williamson, 1997). It is an easily measured and reproducible parameter, and an amide temperature gradient more positive than -5 or -4 ppb/°C is considered as a good indicator of hydrogen bonding (Andersen et al., 1997; Baxter and Williamson, 1997; Cierpicki and Otlewski, 2001). Because there are multiple factors determining both the hydrogen exchange rate and amide temperature gradient, it has been suggested that these two kinds of measurements can be combined together to more accurately estimate hydrogen bonds (Baxter and Williamson, 1997). For the triple helical peptides, amide protons of Gly residues are buried and form standard Rich-Crick II hydrogen bonds with the carbonyl group of the X residue in the adjacent chain. Hydrogen bonding characterization was performed using the above two measurements along with the model structures to probe the alteration of hydrogen bonding around the Gly to X substitutions and breaks in the triple helix, and to contribute to understanding the relationship between triple helix stability and function.

1. 3. Scope of the dissertation

This thesis focuses on the conformation and dynamic characterization of triple helical peptides, and how these properties are affected by collagen sequence variations, including breaks in the $(Gly\text{-}X\text{-}Y)_n$ triple helical repeats in non-fibrillar collagen and by Gly to X substitutions in fibrillar collagen. Mutation-specific equilibrium states of peptides, individual chain assignments and chain stagger were obtained for the first time on selectively $^{13}C/^{15}N$ doubly labeled peptides allowing the development of strategies for the visualization of NMR solution models. Dynamic and hydrogen bonding perturbations at the sequence variation sites were investigated using hydrogen exchange and temperature gradient experiments. Chapter 2 describes the peptide design approach, and their preparation using synthetic peptides and bacteria expressed recombinant peptides. Chapter 3 focuses on the development of NMR methodology for the triple helical peptide system, including structural characterization using NMR and molecular modeling and identification of equilibrium intermediates using translational diffusion measurement. Chapter 4 describes the application of NMR spectroscopy to characterize the conformation and dynamic consequences of natural breaks in the non-fibrillar collagen $(Gly\text{-}X\text{-}Y)_n$ repeat. Chapter 5 describes the application of NMR on peptides that model different Gly to X substitutions, to study the impact of the substitution on the equilibrium states, and on the structural and dynamic properties. The information obtained from these studies helps us understand the molecular basis of the sequence variations in collagen and helps elucidate the molecular mechanism of collagen diseases.

Chapter 2

Peptide design and preparation

2.1. Introduction

Triple helical peptide models of collagen are designed to be approximately 30 residues long, imino acid rich, and need to be $^{15}N/^{13}C$ labeled for NMR studies. Model peptides will form stable homotrimeric triple helices in aqueous solution by spontaneous self-assembly of the three chains. There are two complementary approaches to obtain labeled triple helical peptides: (1) solid-phase peptide synthesis to obtain selectively $^{13}C/^{15}N$ labeled Hyp containing synthetic peptides, and (2) bacterial expression to obtain recombinant uniformly labeled Hyp free peptides. Synthetic peptides are used for most cases, while recombinant peptides can be used when the labeling of all residues becomes important and when Hyp is shown not be biologically important.

Solid-phase synthesis using a peptide synthesizer is a very convenient way for obtaining peptides in a short time and at a low cost. It allows the incorporation of selective ^{15}N or $^{15}N/^{13}C$ labeling and Hyp in the peptides. Hyp containing selectively labeled synthetic peptides have been used as the primary research system of the collagen projects for years (Baum and Brodsky, 1999). One inherent problem of synthetic peptides is that peptide deletions will be produced due to incomplete coupling. The efficiency of the coupling reaction is peptide sequence dependent. Failure to remove impurities results in intense non-associating monomer peaks in the NMR spectra and interferes with quantitative measurement and low-sensitivity triple resonance experiments. In this chapter, I describe the purification of synthetic peptides using liquid chromatography. Under optimal purification conditions, non-associating monomer impurities can be removed and set the basis for implementing low-sensitivity NMR experiments.

The second approach for obtaining labeled peptides is via a recombinant bacterial expression system. Bacterial expression has been widely applied in obtaining high levels of recombinant proteins and in uniformly labeling of isotopic-rich protein for NMR studies (Kay and Gardner, 1997; Wagner, 1993; Wuthrich, 2003). Recently collagen-like domains scl1 and scl2 of group A streptococcus were obtained from bacterial expression (Xu et al., 2002). The adaptation of bacterial expression to collagen peptides will complement synthetic peptides and allow uniform labeling and the high homogeneity of the peptides free from peptide deletions. The recombinant collagen peptides from bacterial expression will not be hydroxylated because of the absence of the prolyl 4-hydroxylase enzyme in bacteria. Though Hyp is important for stabilization of the triple helix, studies have shown that Hyp is not essential for some biological activities such as integrin binding (Humtsoe et al., 2005). The conversion from $(GPO)_n$ to $(GPP)_n$ in the terminal regions of the peptides will not effect the biological significance and the stability loss of 1.8°C in T_m/ triplet (Persikov et al., 2005) can be compensated for by increasing the number of (GPP) triplets relative to (GPO) triplets. In addition, model triple helical peptides with $(GPP)_n$ repeats rather than $(GPO)_n$ repeats in the terminal regions have been used in the biological assays (Raynal et al., 2006). Another concern of the bacterial expression is the overlap problem especially for Gly residues in the NMR spectra arising from repetitive GXY sequences. By carefully designing non-repeptitive peptide sequence, resonances of the central Hyp free region can be well resolved. In this chapter, I describe the application of molecular biology to establish the bacterial expression system for collagen peptides.

2.2. The purification of synthetic collagen model peptides

The peptide purity issue has existed in our lab: the strong monomer peaks in the HSQC spectra at low temperature are not consistent with CD and ultracentrifuge studies that show that a very small amount of monomers exist in solution. In addition, the strong monomer peaks interfere with trimer characterization if a trimer resonance is close to a monomer resonance. Removal of monomer peaks is critical in low-sensitivity NMR triple resonance experiments, in which the relaxation properties of the monomer and trimer are very different and the trimer resonances are completely buried under the strong monomer resonances. Previously, the purification of synthetic collagen peptides had been performed at room temperature and could not remove the strong monomer peaks. Recently, attempts have been made to redevelop the methods of liquid chromatography to purify the synthetic triple helical peptides. Temperature, gradient, sample concentration, and different methods of chromatography were optimized for peptide purification. The peptide $(GPO)_4$-GAAVM-$(GPO)_4$GY (denoted as GAAVM peptide) is used as an example to describe the purification protocol below.

2.2.1. Sample preparation

Peptides were synthesized by Tufts University Core Facility (Boston, MA). The peptides include a Tyr at the C terminus, and the concentration was determined using the molar extinction coefficient 1280 ($M^{-1}cm^{-1}$) at 280 nm on a Beckman model DU640B spectrophotometer. The identity of the peptides was confirmed by Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) mass spectrometry at Tufts University and UMDNJ.

2.2.2. Effect of temperature

Temperature effects of purification were investigated for reversed phase chromatography. For the low temperature chromatography, peptides were purified at 4°C on a GE Amersham AKTA Purifier10 FPLC system. The reversed phase chromatography was applied using a Waters XTerra MS C18 Prep column (5μm, 15 mm i.d. × 190 mm), mobile phase A of 99.9% H_2O / 0.1% trifluoroacetic acid and mobile phase B of 99.9% acetonitrile / 0.1% trifluoroacetic acid. The gradient of elution was optimized and a shallow gradient of 14%-24% B in 12 CV at a flow rate of 8 ml/min was applied. Peaks were detected at 215nm and 280nm to monitor peptide bond and C-terminal Tyr. High temperature chromatography was performed at room temperature (22°C) on a Perkin Elmer PE200 series system using a Waters XTerra MS C18 Ana column (5μm, 4.6 mm i.d. × 150 mm). The chromatography method is scaled down correspondingly for the column to make it comparable with the low temperature method. The elution curves at both temperatures are shown in figure 2.1. At Room Temperature (22°C), there was a single major peak (Figure 2.1A). While at 4°C, two peaks were observed (Figure 2.1B).

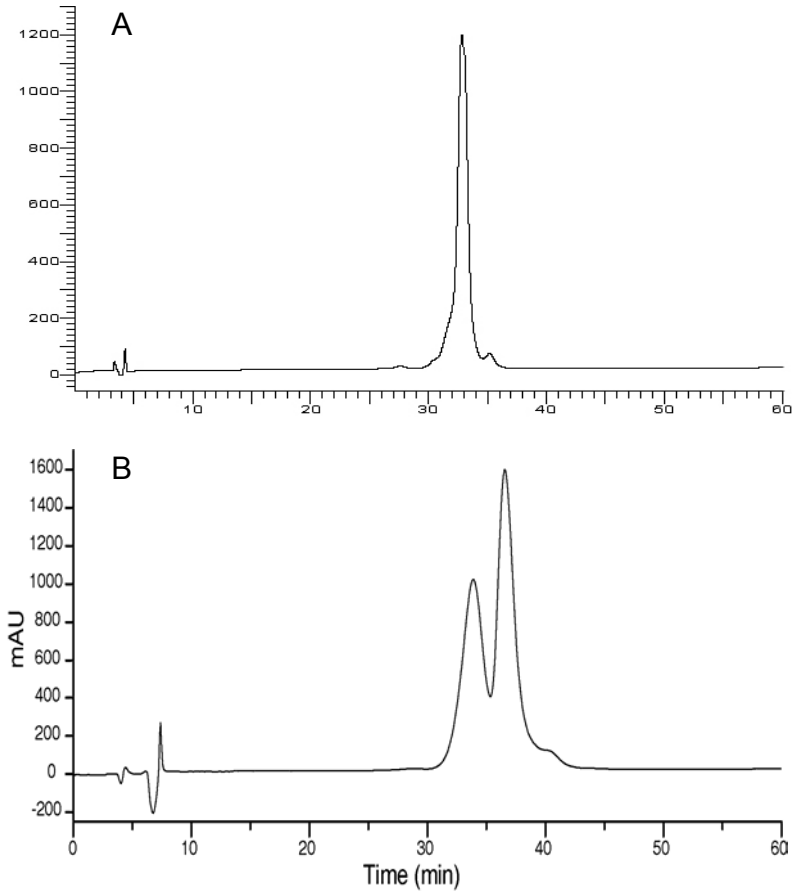

Figure 2.1 Reversed phase chromatography of the GAAVM peptide at 22°C (A) and 4°C (B). At Room Temperature (22°C), there was a single major peak, while at 4°C two peaks were observed.

2.2.3. Identification of two fractions at low temperature

In order to determine the identity of the two peaks at 4°C, fractions of two peaks were collected, directly characterized by ESI ion trap mass spectroscopy or lyophilized for MALDI-TOF mass spectroscopy and NMR characterization. The ESI mass spectroscopy gave similar results to the MALDI-TOF mass spectroscopy and the MALDI-TOF is shown here (Figure 2.2B and 2.2D). The mass spectra for both fractions show a major peak of m/z that is consistent with the correct molecular weight of the peptide. However, fraction 1 (Figure 2.2B) showed many peaks in the region of MW 2000-3000 in the mass spectra, while the fraction 2 showed fewer (Figure 2.2D). There are equal-distanced peaks in both fractions in the lower molecular range of 500~1300 and it could be due to the artifact in the mass spectroscopy of peptides with repeating sequences. The HSQC spectra of the two fractions at equilibrium showed big differences in the monomer resonances as circled in figure 2.2C and 2.2E. The intensities of the monomer resonances in fraction 1 are much stronger than those in fraction 2.

The purification of another peptide $(GPO)_4$-GF-$(GPO)_5$GY (denoted as GFG peptide) at low temperature is also shown (Figure 2.3). It presents a similar feature as for the GAAVM peptide in having multiple peaks in the reversed phase chromatography and in having one fraction with lower molecular weight than expected for the GFG peptide on the MALDI-TOF mass spectroscopy. It confirms that at low temperature peptide deletions can be separated.

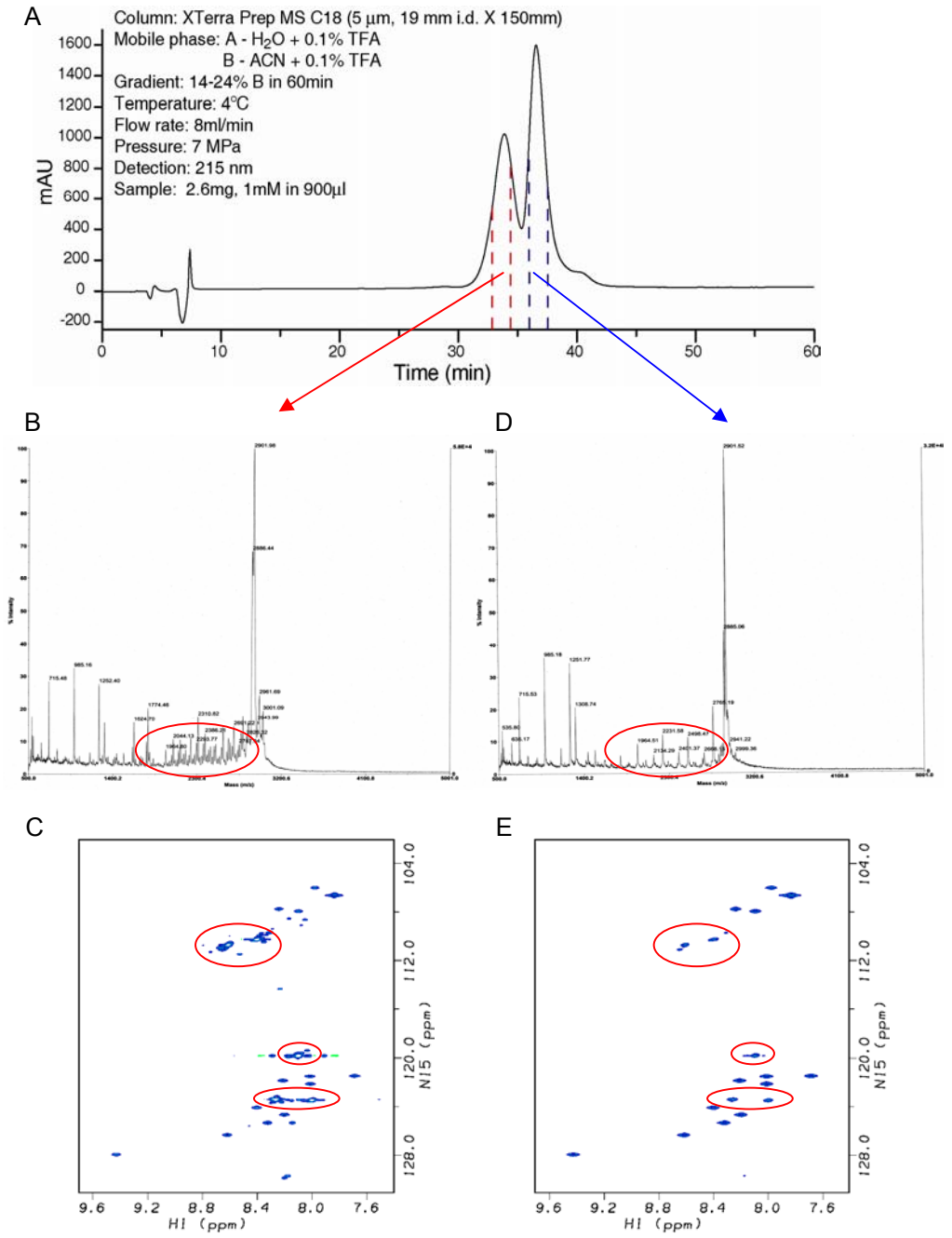

Figure 2.2 Purification of the GAAVM peptide.
A, Reversed phase chromatography of the GAAVM peptide. Fractions collected for further characterizations are shown in dashed lines.
B and D, MALDI-TOF mass spectroscopy of two fractions. The m/z of the major peaks in both spectra is consistent with the GAAVM monomer molecular weight. The differences in two spectra are circled.
C and E, HSQC spectra of two fractions at 5°C.

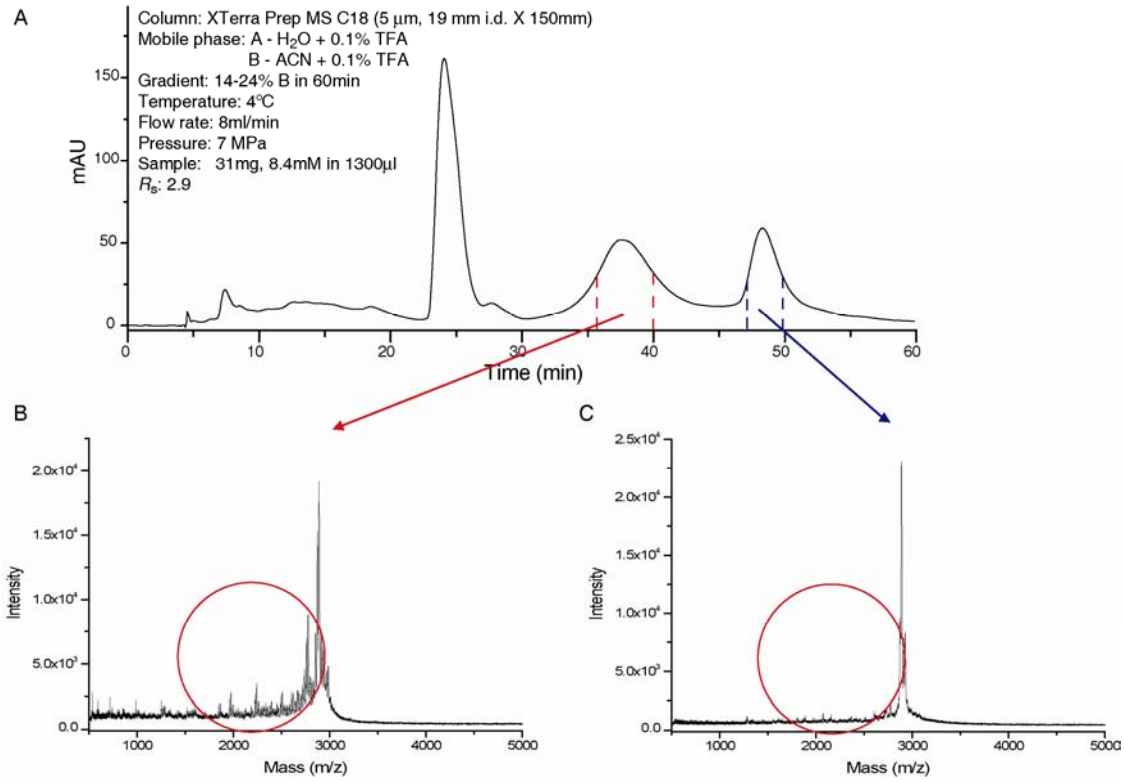

Figure 2.3 Purification of the GFG peptide.

A, Reversed phase chromatography of the GFG peptide. Fractions collected for further characterization are shown in dashed lines.

B and C, MALDI-TOF mass spectroscopy of two fractions. The m/z of the major peaks in both spectra are consistent with the GFG monomer molecular weight. The differences in the two spectra are circled.

For further identification of the two fractions, the GAAVM peptide was heated at 50°C for 15min before loading onto the column at 4°C. The melting temperature of the GAAVM peptide is 29°C and the high temperature treatment denatured the peptide. During loading and running the reversed phase chromatography, the peptide refolded. Comparing the elution curves of heat denatured peptide and the native peptide of the same amount, it showed that the relative peak area of two fractions changed (Figure 2.4). The peak area of fraction 1 significantly increased in the heat denatured sample, therefore the fraction 1 was assigned to monomer peak and the fraction 2 was assigned to trimer peak. This is consistent with the same m/z for the major peak in both fractions in the mass spectra (Figure 2.2B and 2.2D).

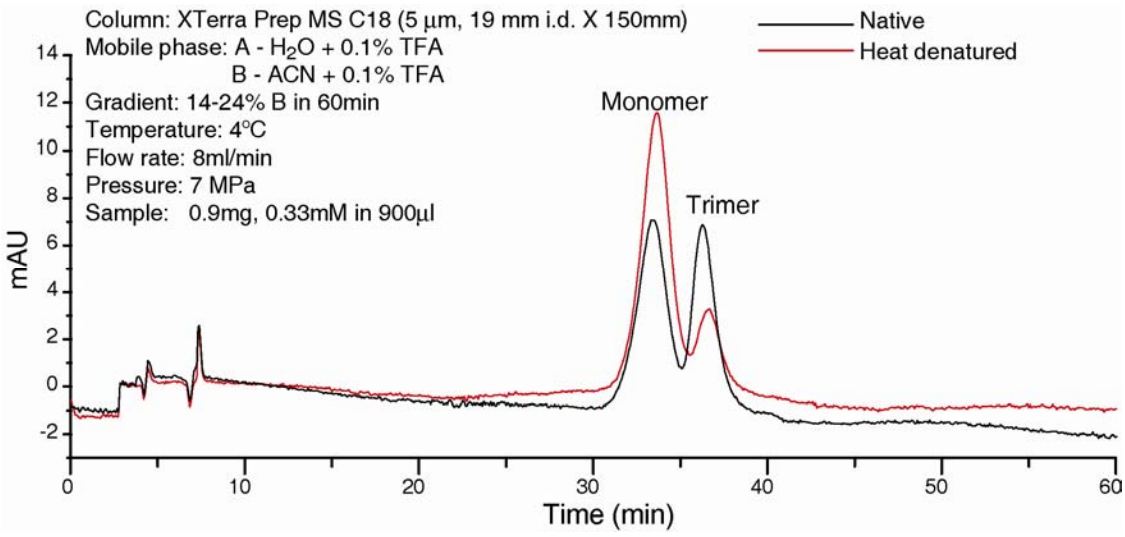

Figure 2.4 Reversed phase chromatography of GAAVM peptide under native (black line) and heat denatured (red line) conditions.

The data above indicated that the strong monomer resonances in fraction 1 (Figure 2.2C) could be from non-associating monomers that are peptide deletions as indicated in the mass spectroscopy (Figure 2.2B). Another possibility is that the strong monomers are from misfolding or partially unfolding species that are trapped by the energy barrier. Molecular dynamics studies showed that there are multiple folding states that adopt local energy minima in collagen model peptides (Stultz, 2002). In order to investigate this possibility, the two fractions were heated to a temperature above its melting temperature to overcome the energy barrier and then refolded at low temperature (Figure 2.5). In both fractions, at 50°C the sample showed only monomers indicating complete unfolding (Figure 2.5C, F). After cooling down to 5°C and reaching equilibrium, the sample showed the same HSQC spectrum as the spectra before heat denaturation, suggesting that the strong monomer resonances in fraction 1 are from non-associating peptide deletion species.

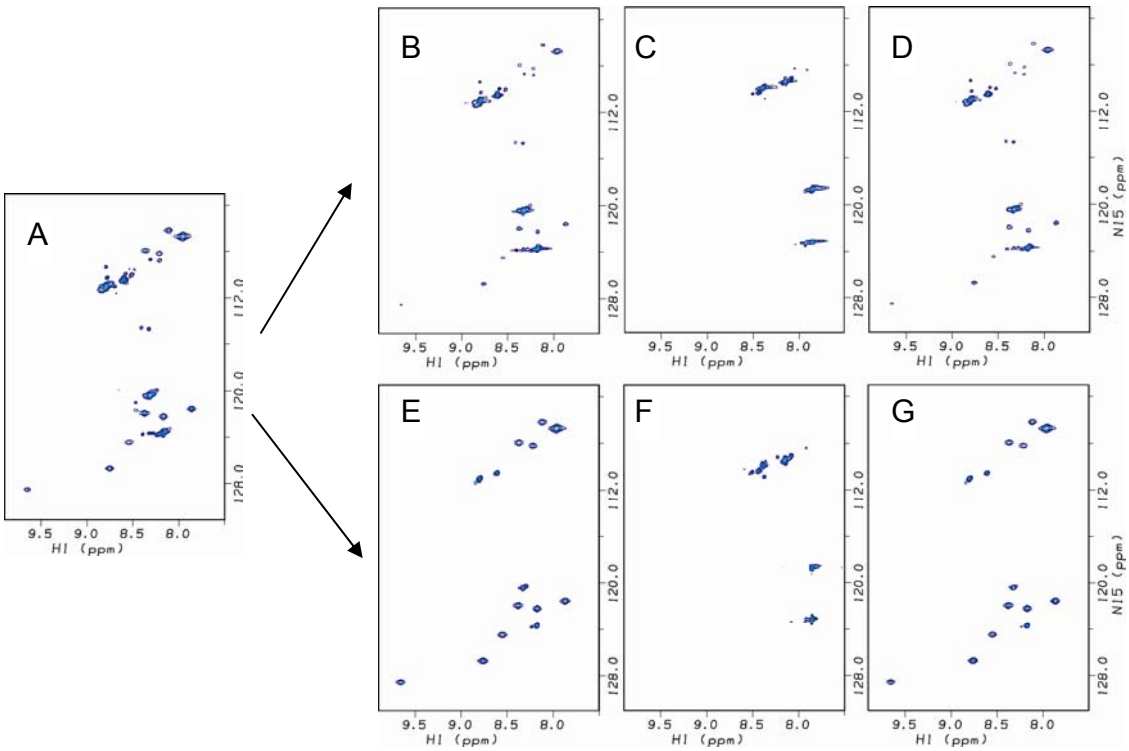

Figure 2.5 HSQC spectra of the ^{15}N labeled GAAVM peptide under conditions: before purification at 5°C (A); Monomer fraction from purification (B-D); Trimer fraction from purification (E-G).

A, before purification at 5°C;

B and E, after purification at 5°C;

C and F, heat denatured at 50°C;

D and G, refold and reached equilibrium at 5°C after heat denaturation.

2.2.4. Effect of concentration

Based on the above data, it can be summarized that the successful purification is based on the different equilibrium properties of pure peptide and peptide deletions. The pure peptide has an equilibrium consisting of monomer:trimer interconverting species, while the peptide deletion can not fold into trimer and elutes mostly as a monomer state under the conditions studied. Therefore, the purification is based on a good separation of monomer and trimer. In order to optimize the purification, it would be necessary to optimize the concentration of samples loaded on the reversed phase column, as the concentration would shift the equilibrium for both pure peptide and peptide deletions. The GAAVM peptide at different concentrations was allowed to reach equilibrium and was loaded onto the reversed phase column. The relative peak areas of monomer and trimer peaks changed with different sample concentrations (Figure 2.6A). HSQC spectra were acquired for samples loaded with 2.6mg, 7.7mg and 15.4mg. Intensity ratios of trimer to monomer resonances were calculated as an estimate of separation efficiency. The 7.7mg sample had a trimer to monomer ratio of 3.2, higher than the trimer to monomer ratio of the 2.6mg and 15.4mg samples, which had ratios of 2.6 and 2.9, respectively. This indicates that the 7.7mg was the optimal loading amount for purifying the GAAVM peptide.

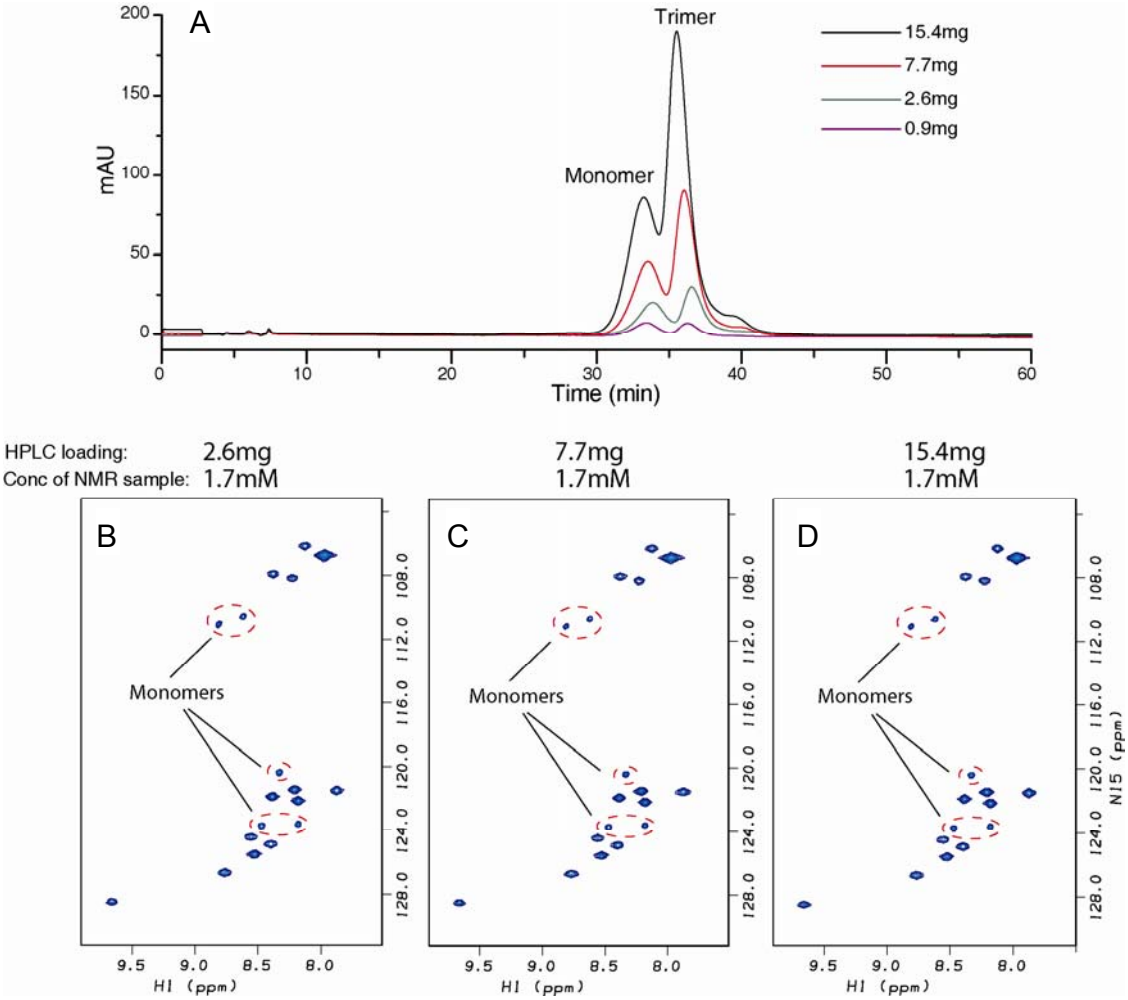

Figure 2.6 Effects of concentration of the GAAVM peptide in reversed phase chromatography.

A, Elution curves for samples with different amounts. All samples were allowed to reach equilibrium for at least 3 days and had a 900μl loading volume.

B-D, HSQC spectra of the trimer fraction for samples with loading amount of 2.6mg (B), 7.7mg (C), and 15.4mg (D).

2.2.5. Effect of peptide sequence and the complementary size exclusion chromatography

The purification is peptide sequence dependent. In the case of peptide $(GPO)_4$-GPV-SPA-GAR-$(GPO)_4$-GY (denoted as peptide GPO(G16S)), reversed phase chromatography failed to achieve good separation for monomer and trimer (Figure 2.7A). Size exclusion chromatography was applied using an Amersham Superdex 75 Hiload Prep column (26 mm i.d. × 600 mm), the solvent was 20% acetonitrile / 79% H_2O/ 1% acetic acid. The elution volume was 1.1 column volume and the flow rate was 2.6 ml/min. The HSQC spectra showed that the strong monomer populations can be removed by the size exclusion chromatography (Figure 2.7B).

Despite the success of purification using size exclusion chromatography, it can be only used on peptides that have already been purified by reversed phase chromatography. For crude product from peptide synthesis, size exclusion chromatography can not achieve good separation. So it is used as a secondary and complementary method to reversed phase chromatography.

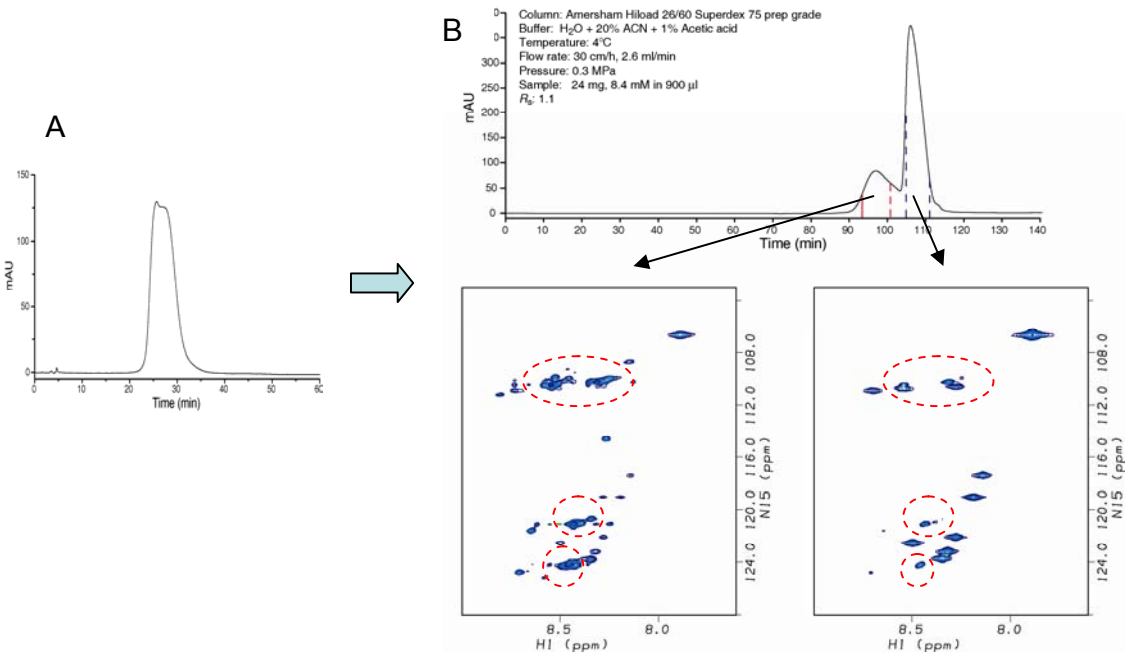

Figure 2.7 Application of size exclusion chromatography on the purification of peptide GPO(G16S).

A, Reversed phase chromatography showed poor separation.

B, size exclusion chromatography showed good separation. Two fractions were characterized by NMR HSQC spectra, and monomer resonances were highlighted in red circles. In the second fraction, most strong monomer peaks were removed.

2.2.6. Conclusion

In conclusion, the newly developed method of reversed phase chromatography and size exclusion chromatography was applied to successfully purify synthetic peptides. In NMR spectra, most of the non-associating monomer resonance can be removed via the purification. It opened the door for the implementation of triple resonance experiments. Without the interference of strong monomer resonances, the low sensitivity NMR experiments can be successfully conducted on trimers. The efficiency of purification is peptide sequence dependent and needs to be optimized for each specific peptide.

2.3. Bacterial expression approaches

2.3.1. Design of bacterial expression system

The design and development of a bacterial expression system for human collagen-like model peptides was performed in collaboration with Dr. Masayori Inouye's group. The recombinant polypeptide construct is shown in Figure 2.8 and features of this construct are shown below.

1) The $(GPP)_4(GXY)_n(GPP)_5GY$ context is designed for collagen model peptides, in which $(GXY)_n$ is variable and taken from real collagen sequences. Four and five GPP triplets are placed N- and C-terminal to the $(GXY)_n$, respectively, to stabilize the peptides. A Tyr residue is added at the C-terminus for UV absorbance to measure the peptide concentration.
2) This peptide is fused at the N-terminal end to proteinS to aid the stable expression of the short collagen-like polypeptides.
3) In order to remove the N-terminal fusion proteinS, a thrombin recognition site (LVPR$^{\downarrow}$GS; an arrow indicates the cleavage site) is added to the N-terminal of the collagen model peptides. As thrombin cleaves between Arg and Gly, a Pro residue is added after GS to satisfy the triple helical GXY pattern. Therefore, the peptide sequence after thrombin cleavage is $GSP(GPP)_4(GXY)_n(GPP)_5GY$.
4) In order to express all proposed polypeptides having different $(GXY)_n$ sequences in the center, *Sma* I and *Apa* I restriction enzyme sites are designed to add any desirable sequences between the N-terminal $GSP(GPP)_4$ region and the C-terminal $(GPP)_5$ region.

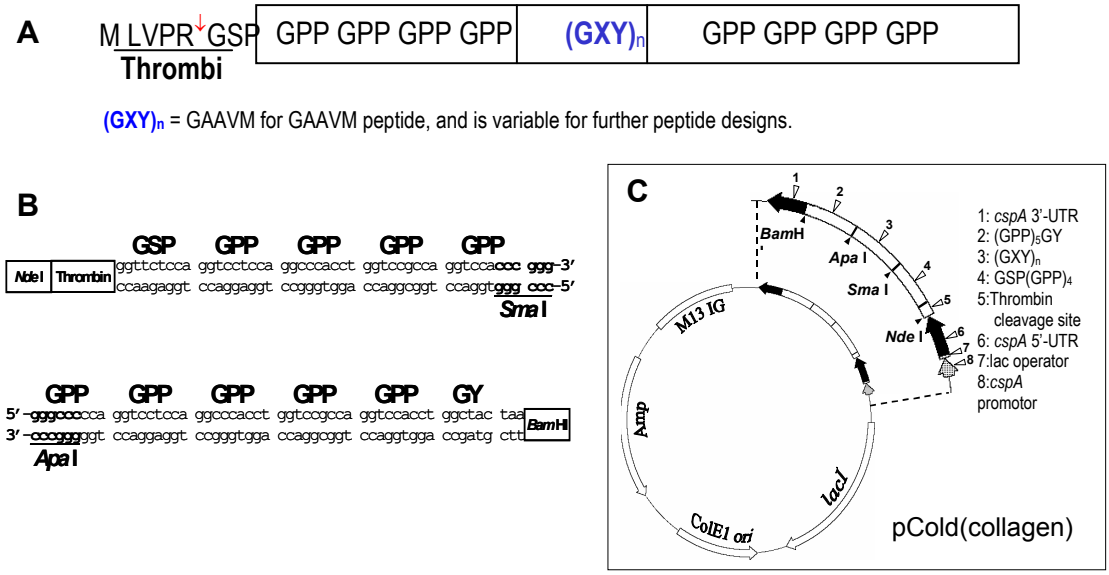

Figure 2.8 Design of bacterial expression system.

A, Schematic representation of the recombinant polypeptides. Collagen model peptides are designed to have a $(GPP)_4(GXY)_n(GPP)_5GY$ context as shown in the box, in which blue $(GXY)_n$ is taken from real collagen sequences, four and five GPP triplets are placed at the N- and C-terminus, respectively to stabilize the peptides. A Tyr is put at the C-terminus to enable UV absorbance to measure the peptide concentration. The N-terminus underlined LVPRGS is the recognition site of thrombin, which cleaves between Arg and Gly as shown with the red arrow. A Pro is added after GS to satisfy the triple helical GXY pattern.

B, DNA sequences for the N-terminal $(GPP)_4$ and C-terminal $(GPP)_5$ region. *Sma*I and *Apa*I are used to add any desirable $(GXY)_n$ sequence in between the N-terminal $(GPP)_4$ and the C-terminal $(GPP)_5$, *Nde*I and *Bam*HI are used to clone the interested DNA fragment into various other vectors.

C, The schematic map of a universal collagen cloning vector [pCold(collagen)]. The zoomed region shows the DNA for collagen model peptide region. The recognition sites for restriction enzymes *Sma*I, *Apa*I, *Nde*I and *Bam*HI are shown.

2.3.2. Development and application of bacterial expression system on the GAAVM peptide

The bacterial expression of the recombinant polypeptides was first applied to the GAAVM peptide, in which $(GXY)_n$ refers to GAAVM. The peptide sequence after thrombin cleavage is GSP-$(GPP)_4$-GAAVM-$(GPP)_5$-GY. The DNA fragments of N-terminal GPP region, central GAAVM region, and C-terminal GPP region were synthesized by Integrated DNA Technologies. Three DNA fragments were purified using PAGE gel, and cloned into the vector. Then the DNA of the full length collagen polypeptide was subcloned into a pCold III expression vector, forming the construct of proteinS-GAAVM fusion protein. The pCold III is a novel cold-shock inducible expression vector developed in Dr. Inouye's laboratory (Qing et al., 2004). The uniformly ^{15}N labeled proteinS-GAAVM fusion protein was expressed in *E. coli* BL21 strain by growing cells in minimum media at 37°C until O.D. reaches 0.6~0.7, then adding 1mM IPTG and growing cells at 15°C for 24 hours. The cells were harvested and cell lysis was conducted using French Press. The fusion protein was purified by ultracentrifugation to remove membrane proteins, 45% ammonium sulfate precipitation and DEAE Sephadex anion exchange column chromatography. The ProteinS-GAAVM was then treated with thrombin (Amersham) at 15°C overnight to cleave the GAAVM peptide. After thrombin digestion, the ProteinS and GAAVM mixture was dialyzed against Tris buffer (pH 9) to remove NaCl, and then loaded onto a Q-sepharose ion exchange column. Under no salt and high pH conditions, ProteinS bound to the column and the GAAVM peptide was collected as flow-through. The GAAVM peptide was further purified to remove traces of thrombin and desalted using reversed phase chromatography with Waters XTerra MS

C18 Prep column. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry (Figure 2.9).

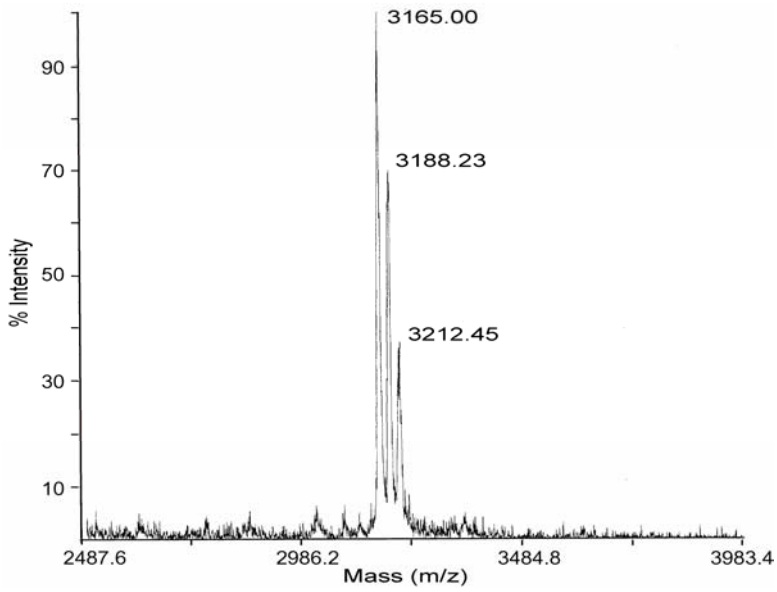

Figure 2.9 MALDI-TOF Mass spectrometry of proteinS/GAAVM mixture after treatment with thrombin. The mass of the major peak was consistent with the molecular weight of the GAAVM peptide.

The HSQC spectrum of the ^{15}N labeled GAAVM peptide in H_2O shows strong monomer resonances and a distinct weak trimer resonance for the Gly residues in the GPP repetitive sequences (Figure 2.10A). It indicates a momoner/trimer equilibrium with a low trimer population. In order to obtain the assignments for monomer resonances of G16 through G21, 3D ^{15}N edited TOCSY-HSQC (Fesik and Zuiderweg, 1988; Messerle et al., 1989) with a mixing time of 70ms and 3D ^{15}N edited NOESY-HSQC (Fesik and

Zuiderweg, 1988; Marion et al., 1989; Messerle et al., 1989) with mixing times of 350 ms were performed at 10°C. Monomers of S2 and Y37 are assigned based on random coil chemical shifts of Ser and Tyr.

In order to increase the stability of the peptide and increase the trimer population, a small amount of trimethylamine N-oxide (TMAO) was added to the peptide solution to make a final sample condition of 1M TMAO with 10% D_2O at pH 3. The HSQC spectrum of the peptide with 1M TMAO (Figrue 2.10B) shows distinct monomer and trimer resonances for all labeled residues. It suggests that the monomer/trimer equilibrium has been shifted by TMAO and the peptide is forming a stable triple helix. Comparing the HSQC of recombinant GAAVM with TMAO and the HSQC of synthetic GAAVM in the GPO environment, it shows a similarity in chemical shifts of trimers for those two peptides. In particular, the characteristic downfield shifted Val NH is clearly seen in both peptides. It suggests a similar triple helical conformation at the interruption site.

2.3.3. Conclusion

In conclusion, the bacterial expression system was established and used to successfully obtain the ^{15}N uniformly labeled GAAVM peptide. By carefully designing and choosing a peptide of high stability, the bacterial expression method shows promising prospect for obtaining uniformly labeled peptides.

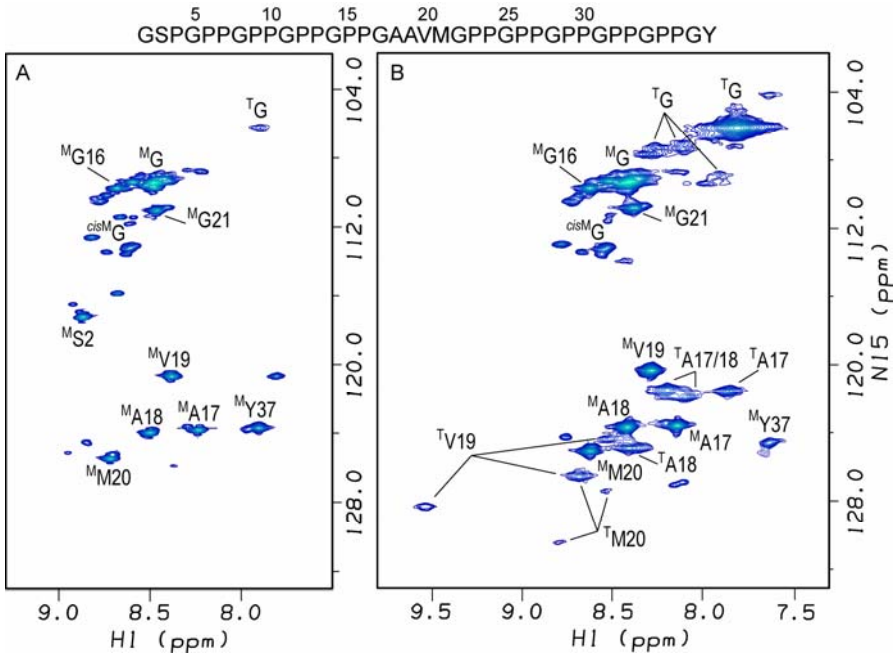

Figure 2.10 ^{1}H-^{15}N HSQC spectra of ^{15}N-uniformly labeled GAAVM peptide from bacterial expression system.

A, GAAVM peptide in 9:1 H_2O/D_2O, pH 3.

B, GAAVM peptide in 1M TMAO with 10% D_2O, pH 3. Triple helical resonances were given preliminary assignments by comparison with the spectrum of synthetic ^{15}N labeled GAAVM peptide. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T, respectively. The spectra were acquired at 0°C.

Chapter 3

NMR methods applied to triple helical peptides

3.1. Introduction

NMR provides a wealth of molecular information about the structural properties and conformational dynamics of proteins (Bax et al., 2001; Cavanagh et al., 1996; Dobson and Hore, 1998; Kay, 2005; Palmer, 2004; Wuthrich, 2003). NMR on collagen model peptides allows the residue-specific characterization of molecular features of the triple helix (Baum and Brodsky, 1999; Buevich and Baum, 2001; Goodman et al., 1998; Mayo, 1996). The NMR conformation and dynamics studies are contributing to understanding the molecular details of the unique triple helix system, and to understanding the relationship between structure, function and disease in collagen. Collagen, the most abundant protein in the human body, is a primary constituent in skin, bone, tendon, ligament, cartilage and membranous tissues (Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004). Collagen has a Gly-X-Y repeating sequence pattern, yet sequence variations that do not satisfy the Gly-X-Y requirement have been found in collagens (Kielty and Grant, 2002). Sequence variations include non-pathological breaks in the Gly-X-Y sequence in non-fibrillar collagen and the Gly-X substitutions of connective tissue disease in fibrillar collagen (Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004). NMR methodology is developed and applied to investigate the structural and dynamic consequences of sequence variations and how these are correlated to function and disease.

Collagen model peptides have a triple helical conformation that is very different from normal globular proteins. In the triple helical conformation, three left-handed polyproline II chains are supercoiled around a central axis and form a long-rod-shape trimer (Bella et al., 1994; Ramachandran, 1967; Rich and Crick, 1961). There is a one

residue stagger in three chains that can be named as leading, middle and trailing chains (Figure 3.1A). Gly residues, presenting as a every third residue in the Gly-X-Y repeating pattern, are closely packed in the center of the molecule. Residues at X and Y positions are exposed to solvent (Figure 3.1B). Collagen model peptides are a challenging system for NMR studies because of their very anisotropic shape, Gly-X-Y repeating sequence, and three homotrimeric chains. The broad linewidths of trimers arising from their anisotropic shape poses a requirement for high sensitivity of NMR experiments. The Gly-X-Y repeating sequence brings resonance overlap and precludes the traditional NMR structure determination. Three homotrimeric chains require chain specific sequential assignment and chain stagger identification, and it is very challenging in cases where a high degree of chemical shift degeneracy is present in three chains. Therefore, alternative strategies of NMR incorporated molecular modeling are developed for triple helix peptides other than those for globular proteins. Here I describe the development and application of NMR methodology to obtain chain identification, visualize the structure by molecular modeling incorporating NMR data, and to identify partially disordered intermediates in equilibrium states.

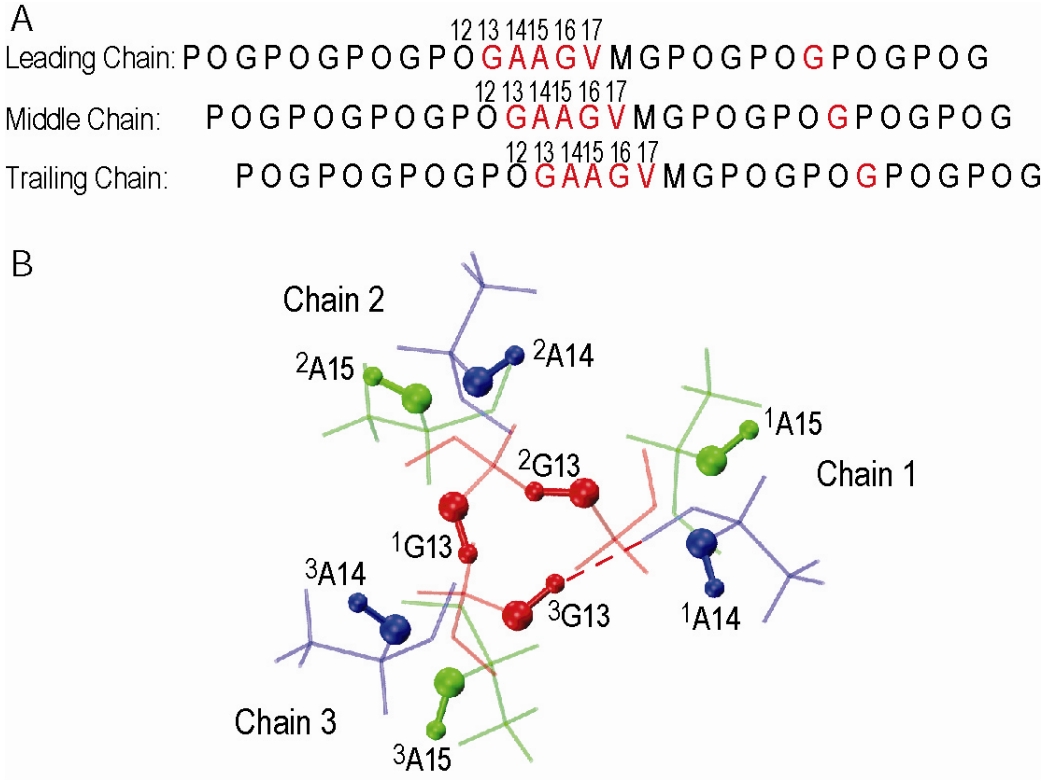

Fig. 3.1 Representation of the peptide sequence GAAGVM and the cross section view of the one residue stagger.

A, Sequence diagram of peptide GAAGVM shows the characteristic one residue stagger. The corresponding isotope labeled residues are colored in red.

B, Cross section view from the model structure of GAAGVM showing the GAA segment. GAA triplets are colored in red, blue, and green for G13, A14, A15, respectively. NH vectors of GAA are shown in ball and stick model and labeled. One example of the Rich and Crick II interchain hydrogen bonding is shown between the G13 NH of chain 3 and the A14 CO of chain 1.

3.2. Development of NMR methodology to characterize structural properties of triple helical peptides

3.2.1. Introduction

In order to perform detailed structural studies on triple helical peptides, chain identification needs to be obtained first for trimers. The application of NMR to obtain chain identification in triple helical peptides can be divided into two steps: chain specific sequential assignment and chain stagger identification. The incorporation of ^{13}C, ^{15}N doubly labeled residues in the triple helical peptides allows the application of NMR triple resonance experiments to obtain chain specific sequential assignments. A variety of triple resonance experiments have been developed and applied on globular proteins, including traditional experiments such as HNCA/HNCOCA and HNCO/HNCACO (Clubb et al., 1992; Ikura et al., 1990; Yamazaki et al., 1994), and newer approaches such as G-matrix Fourier Transform (GFT) experiments (Kim and Szyperski, 2003, 2004). Because of the anisotropic shape of the trimer molecule, triple helical peptides have low efficiency of magnetization transfer in triple resonance experiments resulting in low sensitivity. The homotrimeric nature of the triple helical peptides poses assignment problems when the chemical environments of three chains are similar, such as in a repetitive sequence. It results in high overlap of resonances and requires the acquisition of a suite of triple resonance experiments. After obtaining chain specific sequential assignment, the one residue stagger in three chains needs to be characterized in order to identify leading, middle and trailing chains. The stagger produces non-equivalent neighbors for a residue in three chains, and these subtle differences in the residue to residue contacts can be investigated using NOESY experiments (Figure 3.1A). The establishment of chain

assignments allows the detailed characterization of structural properties in triple helix peptides.

In contrast to globular proteins, solution structures of triple helical peptides can not be solved by NMR because of the very anisotropic and homotrimeric nature of the triple helix and therefore the lack of enough long range NOE distances. Alternative strategies of NMR incorporated molecular modeling are developed to incorporate NMR structural parameters and visualize the structure of triple helical peptides. A number of x-ray crystal structures are determined for triple helical peptides (Bella et al., 1994; Bella et al., 2006; Kramer et al., 1999; Kramer et al., 1998). This allows the application of molecular modeling to triple helical peptides. NMR structural restraints such as dihedral angles from J coupling measurements and proton distances from NOE experiments are incorporated in the molecular modeling. This procedure allows the generation of structures that are consistent with experimental data. In this section 3.2. I describe the chain identification and the measurements of structural parameters in the subsection 3.2.2., and the establishment of molecular modeling procedure in the subsection 3.2.3.

3.2.2. NMR chain assignments and the measurements of structural parameters

3.2.2.1. Material and methods

3.2.2.1.1. Sample preparation

Peptides Ac-$(GPO)_4$GAAGVM$(GPO)_4$GY-$CONH_2$, designated as the GAAGVM peptide, and Ac-$(GPO)_4$GAAVM$(GPO)_4$GY-$CONH_2$, designated as the GAAVM peptide, were synthesized by Tufts University Core Facility (Boston, MA) (Table 3.1). he peptides were made with selectively $^{13}C/^{15}N$ doubly labeled residues (underlined below):

residues G13, A14, A15, G16, V17, and G25 were labeled in peptide GAAGVM; residues G13, A14, A15, V16, and G24 were labeled in the peptide GAAVM. Peptides were purified using a Waters XTerra Prep C18 column on an Amersham FPLC system at 4°C and the identity of the peptides was confirmed by Matrix-Assisted Laser Desorption Ionization (MALDI) mass spectrometry. Samples for both peptides were prepared in 10% D_2O/90% H_2O at pH 2 with concentrations of 3.0 mM for GAAGVM and 1.8 mM for GAAVM.

TABLE 3.1. Peptide sequences of GAAGVM and GAAVM

Peptide name	Peptide sequence[a]
GAAGVM	Ac-GPO-GPO-GPO-GPO-**GAAGV**M-GPO-GPO-**G**PO-GPO-GY-$CONH_2$
GAAVM	Ac-GPO-GPO-GPO-GPO-**GAAV**M-GPO-GPO-**G**PO-GPO-GY-$CONH_2$

[a] ^{15}N and ^{13}C labeled residues are underlined and bolded.

3.2.2.1.2. Sequential assignments

NMR experiments were performed on a Varian Inova 600 MHz spectrometer equipped with a cryoprobe, and the G-matrix Fourier Transform (GFT) (5,3)D HACACONHN/HACA, CONHN experiments (Kim and Szyperski, 2003, 2004) were recorded on a Bruker 900 spectrometer equipped with a cryoprobe at the New York Structural Biology Center (NYSBC) to improve dispersion and sensitivity. For sequential assignments, ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) and a series of triple-resonance experiments were obtained at 25°C, including HNCO (Ikura et al., 1990)/HN(CA)CO (Clubb et al., 1992; Yamazaki et al., 1994), HNCA (Ikura et al., 1990), HA(CA)NH/HA(CACO)NH (Kay et al., 1991), HNCACB (Wittekind and Mueller, 1993)/CBCA(CO)NH (Grzesiek and Bax, 1992), and GFT (5,3)D HACACONHN/HACA,CONHN (Kim and Szyperski, 2003, 2004). The 2D ^{1}H-^{15}N

HSQC spectrum was recorded with 256(t_1) × 1024(t_2) complex points, and were recorded with spectral widths of 2000.0(F_1), 7022.5(F_2) Hz, using the gradient sensitivity-enhanced approach (Kay et al., 1992a). 3D HNCO and HNCA experiments comprised 32(t_1) × 90(t_2) × 512(t_3) complex points, and were recorded with spectral widths of 2000.0(F_1), 6000.0(F_2), and 7022.5(F_3) Hz. 3D HN(CA)CO experiment comprised 90(t_1) × 36(t_2) × 512(t_3) complex points, and were recorded with spectral widths of 6000.0(F_1), 2000.0(F_2), and 7022.5(F_3) Hz. 3D HA(CA)NH/HA(CACO)NH experiments were acquired with 66(t_1) × 32(t_2) × 512(t_3) complex points, and were recorded with spectral widths of 2000.0(F_1), 2000.0(F_2), and 7022.5(F_3) Hz. 3D HNCACB/CBCA(CO)NH were acquired with 59(t_1) × 48(t_2) × 512(t_3) complex points and were recorded with spectral widths of 10000.0(F_1), 2000.0(F_2), and 7022.5(F_3) Hz. GFT (5,3)D HACACONHN/HACA,CONHN experiments were acquired with 50(t_1) × 25(t_2) × 512(t_3) complex points and were recorded with spectral widths of 11325.0(F_1), 2281.0(F_2), and 10775.9(F_3) Hz at 900 MHz, and utilized 58(t_1) × 40(t_2) × 512(t_3) complex points with spectral widths of 8896.8(F_1), 2000.0(F_2), and 6009.6(F_3) Hz at 600 MHz.

3.2.2.1.3. NOESY and TOCSY experiments

3D ^{15}N edited TOCSY-HSQC (Fesik and Zuiderweg, 1988; Messerle et al., 1989) with a mixing time of 45ms and 3D ^{15}N edited NOESY-HSQC (Fesik and Zuiderweg, 1988; Marion et al., 1989; Messerle et al., 1989) with mixing times of 30-80ms were performed at different temperatures from 15-25°C. Short mixing times (30ms) in NOESY-HSQC were used to eliminate spin diffusion and the data at various temperatures were used to help resolve overlapped resonances. The data sets for the 3D

^{15}N edited TOCSY-HSQC and NOESY-HSQC spectra comprised 100(t_1) × 32(t_2) × 512(t_3) complex points and were recorded with spectral widths of 7000.0(F_1), 2000.0(F_2), and 7022.5(F_3) Hz.

3.2.2.1.4. Measurement of $^3J_{HNHa}$ coupling constants

3D HNHA experiments (Vuister and Bax, 1993) were performed to measure homonuclear $^3J_{HNHa}$ coupling constants at 25°C with a H-H coupling period of 25ms, and recorded 40(t_1) × 128(t_2) × 512(t_3) complex points and spectral widths of 2000.0(F_1), 7000.4(F_2), and 7022.5(F_3) Hz. The correction factor for the $^3J_{HNHa}$ coupling constants was obtained by performing T_{1zz} measurements (Kay et al., 1992b) of amide protons on a ^{15}N labeled GAAVM sample at 25°C.

3D ^{15}N/^{13}C based ECOSY type HNCA-J experiments (Weisemann et al., 1994) were performed to measure homonuclear $^3J_{HNHa}$ coupling constants at 15°C, and recorded 50(t_1) × 128(t_2) × 512(t_3) complex points and spectral widths of 2000.0(F_1), 6000.0(F_2), and 7022.5(F_3) Hz.

3.2.2.1.5. Data processing

All data were processed using the FELIX 2004 software package (MSI, San Diego, CA), and/or NMRPipe (Deraglio et al., 1995) and analyzed with FELIX 2004 or NMRView (Johnson and Blevins, 1994). In the acquisition dimensions for all experiments, a solvent suppression filter was applied to the data prior to apodization with a 90° sine-bell window function. The data were subsequently zero filled to 1024 complex points and Fourier transformed. The t_1 and t_2 dimensions in all 3D experiments were increased 1.5 times by forward-backward linear prediction (Zhu and Bax, 1992), multiplied by a sine-bell window function, zero filled to 256 complex points, and Fourier

transformed. The final 3D data for each experiment included 256 × 256 × 512 real points.

3.2.2.1.6. Generation of NOE contact map

A computer model structure of GAAGVM was obtained based on the crystal structure of T3-785 (PDB ID: 1BKV) (Kramer et al., 1999) and built using the Molecular Operating Environment 2005.06 (Chemical Computing Group Inc., Montreal, Canada). The GIT-GAR-GLA residues in T3-785 were replaced with the residues GPO-GAA-GVM. The structure was solvated with a standard MOE Water Soak procedure and energy minimized from residue 6 to residue 21 using the Amber99 all-atom force field (Ponder and Case, 2003) to an RMS gradient of 0.05. This structure was used to generate the background of the NOE map of a standard triple helix motif. Hydrogen atoms were added to the model structure of GAAGVM using the REDUCE program (Word et al., 1999). The predicted background map was generated by calculating NH-H distances equal to or smaller than 5Å and classifying these as NH-NH, NH-H^{α}, and NH-side chain (H^{β}, H^{γ}, H^{δ}). NOE contact maps for peptides GAAGVM were made from observed NH-H NOEs in the 3D ^{1}H-^{15}N NOESY-HSQC experiment and classified as NH-NH, NH-H^{α}, and NH-side chain (H^{β}, H^{γ}, H^{δ}).

3.2.2.2. Results and discussion

3.2.2.2.1. NMR Chain assignments in GAAGVM and GAAVM peptides

For both peptides, all trimer resonances could be assigned to specific chains of the triple helix as indicated in the heteronuclear single quantum coherence (HSQC) spectrum by

the superscripted number (Figure 3.2). The sequential assignment is derived from the triple resonance experiments and the chain stagger is derived from NOE experiments that define interchain interactions.

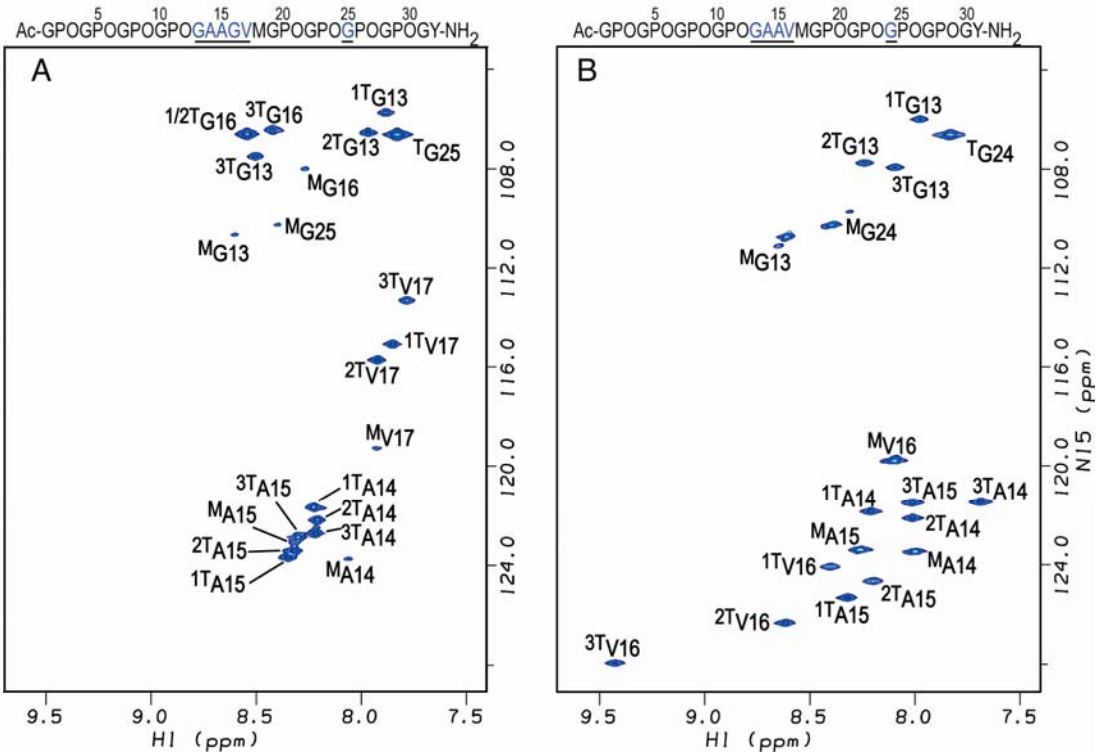

Figure 3.2 ^{1}H-^{15}N HSQC spectra of peptides GAAGVM and GAAVM at 20°C.

A, peptide GAAGVM with a standard Gly-X-Y pattern throughout.

B, peptide GAAVM containing a type IV collagen break. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T, respectively. Assignment of each resonance to chain 1, 2, or 3 is indicated as a number in front of the superscript T.

To obtain the sequence specific assignments, a series of triple resonance experiments were performed on the control GAAGVM peptide, where each residue in the

GAAGV segment of the peptide is double labeled (Table 3.1). For the GAAGVM peptide, traditional triple resonance experiments including HNCO (Ikura et al., 1990)/HN(CA)CO (Clubb et al., 1992; Yamazaki et al., 1994), HNCA (Ikura et al., 1990), HNCACB (Wittekind and Mueller, 1993)/CBCA(CO)NH (Grzesiek and Bax, 1992) experiments were performed (Table 3.2A). Because of the different efficiency of magnetization transfer for trimer and monomer in triple resonance experiments, the monomer resonance will strongly interfere with trimer characterization if monomer and trimer are close in chemical shifts. Therefore, the requirement of removing non-competent monomers or peptide deletions is essential to ensure the good performance of triple resonance experiments. The signal of trimers in the triple resonance experiments are good except in the HN(CA)CO experiment, yet acceptable signal to noise ratio can still be obtained by increasing signal averaging.

From the traditional triple resonance experiments listed above, unambiguous correlations were obtained for residues G13-A14, A15-G16 and G16-V17. The resonances arising from the A14 residue were highly overlapped in the proton and carbon dimensions, which is likely to be a result of the similarity in environment of the two Ala residues in the GAAG segment of the sequence. To resolve the three A14-A15 connectivities, an additional (5,3)D HACACONHN/HACA,CONHN G-matrix Fourier Transform (GFT) experiment (Kim and Szyperski, 2003, 2004) was used and successfully sequentially correlated the chemical shifts of the $^{13}C'$–$^{13}C^{\alpha}$–$^{1}H^{\alpha}$ moieties of residue i - 1 and the NH group of residue i for A14-A15 (Figure 3.3). GFT experiments are good for systems that have a high degree of chemical shift degeneracy, and can be used to resolve the correlations that can not be obtained in traditional experiments, yet

they are not suitable as a routine experiment for triple helical peptides because of the poor sensitivity in the HACA,CONHN experiment even at 900MHz spectrometer equipped with a cryoprobe. For the GAAVM peptide complete sequential assignments were obtained for the doubly labeled GAAV segment in the GAAVM peptide from a single pair of HA(CA)NH/HA(CACO)NH (Kay et al., 1991) experiments (Table 3.2B) as there was no chemical shift overlap in the NH or H^{α} dimensions (Figure 3.4). Assignments were confirmed by the HNCA (Ikura et al., 1990) experiment (Table 3.2B).

TABLE 3.2 Assignment summary of peptide GAAGVM and GAAVM

A. Peptide GAAGVM

Experiment / Correlation	HNCO/ HN(CA)CO	HNCA	HNCACB/ CBCA(CO)NH	GFT (5,3)D HACACONHN/ HACA, CONHN[a]			
				1	2	3	4
^{1}G13-^{1}A14	X[b]	X	X				
^{2}G13-^{2}A14	X			X			
^{3}G13-^{3}A14	X				X		X
^{1}A14-^{1}A15							X
^{2}A14-^{2}A15							X
^{3}A14-^{3}A15						X	X
$^{1/2}$A15-$^{1/2}$G16			X			X	X
^{3}A15-^{3}G16			X			X	X
$^{1/2}$G16-$^{1/2}$V17	X	X					
$^{1/2}$G16-$^{1/2}$V17	X	X				X	
^{3}G16-^{3}V17	X	X				X	

B. Peptide GAAVM

Experiment / Correlation	HA(CA)NH/HA(CACO)NH	HNCA
^{1}G13-^{1}A14	X	
^{2}G13-^{2}A14	X	
^{3}G13-^{3}A14	X	X
^{1}A14-^{1}A15	X	X
^{2}A14-^{2}A15	X	X
^{3}A14-^{3}A15	X	X
^{1}A15-^{1}V16	X	X
^{2}A15-^{2}V16	X	
^{3}A15-^{3}V16	X	

a. In GFT (5,3)D experiment, 1,2,3,4 represents CO+CA+HA, CO-CA+HA, CO+CA-HA, CO-CA-HA, respectively.

b. Sequential correlations that can be obtained from the experiment are marked with "X".

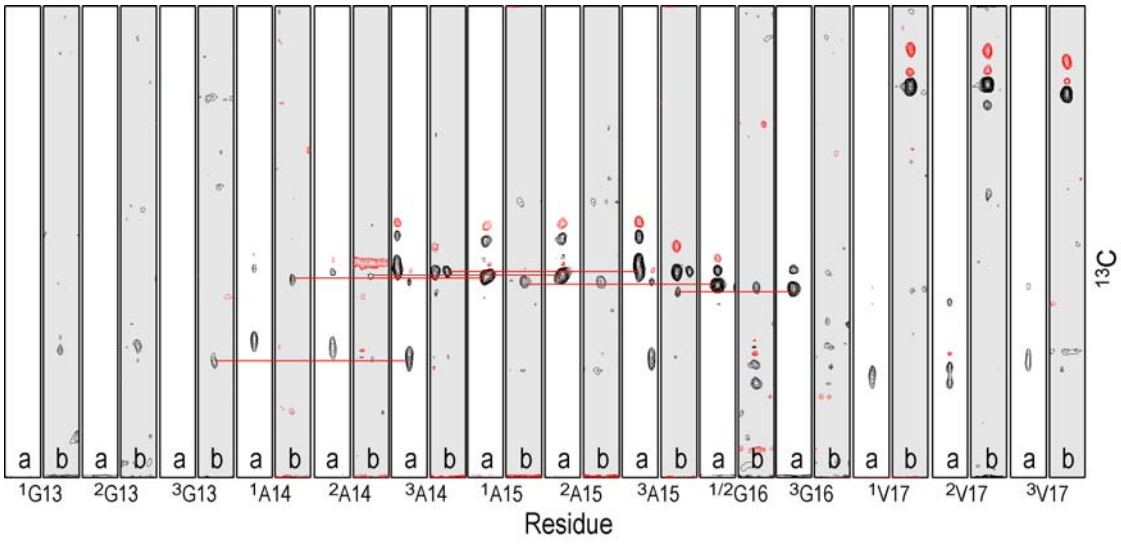

Figure 3.3 Strips from the subset 4 of (5,3)D HACACONHN (strip a) /HACA,CONHN (strip b) GFT experiment of GAAGVM peptide. Partial chain specific sequential correlations of three chains are traced out as shown in red lines.

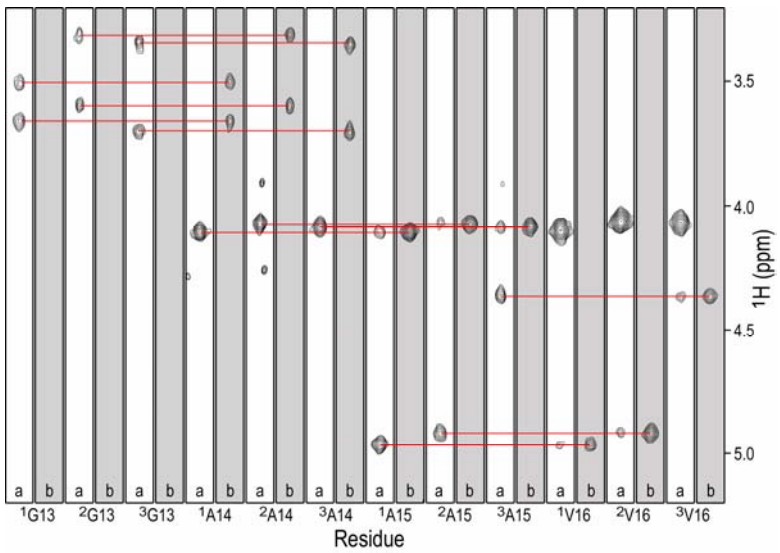

Figure 3.4 Strips from the HA(CA)NH (strip a)/HA(CACO)NH (strip b) spectra of GAAVM peptide. Complete chain specific sequential correlations of three chains are traced out as shown in red lines.

NOE experiments were used to obtain interchain distances which led to the identification of the three chains in terms of their 1-residue stagger. For example, NOEs observed between chain 3 ^{3}G13 NH and chain 1 ^{1}A15 H^{α} in GAAGVM and between ^{1}A14 NH and ^{2}G13 H^{α} in GAAVM indicated the relationship between chains and allowed identification of the leading (chain 1), trailing chains (chain 3) and middle chains (chain 2) of the peptides. The details of chain stagger identification from NOE contacts are described in the section below. Chain specific assignments of peaks set the stage for using distance and dihedral angle constraints in structure determination.

3.2.2.2.2. NOE contacts to identify chain stagger and characterize interchain distances

NMR Prediction of short range distances in a classic triple helical conformation from the GAAGVM model structure - A predicted contact map in a classic triple helical conformation is generated from a GAAGVM model structure. The high resolution structures of a number of collagen-like peptides have been solved by x-ray crystallography (Kramer et al., 1999; Kramer et al., 2000; Kramer et al., 1998; Nagarajan et al., 1999), and these structures have confirmed and provided molecular coordinates for the triple-helical structure. A model structure of the GAAGVM peptide was generated by molecular modeling using the crystal structure of the peptide T3-785 ($(POG)_3$-ITGARGLAG-$(POG)_4$ sequence) (Kramer et al., 1999), since it contains a central sequence with no imino acids surrounded by Gly-Pro-Hyp (GPO) triplets. The central GIT-GAR-GLA residues were replaced with the residues GPO-GAA-GVM, and the structure was energy minimized (see Methods for details). The cross section of this

minimized GAAGVM model structure shows the NH groups of the Gly residues are located near the center of the triple helix, and form hydrogen bonds with the CO of the X residue of a neighboring chain (Figure 3.1). The NH vectors of the Y residues point outward toward the solvent, while the NH vectors of the X residues point more tangentially toward a neighboring chain within the molecule making a water bridged hydrogen bond with the C=O of the Gly of the neighboring chain.

A predicted contact map is generated from the computer model structure for the GAAGVM peptide and allows computation of distances that are within 5Å for the backbone NH to backbone NH, H^{α} and sidechain protons in this structure (Figure 3.5A; shaded squares indicate that there is one or more predicted contact within 5Å and circles indicate the individual predicted NOE). A number of salient features can be seen from the contact map and include:

a) Intrachain sequential distances (indicated in gray highlighted boxes) are not symmetric. The NH atoms of residue i (indicated by ^{j}NH(i), where j refers to the chain number and i refers to the residue) have a larger number of NOEs to the NH, H and side chain residues of i-1 than to the NH, H and side chain residues of i+1 (individual predicted NOEs are indicated by circles which are superimposed on the gray highlighted boxes).

b) Interchain short distances directly across the chain (indicated in yellow highlighted boxes) can be seen for all residues primarily from ^{j}NH(i) to $^{j-1}$NH(i+1) or to $^{j-1}H^{\alpha}$(i+1). NOEs are also predicted for the ^{j}NH(i) to $^{j+1}$NH(i-1). These short distances arise directly from the one residue stagger of the triple helix.

c) There are additional interchain contacts between Gly residues, where the corresponding Gly residues in the three chains are packed in the center in a staggered manner (Figure 3.1B). This staggered packing results in additional interchain ^{j}NH(i) to $^{j+1}$NH(i) and $^{j+1}H^{\alpha}$(i), and ^{j}NH(i) to $^{j-1}$NH(i) and $^{j-1}H^{\alpha}$(i) contacts. Additionally, unique NOEs between Gly ^{j}NH(i) to $^{j-2}$NH(i+3) and $^{j-2}H^{\alpha}$(i+3), and between Gly ^{j}NH(i) to $^{j+2}$NH(i-3) and $^{j+2}H^{\alpha}$(i-3) are predicted and would correspond to NOEs between ^{3}G13 to ^{1}G16, and ^{1}G16 to ^{3}G13.

Experimental NOE observations in GAAGVM - A contact map (Figure 3.5B) was generated for the experimental NOE data of the GAAGVM peptide, now using circles to denote the experimental values and highlighted squares of intramolecular (gray) and intermolecular (yellow) NOEs to represent predictions from the GAAGVM model structure (Figure 3.5A). The expected intrachain NOES and some additional interchain NOEs consistent with the close packing of the central Gly and the 1-residue stagger between chains (^{2}G13NH-^{3}O12H$^{\alpha}$, ^{3}G13NH-^{1}A15H$^{\alpha}$, ^{1}A14NH-^{3}O12H$^{\alpha}$/side chain, ^{2}A14NH-^{1}A15H$^{\alpha}$) are observed. The chain stagger of ^{1}V17 and ^{2}V17 was confirmed from two characteristic NOEs: ^{3}G16NH-^{1}V17 side chain, and ^{1}V17NH-^{3}A15 side chain. There is high degree of chemical shift degeneracy in the GAAGVM peptide arising from the redundant GAAG central sequence, especially for G13H$^{\alpha}$, A14NH, A14H$^{\alpha}$, A14H$^{\beta}$, A15NH, A15H$^{\beta}$, G16NH and G16H$^{\alpha}$ (see half circles in the contact map), and as a result there are some NOEs that are not observed in the experiments when compared to the predicted contact map. For example, due to the direct overlap of the NH and H$^{\alpha}$ resonance of chains 1 and 2 of G16, the interchain NOEs were not observed. Despite the overlap it is possible to determine the triple helix stagger from the interchain NOEs and

to confirm that the solution conformation of the GAAGVM peptide is consistent with a typical triple helix similar to the model structure.

In summary, NOE contact map, developed as a structural strategy in the specialized triple helix system, can be used in identifying 1-residue stagger from interchain contacts, and in characterizing structural deviation from a typical triple helix conformation.

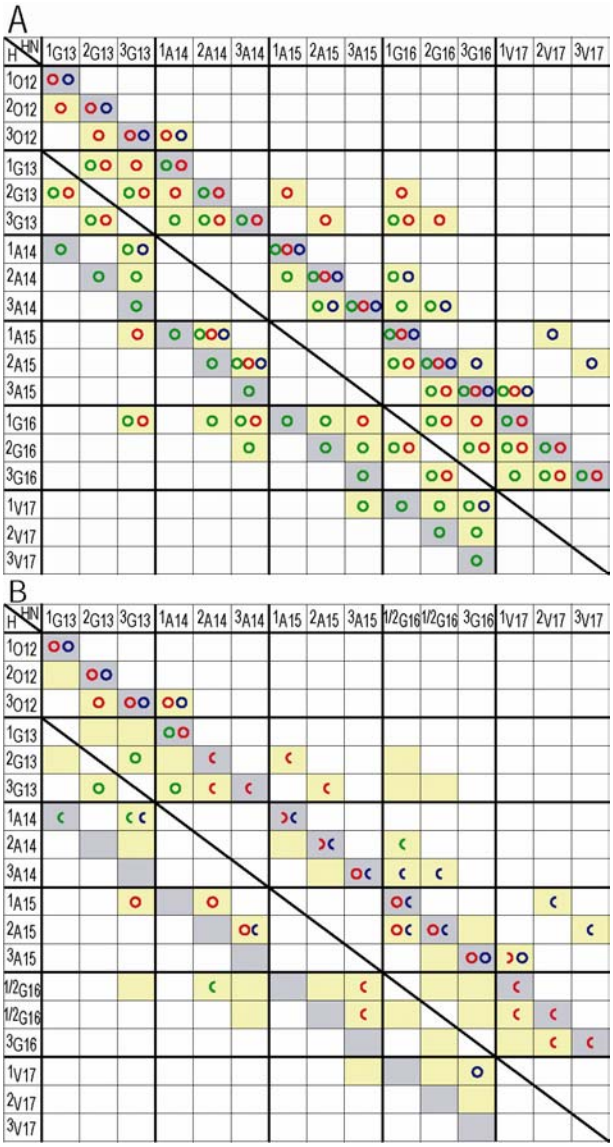

Figure 3.5 Comparison of experimental NOEs from the NOESY-HSQC experiment for the GAAGVM with predicted NOEs from a standard triple helical conformation.
A, Predicted contact map obtained from GAAGVM molecular model structure. Contacts are shaded in gray for intrachain distances less than 5Å and in yellow for interchain distances less than 5 Å. Predicted HN-HN(O), HN-H^{α}(O), and HN-side chain protons(O) interactions less than 5 Å are overlaid on the shading so that one can see that they come from both intra- and intermolecular contacts in a complex manner because of the supercoiled trimeric structure of the triple-helix.
B, Experimental contact map of peptide GAAGVM (represented by circles), overlaid on the shaded contacts of the GAAGVM model, indicating that some of the predicted interchain as well as intrachain NOEs are observed. The fact that many predicted NOEs are not observed is largely attributable to overlap. Half circles( ) represent overlapped NOEs consistent with the GAAGVM model.

3.2.2.2.3. Determination of J coupling constants

The homonuclear three-bond $^{3}J_{HNHa}$ coupling constant is related to the dihedral angle ϕ in peptide and proteins. This relationship can be approximated by the Karplus equation. One of the most widely used experiments to measure $^{3}J_{HNHa}$ is the 3D HNHA experiment (Vuister and Bax, 1993). In the HNHA experiment, $^{3}J_{HNHa}$ can be calculated by measuring the diagonal-peak to cross-peak volume ratio and the dihedral angle ϕ can be obtained from the parameterized Karplus equation (Vuister and Bax, 1993): $^{3}J_{HNHa} = 6.51 \cos^2 (\phi\text{-}60°) - 1.76 \cos (\phi\text{-}60°) + 1.6$. For the error analysis of $^{3}J_{HNHa}$ measurement, experimental error was calculated based on the experimental uncertainty in volume measurement. The experimental uncertainty was estimated as the average of the absolute values of volume integrations in the regions free of signals. The average value and the measurement error of the apparent $^{3}J_{HNHa}$ coupling constants was then calculated based on the maximum and minimum J coupling constants from the maximum and minimum volume ratio of cross/diagonal peaks with experimental uncertainty included. For example, volumes of 1.2E8 and -1.7E7 were measured for the diagonal- and cross-peak of a residue in the HNHA spectrum, respectively. Experimental uncertainty in volume measurement was determined to be 2.7E6 by measuring volumes in the noise region of the spectrum. So the upper and lower bound of volume ratio of cross/diagonal peaks is (-1.7E7+2.7E6)/(1.2E8+2.7E6) and (-1.7E7-2.7E6)/(1.2E8-2.7E6), respectively. From these the maximum and minimum $^{3}J_{HNHa}$ coupling constants were calculated to be 5.0Hz and 4.2Hz, respectively, and the apparent $^{3}J_{HNHa}$ was represented as 4.6±0.4Hz.

Apparent $^{3}J_{HNHa}$ coupling values from HNHA experiments are underestimated for triple helical peptides. During the de- and rephasing delays in the HNHA experiment, the

antiphase magnetization giving rise to cross peaks relaxes at a faster rate than the in-phase magnetization giving rise to diagonal peaks (Vuister and Bax, 1993). The effect of relaxation is to reduce the cross peak intensity relative to the diagonal resonance intensity, and to reduce the apparent $^3J_{hnha}$ coupling constant. The relaxation of the antiphase magnetization (T_{2ap}) can be approximated as in $1/T_{2ap} = 1/T_{2HN} + 1/T_{1sel}$, where T_{2HN} is the transverse relaxation time of the in-phase magnetization and T_{1sel} is the apparent selective T_1 of the Hα spin (Cavanagh et al., 1996; Vuister and Bax, 1993). For proteins in the slow tumbling limit, the T_{1sel} values scale with the inverse of the protein rotational correlation time τ_c (Cavanagh et al., 1996; Vuister and Bax, 1993). In order to correct the $^3J_{HNHa}$ coupling values of triple helical peptides, T_{1sel} was estimated by measuring the rate of NH-^{1}H spin flip from a separate T_{1zz} measurement experiment (Kay et al., 1992b). By measuring the rate of decay of NH-^{15}N heteronuclear J order, $2A_zX_z$, the rate of proton spin flips can be obtained. An average T_{1sel} value of 84ms for amide protons of ^{15}N labeled residues was found from the T_{1zz} measurement on the ^{15}N labeled GAAVM peptide at 25°C. The equation below (Cavanagh et al., 1996) was solved to determine accurate values of $^3J_{HNHa}$. Corrected $^3J_{HNHa}$ values are approximately 116% larger than the apparent $^3J_{HNHa}$ values without T_{1sel} term. Therefore the corrected J coupling constants of doubly labeled GAAGVM and GAAVM peptides were obtained by multiplying the apparent coupling by 1.16.

$$\frac{S_{cross}}{S_{diagonal}} = -\left\{ \frac{\tan[(1-\chi^2)^{1/2}\zeta]}{(1-\chi^2)^{1/2} + \chi \tan[(1-\chi^2)^{1/2}\zeta]} \right\}^2$$

$\chi = 1/(2\pi * T_{1sel} * {}^3J_{HNHa})$
$\zeta = \pi * {}^3J_{HNHa} * \delta$
δ = H-H coupling period

For Gly the issue is more complicated. Because of the strong dipolar interaction between two Ha protons of glycines, it has a shorter T_{1sel} than other residues (Kay et al.,

1992b; Vuister and Bax, 1993; Vuister et al., 1993). In addition, a Ha_2-Ha_3 spin flip causes a decrease in the value for the larger of the two $^3J_{HNHa}$ and an increase for the smaller $^3J_{HNHa}$ (Vuister and Bax, 1993; Vuister et al., 1993). Therefore the $^3J_{HNHa}$ values for Gly are not accurate even after the T_{1sel} correction, and arbitrary larger error bars were estimated for the $^3J_{HNHa}$ couplings of Gly than other residues (see the example in section 3.2.3.2.1 for details).

In addition to the HNHA experiment, $^{15}N/^{13}C$ based ECOSY type HNCA-J experiments (Weisemann et al., 1994) was also carried out to measure $^3J_{HNHa.}$ It measures peak splitting in 3D spectra (Figure 3.6). This measurement gave reasonable $^3J_{HNHa}$ values for monomers (6 Hz, Figure 3.6B), but produced abnormal values for trimers (0.8 Hz, Figure 3.6A). The reason for the failure of HNCA-J measurement in the trimer is not very clear. It could arise from the broad linewidths of trimers due to the anisotropic shape.

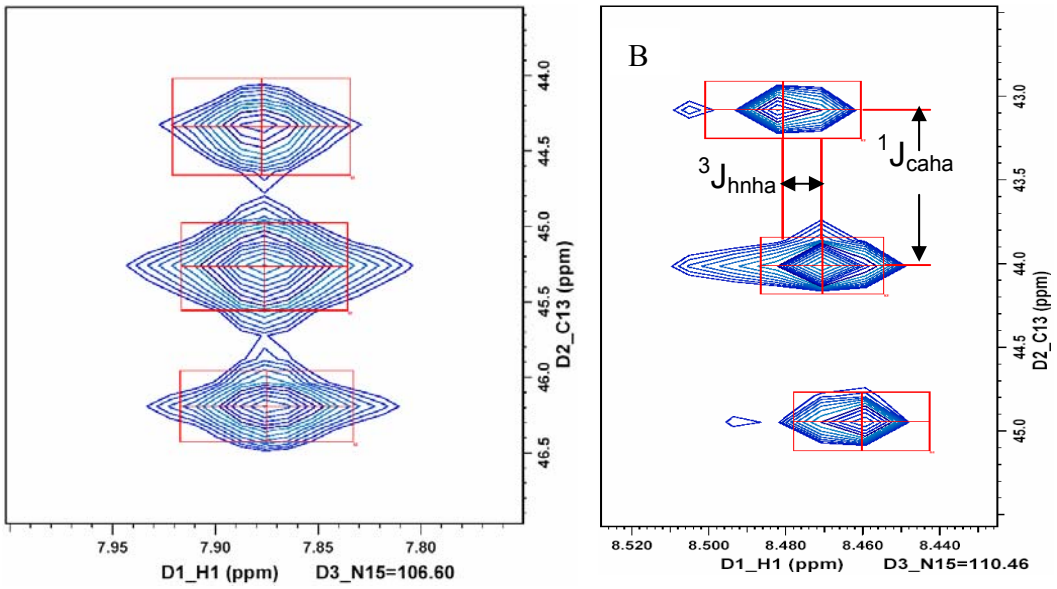

Figure 3.6 Stripes from HNCA-J experiment of ^{T}G24 (A) and ^{M}G24 (B) in GAAGVM peptide. Peak boxes are shown in red. The peak splitting that indicates $^3J_{HNHa}$ is shown in panel B.

3.2.3. Molecular modeling

3.2.3.1. The procedure of molecular modeling

Because of the very anisotropic and homotrimeric nature of triple helix, there is no long range distance information about the molecule from NOEs and the number of localized interchain NOEs is very limited. Therefore, the solution structure of the triple helical peptide can not be solved with the current experimental data. Alternative strategies of NMR incorporated molecular modeling are developed to incorporate NMR structural parameters and visualize the structure of triple helical peptides. NMR incorporated molecular modeling was performed based on x-ray crystal structures of homologous triple helical peptides (Bella et al., 1994; Bella et al., 2006; Kramer et al., 1999; Kramer et al., 1998)(Figure 3.7). From the starting x-ray structure, residues were mutated into the target residues. NMR restraints including dihedral angle ϕ from $^{3}J_{HNHa}$ measurements and interchain distances from NOE experiments were used as restraints in the energy minimization. For the validation of the models, back calculation of NMR parameters of models were performed and compared to the experimental data. If the model is not consistent with the experimental data, the model is eliminated. Actually energy minimization is very sensitive to the initial starting structure, therefore a number of different input structures with different mutation schemes were tried. For the dihedral angle ϕ restraints, four possible solutions can be derived from the Karplus equation $^{3}J_{HNHa} = 6.51 \cos^{2} (\phi\text{-}60°) - 1.76 \cos (\phi\text{-}60°) + 1.6$ for $^{3}J_{HNHa}$ of 5-6 Hz (Vuister and Bax, 1993). Varying restraint ranges were used to either cover all the possible ϕ angles or cover all combinations of ϕ angles of involved residues. For Gly, a wider range of the ϕ angles (expand 5° to upper and lower limits of the ϕ restraints) were used because of the

experimental artifact for Gly as described in the previous section 3.2.2.2.3. A number of different output structures could be produced from those procedures. Back calculation of $^{3}J_{HNHa}$ values and NOEs indicated that most structures were not consistent with all of the experimental data. Therefore, although the final structure is not unique, it provides a representative example of the structure that has a good fit to the experimental data.

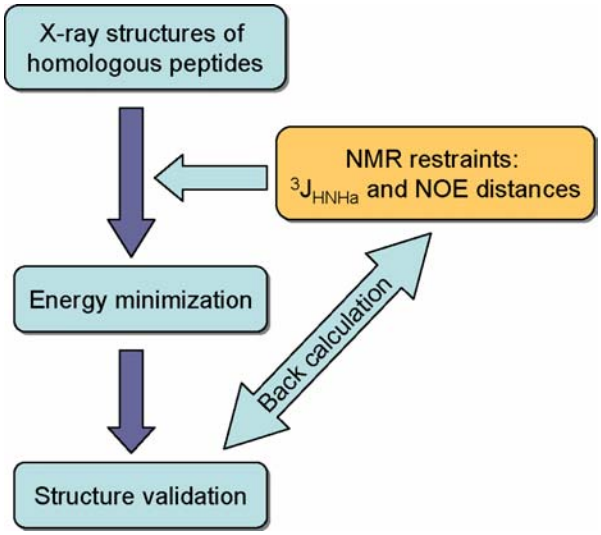

Figure 3.7 The flowchart of molecular modeling procedure incorporating NMR restraints.

3.2.3.2. An example of the molecular modeling: GPO(G16S) peptide

3.2.3.2.1. Starting structures and NMR restraints

Here I describe the molecular modeling procedure using an example of the peptide $(GPO)_4$-GPV-SPA-GAR-$(GPO)_4$ (designated as GPO(G16S)). The GPV-SPA-G segment (the residue numbers are G13, P14, V15, S16, P17, A18, and G19) in the GPO(G16S) peptide is $^{13}C/^{15}N$ labeled, and J coupling constants and distances information could be obtained for 5 residues: G13, V15, S16, A18, and G19. The

computer models of the GPO(G16S) peptide were generated based on the x-ray crystal structure of Gly→Ala (PDB ID: 1CAG) (Bella et al., 1994) using the Molecular Operating Environment 2006.08 (Chemical Computing Group Inc., Montreal, Canada). Residues GPO-APO-GPO were substituted to GPV-SPA-GAR. A different starting structure of T3-785 (PDB ID: 1BKV) (Kramer et al., 1999) was also tried and residues GIT-GAR-GLA were substituted to GPV-SPA-GAR. The structures were solvated with a standard MOE Water Soak procedure. The input NMR restraints included 15 ϕ angle restraints (residues of G13, V15, S16, A18, and G19 in three chains) and 4 NOE distances (^{1}S16NH-^{3}S16NH, ^{1}S16NH-^{3}G13NH, ^{3}S16NH-^{1}G19NH, ^{1}S16NH-^{1}V15NH). These 4 NOE cross peaks are of weak intensities by qualitative estimation. A highly precise distance can not be derived from the NOE peak intensity because there are inherent errors associated with the accuracy of volume measurement, spin diffusion and dynamics. Therefore a lower limit of 1.8Å and a higher limit of 5.0Å were set for distance restraints, same bounds as in the weak NOE category in normal NMR structure determination.

For the ϕ angle restraints, a single experimental $^{3}J_{HNHa}$ coupling measurement could give four solutions of ϕ angle restraints from the karplus equation (Vuister and Bax, 1993), and the following procedure was applied. For non-Gly residues, the range of ϕ angle restraint was set to cover all negative possible ϕ angles because positive ϕ angles are found to be not energetically favorable for residues other than Gly. For example, for a $^{3}J_{HNHa}$ of 5.3 ± 0.7 Hz, there are four possible ranges from Karplus equation: -176° to -166°, -74° to -64°, 26° to 46°, and 74° to 94°. A lower and higher limit of -176° and -64° were used for the ϕ angle restraint in MOE, respectively. For Gly, a wider range of the ϕ

angles was used to also cover the positive ϕ angles. In addition, a 5° expansion was used for upper and lower limits of the ϕ restraints to compensate for experimental artifact of Gly (details see section 3.2.2.2.3.). So if a Gly has a $^{3}J_{HNHa}$ of 5.3 ± 0.7 Hz, -180° to 99° would be used for the ϕ restraints.

The resulting structures were energy minimized with NMR restraints, from residue 7 to residue 27 using the Amber99 all-atom force field (Ponder and Case, 2003) to an RMS gradient of 0.05. The number of residues undergoing energy minimization was determined arbitrarily to include two additional triplets on both sides of the substitution residues (G13 and R21 in this case), to allow efficient space for conformational freedom.

3.2.3.2.2. Iterative procedures of the back calculation and structure validation

The iterative procedure of model structure validation can be divided into three steps: 1) back calculation of ϕ angles; 2) evaluation of the model structures using Molprobity program (Lovell et al., 2003); 3) back calculation of NOE contact maps.

The first step for the structure validation is to compare the ϕ angle in the model structure with ϕ angle from the experimental $^{3}J_{HNHa}$ constants. Taken the example in the previous section again, there are four possible ranges from the $^{3}J_{HNHa}$ of 5.3 ± 0.7 Hz: -176° to -166°, -74° to -64°, 26° to 46°, and 74° to 94°. An initial trial (designated as trial 1) of ϕ angle restraint of -176° to -64° was used for residue 1, take ^{1}S16 for example, and the ϕ angle in the energy minimized structure was compared with the experimental data. If the ϕ angle in the model is -70°, then it is consistent with one of the ϕ solution from the experimental $^{3}J_{HNHa}$ coupling and no further action needs to be taken. If the ϕ angle in the model is -90°, it is not consistent with any of the four solutions. Two additional trials

(designated as trial 2 and trial 3) of energy minimization need to be performed with -176° to -166° for ϕ restraint of ^{1}S16 in trial 2, and -74° to -64° for ϕ restraint of ^{1}S16 in trial 3, while keeping restraints for other residues the same as in trial 1. If the ϕ angles of two residues, ^{1}S16 as well as another residue, are not consistent with the experimental data after trial 1, then an combination number of trials needs to be taken additionally (2*2 in this case). Furthermore, if the inconsistency comes from a non-Gly residue and a Gly residue, more trials are performed (2*4 in this case). One thing that needs to be noticed in this procedure is that not all trials of energy minimization can produce a structure; there are cases where the energy minimization failed. It is probably because the restraints define an unfavored structure that could not satisfy the force field constraints at the same time in the local energy minimum search during energy minimization. Using this iterative procedure, 28 trials were performed starting from the Gly→Ala x-ray crystal structure and 4 structures that have ϕ angles consistent with experimental $^{3}J_{HNHa}$ constants were produced. 34 trials were carried out starting from the T3-785 structure and 10 structures were produced. These 14 structures were then taken to the next step.

The second step in the structure validation is to analyzing all-atom contacts and geometry using the Molprobity program (http://molprobity.biochem.duke.edu)(Lovell et al., 2003). This program from Dr. Richardson's laboratory can add hydrogen to the molecule using the REDUCE program (Word et al., 1999), and evaluates clashes, rotamer, and ramachandran plot. 5 out 14 structures from the previous step were eliminated because of bad rotamers and/or dihedral angles in the unfavored region of ramachandran plot.

The remaining 9 structures were taken to the third step of the structure validation to back calculate the NOE contact maps from the models and compare with the experimental NOEs. The criteria of elimination can be divided into two parts: 1) the experimental NOE contact map should be a subset of the predicted contact map from the model; 2) the extra predicted NOEs from the model should be plausible by carefully checking the raw NOESY-HSQC data. Figure 3.8 represents one example to eliminate a structure by using the criteria 2. Red circle in Figure 3.8B highlights the NOE between ^{1}S16NH and ^{2}S16NH that are not consistent with the experimental data. This distance in the model structure is very short (<2.8Å), and should give a strong NOE peak if it represents the solution structure. But the raw NOESY-HSQC spectrum did not show any NOE between ^{1}S16NH and ^{2}S16NH, and the chemical shifts of those two amides are in the non-overlap region in the NOESY spectrum and thus exclude the experimental uncertainty, therefore this model structure was eliminated. Using this procedure, 6 out 9 model structures were eliminated, and finally 3 structures were consistent with all experimental data. Among the final 3 structures, 2 structures were from the Gly→Ala starting structure and 1 structure was from T3-785. The structure based on the T3-785 peptide was similar and comparable to one of the structure from Gly→Ala.

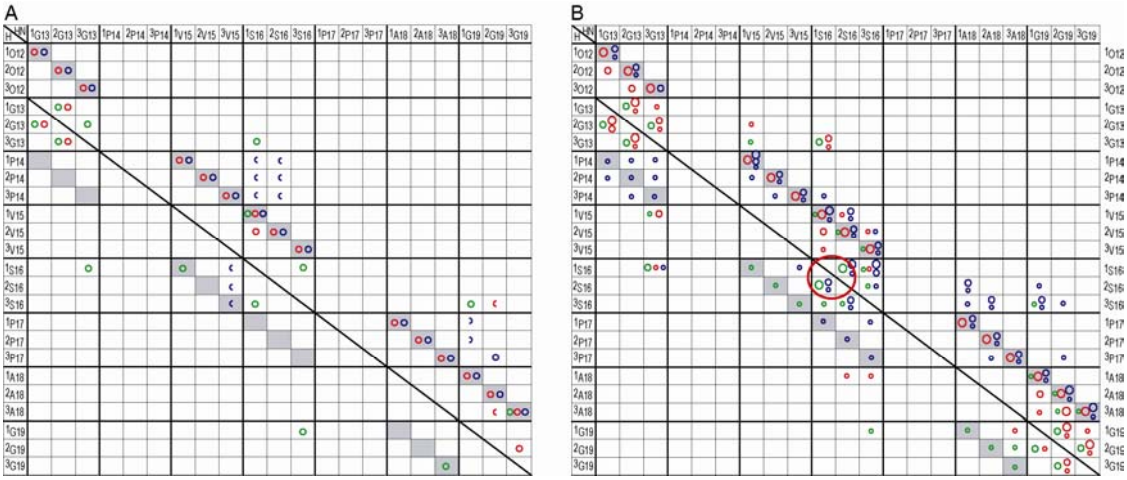

Figure 3.8 Back calculation of the predicted NOEs from model structure of GPO(G16S) and compare it with the experimental NOEs from the NOESY-HSQC experiment.

A, Experimental contact map obtained from NOESY-SHQC experiment are represented by circles (HN-HN(**O**), HN-H$^{\alpha}$(**O**), and HN-side chain protons(**O**)). Half circles( ) represent overlapped NOEs. Gray shaded boxes indicate intrachain contacts.

B, Predicted contact map obtained from one of the GPO(G16S) model structures. Predicted HN-HN(**O**), HN-H$^{\alpha}$(**O**), and HN-side chain protons(**O**) interactions less than 5 Å Contacts are indicated. The big, middle and small sizes of circles correlate with distance of < 2.8 Å, 2.8 Å-3.7 Å, and 3.7 Å-5.0 Å, respectively. The predicted ^{1}S16NH-^{2}S16NH NOE that is not consistent with the experimental data is highlighted in red circle.

3.2.4. Conclusions

In summary, NMR triple resonance studies and structural measurements of ^{13}C/^{15}N doubly labeled peptides allowed the identification of chain stagger of triple helix, and provided the detailed information about the distances and dihedral angle ϕ. These

structural parameters were incorporated in the molecular modeling based on the x-ray crystal structures of homologous collagen peptides, and allowed the visualization of the solution structures of the collagen model peptides. These strategies opened the door for the detailed structural characterization of collagen model peptides in the next level.

3.3. Identification of partially disordered peptide intermediates through residue specific NMR diffusion measurements

3.3.1. Introduction

It is technically challenging to detect low population partially disordered species that are in equilibrium with the folded and unfolded states. Characterization of these states is critical to understand protein folding processes and equilibrium properties. The ability of NMR to probe the exchange dynamics of interconverting species of proteins at specific residues allows the design of sophisticated experiments to approach the identification of these intermediates. Recently, relaxation dispersion NMR was used to define folding intermediates in equilibrium with its unfolded and fully folded states for an SH3 domain (Korzhnev et al., 2004). Here, residue specific translational diffusion coefficients are used to detect equilibrium intermediates in self associating systems. This methodology is applied to equilibrium solutions of related triple helical peptides that model a small region of type I collagen with and without a mutation known to cause Osteogenesis Imperfecta (Baum and Brodsky, 1999). The peptides are composed of interconverting trimer and monomer species, where the trimer folded form has significantly different hydrodynamic parameters from the unfolded state allowing residue specific diffusion coefficients to be sensitive to the existence of partially disordered intermediates.

Triple helical peptides of varying lengths and sequences have been used to model folding, stability and binding regions of collagen (Baum and Brodsky, 1999; Fields and Prockop, 1996). NMR spectra of such peptides always show distinct monomer and trimer resonances, indicating slow interconversion of these two populations on the NMR

timescale (Bhate et al., 2002; Buevich et al., 2004; Liu et al., 1998). A peptide (T1-892c) designed to model type I collagen, with a collagen G-X-Y sequence (α1(I) chain residues 892-909) and a $(GPO)_4$ sequence on the C terminus was found to have trimer and monomer peaks at four labeled positions along the chain, and relaxation studies indicated rigid trimers are in equilibrium with flexible monomers (Liu et al., 1998). Similar results were found for a related peptide of identical sequence but with the $(GPO)_4$ sequence located at the N terminus (T1-892n) (Buevich et al., 2004). A peptide homologous to T1-892c with a G to A change modeling a known OI mutation site (T1-892c[G10A]) showed trimer peaks only C terminal to the mutation site at A10 suggesting the mutation prevented complete folding and resulted in a partially folded molecule that is triple helical C terminal to the mutation site and disordered at the N terminal end (Bhate et al., 2002).

Translational diffusion coefficients have been largely used to investigate hydrodynamic properties of folded and unfolded globular proteins as well as having been applied to monitor kinetic folding processes in a globular protein and in collagen model peptides (Balbach, 2000; Buevich and Baum, 2002; Jones et al., 1997; Mansfield et al., 1998; Wilkins et al., 1999) NMR pulsed field gradient diffusion measurements were obtained for peptides T1-892c, T1-892n and T1-892c[G10A]. Specific ^{15}N labeling of individual residues within each peptide chain allows residue and position specific diffusion measurements to be obtained.

3.3.2. Materials and Methods

3.3.2.1. Sample Preparation

Peptide T1-892c, T1-892c[G10A], and T1-892n were synthesized by Synpep Corporation (Dublin, CA) and purified and characterized as previously described (Bhate et al., 2002; Buevich and Baum, 2002; Buevich et al., 2004). Sequences of peptide T1-892c, T1-892c[G10A] and T1-892n are shown below. ^{15}N labeled residues are underlined, with numbering shown on top of each sequence. A G→A substitution is contained in T1-892c[G10A] at position 10 and shown in the box. For clarity, T1-892[G10A] in reference 2 is being referred to as T1-892c[G10A] in this paper.

	^{7} ^{18} ^{25}
T1-892n	**Ac-GPOGPO<u>G</u>POGPOGPAGP<u>A</u>GPVGPA<u>G</u>ARGPAGV-CONH$_2$**
	^{6} ^{10} ^{13} ^{25}
T1-892c	**Ac-GPAGP<u>A</u>GPV<u>G</u>PA<u>G</u>ARGPAGPOGPO<u>G</u>POGPOGV-CONH$_2$**
	^{6} ^{10} ^{13} ^{25}
T1-892c[G10A]	**Ac-GPAGP<u>A</u>GPV<u>A</u>PA<u>G</u>ARGPAGPOGPO<u>G</u>POGPOGV-CONH$_2$**

3.3.2.2. Diffusion Experiments

NMR experiments were performed on a Varian INOVA 600 MHz spectrometer. Convection compensated LED followed by HSQC (CCLED-HSQC) experiments (Figure 3.9) were used to measure the residue-specific diffusion coefficients of the selectively labeled triple helical peptides. The pulse sequence is similar to the convection compensation CC-LED-HSQC experiment described previously, with an improved convection compensation in which all gradient pulses are self-compensating composites (Buevich and Baum, 2002; Jershow and Muller, 1997). The CCLED-HSQC experiment was performed with the following g3-gradient

strengths (G/cm): 1.88, 13.28, 18.77, 22.99, 26.55, 29.67, 32.48, 35.11, 37.55. The gradient strength was calibrated on a standard sample, doped 1% H_2O in 99% D_2O, using the value of 1.90×10^{-9} m^2s^{-1} for the diffusion coefficient of HDO at 25°C (Holz and Weingaertner, 1991; Mills, 1973). For the calibration of gradient strengths, 18 data points were acquired in the 1D CCLED experiment using g3-gradient strengths (G/cm) from 2.15 to 38.65 with a 2.15 G/cm interval. One dimensional data were processed and spectrum was integrated following baseline correction using VNMR. For residue specific diffusion measurement, all data had 1024 complex data points in the *t2* dimension and 64 increments in the *t1* dimension. A 3s recycle delay was used. Temperatures were calibrated with methanol at low temperature and with ethylene glycol at high temperature. Two-dimensional data were processed on a Silicon Graphics workstation using the FELIX 97 software package (MSI, San Diego, CA).

There are some potential artifacts in the diffusion measurement, such as cross relaxation (Chen and Shapiro, 1999; Chou et al., 2004; Dvinskikh and Furo, 2000; Lucas et al., 2003) and exchange of the amide protons with water (Chen et al., 1998; Derrick et al., 2002; Tillett et al., 1998). Using the bipolar pulse pairs in the CCLED-HSQC can suppress the oscillatory modulation and signal attenuation due to NOE cross relaxation and chemical exchange (Chen and Shapiro, 1999; Chen et al., 1998; Derrick et al., 2002)}(Chou et al., 2004; Dvinskikh and Furo, 2000; Lucas et al., 2003; Tillett et al., 1998). A Control experiment with 50ms diffusion delay (refers to Δ in Figure 3.9) revealed essentially the same diffusion coefficients as those obtained from a 280ms (Δ) delay experiment, indicating that these potential artifacts did not affect the measurements. Another potential problem in the diffusion experiment is the sample heating during ^{15}N

decoupling in the acquisition time. Abnormally high diffusion coefficients due to sample heating were observed when decoupling power was high. A low decoupling power of 33dB was used and did not cause any increase in the diffusion coefficients of standard samples.

3.3.2.3. Data Analysis

The signal attenuation in the presence of pulsed field gradients is given by $I=I_{(0)}\exp[-2D\gamma^2G^2\delta^2(\Delta+2\delta/3+3\tau/4)]$, where D is the translational diffusion coefficient, G is the gradient strength, γ is ^1H gyromagnetic ratio, δ, Δ, and τ are 2ms, 280ms (low temperature) or 200ms (high temperature), and 0.4ms, respectively. The apparent diffusion coefficients were estimated by fitting experimental peak volumes into the above equation by a least-squares minimization algorithm. Figure 3.10B shows the fittings of A6 trimer and monomer peaks in T1-892c in semi-logarithmic scale. The diffusion coefficients at high temperature were normalized to values at 10°C by using the Stokes-Einstein equation, $D_{normalized}=D_T*283.15*\eta_T/(T*\eta_{10°C})$, where D_T is the directly measured diffusion coefficient at high temperature T, η is the viscosity calculated for the solution of 90% H_2O and 10% D_2O at a specific temperature (Cho et al., 1999). Measurements were repeated twice or three times. The errors from repeated experiments were always less than the exponential fitting errors. Therefore, errors were calculated using the average of the fitting errors from the variance-covariance matrix of repeated experiments.

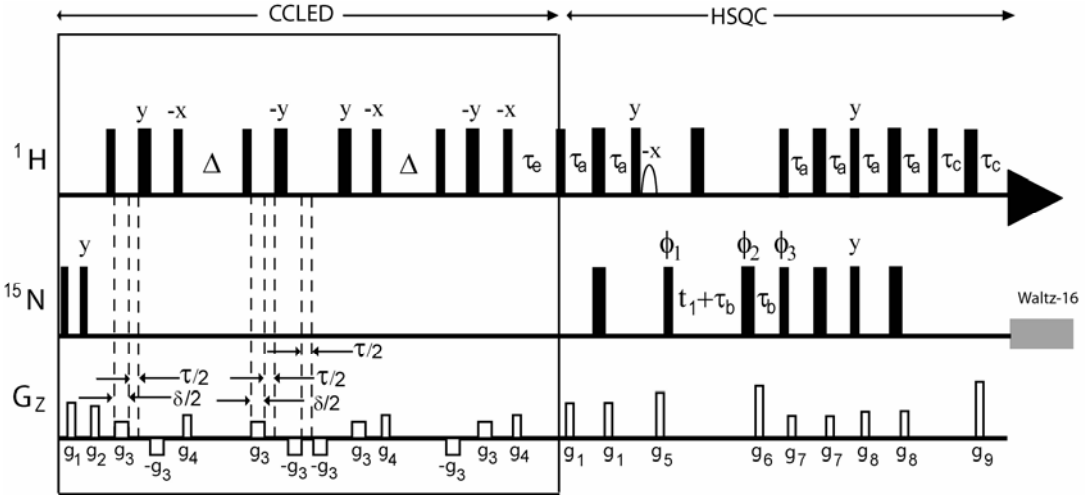

Figure 3.9 CCLED-HSQC pulse sequence with the convection compensated diffusion encoded CCLED part shown in box. Solid narrow and wide black rectangles represent 90° and 180° RF pulses, respectively. Open rectangles represent gradient pulses. The open semi-sphere is the low power 90° selective pulse on the water resonance. The phase cycle was ϕ_1 = x, -x; ϕ_2 = x,x,y,y,-x,-x,-y,-y; ϕ_3 = x; ϕ(receiver) = x,-x,-x,x. The phases of pulses were x if not indicated. The gradient durations and strengths were g_1 = (0.5ms, 16.80 G/cm), g_2 = (0.5ms, 11.76 G/cm), g_3 = (1ms, variable), g_4 = (1ms, 4.20 G/cm), g_5 = (1ms, 23.10 G/cm), g_6 = (2.5ms, 27.30 G/cm), g_7 = (0.5ms, 4.20 G/cm), g_8 = (0.5ms, 6.30 G/cm), g_9 = (0.25ms, 27.47 G/cm). The delays were Δ = 280ms for low temperature and 200ms for high temperature, δ = 2ms, τ = 0.4ms, τ_e = 1.15ms, τ_a= 2.5ms, τ_b = 2.7ms, τ_c= 0.35ms.

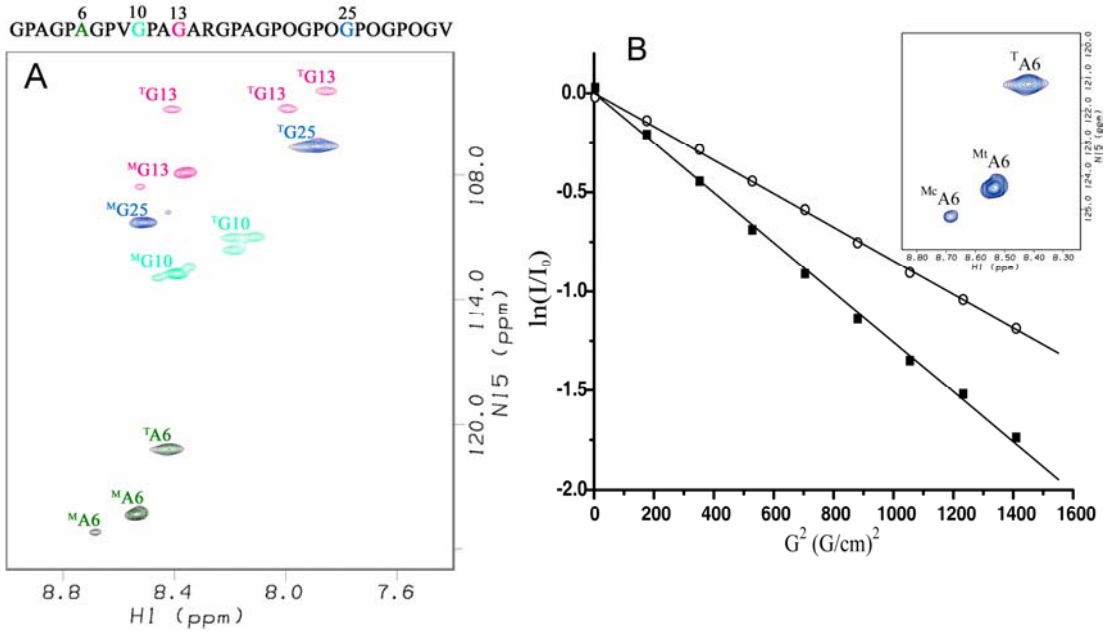

Figure 3.10 Translational diffusion spectrum and data fitting of the T1-892c peptide.

A, ^{1}H-^{15}N CCLED-HSQC spectrum of T1-892c was obtained at 10°C, with a gradient strength of 1.88 G/cm. Peptide sequence is shown above the spectrum, with labeled residues shown in color. Resonances are colored for different residues. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T, respectively.

B, Logarithmic normalized intensities of A6 trimer (○) and monomer (■) of T1-892c versus the square of the gradient strength. The solid lines are mono-exponential decay fitting of the data points. The inset shows a portion of the ^{1}H-^{15}N CCLED-HSQC spectrum of T1-892c. A6 resonances are shown. The peaks corresponding to the *trans*-monomer, *cis*-monomer or trimer state are denoted with a superscript Mt, Mc or T, respectively.

3.3.3. Results and Discussion

3.3.3.1. Residue specific diffusion measurements

Residue specific diffusion measurements were conducted using the CCLED-HSQC pulse sequence (Figure 3.9) (Buevich and Baum, 2002). Translational diffusion coefficients were obtained by single exponential fitting for monomer and trimer resonances of all labeled positions within the three peptides (T1-892n, T1892c[G10A] and T1-892c) at 10ºC and 40ºC (Figure 3.11).

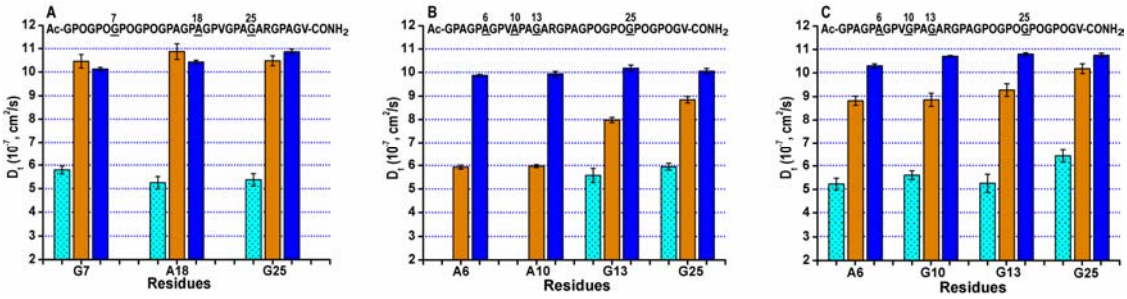

Figure 3.11 Histogram of the residue specific diffusion coefficients versus ^{15}N-labeled residues for T1-892n (A), T1-892c[G10A] (B), and T1-892c (C). $D_T^{10°C}$, $D_M^{10°C}$, and $D_M^{40°C}$ are shown in dotted cyan, orange, and blue, respectively.

At 40ºC, where the three peptides exist as unfolded monomer species, the corrected diffusion coefficients to 10ºC (denoted as $D_M^{40°C}$) for all labeled positions within a given peptide are similar to one another (9.88-10.87×10^{-7}cm^2/s). The uniformity for residues at different positions is consistent with sampling a similar averaged denatured shape. Comparison among the three peptides shows comparable $D_M^{40°C}$,

reflecting the similar molecular weights and hydrodynamic properties. At 10ºC, the monomer is in equilibrium with the trimer and the diffusion coefficients of monomer peaks (denoted as $D_M^{10°C}$) as well as trimer peaks (denoted as $D_T^{10°C}$) can be measured within a single experiment. The $D_T^{10°C}$ (5.23-6.43×$10^{-7}$$cm^2$/s) are significantly smaller than $D_M^{10°C}$ and $D_M^{40°C}$ as expected for the threefold increase in molecular weight and the acquisition of the rodlike shape.

The relationship between the $D_M^{10°C}$ and $D_M^{40°C}$ differs among the three peptides (Figure 3.11). In T1-892n the $D_M^{10°C}$ for all labeled residues (G7, A18, G25) are similar to the corrected $D_M^{40°C}$ values (10.13-10.88×$10^{-7}$$cm^2$/s) (Figure 3.11A). For T1-892n, the slow interconversion between monomer and trimer species and the uniform diffusion coefficients for all residues in each state support a simple two state equilibrium of unfolded monomer and fully triple-helical trimer at low temperature (Figure 3.12A).

For T1-892c and T1-892c[G10A], non uniform diffusion coefficients are observed along the peptide chain for the monomer and trimer resonances. For T1-892c[G10A], the $D_M^{10°C}$ show higher values at the C terminus, in contrast to the uniform $D_M^{40°C}$ values for all residues (Figure 3.11B). In particular, the two N terminal monomer positions (A6, A10) show very low $D_M^{10°C}$ values ($D_M^{10°C}$=5.93, 5.97×$10^{-7}$$cm^2$/s, respectively) that are similar to the trimer $D_T^{10°C}$ values for residues G13 and G25 ($D_T^{10°C}$=5.58, 5.95×$10^{-7}$$cm^2$/s, respectively). No trimer resonances are seen for A6 and A10 since the native state of T1-892c[G10A] is a partially folded conformation with no triple-helix N-terminal to the G10A position. To interpret the non-uniform diffusion coefficients of the monomer resonances, the populations, chemical shifts, and interconversion rates of the partially disordered trimer and the unfolded state need to be

considered (Johnson, 1993). Previous NMR data indicates that the partially disordered T1-892c[G10A] trimer is highly populated and that the monomer resonances N terminal to A10 arise primarily from the disordered regions of the partially disordered final state (Bhate et al., 2002). Interconversion rates for T1-892c[G10A] between monomer and partially disordered trimer are slow compared to the NMR diffusion experiment timescale resulting in distinct diffusion coefficients for both species (Bhate et al., 2002; Buevich et al., 2004; Liu et al., 1998). The non uniform diffusion coefficients of the monomer resonances can therefore be explained by the chemical shift overlap of the pure monomer and the partially disordered trimer; the low N terminal monomer diffusion coefficients reflect a superposition of a low population of pure monomer species and the highly populated partially disordered trimers (Figure 3.12B).

For T1-892c, a two-state model of native trimer and unfolded monomer cannot explain the differences among the $D_M^{10°C}$ values for different residues. The $D_M^{10°C}$ value for G25 ($10.16\times10^{-7}cm^2/s$) is similar to the G25 $D_M^{40°C}$ value, while $D_M^{10°C}$ for A6, G10 and G13 are smaller than $D_M^{40°C}$ values. As opposed to T1-892c[G10A], the final folded form of T1-892c has been shown to be a rigid, fully folded triple helix along the entire chain, therefore, the different $D_M^{10°C}$ values here need to be interpreted in terms of the existence of low populations of partially folded species. The similarity of the $D_M^{10°C}$ and $D_M^{40°C}$ of G25 (Figure 3.11C) indicate the G25 $D_M^{10°C}$ is not modulated by species of molecular weight higher than the monomer and thus there is no significant population of intermediates with partial unfolding at the C terminus. The decreased values for A6, G10, and G13 compared with $D_M^{40°C}$ indicate sampling by additional species of higher molecular weight. It is probable that the lower monomer diffusion coefficients reflect the

presence of trimer species that are partially disordered at the N terminus, in equilibrium with the fully folded trimer and the unfolded monomer. In this way, the diffusion coefficient of G25 reflects pure monomer species whereas the diffusion coefficients for the more N terminal residues reflect a population average of monomer species and partially disordered trimers at the N terminal end (Figure 3.12C).

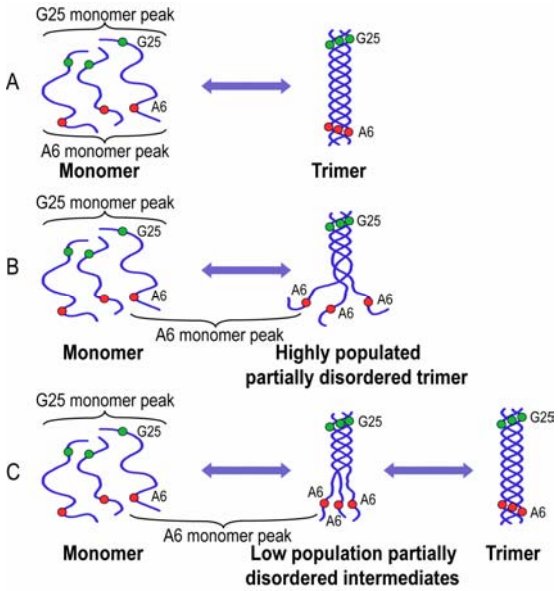

Figure 3.12 Schematic representation of the equilibrium states of T1-892n (A), T1-892c[G10A] (B) and T1-892c (C). The green circle and red circle represents the position of G25 and A6, respectively.

3.3.3.2. Relative populations of the equilibrium intermediates of the T1-892c peptide

For the model proposed above, it will be interesting to learn the time scale for the species interconversion in T1-892c based on the effect of chemical exchange on diffusion measurements. Experiments were also carried out at a different diffusion delay of 100ms, and gave essentially the same diffusion coefficients as values from 560ms delay

experiments. Therefore the interconversion could not be the intermediate time scale relative to the diffusion delay of 560ms. We have assumed that chemical shifts of the open end of partially disordered intermediates overlap with monomers, and those of the folded end overlap with trimers, which is supported by the spectrum of T1-892c[G10A]. Then the exchange rate between monomer and intermediate can be estimated from interconversion between the monomer and the folded-end of intermediates. The exchange between monomer and intermediates has to be slow on the chemical shift time scale to give two well-separated resonances. The slow interconversion rate between monomer and intermediate is also supported by the very slow monomer: trimer exchange rate. So summarizing the above discussion, it is reasonable to assume that the exchange between partially disordered intermediates (pd) and monomer (m) is slow on the diffusion time scale. Two diffusion coefficients are reflected in a single monomer resonance of the N terminal residues in T1-892c, and a bi-exponential decay will be reflected on the monomer diffusion (Johnson, 1993). Bi-exponential fits for A6, G10, G13, and G25 were performed using $I = P_m*\exp[-D_m*K] + P_{pd}*\exp[-D_{pd}*K]$, where $K=[-2\gamma^2G^2\delta^2(\Delta+2\delta/3+3\tau/4)]$ (Figure 3.13A). $D_M^{10°C}$ of G25 was used for D_m and the $D_T^{10°C}$ of G25 for D_{pd}; P_m and P_{pd} reflect relative populations of monomer and partially disordered intermediates, respectively. Residual error plots of mono- and bi-exponential fits are shown for G25 and A6 (Figure 3.13B and 3.13C). Only two floating fitting parameters P_m and P_{pd} are searched by the bi-exponential fitting algorithm as D_m and D_{pd} are fixed, thereby making the number of floating parameters the same for bi- and mono-exponential fits in this case. G25 fits equally well to mono- and bi-exponential whereas A6 has a smaller residual error with a bi-exponential versus a mono-exponential fit.

Based on the proposed model, the $D_M^{10°C}$ of G25 reflects the pure monomer species and is expected to follow mono-exponential behavior. However, the $D_M^{10°C}$ of A6, which is reduced relative to the $D_M^{10°C}$ of G25 contains a contribution from intermediate species and is expected to exhibit bi-exponential behavior. Therefore, the difference in the fitting behavior of residues G25 and A6 strongly supports the proposed model and slow interconversion rate between the equilibrium states.

Relative populations of partially disordered intermediates were obtained by bi-exponential fitting of the monomer resonance decay data (Table 3.3). The errors were estimated from the lower and upper limits of $D_M^{10°C}$ and $D_T^{10°C}$ of G25. Values of P_{pd}/P_m range from 0.13 (G13) to 0.26(A6). Relative ratios of partially disordered intermediates to trimer are not calculated because the trimer diffusion coefficient is very similar to diffusion coefficient of partially disordered intermediates due to their similar molecular weights.

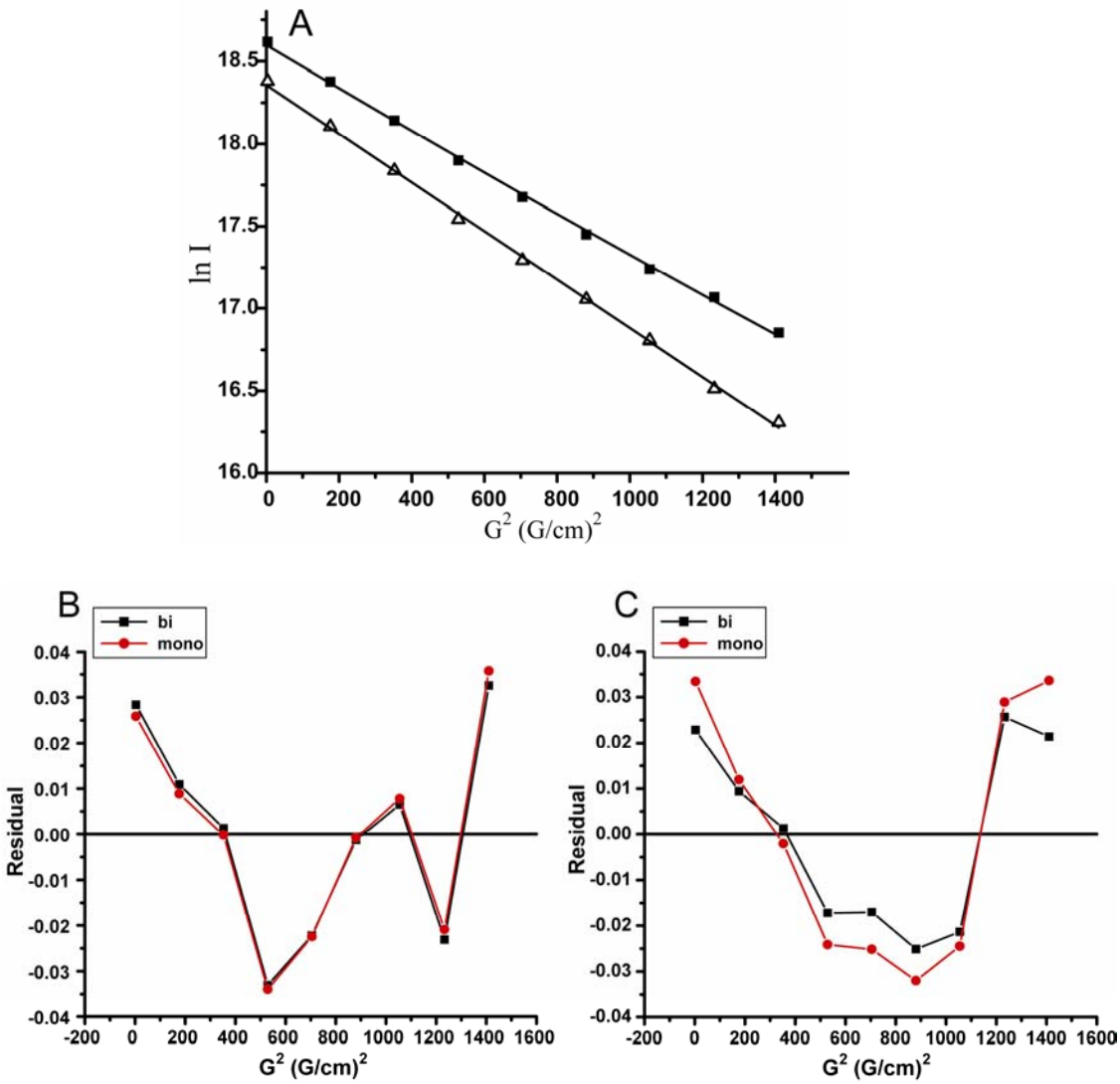

Figure 3.13 Bi- and mono-exponential fitting of T1-892c monomer diffusion decay.

A, Logarithmic intensities of A6 monomer (■) and G25 monomer (Δ) of T1-892c versus the square of the gradient strength. The solid lines are bi-exponential fitting of the data points, as described in methods section.

B and C, The residual plots of the fittings for G25 (B) and A6 (C). Residuals from bi-exponential fitting (Figure 3.13A) are shown in black, and residuals from mono-exponential fitting are shown in red.

Table 3.3 The relative population ratio of partially disordered intermediates (pd) to monomer (m) in T1-892c at 10°C.

	P_{pd}/P_m
A6	**0.26±0.08**
G10	**0.24±0.07**
G13	**0.13±0.06**
G25	**0**

3.3.3.3. Implications on the folding

The existence and the nature of the equilibrium intermediates derived from the residue specific diffusion experiments described above may be related to folding pathways of triple helical peptides. Previous kinetic real time NMR folding experiments have shown that C- to N-terminal directional folding is observed in peptides where the $(GPO)_4$ nucleation domain is at the C terminus e.g. T1-892c (Buevich and Baum, 2002; Buevich et al., 2000). T1-892n where the $(GPO)_4$ nucleation domain was placed at N-terminus has uniform diffusion along the peptide, and shows no preferential folding direction (Buevich et al., 2004). This supports that the partially disordered intermediate modulating the diffusion of T1-892c is arising from C- to N-terminus zipper-like directional kinetic pathway. It is consistent with real time diffusion measurements during the folding kinetics showing a larger dispersion of diffusion coefficients of monomers in T1-892c (Buevich and Baum, 2002). The residue specific diffusion showed here for T1-892c in the equilibrium state indicated that the intermediates from directional folding do not diminish after reaching equilibrium.

3.3.3.4. Conclusions

In summary, the appearance of non-uniform diffusion coefficients within a given peptide suggests that the method can be used to detect the presence and define the conformation of partially disordered species that are in equilibrium with the folded and unfolded states. This method allows discrimination between a simple two state model and more complex multi state models. Thus, the use of residue specific diffusion coefficients offers an approach for detection of low population intermediates in self associating systems. Such experiments may be applicable to aggregation as found in protein folding diseases.

Chapter 4

Structural and dynamic consequences of natural breaks in the non-fibrillar collagen $(Gly\text{-}X\text{-}Y)_n$ repeat

4.1. Introduction

The repeating $(Gly\text{-}X\text{-}Y)_n$ sequence of the collagen triple helix allows for easy identification of this structural motif from amino acid sequences. The collagen triple helix structure consists of three supercoiled polypeptide chains near a central axis, each in a polyproline II-like conformation (Bella et al., 1994; Brodsky and Persikov, 2005; Ramachandran and Kartha, 1955; Rich and Crick, 1961). The three chains are staggered by 1 residue, and Gly residues present as every 3rd residue in each chain so that the three chains can pack very tightly, burying the Gly residues and forming standard Rich-Crick II hydrogen bonds between the amide proton of Gly to the carbonyl group of the X residue in the neighboring chain (Bella et al., 1994; Ramachandran, 1967; Rich and Crick, 1961). The collagen triple helix is the defining motif of all extracellular matrix proteins classified as collagens and is also found as a domain in various other proteins, including C1q, macrophage scavenger receptor, collagenous tail of asymmetric acetylcholinesterase, and the bacterial proteins *Streptococcus pyogenes* scl and *Bacillus anthraces* bcl1 (Boydston et al., 2005; Deprez et al., 2000; Hoppe and Reid, 1994; Kielty and Grant, 2002; Xu et al., 2002).

Although the collagen triple-helix can be considered as a well-characterized protein motif, there are still surprises and undefined molecular features. The $(Gly\text{-}X\text{-}Y)_n$ repeating pattern is perfectly maintained in some collagen domains but not in others (Kielty and Grant, 2002). The most abundant collagens are those found in fibrils with a 670 Å axial period (types I, II, III, V, XI), and these all maintain a precise Gly-X-Y repeat throughout their ~1000-residue triple helix domain. The replacement of even one Gly by another residue as a result of a single base change results in a pathological condition (Kuivaniemi

et al., 1997; Myllyharju and Kivirikko, 2001, 2004). For instance, missense mutations leading to osteogenesis imperfecta have been identified at more than 200 Gly sites along the α1 and α2 chains of type I collagen (Byers and Cole, 2002). However, such a strict requirement of Gly as every 3rd residue throughout the triple helix domain is not followed in non-fibrillar collagens. All types of non-fibrillar collagens contain breaks in the Gly-X-Y repeating pattern (Kielty and Grant, 2002). These non-fibrillar types include collagens found in networks, such as type IV collagen; membrane proteins, such as type XVII and type XIII; FACIT collagens (types IX, XII, and XIV) found on the surface of periodic collagen fibrils; and type VII collagen in anchoring fibrils (Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004). Evidence suggests that these breaks are of functional importance, playing a role in molecular or higher order structure, or serving as recognition sites for interactions (Kielty and Grant, 2002; Miles et al., 1995).

Many breaks in the repeating $(Gly\text{-}X\text{-}Y)_n$ pattern are found in the biologically critical type IV collagen, which forms a network-like array in basement membranes of all multicellular animals. The network architecture of type IV collagen contributes to the essential function of basement membranes in providing mechanical support to cells, serving as semi-permeable barriers between tissues, and providing signals for differentiation (Hudson et al., 1993; Hudson et al., 2003; Schittny and Yurchenco, 1989; Yurchenco and Ruben, 1988). Type IV collagen is a family of six chains,α1(IV), α2(IV), α3(IV), α4(IV), α5(IV), and α6(IV), which can form three distinct heterotrimeric collagen molecules with tissue-specific locations: (α1(IV))2α2(IV) in all mammalian basement membranes; α3(IV)α4(IV)α5(IV) in renal glomerular basement membrane; and (α5(IV))2α6(IV) in smooth muscle basement membrane (Borza et al., 2001). Each

molecule has a long triple helix domain, with a globular C terminus involved in chain selection and dimerization of molecules. There are more than 20 breaks in the ~1350 residues in each chain of type IV collagen. These chains can form three distinct types of heterotrimeric collagen molecules where the breaks are often, but not always, in alignment (Zhou et al., 1992). Another non-fibrillar collagen with many breaks is type VII collagen, which forms anchoring fibrils that mediate attachment of the dermis to the epithelial basement membrane in skin {Kielty, 2002 #148}. Unlike type IV collagen, type VII collagen is a homotrimer, with all 20 breaks identical and in alignment within the molecule. Despite the presence of these breaks in native type IV and type VII collagens, the replacement of one Gly residue in the Gly-X-Y repeat can lead to a clinical phenotype. More than 150 Gly substitution mutations leading to Alport Syndrome with progressive kidney failure have been defined within the X-linked α5(IV) chain (Hudson et al., 1993; Hudson et al., 2003; Lemmink et al., 1997), while Gly substitution mutations in type VII collagen lead to the dystrophic form of epidermolysis bullosa, with scarring and blistering skin (Hudson et al., 1993; Hudson et al., 2003; Lemmink et al., 1997).

A number of possible structural perturbations have been proposed to occur at interruptions in the Gly-X-Y repeating pattern of a triple helix. A kink in the triple helix is known to occur in C1q at a site where the A chain has an extra Thr, the C chain has a Gly to Ala replacement, and the B chain has an uninterrupted Gly-X-Y sequence (Brodsky-Doyle et al., 1976; Kilchherr et al., 1985). Alternatively, an interruption could lead to a flexible site within the triple helix (Brazel et al., 1987; Hofmann et al., 1984). Visualization of type IV collagen chains by rotary shadowing shows some flexible sites, and the positions of flexible sites within the molecule correlate with several sequence

interruptions. A peptide with a break in the very stable peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$, $(POG)_3POGPGPO(GPO)_4G$ was previously observed to be very destabilizing (Long et al., 1993), and its high resolution crystal structure was reported recently (Bella et al., 2006). This GPG containing peptide showed a standard triple-helical domain on both sides of the break, with a highly localized perturbation with altered hydrogen bonding and non-equivalent chain phi, psi angles at the GPG site. The superhelix conformation on both ends is out-of-register along the axis, since the break only has 2/3 of the axial rise of the standard Gly-Pro-Hyp repeats.

In addition to affecting molecular and higher order structure, breaks in the $(Gly\text{-}X\text{-}Y)_n$ pattern can also serve as interaction sites of biological importance. For example, the sequence of the $\alpha 1$ chain of type IV collagen contains a break GAKGEPGEFYFDLRLKGDKGDP, which was shown to promote the adhesion of melanoma, ovarian carcinoma, and Jurkat cells (Furcht et al., 1994; Miles et al., 1994). In the mannose-binding protein, the serine proteinases are shown to bind near a kink in the triple helix domain (Wallis et al., 2004). A kink, flexible site, or loss of helix registration would introduce a structurally distinct site that could be involved in recognition.

Peptide models provide an approach to characterizing the effects of breaks on triple helix structure and stability (Baum and Brodsky, 1999; Bella et al., 1994; Long et al., 1993). Breaks in the form $Gly\text{-}X\text{-}Y\text{-}Gly\text{-}(AA)_n\text{-}Gly\text{-}X\text{-}Y$ are found of varying lengths in non-fibrillar collagens, where n=1, to long breaks where n=26 (Brazel et al., 1987; Zhou et al., 1992). To begin to define the structural and dynamic effects that small breaks have on triple-helix structure, NMR studies are presented here on model peptides of two types of breaks where n=4 and n=1, designating these as G4G and G1G breaks, respectively.

In the first section of the results 5.3.1., I present the studies of a G4G break on a homotrimer peptide model for a natural break found in the α5(IV) chain of type IV collagen, GPOGAAVMGPOGPO (residues 386–399), designating as GAAVM peptide. A peptide in which the Gly as every 3rd residue is restored was designed as a control, designating as GAAGVM peptide. The GAAVM peptide forms a stable triple helix, with Tm =29 °C. CD spectroscopy and differential scanning calorimetry (DSC) showed that the break decreases the triple-helix content as shown in a marked decrease in the 225 nm maximum of its CD spectrum, destabilizes the triple-helix by 10°C, and reduces the enthalpy substantially which may result from disruption of ordered water structure anchored by regularly placed backbone carbonyls (Mohs et al., 2006). NMR studies are presented here to investigate detailed molecular information of whether there is increased conformational flexibility at the break site, whether there is a way to continue a modified triple-helix through this the break, or whether a new local structure, such as a β-bend, is introduced. Here NMR triple resonance experiments are presented on a set of doubly labeled GAAVM and GAAGVM peptides (Table 3.1), allowing the assignment and tracking of individual chains, the identification of intra and intermolecular NOEs, and the determination of ϕ dihedral angles through $^{3}J_{HNHa}$ coupling constants. These NMR parameters are used as restraints in energy minimization of molecular models derived from high resolution x-ray structures of several homologous peptides, to allow the first visualization of the solution molecular conformation of a standard triple-helix and of a triple helix with a naturally occurring type IV collagen break.

In the second section of the results 5.3.2., I describe the NMR studies on a G1G break peptide. Analysis of G1G breaks in all non-fibrillar collagens show them to be the

most common kind of interruptions, and to have some relationship with G4G breaks. Peptides were studied which varied residues within the break as well as distinctive sequences surrounding the break (Geetha Thiagarajan, personal communication). The degree of destabilization is found to depend on the residue in the middle of the break as well as its environment. Using a GFG peptide (GPOGPOGPOGPOGFGPOGPOGPOGPOGY) as a prototype of the G1G breaks, NMR structural and dynamic studies were conducted and the molecular models are presented. The effects of the GFG break are compared with the G4G break in the GAAVM peptide.

4.2. Materials and methods

4.2.1. Sample Preparation

Peptides Ac-$(GPO)_4$GAAGVM$(GPO)_4$GY-$CONH_2$, designated as the GAAGVM peptide, and Ac-$(GPO)_4$GAAVM$(GPO)_4$GY-$CONH_2$, designated as the GAAVM peptide, were synthesized by Tufts University Core Facility (Boston, MA). For hydrogen change experiments and amide temperature gradient measurements of GAAVM peptides, two samples with identical sequence were selectively ^{15}N-labeled: one peptide at positions G13, A15, G24, whereas a second peptide is labeled at positions G7, G13, A14, and V16. Samples for both peptides were prepared in 10% D_2O/90% H_2O at pH 1.9 with concentrations of 5.3–5.6 mM. For all other experiments, the peptides were made with selectively ^{13}C/^{15}N doubly labeled residues: residues G13, A14, A15, G16, V17, and G25 were labeled in peptide GAAGVM; residues G13, A14, A15, V16, and G24 were labeled in the peptide GAAVM. Sample preparation and conditions are described in chapter 3.2.2.1.1.

Peptide NH_3-$(GPO)_4$GFG$(GPO)_4$GY-COOH, designated as the GFG peptide was synthesized with ^{15}N amino acids labeled at positions G13, P14, and G24. The NMR sample was prepared in 0.4M GuHCl in 10% D_2O/90% H_2O at pH 2.0 with concentration of 4.7 mM.

4.2.2. NMR chain assignments and structural measurements

NMR experiments were performed on a Varian Inova 600 MHz spectrometer equipped with a cryoprobe, and the G-matrix Fourier Transform (GFT) (5,3)D HACACONHN/HACA, CONHN experiments on GAAGVM peptide(Kim and Szyperski, 2003, 2004) were recorded on a Bruker 900 spectrometer equipped with a

cryoprobe at the New York Structural Biology Center (NYSBC) to improve dispersion and sensitivity. For GAAVM and GAAGVM peptide, HSQC, triple resonance, NOESY, HNHA experiments were performed as described in chapter 3.2. For GFG peptide, HSQC was obtained at 10°C. 3D ^{15}N edited TOCSY-HSQC (Fesik and Zuiderweg, 1988; Messerle et al., 1989) with a mixing time of 45ms and 3D ^{15}N edited NOESY-HSQC (Fesik and Zuiderweg, 1988; Marion et al., 1989; Messerle et al., 1989) with mixing times of 30-50ms were performed at 5°C and 10°C. 3D HNHA experiments (Vuister and Bax, 1993) were performed to measure homonuclear $^{3}J_{HNHa}$ coupling constants at 10°C, with a H-H coupling period of 25ms. The correction factor for the $^{3}J_{HNHa}$ coupling constants was obtained as described earlier in chapter 3.2.2.1.4.

The NOE contact map for peptide GAAVM and GFG was made from observed NH-H NOEs in the 3D ^{1}H-^{15}N NOESY-HSQC experiment and classified as NH-NH, NH-H^{α}, and NH-side chain (H^{β}, H^{γ}, H^{δ}). The generation of predicted background map and GAAGVM experimental contact map was described in chapter 3.2.2.1.6.

4.2.3. Hydrogen exchange and amide temperature gradients

Hydrogen exchange experiments were carried out at 10°C, $pD_{correct}$ 2.6 for GAAVM, $pD_{correct}$ 2.5 for GAAGVM, and $pD_{correct}$ 2.5 for GFG peptide, where the pD includes the correction for the glass electrode solvent isotope artifact (Glasoe and Long, 1960). The sample was equilibrated in H_2O at 10°C for a minimum of 72 hours to ensure that the monomer: trimer interconversion had reached equilibrium. The sample was then lyophilized and dissolved in D_2O. The HSQC spectra were acquired with a 2.5 min acquisition time for the first 34 spectra and a 10 min acquisition time for the following

spectra to allow the measurements of both fast- and slow- exchanging amide protons. Peak volumes were measured as a function of time. The volumes were normalized by the integration of the non-exchangeable resonance of DSS in 1D spectra collected before and after the HSQC experiments. The HD exchange rates were determined by a non-linear least squares fit of a single exponential with a non-zero asymptote I_∞ in equation $I(t) = (I_0 - I_\infty) \exp(-k_{ex}*t) + I_\infty$. The error in k_{ex} was the fitting error from the variance-covariance matrix. The protection factors (P) were calculated by equation $P = k_{int}/k_{ex}$, where k_{int} is the monomer exchange rate for the amide proton at a specific pH and temperature (Bai et al., 1993). The error in the intrinsic rate was estimated by assuming an experimental uncertainty of ±0.2°C and ±0.1 pH units. The error in the protection factor σ_P was then calculated using the equation

$\sigma_P = P * \sqrt{(\sigma_{kex} / k_{ex})^2 + (\sigma_{kint} / k_{int})^2}$, where σ_{kex} is the error in the measured exchange rate and σ_{kint} is the error in the intrinsic rate.

For measurements of amide proton temperature gradients, ^1H-^{15}N HSQC spectra were obtained at 0-30°C with an interval of 5°C for GAAVM and GAAGVM peptides, and at 0-20°C with an interval of 5°C for GFG peptide. The sample was equilibrated at each temperature for at least 3 hours. Amide proton temperature gradients were obtained by linear regression analysis of the amide proton chemical shifts versus temperature.

4.2.4. Molecular modeling

Computer model structures of GAAGVM without and with the incorporation of NMR data were obtained based on the crystal structure of T3-785 (PDB ID: 1BKV) (Kramer et al., 1999) and built using the Molecular Operating Environment 2005.06 (Chemical Computing Group Inc., Montreal, Canada) as described in chapter 3.2.2.1.6.

and 3.2.3. A model of GAAVM was generated based on the x-ray crystal structure of Hyp- (PDB ID: 1EI8) (Bella et al., 2006). Residues GPO-GP were substituted to GAA-VM. The resulting model was energy minimized with $^{3}J_{HNHa}$ coupling restraints using the procedures as described in section 3.2.3. A different mutation scheme was also used to substitute residues GP-GPO in the Hyp- structure to GA-AVM. Back calculation of $^{3}J_{HNHa}$ values and NOEs indicated that most structures were not consistent with all of the experimental data. A representative structure that was consistent with experimental $^{3}J_{HNHa}$ values and all ^{1}H-^{1}H NOEs was selected for GAAVM.

For the GFG peptide, computer model structures were generated based on the x-ray crystal structure of Hyp- (PDB ID: 1EI8) (Bella et al., 2006) using the same procedure as in section 3.2.3. Residues GP in the Hyp- structure were substituted to GF. The resulting model was energy minimized with dihedral angle ϕ and NOE distance restraints, and back calculation was conducted for structure validation. A set of three representative structures that was consistent with experimental $^{3}J_{HNHa}$ values and all NOEs was selected for GFG.

4.3. Results and Discussion

4.3.1. NMR studies on a G4G break

4.3.1.1. Peptide Design

In the α5 chain of human type IV collagen, there are 22 breaks in the Gly-*X*-*Y* repeating sequence distributed all along the triple helix (Figure 4.1) (Mohs et al., 2006; Zhou et al., 1992). A set of peptides was synthesized to investigate one of these naturally occurring breaks, by incorporating the sequence GPOGAAVMGPOGPO (residues 386–399) into a peptide model. Additional Gly-Pro-Hyp sequences were added on each side to ensure stability and to create a 30-mer peptide that could be compared with earlier peptides of this length (Bella et al., 1994; Buevich et al., 2004; Fan et al., 1993; Persikov et al., 2005). A peptide in which the Gly as every 3rd residue is restored was designed as a control, designating as GAAGVM peptide.

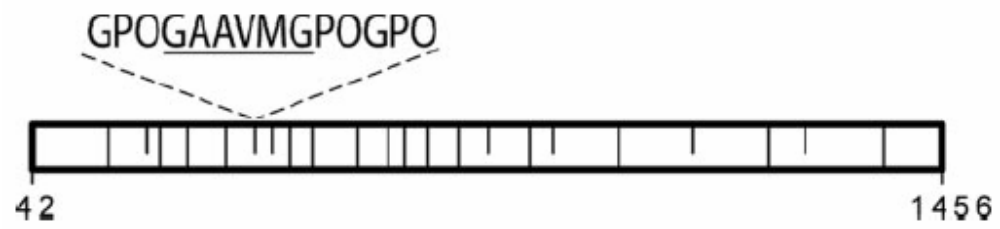

Figure 4.1 Schematic illustration of the locations of the 22 breaks within the 1414-residue Gly-*X*-*Y* repeating region (residues 42–1456) of the triple helix in the α5 chain of human type IV collagen. Breaks of the G4G type, Gly-AA_1-AA_2-AA_3-AA_4-Gly, are indicated by half-ticks, whereas other breaks are denoted by full lines. The location of the GPOGAAVMGPOGPO in the chain is indicated.

4.3.1.2. NMR Chain assignments and chemical shifts

For both peptides, all trimer resonances could be assigned to specific chains of the triple helix as indicated in the HSQC spectrum by the superscripted number (Figure 3.2 in chapter 3). The sequential assignment is derived from the triple resonance experiments and the chain stagger is derived from NOE experiments that define interchain interactions as described in chapter 3.2.2.

The HSQC spectrum of the GAAGVM peptide gives a typical pattern expected for the triple-helix. Each labeled residue shows at least one monomer peak and one or more trimer peaks, consistent with the presence of the triple-helix structure. G25 shows only a single trimer resonance due to the repetitive Gly-Pro-Hyp environment (Mohs et al., 2006), while residues G13, A14, A15, G16 and V17 all have multiple trimer peaks (two peaks for G16, and three peaks for all the other residues) due to the non-repeating sequence environment (Li et al., 1993; Liu et al., 1998; Mohs et al., 2006). A14 and A15 have similar chemical shifts to each other which may reflect their similar environments in the GAAG sequence. The HSQC spectrum of the GAAVM peptide again shows three trimer peaks as well as monomer peaks for each labeled residue. This indicates that a well-defined structure with 3-non-equivalent chains is present in the region where the Gly-X-Y pattern is broken. G24, within the Gly-Pro-Hyp repeating region at the C-terminus, shows only a single trimer resonance at the typical triple-helix position similar to G25 above in GAAGVM, while three trimer peaks are seen for residues G13, A14, A15 and V16. Comparison of the chemical shifts of the trimer resonances shows changes in chemical shift for most resonances in GAAVM compared with GAAGVM indicating that there is a difference in the local electronic environment of the central region. Most

notably, the Val resonances in the GAAVM peptide have shifted downfield very significantly in both the ^{1}H and ^{15}N dimensions with $^{3}V16$ shifting the most from 7.8ppm down to 9.4ppm. The large downfield shift may reflect a change in the hydrogen bonding behavior of Val in GAAVM relative to GAAGVM.

4.3.1.3. Experimental NOE observations

In order to compare the interchain contacts in GAAVM and GAAGVM peptides, first a predicted contact map in a classic triple helical conformation is generated from a GAAGVM model structure as described in chapter 3.2.2.2. (Figure 3.5A and 4.2B). A contact map (Figure 3.5B and 4.2B) was generated for the experimental NOE data of the GAAGVM peptide, now using circles to denote the experimental values and highlighted squares of intramolecular (gray) and intermolecular (yellow) NOEs to represent predictions from the GAAGVM model structure (Figure 3.5A and 4.2A) (details see chapter 3.2.2.2.). The superimposition of GAAGVM experimental and predicted contact indicates that the solution conformation of the GAAGVM peptide is consistent with a typical triple helix similar to the model structure.

A contact map diagram was constructed for the experimental NOE data for GAAVM in order to compare the solution conformation of the peptide with the break to the predicted intra- and intermolecular contacts in the GAAGVM model, which are again represented as highlighted squares (Figure 4.2C; Highlighted boxes are taken from Figure 4.2A). The experimental NOEs of GAAVM, represented as circles, were placed on the background of the predicted NOEs from the GAAGVM model and show a 1- residue stagger of the triple-helix throughout the GAAVM break even though a standard triple-helix is not possible due to the absence of the Gly residue. The diagnostic interchain NH-

NH, NH-H^{α} NOEs between the three chains of the G13 residues (^{1}G13NH to ^{2}G13NH, ^{1}G13NH to ^{2}G13H^{α}, ^{2}G13NH to ^{1}G13H^{α}, ^{2}G13NH to ^{3}G13NH) are seen as well as interchain backbone NOEs from ^{1}A14 to ^{2}G13 and ^{2}A15 to ^{3}A14 supporting the one residue stagger. In addition, new interchain NOEs between the three chains of the V16 residues at the interruption site as well as between the V16 residues and the backbone protons of residues G13 and A14 are observed in the GAAVM peptide. These new NOEs observed for V16 at the break are similar to the predicted NOEs for G16 in the Gly-X-Y repeating sequence of the GAAGVM model. This suggests that the amide group of V16 is packing into the center of the triple helix in this break region, attempting to mimic the features of the much smaller Gly residue found at the same position as the Val in the GAAGVM peptide.

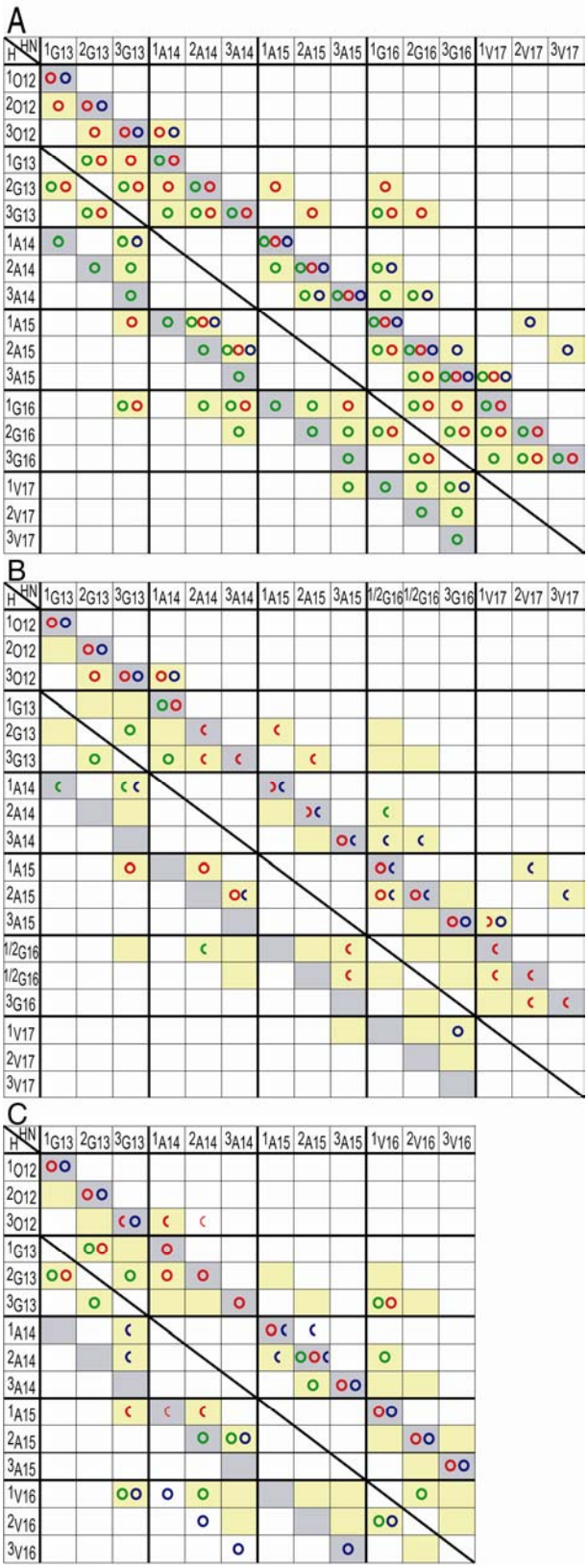

Figure 4.2 Comparison of experimental NOEs from the NOESY-HSQC experiment for the GAAGVM and GAAVM model peptides.

A, Predicted contact map obtained from GAAGVM molecular model structure. Contacts are shaded in gray for intrachain distances less than 5Å and in yellow for interchain distances less than 5 Å. Predicted HN-HN(O), HN-H^{α}(O), and HN-side chain protons(O) interactions less than 5 Å are overlaid on the shading so that one can see that they come from both intra- and intermolecular contacts in a complex manner because of the supercoiled trimeric structure of the triple-helix.

B, Experimental contact map of peptide GAAGVM (represented by circles), overlaid on the shaded contacts of the GAAGVM model, indicating that some of the predicted interchain as well as intrachain NOEs are observed. The fact that many predicted NOEs are not observed is largely attributable to overlap. Half circles( ) represent overlapped NOEs consistent with the GAAGVM model.

C, Experimental contact map for GAAVM. The observed GAAVM NOEs (circles) are overlaid on the background of the predicted intra- and intermolecular contacts in the GAAGVM model, showing the expected intrachain and interchain NOEs and some new interchain NOEs consistent with the packing of the Val residues at the break and the 1-residue stagger between chains. In panel *C*, half circles ( ) and thinner half circles ( ) represent overlapping NOEs.

4.3.1.4. Determination of J coupling constants

Additional NMR measurements were carried out to obtain $^{3}J_{HNHa}$ coupling constants for both the GAAGVM and GAAVM peptides in order to compare the range of dihedral angles that can be adopted in the break with those in a standard triple-helix with a Gly-X-Y repeat (Figure 4.3). For GAAGVM the $^{3}J_{HNHa}$ coupling constant values of A14, A15 and one of the H^{α} of Gly13, Gly 16 and Gly 24 are relatively uniform in the range of 4-5 Hz. Val appears to have a slightly larger value in the range of 6-7 Hz. Typically $^{3}J_{HNHa}$ values around 5Hz pose a multiple minimum problem and experimental determination of a 5Hz coupling constant generates four possible ϕ angles from the parameterized Karplus equation (Vuister and Bax, 1993) $^{3}J_{HNHa} = 6.51\cos^{2}(\phi\text{-}60°) - 1.76\cos(\phi\text{-}60°) + 1.6$. One of the solution sets gives values of approximately -67° which is consistent with average Gly dihedral angles ϕ of –69.1° for T3-785, suggesting that the ϕ angles for the labeled residues in the GAAGVM peptide are consistent with the ϕ angles found in the polyproline II (PPII) conformation (Kramer et al., 1999). In contrast, the GAAVM peptide does not have uniform coupling constants along the entire peptide chain; values of G13 and A14 as well as the value of G24 at the C terminal end are very similar to those in GAAGVM. The coupling constant values of A15 and V16 in the interrupted peptide are now in the range of 8.5-9.5 Hz, shifted from the 5-7Hz values seen in GAAGVM. There are two solutions to the Karplus equation for values of around 9Hz that include ϕ angles of -100° and -140° respectively. The comparison of the values of the coupling constants between GAAGVM and GAAVM indicate that the distortion of the ϕ angles is very local and that A15 and V16 adopt conformational angles outside the PPII region.

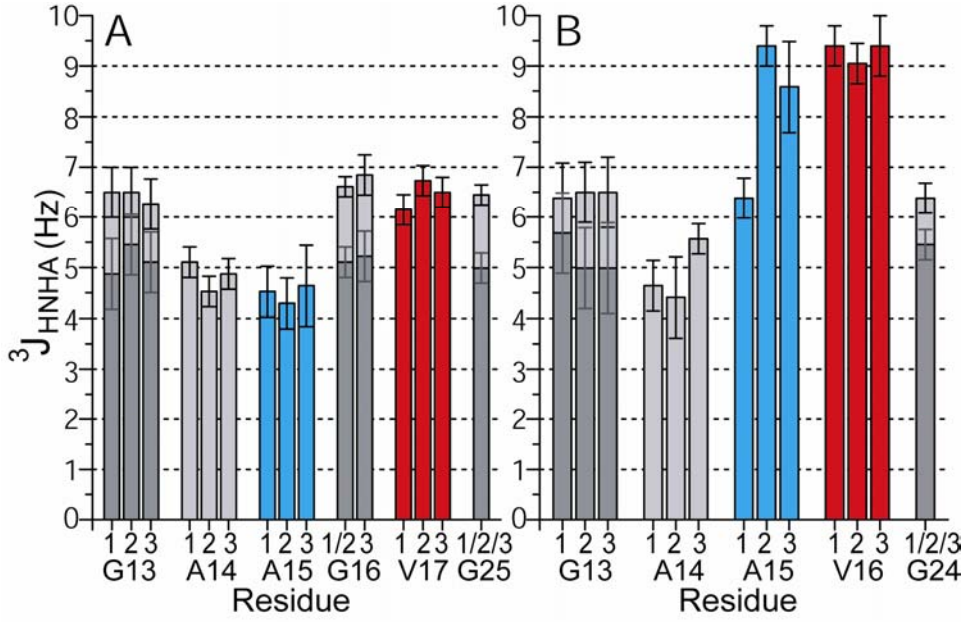

Figure 4.3 Experimental $^3J_{HNHa}$ coupling constants of peptides GAAGVM (A) and GAAVM (B). A15 and V17/V16 residues are colored in blue and red, respectively. $^3J_{HNHa}$ coupling constants from the two H^{α} Gly residues are shown in dark gray and light bars.

4.3.1.5. Hydrogen exchange and amide temperature gradients

Evaluating the dynamics at the interruption site and predicting the existence of hydrogen bonding is important to understanding the nature of the interruption. In the hydrogen exchange experiment of triple helical peptides, the effect of complete unfolding can be eliminated as NMR folding data showed that the triple helical peptide folding is a very slow process, taking place within minutes or even hours (Liu et al., 1996). Therefore on the millisecond to second time scale of the hydrogen exchange experiment, complete unfolding is too slow to have effects on exchange rates. Yet the local unfolding of the triple helix, can play a role in the hydrogen exchange. The solvent accessibility and hydrogen bonding also affects the exchange. The central Gly exchanges

slower than residues at the X and Y positions because the amide proton of Gly is hydrogen bonded and buried in the center of triple helix (Fan et al., 1993). Therefore hydrogen exchange experiments are useful to estimate the hydrogen bonding and to analyze the sequence-dependent mobility.

For the GAAVM and GAAGVM peptides, rates of hydrogen/deuterium exchange for the individual amide protons were determined at 10 °C, and protection factors were calculated from these exchange rates by taking the ratio of the theoretical monomer rate to the trimer rate for the same residue. In the GAAGVM peptide, a very high degree of protection from exchange was seen for all Gly residues at 13, 15, and 25 positions. Residues at X and Y positions have very low protection factors, indicating that these residues are fully exposed to solvent and are not hydrogen bonded (Figure 4.4A). For the GAAVM peptide, a similarly high degree of protection (protection factors of 346 and 461) was seen at the C- and N-terminal ends of the peptides as in the GAAGVM peptide (Figure 4.4B). The central Gly residues at the beginning of the break, G13, show a 2–3-fold reduction in its protection factor (average 155) relative to the terminal G7 and G24, which could relate to decreased strength of hydrogen bonding, shielding from solvent, or increased flexibility in the hydrogen exchange time scale. The 2 Ala residues in the break, A14 and A15, have protection factors in the range of 2–18, indicating that these 2 residues are fully exposed to solvent and are not hydrogen bonded. The three resonances assigned to V16 residue, with their atypical downfield-shifting chemical shifts, show a protection factor of ~102, indicative of significant shielding from solvent or hydrogen bonding but somewhat lower than the G13 value.

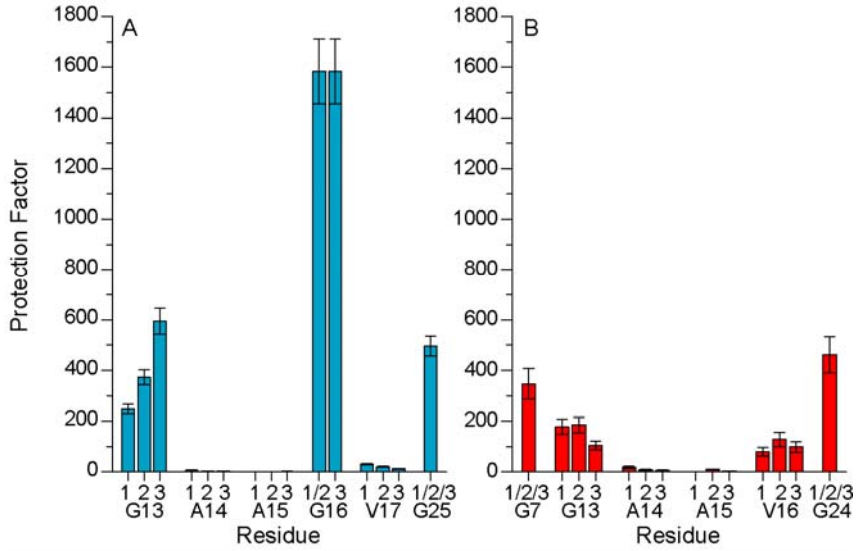

Figure 4.4 Histogram of H/D protection factors for the labeled residues in GAAGVM (A) and GAAVM (B) peptide.

Complementary amide proton temperature gradients were measured in the GAAGVM and GAAVM peptide in order to resolve the existence of hydrogen bonds in the peptides (Figure 4.5). In the GAAGVM peptide, G13, G16 and G25 residues have NH temperature gradients that are consistent with hydrogen bonding as expected (Figure 4.5A) (Baxter and Williamson, 1997; Cierpicki and Otlewski, 2001). The A14, A15 and V17 residues show more negative temperature gradients with the X position residues A14 and V17 closer to the cutoff line, supporting the fact that these NH vectors of the X and Y residues point towards the solvent.

In the GAAVM peptide, as expected the NH temperature gradient for G7 and G24 is more positive than -4.6 ppb/°C indicating the existence of hydrogen bonds at the GPO rich N and C terminal end (Baxter and Williamson, 1997; Cierpicki and Otlewski, 2001). Similarly, G13 at all three chains has a more positive temperature gradient than -4.6 ppb/°C indicating the formation of a hydrogen bond at the G13. This is consistent with a

degree of protection from solvent at G13 in the hydrogen exchange experiment, and the combination of temperature gradient and hydrogen exchange data indicates an increased flexibility at G13. The negative values of NH temperature gradient (-9 to -6 ppb/°C) for A14 and A15 suggest that they are not involved in hydrogen bonding and are consistent with their very low protection factors. The NH temperature gradients of three V16 residues are non-uniform with gradients of ^{1}V16 and ^{2}V16 close to the cutoff line while ^{3}V16 shows a very negative gradient (-11.1 ppb/°C). It supports non-equivalence in hydrogen bonding for three chains at V16. Interestingly, the temperature gradient of the most downfield-shifted ^{3}V16 is more negative than that of monomers (~ -10 to -6 ppb/°C). One caution that needs to be taken when interpreting the amide temperature gradient in peptide is the conformational averaging problem (Andersen et al., 1997). If the peptide is undergoing thermal unfolding to a coil state during the measurement of amide temperature gradients, the observed amide chemical shifts would be population weighted averages of the folded and unfolded states and the temperature gradient would correlate with chemical shift deviation from the random coil states and no longer be a simple indication of hydrogen bonding by the -4.6 ppb/°C cutoff (Andersen et al., 1997). In order to investigate this possibility, the ^{15}N relaxation measurement was performed at 25°C, the T_m of the GAAVM peptide and the highest temperature data point for the temperature gradient measurement. ^{15}N R_2 of ^{3}V16 is very similar to the rigid and hydrogen-bonded control residues of G24 in the GPO repeating region and of G25 in the $(GPO)_{10}$ peptide (Table 4.1), indicating that there is no conformational average observed at ^{3}V16 resonance even at high temperature. It is not clear why the ^{3}V16 amide has a

very negative temperature gradient but a similar protection factor to the other two V16. It could be related to hydrogen bonds that are mediated by water.

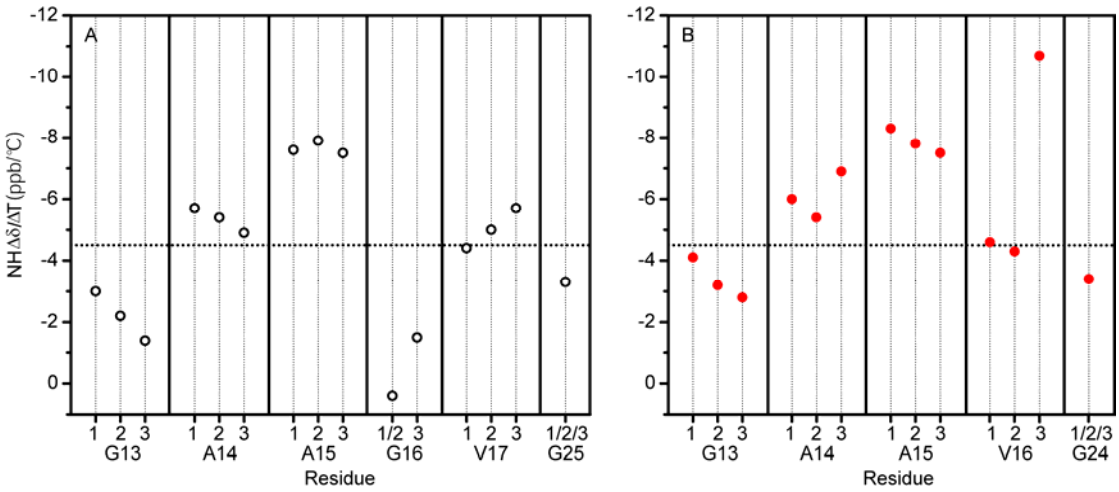

Figure 4.5 Amide NH Δδ/ΔT plot for GAAGVM (A, black open circles) and GAAVM (B, red filled circles) peptide. Staggering chain numbers are indicated for each residue. The dotted horizontal line corresponds to Δδ/ΔT = -4.5 ppb/°C which provided the cutoff line of hydrogen bonding.

Table 4.1. ^{15}N R_2 and NOE of the GAAVM peptide

Residues	^{15}N R_2 at 25°C	^{15}N NOE at 0°C
^{1}G13	10.09 ± 0.48	0.53 ± 0.06
^{2}G13	9.85 ± 0.41	0.81 ± 0.04
^{3}G13	10.45 ± 0.43	0.68 ± 0.13
^{1}A14	9.54 ± 0.65	0.82 ± 0.08
^{2}A14	8.62 ± 0.13	0.62 ± 0.08
^{3}A14	8.51 ± 0.46	0.53 ± 0.08
^{1}A15	8.78 ± 0.61	0.75 ± 0.11
^{2}A15	9.34 ± 0.86	0.81 ± 0.06
^{3}A15	8.33 ± 0.64	0.58 ± 0.11
^{1}V16	9.01 ± 0.30	0.49 ± 0.08
^{2}V16	8.67 ± 0.81	0.53 ± 0.08
^{3}V16	8.46 ± 0.63	0.60 ± 0.08
G24	9.53 ± 0.18	0.57 ± 0.06
G25 in the $(GPO)_{10}$ control peptide	9.34 ± 0.13	

In summary, the combination of the hydrogen exchange data and the amide temperature gradients suggests that the dynamics and the nature of the hydrogen bonding around the break site in the GAAVM peptide are different from the control peptide. The GAAVM peptide is more flexible on the slow hydrogen exchange timescale at the break as well as at the G13 just N-terminal to it.

4.3.1.6. Molecular modeling

The ϕ angle restraints from $^{3}J_{HNHa}$ measurements were used to refine the model of GAAGVM derived from an x-ray crystal structure as described in the previous section 4.3.1.3. The refined model did not differ significantly from the original model, indicating that the original model derived from x-ray structure is in good agreement with NMR (Figure 4.6A,B). In the GAAGVM model, all ϕ,ψ angles fall into the polyproline II (PPII) region of the Ramachandran plot, consistent with a well-behaved triple-helix (Figure 4.6C).

The agreement between the model obtained from an x-ray crystal structure and the NMR solution data for the GAAGVM peptide established a strategy to use NMR data to define the conformation in solution for a GAAVM break. For this peptide, a model was generated based on a recent x-ray structure with a break in the Gly-X-Y repeating sequence (Bella et al., 2006). Our peptide differs from the peptide solved by x-ray in having a 4-residue break compared with a 1-residue break and in having a natural type IV collagen sequence compared with the repeating $(POG)_n$ sequence. The central GPOGP sequence of the crystal structure peptide $(POG)_2$-POGPOGPG-$(POG)_5$ was replaced by GAAVM, and the structure was energy minimized incorporating ϕ angle restraints from

experimental GAAVM NMR $^{3}J_{HNHa}$ couplings. The structure obtained gives good agreement with NOE NMR data.

The principal features of the model preserves the 1-residue stagger between the 3 chains and introduces close packing of Val near the central axis (Figure 4.6D, E). Although, the larger Val residues are unable to fit into the center like the small Gly residues. The three Val residues form a small localized hydrophobic core in GAAVM (Figure 4.6E). As a result of localized changes in the dihedral angles of 8 residues (Figure 4.6F), the three Val can be placed in this central location, maintaining the rod-like structure without creating a bulge. The GAAVM region may be denoted as a "pseudo-triple-helix" because it retains the 1 residue stagger and the close packing of the 3 chains, and differs only in the dihedral angles of a small number of residues. The absence of one residue in GAAVM versus GAAGVM creates a loss of axial register of the triple-helix on both sides of the break, as seen in the crystal structure for peptides with breaks (Bella et al., 1994; Bella et al., 2006).

The Ramachandran plot of the GAAVM model indicates a local disruption of the PPII dihedral angles relative to the model of GAAGVM. The V16 dihedral angles are the most perturbed with all three chains falling outside of the PPII range (Figure 4.6F). In addition, the dihedral angles of ^{2}A15 and ^{3}A15 (just N terminal to V16), ^{2}A14, ^{1}M17, and ^{1}G18 all are outside the PPII region. The highly localized nature of the perturbation is recognized by the observation that only 8 residues in the 3 chains are altered from standard collagen dihedral angles, suggesting that the well defined deformation in structure is primarily absorbed by a subset of residues in the break so that one can optimize the extent of normal triple-helix structure on both sides of the break.

A number of lines of argument support this GAAVM model obtained by molecular modeling with NMR J-coupling restrictions. Back calculation of the $^{3}J_{HNHa}$ and NMR distances from the model shown in Figure 4.6D and comparison to the experimentally observed $^{3}J_{HNHa}$ and distances indicate that the proposed model has a good fit to the experimental data (data not shown). Other models tested did not show good agreement with the experimental data. Therefore, although the GAAVM model is not unique, it provides a representative example of the structure that can be adopted by the peptide containing the break that fits the experimental data and, as described below, it explains the observed consensus sequence of 4 residue breaks. The NMR data on the GAAVM peptide in solution is consistent with the rod-like nature of the triple-helices and the highly localized nature of the structural perturbation of previous x-ray structures on two peptides with small breaks (Bella et al., 1994; Bella et al., 2006). The increased conformational flexibility at the break site reported for the crystal structure is consistent with the faster hydrogen exchange rates at G13 in GAAGVM versus GAAVM (Mohs et al., 2006).

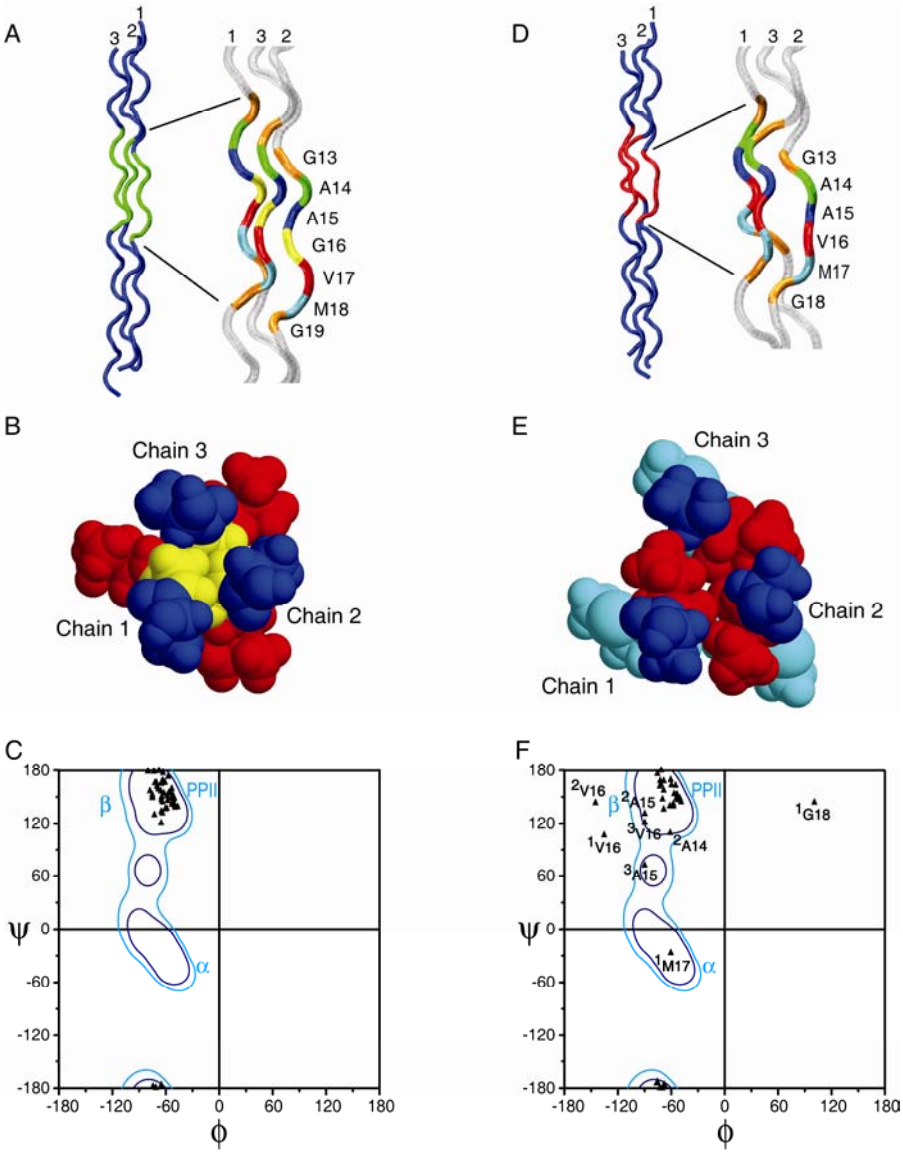

Figure 4.6 Model structures of peptide GAAGVM with a standard Gly-X-Y pattern and peptide GAAVM containing a type IV collagen break.

A, D, Ribbon diagram of GAAGVM (A) and GAAVM (D) represented with the N-terminal at the top and chain numbers. Leading, middle and trailing chains are represented as 1, 2 and 3. For ease of comparison both peptides are shown with comparable lengths by selecting residues 4 to 30 for GAAGVM and residues 1 to 26 for GAAVM. The GAAGVMG segment in peptide GAAGVM and the GAAVMG segment in peptide GAAVM are colored in green and red, respectively. Zoomed views of those two segments are shown on the right of their ribbon diagram. For GAAGVM, residues are color coded as G13 (■), A14(■), A15(■), G16 (■), V17 (■), M18(■), and G19 (■). For GAAVM, residues are color coded as G13 (■), A14(■), A15(■),V16 (■), M17(■), and G18 (■).

B and E, Space filling model of the cross section view from the N terminus to the C terminus of GAAGVM (B) and GAAVM (E). Views show that the GAAVM region adopts a "pseudo-triple-helix" which preserves the 1-residue stagger between the 3 chains and introduces close packing and hydrophobic interactions of the three Val residues at the site of the break. Residues are color coded similarly to the zoomed ribbon diagrams. The AGV segment in GAAGVM is colored as A15(■), G16 (■), and V17 (■), and the AVM segment in GAAVM is colored as A15(■),V16 (■), and M17(■).

C and F, Ramachandran plots for model structures of GAAGVM (C) and GAAVM (F). To highlight the central region of the peptides, only residues 6 to 21 in GAAGVM and residues 6 to 17 in GAAVM are plotted and shown in black triangles. Ramachandran contour map for Pro residues is shown in the background and typical secondary structures are indicated (α, α-helix; β, β-sheet; PPII, polyproline II and collagen).

4.3.2. NMR studies on a prototypical G1G break

4.3.2.1. The chain assignment and conformational characterization

NMR studies were carried out on the GFG peptide, which represents one prototype of the G1G breaks. In the HSQC spectrum of the GFG peptide (Figure 4.7A), each labeled residue shows at least one monomer peak and one or more trimer peaks, consistent with the formation of trimers at all positions. Residues G13 and F14 have three well separated trimer peaks, indicating the presence of a well-defined conformation with 3-non-equivalent chains at the G1G break site. Residue G24 in the Gly-Pro-Hyp repeating region at the C-terminus shows only a single trimer resonance with the typical triple-helix chemical shifts for Gly in the Gly-Pro-Hyp repeating region (Mohs et al., 2005; Mohs et al., 2006). Trimer resonances for G13 and F14 were assigned to specific chains of the triple helix from the strong sequential NH_i-$H\alpha_{i-1}$ NOEs of the NOESY-HSQC experiment and the chain stagger was derived from NOEs that define interchain interactions. For example, the observation of an NOE between ^{2}G13 NH of chain 2 and ^{3}O12 H^{α} of chain 3 indicates a 1-residue stagger between these chains.

An NOE contact map diagram (Figure 4.7B) was constructed for the experimental NOE data for GFG in order to characterize interchain contacts and chain stagger. The experimental NOEs of GFG are represented as circles. Intrachain NOEs are shaded in gray. Additional interchain backbone NOEs (^{2}G13 NH to ^{3}O12 H^{α}, ^{1}F14 NH to ^{3}O12 H^{α}, ^{3}G13NH to ^{2}G13H^{α}, ^{3}G13NH to ^{1}F14NH) are observed, supporting the one residue stagger of the triple-helix throughout the GFG break even though a standard triple-helix is not possible due to the absence of one residue in the Gly-X-Y triplet. The limited number of NOEs is due to resonance overlapping in the NOESY spectrum.

$^{3}J_{HNHa}$ coupling constants which can be related to the dihedral angle ϕ are obtained for the trimers of GFG peptide from HNHA experiments (Figure 4.7C). All $^{3}J_{HNHa}$ coupling constants are uniform in the range of 5-6 Hz except ^{2}F14, which has a large J coupling value of 8 Hz. Although multiple solutions of angle ϕ are possible from the parameterized Karplus equation (Vuister and Bax, 1993), the precise one can not be determined from the $^{3}J_{HNHa}$ measurement alone. It indicates that the non-equivalent distortions of backbone conformations are adopted by three chains at the G1G break site.

4.3.2.2. Hydrogen bonding studies of the GFG peptide: hydrogen exchange and amide temperature gradients

Hydrogen exchange experiments were performed on the GFG peptide to explore the protection of labile amide protons from solvent (Figure 4.7D). The protection factors were calculated by taking the ratio of the theoretical monomer rate to the trimer rate for the same residue. The experimental monomer rate is too fast to be measured accurately, therefore the theoretical monomer rate is used. A very high degree of protection from exchange (protection factors of 457) was seen for G24 at the stable C- terminal Gly-Pro-Hyp regions. The G13 show a 15-25 fold reduction in their protection factors (average 27) relative to G24, and residues F14 shows an even lower protection factor than G13 (average 11). Low protection factors could be related to decreased strength of hydrogen bonding, shielding from solvent, or local unfolding/breathing of the trimers.

Complementary amide proton temperature gradients are performed on the G1G peptide in order to resolve the existence of hydrogen bonds in the peptide (Figure 4.7E). As expected, the NH temperature gradient for G24 is more positive than -4.5 ppb/°C indicating the existence of hydrogen bonds at the GPO rich C terminal end (Baxter and

Williamson, 1997). Similarly, ^{3}G13 has a more positive temperature gradient than -4.5 ppb/°C indicating the formation of a hydrogen bond at the Gly of chain 3. The NH temperature gradients of the G13 residues in chains 1 and 2 are more negative than -4.5 ppb/°C supporting the non-equivalence in hydrogen bonding for three chains at G13. The more negative values of NH temperature gradient for three F14 residues suggest that they are not involved in hydrogen bonding.

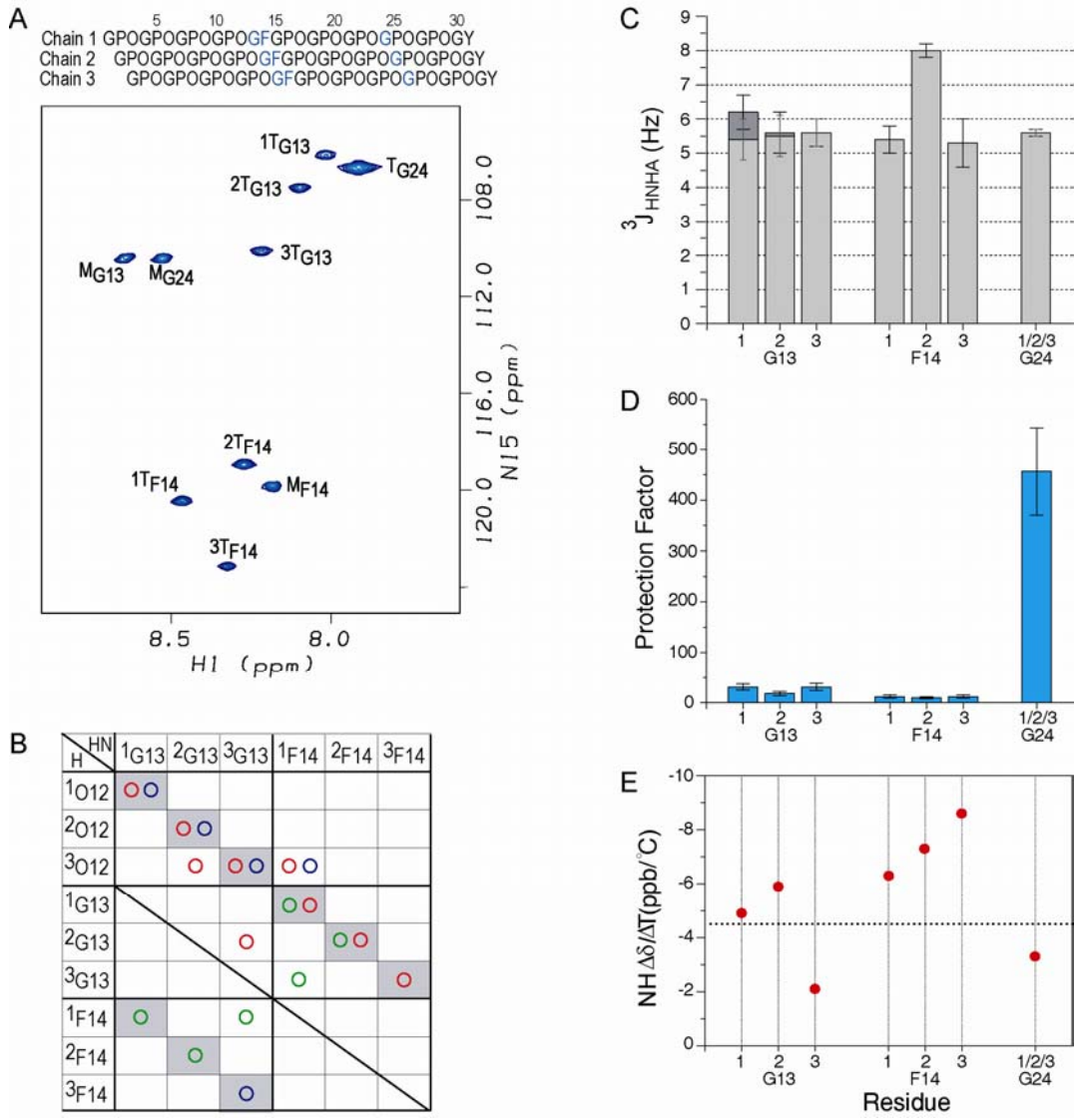

Figure 4.7 NMR structural, dynamic and hydrogen bonding characterization of the GFG peptide.

A, ^{1}H-^{15}N HSQC spectrum of peptide GFG at 10°C. Sequence diagram of the peptide above the HSQC shows the characteristic one residue stagger. The corresponding isotope labeled residues are colored in blue. All labeled residues GF and G have trimer peaks as well as monomer peaks, showing the GFG break is incorporated into a trimer. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T, respectively. Leading, middle, or trailing chain stagger assignment is indicated as chain 1, 2, or 3 as a number in front of the superscript T, respectively.

B, Comparison of experimental NOEs from the NOESY-HSQC experiment for the GFG peptide with a standard triple helix conformation. Predicted contact map obtained from a standard triple helix model structure are shown in shaded boxes as background (Li et al., 2007). Contacts are shaded in gray for intrachain distances less than 5Å and in yellow for interchain distances less than 5 Å. Experimental NOEs for GFG peptide are represented by circles (HN-HN(**O**), HN-H$^{\alpha}$(**O**), and HN-side chain protons(**O**)) and are overlaid on the shaded contacts, showing the interchain contacts are consistent with one residue staggering of triple helix.

C, Experimental $^{3}J_{HNHa}$ coupling constants of peptide GFG. $^{3}J_{HNHa}$ coupling constants from the two H$^{\alpha}$ Gly residues are shown in dark gray and light bars.

D, Histogram of hydrogen/deuterium protection factors for the labeled residues in GFG peptide, indicating that the G13 and F14 are more exposed, less hydrogen bonded, or more flexible than the control G24 in the GPO region.

E, Amide NH $\delta\Delta/\Delta T$ plot for GFG peptide. The dashed horizontal line corresponds to $\delta\Delta/\Delta T$ = -4.5 ppb/°C which provided the cutoff line of hydrogen bonding. G24 and ^{3}G13 appear to have NH temperature gradients that are consistent with hydrogen bonding.

4.3.2.3. Molecular modeling

In order to define the conformation in solution for a GFG break, molecular modeling was performed with the incorporation of NMR data using the strategy described previously (Li et al., 2007). Models were generated based on a x-ray structure of a peptide where the central GPG sequence of the crystal structure peptide $(POG)_3$-PO-GPG-$(POG)_5$ was replaced by GFG (Bella et al., 2006). The structure was energy minimized incorporating ϕ angle restraints from $^3J_{HNHa}$ measurements and distance restraints from NOESY experiments. Back calculation of NMR parameters from the resulting structures were performed to eliminate structures that were not consistent with experimental data. A set of three structures were obtained that gave good agreement with NMR data.

All three models present a rod like structure without a kink or bulge at the break site and with standard triple helical structures on both sides of the break. The 1-residue stagger between the 3 chains is preserved at the break region (Figure 4.8A). The G13 residues in all 3 chains are closely packed near the central axis although a standard triple helix can not be formed because of the residue deletion. The three F14 residues are located on the outside of the triple-helix, similar to residues found at X and Y positions, but with a certain deformation (Figure 4.8B). The Ramachandran plot of the GFG models indicates a local disruption of the PPII dihedral angles. In all three models, the F14 residues are the most distorted with two chains falling outside of the PPII region and in two models the G13 have two chains outside the PPII region (Figure 4.8C). In addition, the dihedral angles of ^{3}P11, ^{2}O12, ^{3}O12, ^{1}G15, and ^{2}G15 are non-PPII. The deformation of the break is highly localized, and only residues 11 to 15 have dihedral angles which

deviate from the standard triple helical conformation. These models indicate a well defined perturbation which is absorbed by a small number of residues. A significant variation among the three acceptable model structures can be observed around residue G15 in chain 1, as seen for the red chains in figure 4.8A and in the Ramachandran plot. There is insufficient NMR data on the two labeled residues G13 and F14 to determine whether these variations arise from flexibility at that site.

An examination of hydrogen bonding in the three models indicates that the G13 residue in the three chains adopts different and non-standard hydrogen bonds. The amide proton of ^{3}G13 is hydrogen bonded to the carbonyl group of ^{2}G13 in all three models, while amide protons of ^{1}G13 and ^{2}G13 are not hydrogen bonded, or have standard Rich-Crick hydrogen bonds to the carbonyl group of P11 in the neighboring chain. The non-equivalence in hydrogen bonds of three G13 residues in the three models is consistent with the amide proton temperature gradients of G13, with ^{3}G13 showing a more positive temperature gradient than G13 residues in the other two chains. All three G13 residues have low protection factors compared with G24 in the C-terminal stable region indicating an increased flexibility that may be due to local breathing at the break site. The Phe chains are exposed on the outside of the helix which is consistent with their very low protection factors and some flexibility.

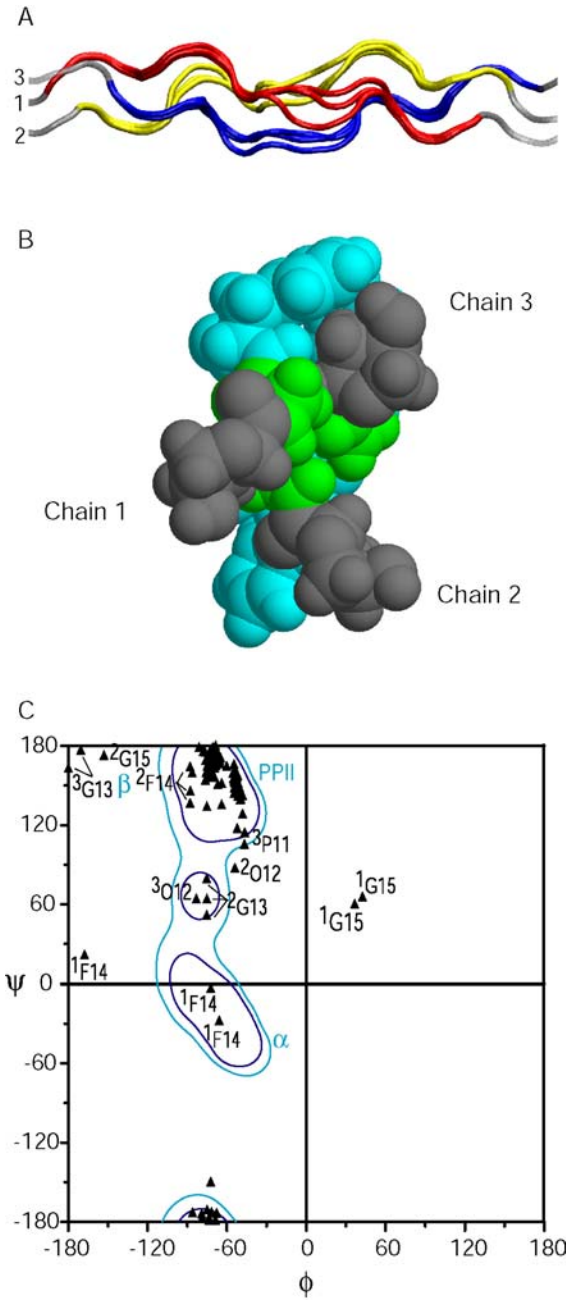

Figure 4.8 Model structures of peptide GFG.

A, Ribbon diagram of GFG models represented with the N-terminal at the left and chain numbers. The NMR parameters are used as the restraints in the energy minimization, including the phi angle restraints from J coupling constants and distance restraints from NOEs. A set of three model structures that are consistent with all NMR experimental data is shown. Leading, middle and trailing chains are represented as 1, 2 and 3, and are color in red, yellow, and blue, respectively, for energy minimized segment of residues 6 to 19.

B, Space filling model of the cross section view from the N terminus to the C terminus of one model structure of GFG. Views show that at the GFG region the G13 residues are closely packed at the center while the F14 residues are outward. The OGF segment is colored as O12(■), G13(■), and F14(■). **C,** Ramachandran plot for three possible model structures of GFG. To highlight the central region of the peptides, only residues 6 to 19 in GFG are plotted and shown in black triangles. Ramachandran contour map for Pro residues is shown in the background and typical secondary structures are indicated (α, α-helix; β, β-sheet; PPII, polyproline II and collagen).

4.3.3. Implications for collagen

The natural presence of breaks in the $(Gly\text{-}X\text{-}Y)_n$ triple helix repeats in non-fibrillar collagens is indicated in playing a role in molecular or higher order structure and biological function (Kielty and Grant, 2002; Miles et al., 1995). For example, it has been suggested that breaks could play an important role in the network-like structure of type IV collagen in basement membranes, which is required for its filtration function, and in the large fanned arrays of antiparallel type VII collagen molecules that make up anchoring fibrils (Hudson et al., 2003; Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004). The NMR model structures of a G4G and a G1G break described here provided the characterization of the structural consequences of two most common types of breaks in the triple helix.

The GAAVM break studied here represents a prototype of the G4G breaks, with 4 non-Gly residues in a row, Gly-X-Y-Gly-AA_1-AA_2-AA_3-AA_4-Gly-X-Y. The consensus pattern for G4G breaks includes a small residue at the AA_2 position and a hydrophobic residue at the AA_3 position (Mohs et al., 2006). The GAAVM break contains the consensus sequence, with an Ala residue at the AA_2 position and a Val residue at the AA_3 position. The model structure derived from NMR data suggests that the placement of a hydrophobic residue at the AA_3 position of the break promotes novel hydrophobic packing near the central axis of the superhelix, while the presence of a smaller AA_2 residue is required just before the hydrophobic residue to permit the distortion required for this hydrophobic packing. The presence of a small hydrophobic core formed by the Val residues at position AA_3 suggests that hydrophobic stabilization can partly compensate for the Gly packing normally found at the same position and helps explain

why this and similar breaks can be present in non fibrillar collagens with melting temperatures similar to those of fibrillar collagens.

The GFG break studied here represents a prototype of the G1G breaks, the most common type of break in all non-fibrillar collagens where the Gly-X-Y repeat is interrupted by the absence of one residue. Analysis suggests that there is a high occurrence of hydrophobic residues within these breaks suggesting functional significance and implies G1G breaks will introduce the same structural perturbation wherever they are found (Geetha Thiagarajan, personal communication). The models obtained from NMR data on the GFG peptide, with a hydrophobic residue at the break, indicate a localized change in dihedral angles, with the Gly preceding the Phe and the Phe itself adopting a non-polyproline II conformation and non-standard hydrogen bonding. The fast H-exchange rates for both GF residues suggesting increased flexibility or exposure. The NMR results on the GFG peptide are in general agreement with the GPG high resolution structure solved by x-ray crystallography, in showing a straight triple-helical molecule, but localized changes in H-bonding and dihedral angles, and flexibility in the break region (Bella et al., 2006). It is interesting to note that GPG and GFG structures show no indication of a kink, which has been proposed to occur at a GQG break in mannose binding lectin on the basis of electron microscopy.

The NMR structural information for the G1G break can be compared with that obtained for a G4G break. We propose that G1G and G4G breaks form a subset of similar interruptions in the triple-helix, that can replace each other and be opposite each other within a triple-helix. Both allow continuation of rod-like helices and maintenance of the 1-residue stagger throughout the break, and it appears that G1G or G4G sites in the human

α1(IV) chain can be occupied by each other or by uninterrupted Gly-X-Y sequences. Both will lead to a loss of axial register of the superhelix on both sides of the break, because they place the Gly-X-Y repeat out of axial phase. In both cases, there is a highly localized perturbation in hydrogen bonding and dihedral angles, but in the G4G case studied, the hydrophobic residue (Val) in the AA_3 position packed near the central axis of the superhelix while in the GFG case, the Phe is located on the outside of the helix. The local conformational flexibility in both G1G and G4G breaks could serve as specific recognition sites for binding, and it is likely that disruption of the superhelix register caused by G1G and G4G breaks will affect molecular association properties. These common short breaks may introduce the minimum perturbation to a triple helix, and one might expect increased flexibility or kinks between the triple-helix domains on each side as breaks contain more residues.

The results here all deal with homotrimer G4G and G1G breaks, such as type VII collagen in anchoring fibrils and type X collagen in hypertrophic cartilage (Kielty and Grant, 2002). The G4G and G1G breaks in these homotrimer collagens are predicted to resemble the locally perturbed triple-helix structure reported here for the model peptides. Other non-fibrillar collagens, such as type IV collagen in basement membranes, are heterotrimeric, consisting of two or three distinct chain types. In most cases one type of break in one type IV collagen chain is opposite another kind of break or an uninterrupted sequence in the other chains. It is not clear whether such "mixed breaks" will have larger or different effects on the triple-helix structure, but the relationship between G1G and G4G breaks studied here may shed light on heterotrimer breaks when these types are opposite each other.

Chapter 5

Conformational and dynamic consequences of Gly to X substitutions of collagen disease in triple helical peptides

5. 1. Introduction

A number of hereditary connective tissue diseases have been associated with mutations in the triple helix of fibrillar collgens (Byers and Cole, 2002; Kuivaniemi et al., 1997; Myllyharju and Kivirikko, 2001). The most common mutation is a single base change that leads to the replacement of a Gly by another amino acid, breaking the repeating $(Gly\text{-}X\text{-}Y)_n$ pattern. The first and best characterized collagen disease was osteogenesis imperfecta (OI), a brittle bone disease characterized by fragile bones and easy susceptibility to fracture. Most cases of OI are caused by mutations in type I collagen, the major collagen in bone (Byers and Cole, 2002; Kuivaniemi et al., 1997). 682 of such Gly to X substitutions have been found in the α1(I) and α2(I) chains of type I collagen, consisting of the substitutions of Gly to one of 8 bulkier residues (Ser, Ala, Cys, Val, Arg, Asp, Typ) (Marini et al., 2007). The severity of the disease varies widely from mild cases with multiple fractures to perinatal lethal cases, depending on the type of amino acid substitution and the site of the mutation (Byers and Cole, 2002; Kuivaniemi et al., 1997; Marini et al., 2007). The structural and dynamic consequences of Gly substitutions and its relationship with the disease phenotypes are not understood.

A standard collagen triple helix motif has three extended polyproline II-like chains supercoiled about a common axis. The typical triple-helix is characterized by a $(Gly\text{-}X\text{-}Y)_n$ repeating sequence. The three chains are staggered by 1 residue, and Gly residues must be present as every 3rd residue in each chain so that the three chains can pack very tightly, burying the Gly residues and forming standard Rich-Crick II hydrogen bonds between the amide of Gly and the carbonyl group of the X residue in the adjacent chain (Bella et al., 1994; Ramachandran, 1967; Rich and Crick, 1961). Fibrillar

collagens, such as type I collagen, can self-assemble and cross-link between triple helix molecules and form 67nm periodic fibrils (Kielty and Grant, 2002). Collagen fibrils further interact with other molecules in the extracelluar matrix, such as integrins, matrix metalloproteinases (MMPs), fibronectin, and cartilage oligomeric matrix protein (COMP) (Di Lullo et al., 2002; Kielty and Grant, 2002; Myllyharju and Kivirikko, 2004).

A single base change in a Gly codon can lead to eight residues (Ser, Ala, Cys, Val, Arg, Asp, Trp) or a missense mutation. Residues with branched or charged side chains such as Val, Arg, Asp and Glu are all associated largely with lethal OI phenotypes. Among the eight residues that Gly is substituted to, Ser is the most common residue and accounts for around 40% of all type I collagen substitutions (Marini et al., 2007). In addition, the occurrence of Gly to Ser substitutions among all Gly to X mutations is over 3 times higher than the theoretical expectation calculated from random mutations (Marini et al., 2007). Of the 152 Gly$\rightarrow$ Ser mutations at different sites in the α1(I) chain, 115 are associated with non-lethal phenotypes and 37 are associated with lethal phenotypes (Marini et al., 2007). It has been suggested that the environment of a Gly to Ser replacement must be important in determining its phenotype, or its location at a binding site. In contrast with the high occurrence of Ser substitutions, Gly to Ala substitutions are highly underrepresented, constituting only 22 cases and 6% of all substitutions in the α1(I) chain (Marini et al., 2007).

Synthetic collagen peptides offer an approach to characterizing the effect of mutations on the triple-helix (Baum and Brodsky, 1999; Beck et al., 2000; Bella et al., 1994). The replacement of a single Gly in a host-guest $(Gly\text{-}Pro\text{-}Hyp)_8$ peptide system by the set of residues in the OI substitutions leads to a dramatic loss of stability, with the

degree of destabilization dependent on the identity of the residue replacing Gly: Ala < Ser < Cys < Arg < Val < Glu, Asp < Trp, and there is some correlation between disease severity and destabilization (Beck et al., 2000; Persikov et al., 2004). The sequence surrounding the replacement can also affect the degree of destabilization. Peptides that model two Gly to Ser OI sites both show a significant decrease in triple-helix stability, but the change in T_m is 11 °C in one case and 22 °C for the other case (Yang et al., 1997). A peptide model composed of all Gly-Pro-Hyp triplets can successfully incorporate a Gly to Ala mutation into the triple helix structure. The high-resolution structure of a (Pro-Hyp-Gly)$_{10}$ peptide with a Gly to Ala substitution near the center was solved using x-ray crystallography. It showed a standard triple-helical domain on both sides of the mutation with a highly localized perturbation and untwisting and altered hydrogen bonding mediated by water at the mutation site (Bella et al., 1994).

NMR studies on synthetic collagen peptides offer an approach to characterizing the detailed conformational and dynamic consequences of substitutions on the triple-helix at the molecular level (Baum and Brodsky, 1999; Li et al., 2007). In chapter 3, I described the development of NMR methodology on the selectively ^{15}N labeled peptides to identify the partially disordered equilibrium intermediates (section 3.3) and on the selectively $^{13}C/^{15}N$ doubly labeled collagen peptides to visualize the peptide solution structure by molecular modeling and NMR (section 3.2). In this chapter, I describe the application of these approaches to define the effects of Gly to X substitutions in a physiological sequence context on the equilibrium states, conformation, dynamics, and hydrogen bonding of triple helical peptides.

In section 5.3.1, studies on the equilibrium states of a series of peptide models with different Gly to X substitutions in a physiological collagen sequence contexts are presented. Previous studies reported the design and characterization of a collagen peptide with a $GPO(GAO)_3$ renucleation region on the N-terminal side of the Gly to Ala substitution and $(GPO)_4$ sequence on the C-terminal side of the substitution. It showed the incorporation of the substitution and the formation of a full length trimer (Hyde et al., 2006). Here, the NMR studies are extended on the Gly to Ala model peptide, and explored on other substitutions including Ser, Arg, and Asp (denoted as T1-898(G16A), T1-898(G16S), T1-898(G16R), T1-898(G16D) peptides). It identified multi-conformations in the equilibrium states impacted by the substitution type and a different ability to fold around the substitutions for different mutation residues.

In section 5.3.2, I describe the detailed structural and dynamic studies on the common Gly to X substitution in OI. A Gly to Ser model peptide that models the 901 site of the Gly to Ser mutation in the α1(I) chain resulting in a mild case of OI (Kuivaniemi et al., 1997; Liu et al., 1998) was $^{13}C/^{15}N$ doubly labeled, and NMR studies on this peptide allowed the first visualization of the solution molecular conformation of a Gly to Ser substitution, and clarified the disturbed dynamic and hydrogen bonding at the substitution site. This peptide (denoted as GPO(G16S) peptide) was redesigned on the basis of the T1898(G16S) peptide in section 5.3.1. A $(GPO)_4$ renucleation sequence instead of $GPO(GAO)_3$ was used at the N-terminal end to increase the stability of the peptide, and to ensure that we are only observing the fully folded species. The studies on the GPO(G16S) peptide are compared with the studies on the peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ with a Gly to Ala substitution near the center (denoted as Gly$\rightarrow$Ala peptide), which has an x-ray

crystal structure (Bella et al., 1994). The two peptides are compared to examine the effect of replacing Gly by Ser compared with Ala, and the effect of sequences surrounding the substitution site.

5.2. Materials and methods

5.2.1. Translational diffusion and J coupling measurements on the T1-898 peptide series

T1-898 peptide sets were synthesized with ^{15}N labeled amino acids at selective positions for NMR characterization (Table 5.1). For peptide T1-898, two peptides of identical sequence were synthesized with ^{15}N labels at different positions to optimize NMR assignments. One peptide is selectively ^{15}N-labeled at positions G7, A18, G28, whereas a second peptide is labeled at positions G7, G13, V15, G16, and G28. Peptide T1-898(G16A) was ^{15}N labeled at positions G7, A16, and G28. Peptide T1-898(G16S) was labeled at positions G7, V15, A18, and G28. Both peptide T1-898(G16R) and T1-898(G16D) were labeled at positions G7, G13, and G28. Samples for all peptides were prepared in 10% D_2O/90% H_2O at pH 2.5 with concentrations of 6 mM.

Table 5.1 Peptide sequences of the T1-898 series

Peptide name	Peptide sequence[a]
1. T1-898	Ac-GPO-GAO-**G**AO-GAO-GPV-G[P]**A**-GAR-GPO-GPO-**G**PO-GPO-GY-$CONH_2$
2. T1-898[G16A]	Ac-GPO-GAO-**G**AO-GAO-GPV-**A**[P]A-GAR-GPO-GPO-**G**PO-GPO-GY-$CONH_2$
3. T1-898[G16S]	Ac-GPO-GAO-**G**AO-GAO-GP**V**-S[P]**A**-GAR-GPO-GPO-**G**PO-GPO-GY-$CONH_2$
4. T1-898[G16R]	Ac-GPO-GAO-**G**AO-GAO-**G**PV-R[P]A-GAR-GPO-GPO-**G**PO-GPO-GY-$CONH_2$
5. T1-898[G16D]	Ac-GPO-GAO-**G**AO-GAO-**G**PV-D[P]A-GAR-GPO-GPO-**G**PO-GPO-GY-$CONH_2$

[a] ^{15}N labeled residues are underlined and bolded. Substitutions are boxed.

NMR experiments were performed on a Varian INOVA 500 MHz spectrometer. ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) was performed at 0°C. Residue specific translational diffusion measurements were performed as described in section 3.3. (Li et al., 2005). Convection compensated LED followed by HSQC (CCLED-HSQC) experiments were used to measure the residue-specific diffusion coefficients of the selectively labeled peptides at 0°C and 40°C. The gradient strength was calibrated on a standard doped 1% H_2O in 99% D_2O sample using the value of 1.90×10^{-9} m^2s^{-1} for the

diffusion coefficient of HDO at 25°C (Holz and Weingaertner, 1991; Mills, 1973). 3D HNHA experiments (Vuister and Bax, 1993) were performed to measure homonuclear $^{3}J_{HNHa}$ coupling constants at 15°C (30°C for T1-898 peptide 1), with a H-H coupling period of 25ms. The correction factor for the $^{3}J_{HNHa}$ coupling constants was obtained as described in section 3.2.2.1.4 (Li et al., 2007).

All data were processed using the FELIX 2004 software package (MSI, San Diego, CA). HSQC spectra of two T1-898 peptides were overlaid using Adobe Illustrator. Data analysis of translational diffusion measurements was conducted as described previously (Li et al., 2005). The diffusion coefficients at 40°C were normalized to values at 0°C by using the Stokes-Einstein equation, $D_{normalized}=D_T*273.15*\eta_{40°C}/(313.15*\eta_{10°C})$, where D_T is the directly measured diffusion coefficient at 40°C, η is the viscosity calculated for the solution of 90% H_2O and 10% D_2O at a specific temperature (Cho et al., 1999). Errors in the diffusion measurements were fitting errors from the variance-covariance matrix. The data analysis of HNHA experiments were as described in section 3.2. HNHA measurements for T1-898 peptide 1 and T1-898(G16S) were repeated twice. The errors from repeated experiments were always less than the uncertainty in measuring cross and diagonal peak volumes. Therefore, errors were calculated using the average of the measurement uncertainty from repeated experiments.

5.2.2. Structural and dynamic characterization of the GPO(G16S) peptide

5.2.2.1. Sample preparation

Peptide Ac-$(GPO)_4$GPVSPAGAR$(GPO)_4$GY-$CONH_2$, designated as the GPO(G16S) peptide, was synthesized by Tufts University Core Facility (Boston, MA). The peptide was made with selectively $^{13}C/^{15}N$ doubly labeled residues: residues G13, P14, V15, S16, P17, A18, G19, and G28. Peptide was purified using a Waters XTerra Prep C18 column and then an Amersham Superdex 75 Prep column on an Amersham FPLC system. The identity of the peptides was confirmed by Matrix-Assisted Laser Desorption Ionization (MALDI) mass spectrometry. Peptide $(POG)_4POA(POG)_5$, designated as the Gly→Ala peptide, was synthesized by Tufts University Core Facility (Boston, MA) and purified using a Waters XTerra Prep C18 column on an Amersham FPLC system. The Gly→Ala peptide was made with selectively ^{15}N labeled residues at A15 and G24 positions. Samples for GPO(G16S) and Gly→Ala peptide was prepared in 10% D_2O/90% H_2O at pH 2 with concentrations of 4 mM and 3mM, respectively.

5.2.2.2. NMR assignment, structural and dynamic measurements

NMR experiments for the GPO(G16S) peptide were performed on a Varian Inova 600 MHz spectrometer equipped with a cold probe. For sequential assignment, HSQC and HNCA experiments (Ikura et al., 1990) were obtained at 15°C. 3D ^{15}N edited TOCSY-HSQC (Fesik and Zuiderweg, 1988; Messerle et al., 1989) with a mixing time of 45ms and 3D ^{15}N edited NOESY-HSQC (Fesik and Zuiderweg, 1988; Marion et al., 1989; Messerle et al., 1989) with mixing times of 30-50ms were performed at 10°C, 15°C, and 20°C as described in section 3.2.2. 3D H(CCO)NH experiments were

performed at 20°C (Grzesiek et al., 1993) and paired with NOESY-HSQC to help assign NOE peaks. 3D H(CCO)NH experiments comprised 30(t_1) × 70(t_2) × 512(t_3) complex points, and were recorded with spectral widths of 2000.0(F_1), 7000.0(F_2), and 7022.5(F_3) Hz. 3D HNHA experiments (Vuister and Bax, 1993) were performed to measure $^3J_{HNHa}$ coupling constants at 15°C, with a H-H coupling period of 25ms. The correction factor for the $^3J_{HNHa}$ coupling constants was obtained as described in section 3.2.2.1.4 (Li et al., 2007). Hydrogen exchange experiments were carried out at 10 °C, $pD_{correct}$ 2.4, as described in section 4.2.3 (Mohs et al., 2006). For measurements of amide proton temperature gradients, ^1H-^{15}N HSQC spectra were obtained at 0-25°C with an interval of 5°C in the procedure as described in section 4.2.3.

NMR experiments for the Gly→Ala peptide were performed on a Varian INOVA 500 MHz spectrometer. The HSQC experiment was performed at 15°C. The $^3J_{HNHa}$ coupling constants were measured and corrected, and the amide proton temperature gradients were measured similarly as for the GPO(G16S) peptide.

All data were processed using the FELIX 2004 software package (MSI, San Diego, CA), and/or NMRPipe (Delaglio et al., 1995) and analyzed with FELIX 2004 or NMRView (Johnson and Blevins, 1994). For HC(CO)NH experiments, a solvent suppression filter was applied in the acquisition dimension to the data prior to apodization with a 90° sine-bell window function. The data were subsequently zero filled to 1024 complex points and Fourier transformed. The t_1 and t_2 dimensions were increased 1.5 times by forward-backward linear prediction (Zhu and Bax, 1992), multiplied by a sine-bell window function, zero filled to 256 complex points, and Fourier transformed. The final 3D data for the HC(CO)NH experiment included 256 × 256 × 512 real points.

For the NOE contact map, the predicted background map was generated as described in section 3.2.3 from a classic triple helical conformation of GAAGVM peptide. The NOE contact map for peptide GPO(G16S) was made from observed NH-H NOEs in the 3D ^{1}H-^{15}N NOESY-HSQC experiment and classified as NH-NH, NH-H^{α}, and NH-side chain (H^{β}, H^{γ}, H^{δ}).

5.2.2.3. Molecular modeling

A computer model structure of GPO(G16S) was generated based on the x-ray crystal structure of Gly→Ala (PDB ID: 1CAG) (Bella et al., 1994) using the Molecular Operating Environment 2006.08 (Chemical Computing Group Inc., Montreal, Canada). Residues GPO-APO-GPO were substituted to GPV-SPA-GAR. A different starting structure of T3-785 (PDB ID: 1BKV) (Kramer et al., 1999) was also tried and residues GIT-GAR-GLA were substituted to GPV-SPA-GAR. The resulting model was energy minimized with dihedral angle ϕ and NOE distance restraints as described in section 3.2.3. Back calculation of $^{3}J_{HNHa}$ values and back calculation of an NOE map were used to eliminate models that were not consistent with the experimental data. A set of three representative structures that was consistent with experimental $^{3}J_{HNHa}$ values and all ^{1}H-^{1}H NOEs was selected for GPO(G16S).

5.3. Results and discussion

5.3.1. Effects of residue identity of Gly to X substitution on the equilibrium states of T1-898 peptide series

5.3.1.1. Peptide design

Previous studies reported the design and characterization of a collagen peptide with a $GPO(GAO)_3$ renucleation region on the N-terminal side of the Gly to Ala substitution and $(GPO)_4$ sequence on the C-terminal side of the substitution. It showed the incorporation of the substitution and the formation of a full length trimer (Hyde et al., 2006). Here NMR studies are extended on the Gly to Ala model peptide, and explored on other substitutions including Ser, Arg, and Asp (denoted as T1-898(G16A), T1-898(G16S), T1-898(G16R), T1-898(G16D) peptides). The sequences of this series of peptides are shown in table 5.1. The middle region GPV-XPA-GAR was taken from the natural collagen sequence from position 898-906 in the α1(1) chain, and the X residue near the center correlates with a Ser mutation at 901 position resulting in a mild case of OI (Kuivaniemi et al., 1997; Liu et al., 1998). It allowed the study of the effect of different residue substitutions in a physiological collagen sequence environment.

5.3.1.2. Translational diffusion measurements

NMR studies were performed on the T1-898 peptide set to identify the changes in conformation that result from the different substituting residues in Gly$\rightarrow$X mutations (where X=Ala, Ser, Asp, Val, Arg). The HSQC spectrum of the T1-898 peptide contains the typical features for a triple helical conformation (Figure 5.1A) as found in other triple helical peptides (Buevich and Baum, 2001; Li et al., 2007). The ^{15}N labeled residues of

the T1-898 peptide each show distinct monomer and trimer resonances, supporting a monomer:trimer equilibrium in solution. Residues G28 and G7 show only a single trimer resonance due to their repetitive GPO and GAO environment, respectively. Residues G13, A15, G16, and A18 each have three trimer resonances, indicating that the three chains are in non-equivalent chemical environment (Buevich and Baum, 2001; Li et al., 2007). For the mutant peptides, the HSQC spectra of T1-898(G16A) (Hyde et al., 2006) and T1-898(G16S) (Figure 5.1C) show that the Gly→Ala and Gly→Ser substitutions can be incorporated in the two mutant peptides, reflected in the appearance of three distinct trimer resonances for each residue around the substitution site (A16 in peptide T1-898(G16A) and V15/A18 in peptide T1-898(G16S)). However, for peptide T1-898(G16R) and T1-898(G16D) (Figure 5.1D and E), trimers for G13 that are 3 residues N-terminal to the substitution site are not observed, whereas trimer for G7 and G28 are present and have the same chemical shifts as in other T1-898 peptide series. It raises the question of whether the G7 trimer and G28 trimer are in the same molecule, i.e. whether the equilibrium states contain a molecule with two ends folded and the middle disordered, or molecules with only one end folded and the other end disordered.

Residue specific translational diffusion measurements (Li et al., 2005) were performed on the T1-898 peptide set to investigate the effect of different substitutions on the equilibrium states. Diffusion coefficients of monomers for ^{15}N labeled residues at 0°C and 40°C are shown in figure 5.1F-J. At 40 °C, peptides exist as unfolded monomers except in T1-898 very weak trimer resonances are still observed, and the diffusion coefficients of monomers (denoted as $D_{M40°C}$) for all labeled positions within a given peptide are similar to one another (Figure 5.1F-J blue bars). The uniformity for residues

at different positions is consistent with sampling a similar averaged denatured shape. At 0°C, the monomer diffusion coefficients (denoted as $D_{M0°C}$) in T1-898 are uniform for different residues and are similar to $D_{M40°C}$, supporting a simple monomer:trimer equilibrium model (Figure 5.2A). In T1-898(G16A) and T1-898(G16S), the $D_{M0°C}$ for different residues is not uniform. $D_{M0°C}$ and $D_{M40°C}$ of G28 are similar, indicating that at low temperature the monomer resonances of G28 reflect the same denatured state as at high temperature. In both peptides, $D_{M0°C}$ for the N-terminal G7 and for residues in the middle of the molecule is lower than $D_{M40°C}$, indicating that those monomer diffusion rates are modulated by partially disordered intermediates that have higher molecular weight and thus smaller diffusion coefficients than the monomer (Li et al., 2005). The residue specific monomer diffusion showed that this ensemble of intermediates is folded at the C-terminal GPO rich region and disordered at the N terminal and the middle substitution site (Figure 5.2B). In T1-898(G16R) and T1-898(G16D), $D_{M0°C}$ of all residues are very low and similar to the trimer diffusion coefficients (data not shown). It is consistent with sampling of a low population of pure monomer and a very high population of partially disordered trimers (Figure 5.2C). The presence of this partially disordered species indicates the R and D substitutions disrupt the trimer formation, and the peptides can not fold into full length trimer.

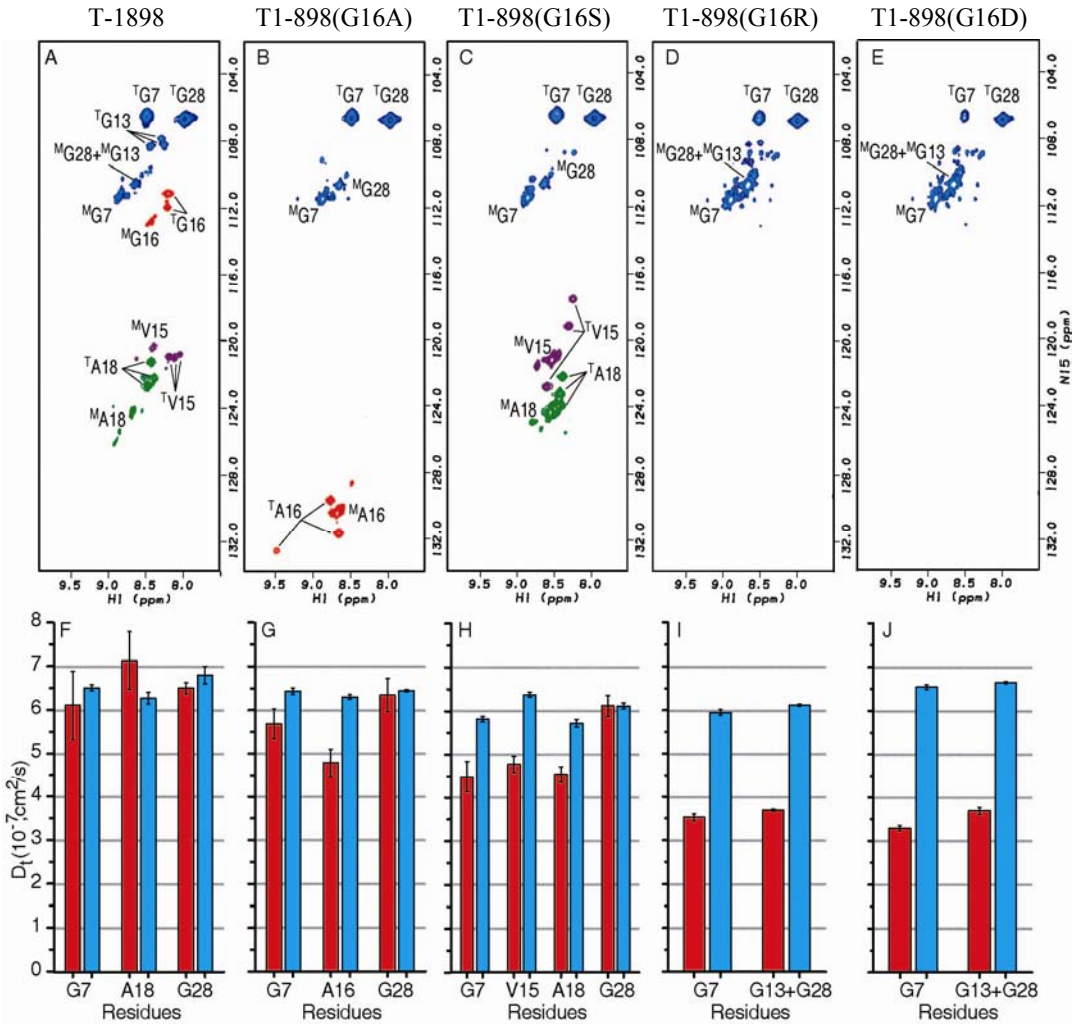

Figure 5.1 HSQC and monomer translational diffusion measurements on the T1-898 peptide series.

Upper panel: ^{1}H-^{15}N HSQC spectra of T1-898 (A), T1-898(G16A) (B), T1-898(G16S) (C), T1-898(G16R) (D) and T1-898(G16D) (E) peptide at 0°C. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T, respectively. For T1-898, spectra of two identical peptides with different labeling positions are overlaid.

Lower panel: Histogram of the residue-specific monomer diffusion coefficients of ^{15}N-labeled residues for peptide T1-898 (F), T1-898(G16A) (G), T1-898(G16S) (H), T1-898(G16R) (I) and T1-898(G16D) (J). Monomer diffusion at 0°C and 40°C are shown in red and blue bars, respectively.

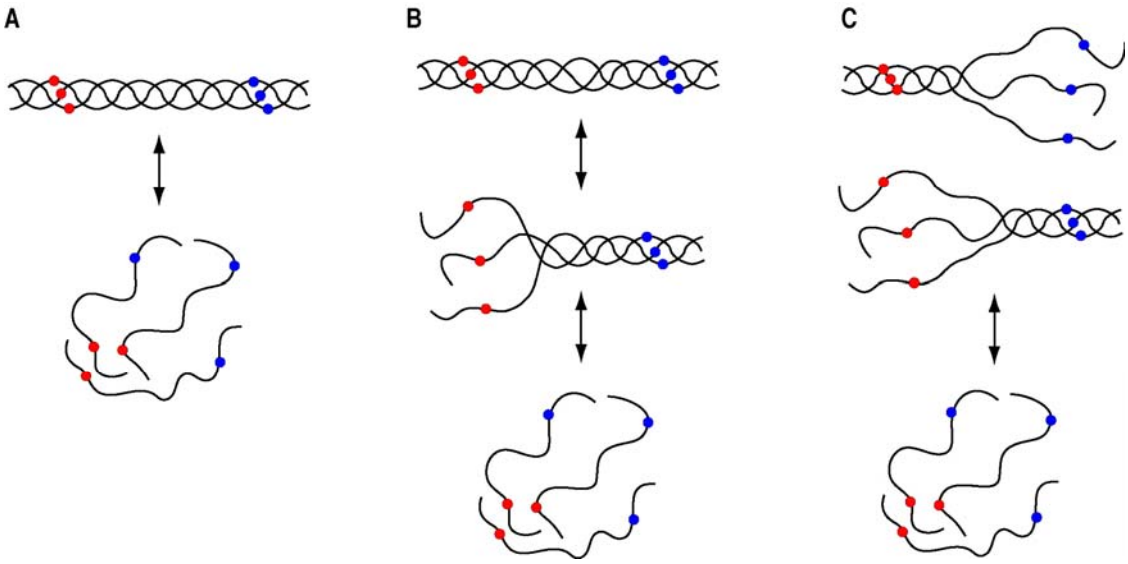

Figure 5.2 Schematic representation of the equilibrium states of T1-898 (A), T1-898(G16A) and T1-898(G16S) (B), T1-898(G16R) and T1-898(G16D) (C). The blue and red circle represents the position of G28 and G7, respectively.

5.3.1.3. J coupling measurements on T1-898(G16A) and T1-898(G16S) peptides

$^{3}J_{HNHA}$ coupling constants were measured to study the trimer conformation in T1-898, T1-898(G16A) and T1-898(G16S) peptides (Figure 5.3). The $^{3}J_{HNHA}$ coupling constants are uniform for different residues in the control T1-898 peptide (Figure 5.3A), indicating a standard triple helical conformation. For T1-898(G16A), one of the A16 J couplings is significantly larger than the G7 and G28 coupling constants suggesting that the PPII conformation is not maintained at the interruption site. In T1-898(G16S), one of the V15 J couplings is slightly larger (around 5.9 Hz) than the V15 in the control T1-898 (around 3.1~4.1 Hz), indicating an altered conformation around the interruption site.

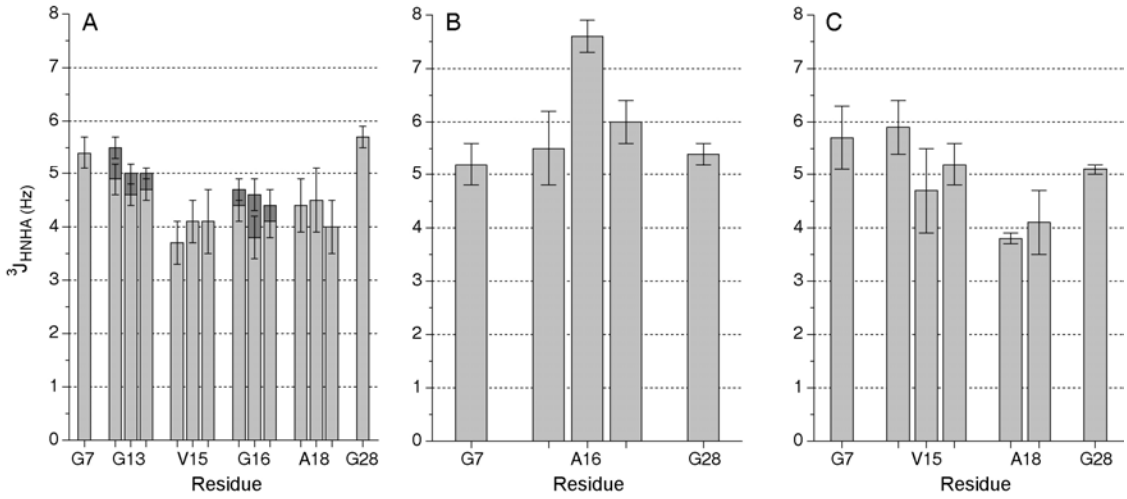

Figure 5.3 $^3J_{HNHA}$ coupling constants of labeled residues in peptide T1-898 (A), T1-898(G16A) (B), and T1-898(G16S) (C). $^3J_{HNHa}$ coupling constants from the two H^{α} Gly residues are shown in dark gray and light gray bars.

5.3.1.4. Discussion

The diffusion studies on the equilibrium states of the T1-898 peptide series indicate a mutation specific ability to fold around the substitution site of different mutant peptides. The T1-898(G16R) and T1-898(G16D) peptides could not incorporate the substitutions into the full length trimers. The incomplete folding in T1-898(G16R) and T1-898(G16D) peptides is consistent with the CD analysis, which showed a decrease of mean residue ellipticity at 225nm (MRE_{225}) from 4337 deg·cm^2·dmol^{-1} in the T1-898 peptide to 2922 deg·cm^2·dmol^{-1} for the Arg substitution and to 1693 deg·cm^2·dmol^{-1} for the Asp substitution (Mike Bryan, personal communication). In addition, very low thermal stabilities were indicated for the T1-898(G16R) and T1-898(G16D) peptides (Mike Bryan, personal communication). This data suggested that Arg and Asp have the largest disruptive effects on the triple helix stability and trimer formation, which

correlates with their high occurrence in the lethal cases of OI in collagen (Kuivaniemi et al., 1997; Marini et al., 2007).

The diffusion studies on the T1-898(G16A) and T1-898(G16S) peptide showed a fully folded trimer as well as partially disordered intermediates. Equilibrium states can be considered as the end point of the triple helix folding process and reflect the folding species in some degree (Buevich and Baum, 2002; Li et al., 2005). Consistent with the presence of partially disordered intermediate, differential scanning calorimetry (DSC) studies probed a population of intermediates during folding for the T1-898(G16A) and T1-898(G16S) peptide (Mike Bryan, personal communication). The $^{3}J_{HNHA}$ coupling measurements on the T1-898(G16A) and T1-898(G16S) peptide showed an alteration of dihedral angle ϕ. The multiple trimer-like conformations in the equilibrium states complicates the interpretation of $^{3}J_{HNHA}$ coupling constants because the $^{3}J_{HNHA}$ coupling reflects a weighed average for the full-length trimer and partially disordered intermediates. However, the population of intermediates relative to trimer is very low because of two reasons: 1) bi-exponential fitting of the diffusion data suggested that the population of intermediates is similar to that of the monomer, 2) CD melting indicated a much lower population of monomer relative to trimer because the low temperature equilibrium state correlates with the flat part in the folding curve (Mike Bryan, personal communication). Therefore the full-length trimer dominates the $^{3}J_{HNHA}$ coupling measurement, and $^{3}J_{HNHA}$ coupling can be taken as a qualitative measurement for the full-length trimer. The non-uniform $^{3}J_{HNHA}$ constants in T1-898(G16A) and T1-898(G16S) indicate that although the peptide is folded around the Gly to Ala or the Gly to Ser

mutation, a standard triple helix could not be formed and that there is conformational distortion around the substitution site.

The data on the T1-898 peptide series provide clues on the nature of renucleation in collagen. In mutant collagen, full length trimer is formed in the presence of all Gly to X substitutions. During the C-terminus to N-terminus directional folding of the mutant collagen (Myllyharju and Kivirikko, 2001), renucleation must occur N-terminal to the substitution site. The equilibrium data on the T1-898(G16R) and T1-898(G16D) peptides showed partially disordered trimers that are folded either at the N-terminus or the C-terminus. It suggests that in mutant collagen folding could occur from both directions towards the substitution site (Buevich et al, 2004; Hyde et al., 2006). Furthermore, the $GPO(GAO)_3$ region taken from collagen could not force T1-898(G16R) and T1-898(G16D) to fold around the Gly to Arg and Gly to Asp mutation sites, whereas in collagen these two kinds of mutations are incorporated in the full-length trimer. It indicates that the collagen length may play a role in enhancing the renucleation ability.

5.3.2. Structural and dynamic consequences of a Gly to Ser substitution in the GPO(G16S) peptide

5.3.2.1. Peptide design

In this study, the peptide Ac-$(GPO)_4$-GPV-SPA-GAR-$(GPO)_4$GY-$CONH_2$ (designated as GPO(G16S)) was designed on the basis of peptide T1-898(G16S) and used as a prototype of Gly to X substitution to define the detailed structural and dynamic consequences of Gly mutations on the triple helix. In the previous section, translational diffusion studies showed the presence of partially disordered intermediates in the equilibrium states of T1-898(G16S) (Figure 5.2) and thus T1-898(G16S) was not amenable for structural characterization. Therefore, the renucleation sequence at the N-terminus was redesigned from $GPO(GAO)_3$ to $(GPO)_4$ to increase the stability of the peptide, and to ensure that all trimeric NMR signals arise from the fully folded species Translational diffusion coefficients were measured for GPO(G16S) and proved that there are no partially disordered intermediates in solution (Figure 5.4). This peptide Ac-$(GPO)_4$-GPV-SPA-GAR-$(GPO)_4$GY-$CONH_2$ includes the collagen sequence (GPV-SPA-GAR) from position 898-906 in the α1(1) chain and a Gly to Ser substitution at the 901 position associated with a mild case of OI. A Tyr is added in the C terminus of the peptide to allow the determination of concentration by UV absorbance. Circular Dichroism (CD) and differential scanning calorimetry (DSC) studies on this peptide showed a high thermal stability with T_m of 28°C and a high enthalpy of 12 kJ/mol that is similar as native peptides without mutation ($(POG)_{10}$: 13kJ/mol, T1-898: 12kJ/mol) (Mike Bryan, personal communication). Two peptides with different labeled positions were synthesized. One peptide was synthesized with selectively ^{15}N labeled at G7, V15,

A18, and G22 positions for the translational diffusion study to confirm the sequence design, and the other peptide was consecutive $^{13}C/^{15}N$ doubly labeled residues at the G13, P14, V15, S16, P17, A18, G19, and G28 positions, to allow the procedure of identifying chain assignment, measuring NMR conformational parameters, and obtaining model structures (Li et al., 2007).

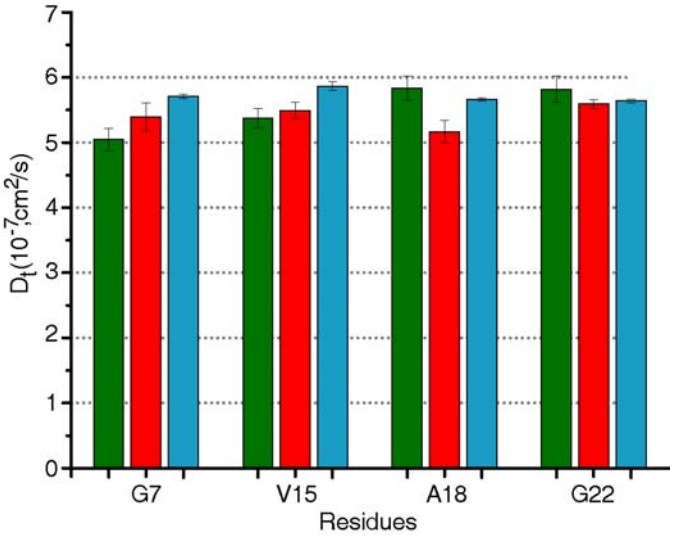

Figure 5.4 Translational diffusion measurements of the GPO(G16S) peptide. The GPO(G16S) peptide was selectively ^{15}N labeled at G7, V15, A18, and G22 positions. The translational diffusion coefficients of monomer resonances for ^{15}N labeled residues were measured at 10°C (green), 20°C (red), and 40°C (blue), and shown in normalized values at 0°C. The diffusion experiments were performed at a Varian INOVA 600 MHz spectrometer. In order to make the values comparable with those in figure 5.1 measured at 500 MHz spectrometer, the diffusion coefficients were multiplied by 1.21 that is the normalization for different gradient strengths at two spectrometers.

NMR studies were carried out on the Gly→Ala peptide, a mutant peptide for which an x-ray structure has been obtained (Bella et al., 1994), to complement the mutation studies in crystals and to compare with the GPO(G16S) peptide. The A15 and G24 residues were ^{15}N labeled in the peptide $(POG)_4POA(POG)_5$. The model studies that were used for the GPO(G16S) peptide could not be applied to the Gly→Ala peptide, because the high content of Pro and Hyp could not provide enough conformational information for modeling due to the lack of amide protons in those residues. Nevertheless, the NMR conformational parameters at the A15 substitution site of the Gly→Ala peptide can be compared with those of the GPO(G16S) peptide, and complement the x-ray crystal structure.

5.3.2.2. NMR Chain assignments and chemical shifts

The HSQC spectrum of the GPO(G16S) peptide shows one or more trimer peaks as well as monomer peaks for each labeled residue that has amide protons (Figure 5.5). G28, within the Gly-Pro-Hyp repeating region at the C-terminus, shows only a single trimer resonance at the typical position for the C-terminal Gly in other triple helical peptides (Li et al., 2007; Mohs et al., 2005). For other residues G13, V15, S16, A18, and G19, three trimer peaks are observed. It indicates that a well-defined structure with 3-non-equivalent chains is present in the substitution region. Most notably, one of the S16 trimers have shifted downfield significantly in the ^{1}H dimension (9.1ppm). The large downfield shift may reflect an altered conformation or hydrogen bonding behavior of S16 compared to standard triple helical peptides.

The three chains of the triple helix are staggered by one residue (Figure 5.5) and need to be identified in order to perform the detailed conformational characterization. The chain assignment of trimer resonances can be divided into two steps: sequential assignment by triple resonance experiment and chain stagger identification by NOE distances (Li et al., 2007). The sensitivity of the HCAN experiment to correlate Pro with its next residue (Powers et al., 1991) was too low for trimers because of the broad linewidth of the trimer peaks, therefore the trial to assign through Pro was not successful. Although the labeling of Pro did not help the sequential assignment by triple resonance experiment, it provided additional NOEs in the NOE contact map to assist the identification of chain stagger. A single HNCA experiment was able to provide the sequential correlation for V15-S16 and A18-G19 segments. The assignment of central labeled residues was interrupted by P14 and P17 and was divided into three segments: G13, V15-S16, and A18-G19. By assuming a 1-residue stagger, the connection between these two segments and to G13 was obtained by NOE experiments. For example, NOEs observed from ^{3}G13 NH to ^{1}S16 NH, and from ^{2}G13 NH to ^{1}G13NH and ^{3}G13 NH indicated the connection between the G13 and V15-S16 segment and allowed the identification of the leading (chain 1), trailing chains (chain 3) and middle chains (chain 2) (See the next section for details). All trimer resonances in the GPO(G16S) peptide could be assigned to specific chains of the triple helix as indicated in the HSQC spectrum by the superscripted number (Figure 5.5).

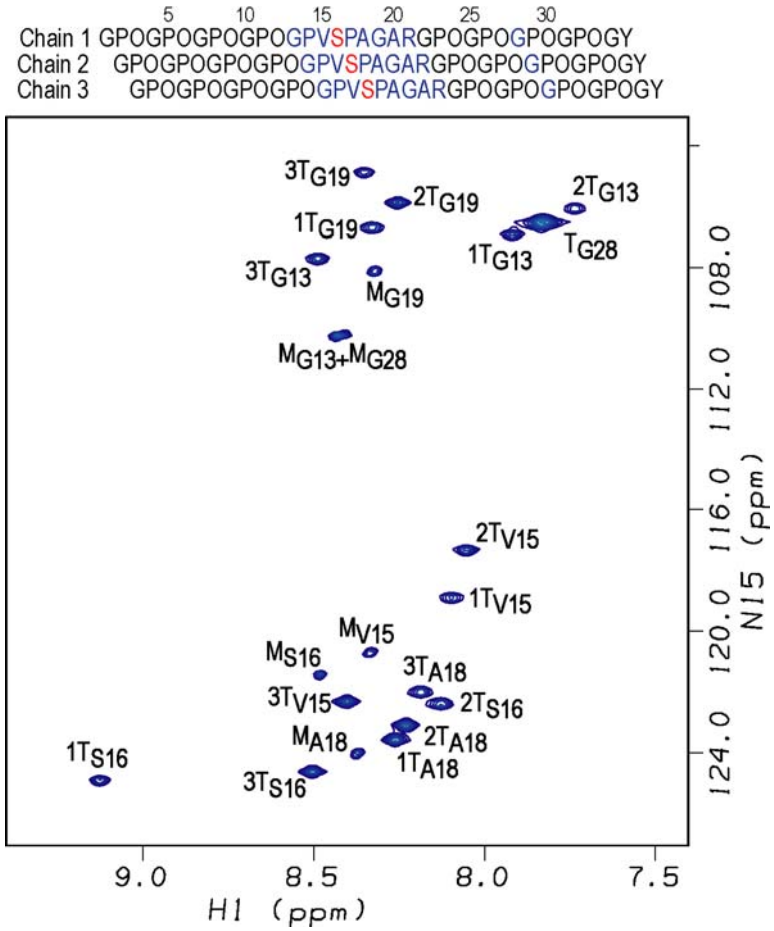

Figure 5.5 The HSQC spectrum of GPO(G16S) peptide at 15°C. Sequence diagram of the peptide above the HSQC shows the characteristic one residue stagger. The $^{13}C/^{15}N$ doubly labeled residues are colored in blue and the Gly to Ser substation site is shown in red. The peaks corresponding to the monomer and trimer state are denoted with a superscript M or T, respectively. Leading, middle, or trailing chain stagger assignment is indicated as chain 1, 2, or 3 as a number in front of the superscript T, respectively.

The HSQC of the Gly→Ala peptide shows trimer and monomer peaks for A15 and G24. Three well resolved trimer peaks for A15 are downfield shifted in the proton dimension relative to the A15 monomer, with one chain shifted to 9.4ppm (Figure 5.6). Due to the lack of stagger information from the single A15 labeling, chain assignment could not be obtained.

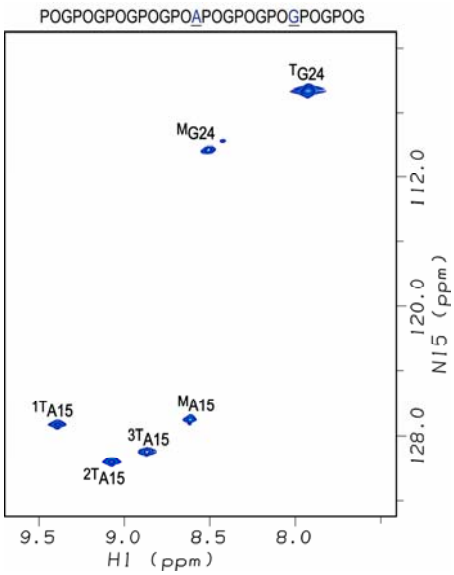

Figure 5.6 The HSQC spectrum of the Gly→Ala peptide at 15°C. Sequence is shown above the spectrum with ^{15}N labeled residues underlined and Gly to Ala substitution site bolded. Superscript chain number indicates tentative chain stagger number from the comparison of NMR data with the x-ray structure.

5.3.2.3. NMR conformational characterization

NMR conformational parameters were measured on the GPO(G16S) peptide, including NOEs to obtain distance information and $^{3}J_{HNHa}$ constants to obtain the dihedral angle ϕ, and to prepare for the molecular modeling procedures described in section 3.2.3. As shown in section 3.2.3, examination of an NOE contact map summarizing the NH-H

distance information provides a way to identify leading, middle, and trailing chains, and to detect any deviation from a standard triple helix conformation. An NOE contact map diagram (Figure 5.7) was constructed for the experimental NOESY-HSQC data for GPO(G16S). In order to compare the conformation of the substitution peptide with a standard triple helical conformation, the experimental NOEs (represented as circles) were overlaid on shaded boxes that represent predicted NOEs from a standard triple helical structure of GAAGVM peptide (Li et al., 2007). The predicted intrachain and interchain contacts for a standard triple helical structure are found to be sequence-independent (data not shown). The superimposition of the experimental NOEs and the background indicates the 1- residue stagger of the triple-helix throughout the GPO(G16S) substitution site. The diagnostic interchain NH-NH, NH-H^{α} NOEs between the three chains of the G13 residues (^{1}G13NH to ^{2}G13NH and $^{2}G13H^{\alpha}$, ^{2}G13NH to ^{1}G13NH and $^{1}G13H^{\alpha}$, ^{2}G13NH to ^{3}G13NH and $^{3}G13H^{\alpha}$, ^{3}G13NH to ^{2}G13NH) and between G19 residues (^{3}G19NH to $^{2}G19H^{\alpha}$) are seen as well as interchain NOEs from ^{2}G19NH to $^{3}P17H^{\beta/\gamma}$ supporting the one residue stagger. In addition, the NOEs normally unique between Gly are observed between the S16 residues and residues G13 and G19 (^{1}S16NH to ^{3}G13NH, ^{3}S16NH to ^{1}G19NH), suggesting that the S16 residues are packed into the center of the triple helix at the substitution site, in a behavior similarly to a Gly at the same position. One new NOE that is not consistent with the predicted background is observed between the V16 residues (^{1}S16NH and ^{3}S16NH) indicating a different packing of S16 residues with Gly.

Figure 5.7 Comparison of experimental NOEs of GPO(G16S) peptide with a standard triple helical conformation. Predicted contact map obtained from a standard triple helix model structure are shown in shaded boxes as background (Li et al., 2007). Contacts are shaded in gray for intrachain distances less than 5Å and in yellow for interchain distances less than 5 Å. Experimental NOEs for GPO(G16S) peptide are represented by circles (HN-HN(**O**), HN-H^{α}(**O**), and HN-side chain protons(**O**)) and are overlaid on the shaded contacts, showing the expected intrachain and interchain NOEs and one new contact (^{1}S16 NH to ^{3}S16 NH) consistent with the one residue stagger between chains and the packing of the Ser residues at the substitution site. Half circles( ) represent overlapped NOEs.

$^{3}J_{HNHa}$ coupling constants which can be related to the dihedral angle ϕ were obtained for the GPO(G16S) peptide from the HNHA experiment (Figure 5.8). Most of $^{3}J_{HNHa}$ coupling constants are similar in the range of 4-6 Hz and the $^{3}J_{HNHa}$ of ^{2}S16 is slightly higher than other residues. Because multiple solutions of the angle ϕ are possible from the parameterized Karplus equation (Vuister and Bax, 1993), the precise one can not be determined from the $^{3}J_{HNHa}$ measurement alone and multiple ϕ solutions are taken as restraints in the molecular modeling.

$^{3}J_{HNHa}$ coupling measurements were also carried out on the Gly→Ala peptide (Table 5.2). One of the A15 trimers shows a very high J coupling constant (8.6Hz), indicating a disruption of triple helical PPII conformation at the substitution site. The multiple ϕ angle solutions were calculated for A15 and compared with the ϕ angles in the x-ray crystal structure (Table 5.2). The chain number from NMR can not be correlated with the chain number from x-ray crystal structure because of lack of experimental data. The ϕ angle of chain 3 in experimental data is different from in the x-ray structure, indicating a deviation of conformation in solution with that in the crystal.

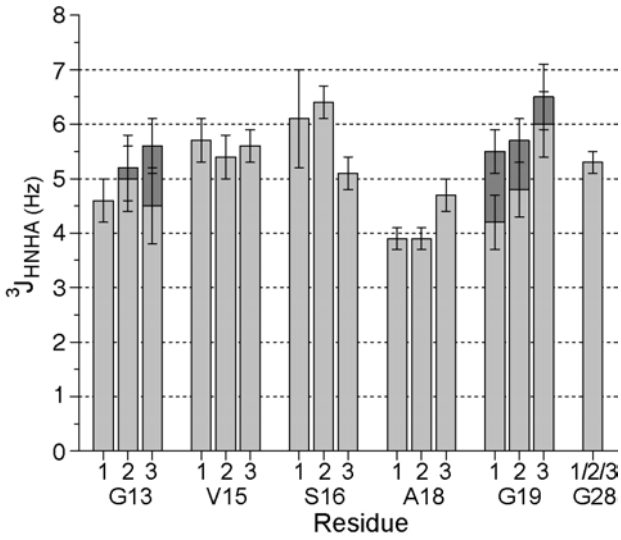

Figure 5.8 The histogram of the $^3J_{HNHa}$ coupling constants are measured for the GPO(G16S) peptide. $^3J_{HNHa}$ coupling constants from the two H^{α} Gly residues are shown in dark gray and light gray bars.

Table 5.2. $^3J_{HNH\alpha}$ of the Gly→Ala peptide and the corresponding multiple dihedral angle solutions from the Karplus equation and from the x-ray crystal structure, and the NH temperature gradients of Gly→Ala.

Residue	$^3J_{HNHa}$(Hz)	Φ(1)	Φ(2)	Φ(3)	Φ(4)	Φ in x-ray structure	NH temperature gradients (ppb/°C)
^{1}A15[a]	5.6 ± 0.8	76 ± 16	-168 ± 6	44 ± 16	-72 ± 6	-81	-12.2
^{2}A15	8.6 ± 0.5		-144 ± 5		-96 ± 5	-104	-10.7
^{3}A15	5.6 ± 0.6	78 ± 11	-168 ± 5	42 ± 11	-72 ± 5	-61	-4.3
G24	4.7 ± 0.3	93 ± 3	-175 ± 2	27 ± 3	-65 ± 2	-72[b]	-3.5

a. Tentative chain numbers are shown in superscript. Chain 1, 2, 3 may not indicate leading, middle and trailing chains.
b. Average Φ angle of G24 in three chains in x-ray structure was used.

5.3.2.4. Hydrogen exchange and amide temperature gradients

Two kinds of experiments, hydrogen exchange and amide temperature gradient measurements, were performed to characterize the hydrogen bonding features and dynamic consequences of the Gly to Ser substitution in the GPO(G16S) peptide. Hydrogen exchange experiments were conducted to study the protection of amide protons from solvent (Figure 5.9A). As expected, the control G28 at the stable C-terminal GPO rich region has a very high degree of protection from exchange (protection factors of ~457), and the two residues in the Y position V15, A18, showed almost no protection. The residues in the Gly/Ser positions in the substitution region were all less protected than Gly 28, but showed interesting variations. The G13 residues are located in the tripeptide N-terminal to the substitution site, and showed a striking difference between the 3 chains, ranging from ^{1}G13 (~330), which is almost as fully protected as Gly28, followed by ^{2}G13, with an intermediate protection (~110), and then ^{3}G13 which shows little protection (~25). Because of the staggering of the 3 chains, chain 1 is surrounded by an almost normal triple-helix environment, while the Gly13 in chain 3 is falling heavily under the influence of the Ser substitution site. The S16 residues show low protection factors (average ~37), with chain 3 again the least protected (~10), indicating all three Ser NH groups are either forming weak hydrogen bonds or are exposed to solvent either permanently or due to local breathing/unfolding. The three G19 residues in the triplet C-terminal to the Ser substitution site show the lowest protection (average ~15) for this Gly position, less the seen at the site itself or in the triplet N-terminal to the substitution site, suggesting an asymmetry in the conformational effects. It is interesting that G19 in chain 3 appears a little more protected, since that chain is the furthest from the Ser perturbation.

Complementary amide proton temperature gradients are performed on the GPO(G16S) peptide in order to resolve the existence of hydrogen bonds in the peptide (Figure 5.9B). As expected the NH temperature gradient for G28 is more positive than -4.5 ppb/°C indicating the existence of hydrogen bonds at the C terminal GPO rich end(Baxter and Williamson, 1997). Similarly, all three chains of G13 have more positive temperature gradients than -4.5 ppb/°C indicating the formation of hydrogen bonds at the G13. Three S16 residues show non-equivalence with ^{1}S16 and ^{2}S16 close to the -4.5 ppb/°C cutoff line and ^{3}S16 more negative than -4.5 ppb/°C, indicating a disturbed hydrogen bonding pattern at the Gly to Ser substitution. The three G19 residues are non-equivalent, with indication of a hydrogen bond for ^{3}G19, the residue again further from the substitution site due to the stagger. In addition, similarly as from protection factors, the G19 residues in the triplet c-terminal to the mutation site show different characteristics from the G13 residues which are one triplet N-terminal to the mutation site. The more negative values of NH temperature gradients for the three V15 and A18 suggest that they are not involved in hydrogen bonding, as expected for residues in the Y positions.

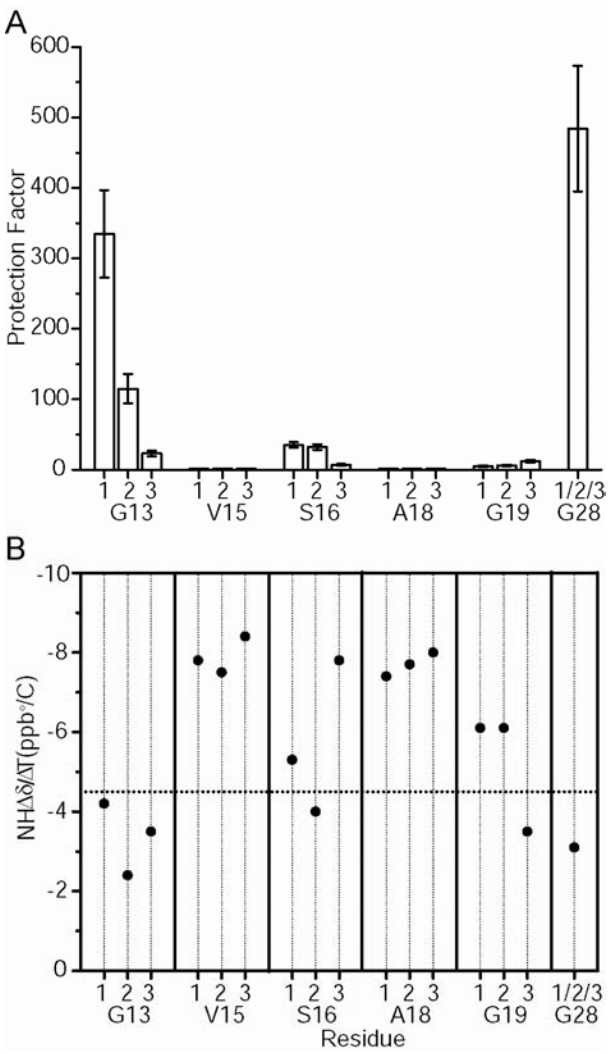

Figure 5.9 Hydrogen exchange and NH temperature gradient studies on the GPO(G16S) peptide.

A, Histogram of hydrogen/deuterium protection factors for the labeled residues in the GPO(G16S) peptide.

B, Amide NH $\Delta\delta/\Delta T$ plot for GPO(G16S) peptide. The dashed horizontal line corresponds to $\Delta\delta/\Delta T$ = -4.5 ppb/°C which provided the cutoff line of hydrogen bonding.

Amide proton temperature gradients were also measured for the Gly→Ala peptide (Table 5.2). For the three Ala residues, one chain has a temperature gradient more positive than -4.5ppb/°C indicating the formation of a hydrogen bond, whereas the other two A15 have extremely negative temperature gradients suggesting that they are not involved in direct interchain hydrogen bonding.

5.3.2.5. Molecular modeling

In order to define the conformation in solution for the GPO(G16S) peptide, molecular modeling was performed with the incorporation of NMR data using the strategy described in section 3.2.3. Models were generated based on x-ray structures with a Gly→Ala substitution in the $(POG)_{10}$ repeating sequence (PDB ID: 1CAG)(Bella et al., 1994) or a standard triple helix structure (T3-785 peptide)(PDB ID: 1BKV)(Kramer et al., 1999). The central residues GPO-APO-GPO were substituted to GPV-SPA-GAR in the Gly→Ala peptide, and residues GIT-GAR-GLA were substituted to GPV-SPA-GAR in the standard triple helix peptide. The structure was energy minimized by incorporating ϕ angle restraints from $^{3}J_{HNHa}$ measurement and distance restraints from NOESY experiments. Back calculation of NMR parameters from the resulting structures were performed to eliminate structures that are not consistent with the experimental data. A set of three representative structures were obtained and gave good agreement with NMR data, two from the Gly→Ala starting structure and one from the T3-785 starting structure. The two starting x-ray structures produced similar models for the GPO(G16S) peptide. The final three models show similar features, and one model from the Gly→Ala starting structure is shown for illustration.

The model presents a rod like structure without a kink or central loop-out. Local deformation is observed around the substitution site and standard triple helical structures are maintained at both sides of the substitution (Figure 5.10A). The 1-residue stagger between the 3 chains is preserved along the whole molecule. The S16 residues are closely packed near the central axis despite the larger size of Ser relative to Gly. The side chain of ^{1}S16 and ^{2}S16 can be packed toward the center of the molecule (Figure 5.10B). Although the standard triple helix can not be formed because of the Gly substitution, the Ser residues can be accommodated inside the triple helix through minor conformational deformation. Examination of the backbone dihedral angles of the model shows that most of the residues have dihedral angles around the polyproline II (PP II) region. It indicates that the conformational distortion is highly localized and absorbed by a small number of residues.

In an alternative model, the residue ^{2}S16 adopts a ϕ angle of -160° that is in the β-strand region in the Ramachandran plot, whereas all other residues fall in the PP II region. A ϕ angle of -160° here and the ϕ of -80° for ^{2}S16 in the illustrated model are two solutions of $^{3}J_{HNHa}$ from the Karplus equation. The NMR experimental data could not distinguish these two models. The backbone trace of the N-terminal side of S16 in the alternative model deviates from the illustrated model, yet both models show many similar features of a straight rod like structure, close packing of S16, and very localized distortion.

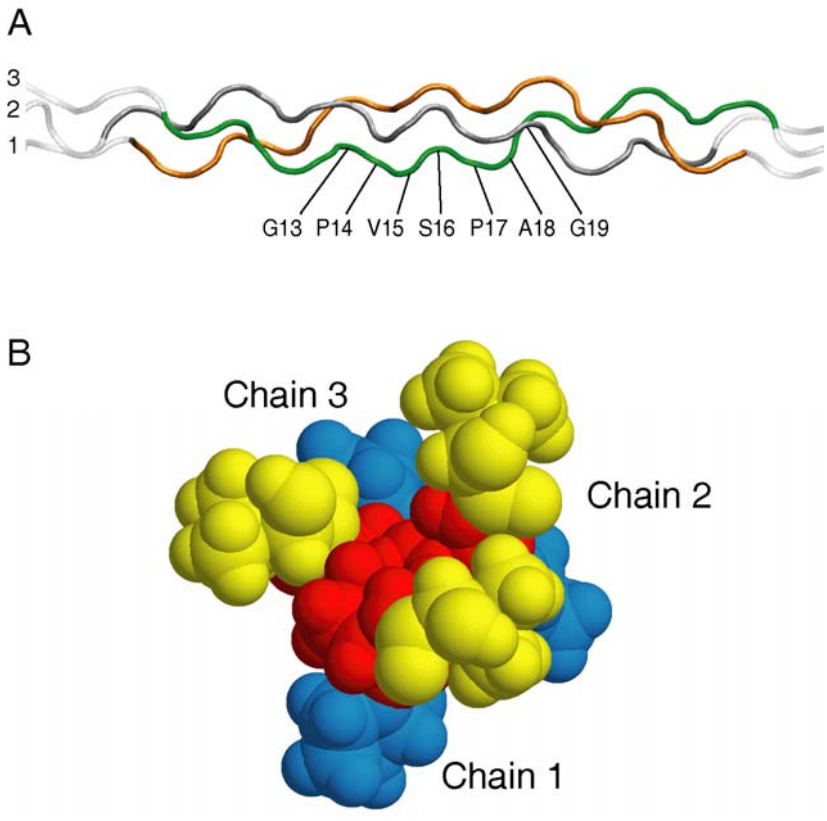

Figure 5.10 The model structure of the GPO(G16S) peptide.

A, Ribbon diagram of GPO(G16S) models represented with the N-terminal at the left and chain numbers. The NMR parameters are used as the restraints in the energy minimization, including the phi angle restraints from J coupling constants and distance restraints from NOEs. Three model structures that are consistent with all NMR experimental data are obtained and show similar features. One model is shown for illustration. Leading, middle and trailing chains are represented as 1, 2 and 3, and are colored in gray, orange, and green, respectively, for the energy minimized segment of residues 6 to 26. The G13 to G19 residues in chain 3 are indicated.

B, Space filling model of the cross section view from the N terminus to the C terminus. Views show that at the S16 residues are closely packed at the center. The VSP segment is colored as V15(■), S16(■), and P17(■).

5.3.2.6. Discussion

5.3.2.6.1. The conformation and hydrogen bonding of the (GPO)G16S model

Mutations replacing one Gly by Ser in the Gly-X-Y repeating sequence of the triple-helix are the most common missense mutations in type I collagen chains leading to OI. Such replacements are known to delay collagen folding (Raghunath et al., 1994) and must perturb the final triple-helix structure formed, since the standard triple-helix requires Gly as every third residue. The model structure was obtained here for a collagen peptide modeling a Gly to Ser substitution within a type I collagen sequence. This visualization of the effect of the Ser substitution on the triple-helix conformation, hydrogen bonding, and dynamic properties represents a first step in clarifying how mutations could impact collagen functions and result in OI disease.

The peptide GPO(G16S) forms a stable triple-helix, and models obtained from NMR data indicate a straight fully helical molecule, with the central region locally distorted by close packing of the three Ser residues where the Gly residues should be. The model suggests that the Gly residues in the triplets N-terminal and C-terminal to the Ser site can form standard interchain Rich-Crick II hydrogen bonds (Gly NH…C=O X) (Figure 5.11A). The amide groups of the three Gly13 residues in the tripeptide N-terminal to the Ser site appear to form hydrogen bonds. The region C-terminal to the Ser mutation site appears to be more distorted than that N-terminal to the site, with ^{1}G19 and ^{3}Gly19 appearing to form a hydrogen bond. Examination of the hydrogen bonding in the alternative models from starting structures of either Gly$\rightarrow$Ala or T3-785 shows similarly asymmetric patterns with all three of G13, ^{3}G19 that is furthest from the Ser site and

either one of ^{1}G19 and ^{2}G19 forming standard hydrogen bonds, consistent with the amide temperature gradient data.

Direct hydrogen bonds between the Ser NH and carbonyl of the Pro in an adjacent chain are not possible, but it is possible that a water mediated hydrogen bond, which is not visible in this modeling, can form as seen for the Gly→Ala structure (Figure 5.11B). However, the replacement of a Gly by a Ser residue introduces the possibility of additional hydrogen bonding through the hydroxyl side chain. Molecular dynamics studies by Mooney and Klein on a range of Gly to Ser OI mutation sites showed that the hydroxyl group of Ser is hydrogen bonded to the backbone of an adjacent chain for a significant fraction of time (Mooney and Klein, 2002). In the models obtained by NMR, hydrogen bonds can be formed between the hydroxyl groups of the S16 side chains and the carbonyl group of Pro residue in the neighboring chain, between the ^{1}S16 OH and the amide of ^{3}S16, or between the ^{1}S16 OH and the ^{3}S16 OH. These Ser hydrogen bonds may replace or augment amide hydrogen bonds at this site to help maintain and stabilize the triple helix assembly.

5.3.2.6.2. The dynamics of the GPO(G16S) peptide

Complementary to the hydrogen bonding analysis from the models, hydrogen exchange data suggest a more dynamic situation. For three G13 residues, despite the presence of hydrogen bonds in the models, different protection factors are shown for three chains (Figure 5.9). The chain furthest from the perturbation shows very slow hydrogen exchange while the chain closest to the Ser perturbation shows fast exchange. It is consistent with a consequence of the 1-residue stagger, and suggests the different

dynamic flexibilities for three G13 residues depending on their relative positions to the Ser mutation. All of the amide groups of the three Gly19 residues in the tripeptide C-terminal to the Ser site show fast hydrogen exchange, indicating an asymmetry in dynamics on N- and C-terminal sides of the mutation. The S16 amides show low protections from solvent, suggesting an increased flexibility at the substitution sites.

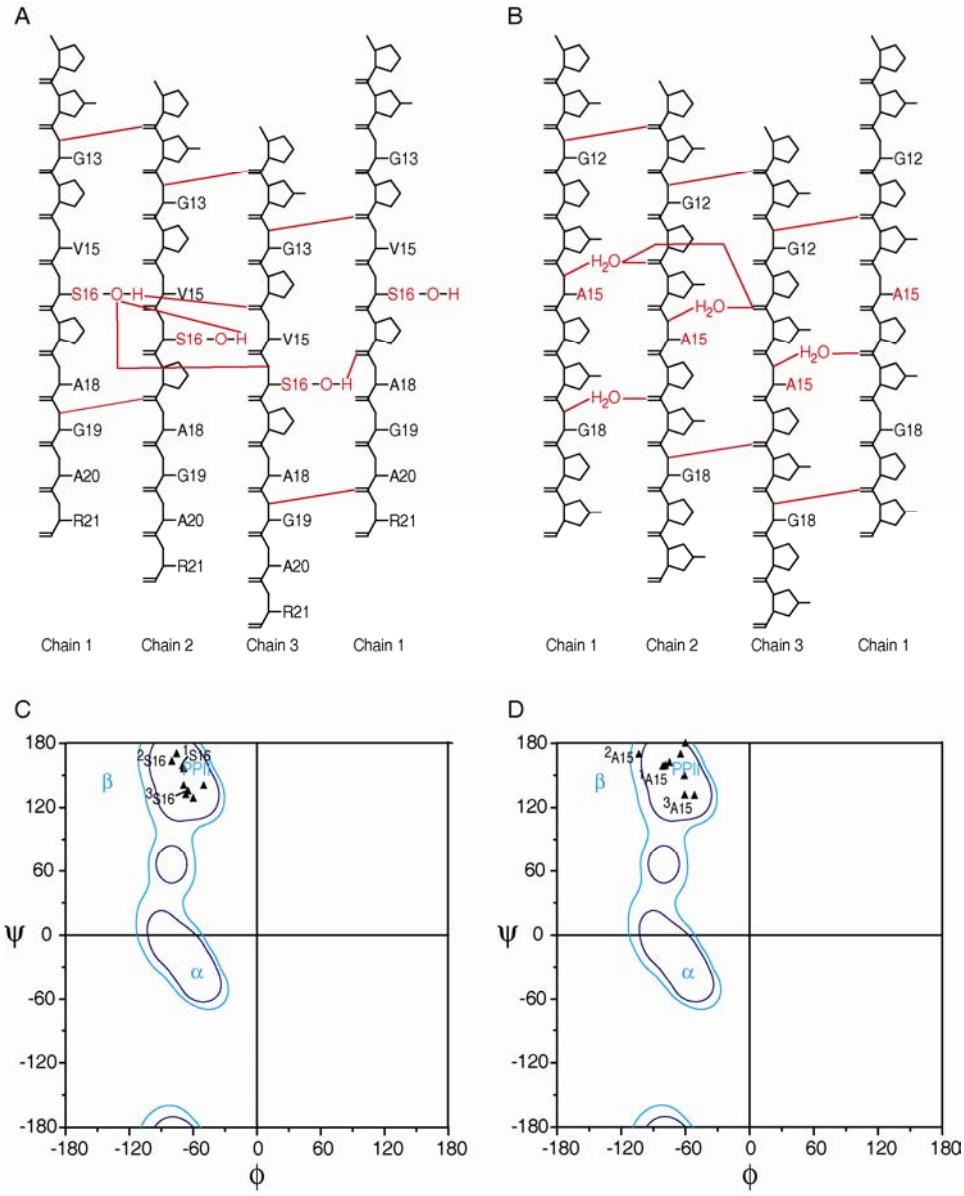

Figure 5.11 Comparison of conformational and hydrogen bonding features of GPO(G16S) peptides to the Gly→Ala peptide.
(A) and (B) Schematic representation of hydrogen bonding topology in the GPO(G16S) model structure around the Gly to Ser substitution region (A), and in the Gly→Ala x-ray crystal structure around the Gly to Ala substitution region (B). The Ser and Ala substituting residues and water molecules in the Gly→Ala structure are colored in red. Hydrogen bonds are shown in red lines.
(C) and (D) Ramachandran plot for V15-S16-P17 residues in the model of GPO(G16S) (C) and for O14-A15-P16 in the Gly→Ala structure (D). Ramachandran contour map for Pro residues is shown in the background and typical secondary structures are indicated (α, α-helix; β, β-sheet; PPII, polyproline II and collagen).

5.3.2.6.3. Comparison of the GPO(G16S) peptide to the Gly→Ala peptide

The NMR model solution structure for the GPO(G16S) peptide, which has the Ser mutation surrounded by the actual α1(I) sequence GPVSPAGAR, is compared with the x-ray structure of the peptide which has an Ala substitution in a GPO environment GPOAPOGPO (Bella et al., 1994) (Table 5.3). Two peptides share many similarities. Peptide studies have indicated a similar destabilization effect for Ala and Ser substitutions (Beck et al., 2000)(Mike Bryan, personal communication). Both the GPO(G16S) and the Gly→Ala peptide show downfield shifted chemical shifts for 1 chain of the substituted residue in the HSQC spectra (Figure 5.5 and 5.6). Similarity is also seen in the experimental NOE contacts of GPO(G16S) and in the predicted contacts of the Gly→Ala x-ray crystal structure, including the non-standard-triple-helical NH-NH contact between chain 1 and chain 3 of the substituted residue. The hydrogen exchange and amide temperature gradient data of the two peptides also indicated similar fast exchange and non-equivalent temperature gradients for three chains at the substitution (Figure 5.9, Table 5.2)(Liu, 1997). Both the (GPO)G16S model and the Gly→Ala x-ray crystal structure indicate in localized conformational distortion and asymmetry of the hydrogen boning at N- and C-terminal sides of the mutation (Figure 5.11).

Despite of these similarities, the GPO(G16S) and the Gly→Ala peptide show differences in some aspects. There are subtle differences in the NOE contacts between the two peptides suggesting different conformations are adopted around the substitution sites (NOEs from ^{1}S16 NH to ^{2}V15 H^{α}, from ^{3}S16 NH to ^{1}G19 NH in the GPO(G16S), while from ^{1}A15 NH to ^{3}O14 H^{α}, from ^{3}A15 NH to ^{1}G18 H^{α}, from ^{1}G18 NH to ^{3}A15 side chain in the Gly→Ala instead). The $^{3}J_{HNHa}$ of ^{2}A15 in the Gly→Ala peptide is very high (Table

5.2) indicating a non-polyproline II conformation, while the $^{3}J_{HNHa}$ of three S16 in the GPO(G16S) are lower that could be consistent with polyproline II (Figure 5.8).

The surprisingly similar features and small number of differences of homotrimeric peptide models of Gly to Ser and Gly to Ala mutations could not be correlated with the high occurrence of Ser substitution and underrepresentation of Ala substitution in the type I collagen and the milder phenotype of OI disease associated with Ala than Ser mutations when they present at the same position. The interpretation could lie in two aspects: 1) the small changes as seen between the two peptides could be significant; 2) the differences of the two types of mutations are in the heterotrimer. Mooney and Klein showed that the Gly to Ala mutation on the three chains causes different degree of destalization of the triple helix in the heterotrimer (Mooney et al., 2001). Here for the Gly to Ser substitution, the asymmetric packing and hydrogen bonding properties of Ser OH in three chains could show more differences in the heterotrimers of Gly to Ser and Gly to Ala mutations.

Table 5.3 Comparison of the (GPO)G16S and Gly→Ala peptide.

a. Similarity between two peptides

	GPO(G16S)	Gly→Ala
Chemical shift	One chain of the substituted residue is downfield-shifted	
NOE contacts	Similar NOE contacts including the non-standard-triple-helical NH-NH contacts between chain 1 and 3 of the substituted residue	
Hydrogen exchange	Low protection factors for the substituted residue	
NH temperature gradient	One chain of the substituted residue shows hydrogen bonding	
Hydrogen bonds in the structure	Asymmetric pattern: C-terminal side of the substitution is more distorted than the N-terminal side of the substitution	
Conformation	Localized deformation, similar cross-view of the Cα atoms around the substitution site	
Thermal dynamic properties[a]	Same in drop in T_m, same mean residue elliplicity and enthalpy	

b. Difference between two peptides

	GPO(G16S)	Gly→Ala
NOE contacts	^{1}S16 NH - ^{2}V15 H$^{\alpha}$ ^{3}S16 NH - ^{1}G19 NH ^{1}G19 NH - ^{3}S16 NH	^{1}A15 NH - ^{3}O14 H$^{\alpha}$, ^{3}A15 NH - ^{1}G18 H$^{\alpha}$, ^{1}G18 NH - ^{3}A15 side chain
$^{3}J_{HNHa}$	All three S16 has values from 5 to 6.5 Hz	One of A15 has high value of 8.6 Hz

[a] Mike Bryan, personal communication.

5.3.2.6.4. Biological implications

It has been suggested that structural alterations arising from Gly substitutions disturb the higher order structure of collagen and thus affect ligand recognition and binding (Di Lullo et al., 2002; Marini et al., 2007). There are two exclusively lethal regions that correlate with major ligand binding regions (MLBRs) in type I collagen (Di Lullo et al., 2002; Marini et al., 2007). Disruption of interactions of collagen with other molecules such as integrins, matrix metalloproteinases (MMPs), fibronectin, and cartilage oligomeric matrix protein (COMP) by the Gly to X substitutions could be part of the pathological process (Kleinman et al., 1978; Lauer-Fields et al., 2002; Marini et al., 2007;

Rosenberg et al., 1998; Xu et al., 2000). The NMR visualization of the effect of the Ser substitution on the triple-helix conformation and hydrogen bonding suggests that it could alter the appearance of the exterior of the triple-helix molecule as it is seen by binding partners. In a standard triple helical conformation, Cα of Gly residues are equal distance from the central axis of the triple helix, and a circle can be drawn to trace all Gly residues (Figure 5.12, the circle denoted as G). Similarly, Cα of residues at X and Y positions can be also placed on their specific circles (Figure 5.12, circles denoted as X and Y). Plotting the cross-section of the GPO(G16S) peptide indicates that the Cα atoms of residues in the X and Y positions surrounding the Ser are located further out from center than a standard triple-helix (Figure 5.12A). A similar deviation from a standard triple helix is also observed for the Gly→Ala x-ray structure (Figure 5.12B). This could alter the recognition or interaction with proteoglycans or other matrix molecules. In addition, the NMR studies on the GPO(G16S) peptide showed a increased flexibility around the substitution site. Studies showed that substrate dynamics play an important role in the enzyme recognition and catalysis (Kern et al., 2005), and the dynamic fluctuation of the triple helix at the substitution site could be involved in the binding and interaction phenomena of collagen.

In homotrimeric collagen with such as type III collagen, Gly to X substitution in one allele of the α1(III) chain in type III collagen leads to Ehlers-Danlos syndrome type IV, a collagen disease indicated by arterial rupture (Kuivaniemi et al., 1997). Because of the self assembly of the three mixed chains, there is a small percentage of collagen molecules with the Gly mutation in all three chains, and a big percentage of collagen molecules in which the Gly mutation is opposite one substitution/one native chain or two native

chains. In a heterotrimeric collagen, such as type I collagen, a single Gly mutation in any one allele of the α1(I) or α2(I) chain in the $[\alpha1(I)]_2\alpha2(I)$ collagen molecule leads to OI (Kuivaniemi et al., 1997). After self-assembly, the Gly to X substitution is opposite one substitution/one native chain or two native chains. The localized alterations in conformation, dynamic, hydrogen bonding shown here for the Gly to Ser homotrimeric molecule may shed light on heterotrimers.

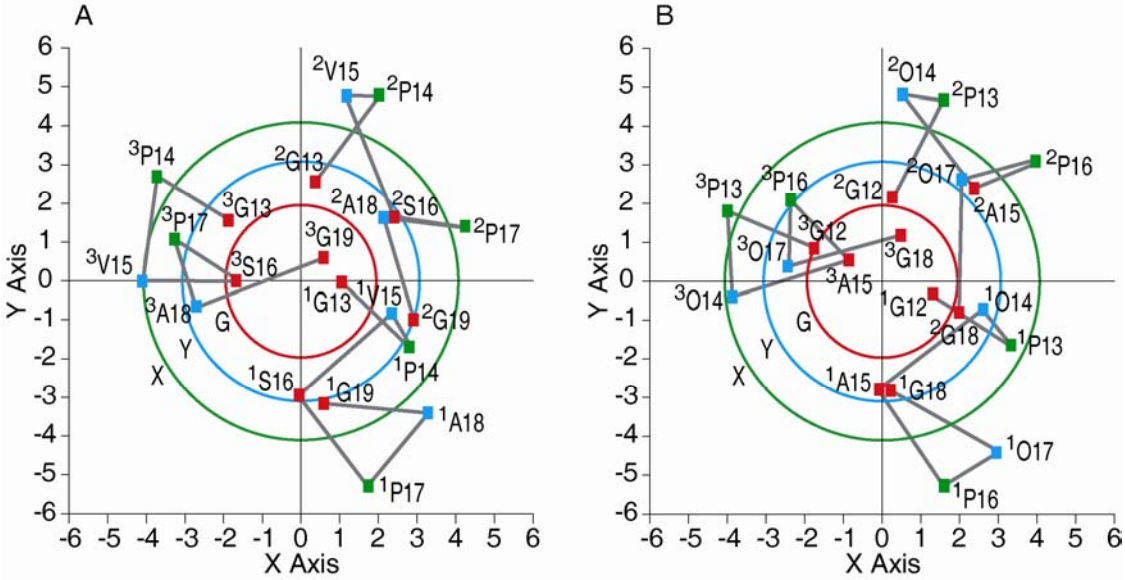

Figure 5.12 Comparison of the cross views of Cα of residues around the substitution region in the GPO(G16S) model (A) and Gly→Ala structure (B). The structures were aligned to the central axis of the triple helix molecule using pdbinertia program (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software/pdbinertia.html). X and Y axis indicates the coordinates in the structure. The projection of Cα atoms of residues G13 to G19 in one GPO(G16S) model structure and residues G12 to G18 in the Gly→Ala structure are shown in squares. Residues at the Gly, X, and Y positions are colored in red, green, and blue, respectively. The substituting residues are colored in red. The Cα traces are shown in gray lines. In a standard triple helical conformation, Cα of Gly, X and Y residues are equal distance from the central axis of the triple helix, respectively. Three circles indicate the Cα traces of Gly, X and Y residues in $(PPG)_{10}$ x-ray crystal structure (Berisio et al., 2002), and colored in red, green, and blue for residues at Gly, X, and Y positions, respectively.

REFERENCES

Andersen, N.H., Neidigh, J.W., Harris, S.M., Lee, G.M., Liu, Z.H., and Tong, H. (1997). Extracting information from the temperature gradients of polypeptide NH chemical shifts .1. The importance of conformational averaging. J Am Chem Soc *119*, 8547-8561.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. Proteins *17*, 75-86.

Balbach, J. (2000). Compaction during protein folding studied by real-time NMR diffusion experiments. J Am Chem Soc *122*, 5887-5888.

Baum, J., and Brodsky, B. (1999). Folding of peptide models of collagen and misfolding in disease. Curr Opin Struct Biol *9*, 122-128.

Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. Methods Enzymol *339*, 127-174.

Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R., and Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. Methods Enzymol *239*, 79-105.

Baxter, N.J., and Williamson, M.P. (1997). Temperature dependence of ^{1}H chemical shifts in proteins. J Biomol NMR *9*, 359-369.

Beck, K., Chan, V.C., Shenoy, N., Kirkpatrick, A., Ramshaw, J.A., and Brodsky, B. (2000). Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. Proc Natl Acad Sci U S A *97*, 4273-4278.

Bella, J., Eaton, M., Brodsky, B., and 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., and Berman, H.M. (2006). Conformational effects of Gly-X-Gly interruptions in the collagen triple helix. J Mol Biol *362*, 298-311.

Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002). Crystal structure of the collagen triple helix model $[(\text{Pro-Pro-Gly})_{10}]_3$. Protein Science *11*, 262-270.

Bhate, M., Wang, X., Baum, J., and Brodsky, B. (2002). Folding and conformational consequences of glycine to alanine replacements at different positions in a collagen model peptide. Biochemistry *41*, 6539-6547.

Borza, D.B., Bondar, O., Ninomiya, Y., Sado, Y., Naito, I., Todd, P., and Hudson, B.G. (2001). The NC1 domain of collagen IV encodes a novel network composed of the alpha 1, alpha 2, alpha 5, and alpha 6 chains in smooth muscle basement membranes. J Biol Chem *276*, 28532-28540.

Boydston, J.A., Chen, P., Steichen, C.T., and Turnbough, C.L., Jr. (2005). Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of Bacillus anthracis. J Bact *187*, 5310-5317.

Brazel, D., Oberbaumer, I., Dieringer, H., Babel, W., Glanville, R.W., Deutzmann, R., and Kuhn, K. (1987). Completion of the amino acid sequence of the alpha 1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. Eur J Biochem *168*, 529-536.

Brodsky-Doyle, B., Leonard, K.R., and Reid, K.B. (1976). Circular-dichroism and electron-microscopy studies of human subcomponent C1q before and after limited proteolysis by pepsin. Biochem J *159*, 279-286.

Brodsky, B., and Persikov, A.V. (2005). Molecular structure of the collagen triple helix. Adv Protein Chem *70*, 301-339.

Buevich, A., and Baum, J. (2001). Nuclear magnetic resonance characterization of peptide models of collagen-folding diseases. Philos Trans R Soc Lond B Biol Sci *356*, 159-168.

Buevich, A.V., and Baum, J. (2002). Residue-specific real-time NMR diffusion experiments define the association states of proteins during folding. J Am Chem Soc *124*, 7156-7162.

Buevich, A.V., Dai, Q.H., Liu, X., Brodsky, B., and Baum, J. (2000). Site-specific NMR monitoring of cis-trans isomerization in the folding of the proline-rich collagen triple helix. Biochemistry *39*, 4299-4308.

Buevich, A.V., Silva, T., Brodsky, B., and Baum, J. (2004). Transformation of the mechanism of triple-helix peptide folding in the absence of a C-terminal nucleation domain and its implications for mutations in collagen disorders. J Biol Chem *279*, 46890-46895.

Byers, P.H., and Cole, W.G. (2002). Osteogenesis Imperfecta. In Connective tissue and its hereditable disorders, P.M. Royce, and B. Steinmann, eds. (New York, Wiley-Liss), pp. 385-430.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G., and Skelton, N.J. (1996). Protein NMR spectroscopy: principle and practice (San Diego, Academic Press).

Chen, A., and Shapiro, M. (1999). Nuclear overhauser effect on diffusion measurements. J Am Chem Soc *121*, 5338-5339.

Chen, A.D., Johnson, C.S., Lin, M., and Shapiro, M.J. (1998). Chemical exchange in diffusion NMR experiments. J Am Chem Soc *120*, 9094-9095.

Cho, C.H., Urquidi, J., Singh, S., and Robinson, G.W. (1999). Thermal offset viscosities of liquid H_2O, D_2O, and T_2O. J Phys Chem B *103*, 1991-1994.

Chou, J.J., Baber, J.L., and Bax, A. (2004). Characterization of phospholipid mixed micelles by translational diffusion. J Biomol NMR *29*, 299-308.

Cierpicki, T., and Otlewski, J. (2001). Amide proton temperature coefficients as hydrogen bond indicators in proteins. J Biomol NMR *21*, 249-261.

Clubb, R.T., Thanabal, V., and Wagner, G. (1992). A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ^{1}HN, ^{15}N, and ^{13}C′ chemical shifts in ^{15}N-^{13}C-labelled proteins. J Magn Reson *97*, 213-217.

Consonni, R., Zetta, L., Longhi, R., Toma, L., Zanaboni, G., and Tenni, R. (2000). Conformational analysis and stability of collagen peptides by CD and by ^{1}H- and ^{13}C-NMR spectroscopies. Biopolymers *53*, 99-111.

Cordier, F., and Grzesiek, S. (2002). Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. J Mol Biol *317*, 739-752.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Deprez, P., Doss-Pepe, E., Brodsky, B., and Inestrosa, N.C. (2000). Interaction of the collagen-like tail of asymmetric acetylcholinesterase with heparin depends on triple-helical conformation, sequence and stability. Biochem J *350 Pt 1*, 283-290.

Derrick, T.S., Lucas, L.H., Dimicoli, J.L., and Larive, C.K. (2002). F^{19} diffusion NMR analysis of enzyme-inhibitor binding. Magn Reson Chem *40*, S98-S105.

Di Lullo, G.A., Sweeney, S.M., Korkko, J., Ala-Kokko, L., and 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*, 4223-4231.

Dobson, C.M., and Hore, P.J. (1998). Kinetic studies of protein folding using NMR spectroscopy. Nat Struct Biol *5 Suppl*, 504-507.

Dvinskikh, S.V., and Furo, I.I. (2000). Cross-relaxation effects in stimulated-echo-type PGSE NMR experiments by bipolar and monopolar gradient pulses. J Magn Reson *146*, 283-289.

Englander, S.W., and Krishna, M.M.G. (2001). Hydrogen exchange. Nat Struc Biol *8*, 741-742.

Englander, S.W., Sosnick, T.R., Englander, J.J., and Mayne, L. (1996). Mechanisms and uses of hydrogen exchange. Curr Opin Struct Biol *6*, 18-23.

Fan, P., Li, M.H., Brodsky, B., and Baum, J. (1993). Backbone dynamics of (Pro-Hyp-Gly)$_{10}$ and a designed collagen-like triple-helical peptide by ^{15}N NMR relaxation and hydrogen-exchange measurements. Biochemistry *32*, 13299-13309.

Fesik, S.W., and 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., and Prockop, D.J. (1996). Perspectives on the synthesis and application of triple-helical, collagen-model peptides. Biopolymers *40*, 345-357.

Frank, S., Kammerer, R.A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bachinger, H.P., and Engel, J. (2001). Stabilization of short collagen-like triple helices by protein engineering. J Mol Biol *308*, 1081-1089.

Furcht, L.T., Skubitz, A.P., and Fields, G.B. (1994). Tumor cell invasion, matrix metalloproteinases, and the dogma. Laboratory investigation; a journal of technical methods and pathology *70*, 781-783.

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

Goodman, M., Bhumralkar, Jefferson, E.A., Kwak, J., and Locardi, E. (1998). Collagen mimetics. Biopolymers *47*, 127-142.

Grzesiek, S., Anglister, J., and Bax, A. (1993). Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in C^{13}/N^{15} Enriched Proteins By Isotropic Mixing of C^{13} Magnetization. J Magn Reson B *101*, 114-119.

Grzesiek, S., and Bax, A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. J Am Chem Soc *114*, 6291-6293.

Grzesiek, S., Cordier, F., and Dingley, A.J. (2001). Scalar couplings across hydrogen bonds. Methods Enzymol *338*, 111-133.

Guntert, P. (1998). Structure calculation of biological macromolecules from NMR data. Q Rev Biophys *31*, 145-237.

Heidemann, E., and Roth, W. (1982). Synthesis and investigation of collagen model peptides. Adv Polym Sci *43*, 143-203.

Hofmann, H., Voss, T., Kuhn, K., and 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*, 325-343.

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

Hoppe, H.J., and Reid, K.B. (1994). Collectins--soluble proteins containing collagenous regions and lectin domains--and their roles in innate immunity. Protein Sci *3*, 1143-1158.

Hudson, B.G., Reeders, S.T., and Tryggvason, K. (1993). Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. J Biol Chem *268*, 26033-26036.

Hudson, B.G., Tryggvason, K., Sundaramoorthy, M., and Neilson, E.G. (2003). Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N Engl J Med *348*, 2543-2556.

Humtsoe, J.O., Kim, J.K., Xu, Y., Keene, D.R., Hook, M., Lukomski, S., and Wary, K.K. (2005). A streptococcal collagen-like protein interacts with the alpha2beta1 integrin and induces intracellular signaling. J Biol Chem *280*, 13848-13857.

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

Hyde, T.J., Bryan, M., Brodsky, B., and Baum, J. (2006). Sequence dependence of renucleation after a GLY mutation in model collagen peptides. J Biol Chem *281*, 36937-36943.

Ikura, M., Kay, L.E., and Bax, A. (1990). A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry *29*, 4659-4667.

Jershow, A., and Muller, N. (1997). Suppression of Convection Artifacts in Stimulated-Echo Diffusion Experiments. Double-Stimulated-Echo Experiments. J Magn Reson *125*, 372-375.

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

Johnson, C.S. (1993). Effects of Chemical Exchange in Diffusion Ordered NMR Spectra. J Magn Reson A *102*, 214-218.

Jones, J.A., Wilkins, D.K., Smith, L.J., and Dobson, C.M. (1997). Characterisation of protein unfolding by NMR diffusion measurements. J Biomol NMR *10*, 199-203.

Kadler, K. (1994). Extracellular matrix. 1: fibril-forming collagens. Protein Profile *1*, 519-638.

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

Kay, L.E., and Gardner, K.H. (1997). Solution NMR spectroscopy beyond 25 kDa. Curr Opin Struct Biol *7*, 722-731.

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

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

Kay, L.E., Nicholson, L.K., Delaglio, F., Bax, A., and Torchia, D.A. (1992b). Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T_1 and T_2 values in proteins. J Magn Reson *97*, 359-375.

Kern, D., Eisenmesser, E.Z., and Wolf-Watz, M. (2005). Enzyme dynamics during catalysis measured by NMR spectroscopy. Methods Enzymol *394*, 507-524.

Kielty, C.M., and Grant, M.E. (2002). The collagen family: strucutre, assembly, and organization in the extracellular matrix. In Connective tissue and its heritable disorders, molecular, genetic and medical aspects, P.M. Royes, and B.U. Steinmann, eds. (New York, Wiley Liss), pp. 159-222.

Kilchherr, E., Hofmann, H., Steigemann, W., and Engel, J. (1985). Structural model of the collagen-like region of C1q comprising the kink region and the fibre-like packing of the six triple helices. J Mol Biol *186*, 403-415.

Kim, S., and Szyperski, T. (2003). GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. J Am Chem Soc *125*, 1385-1393.

Kim, S., and Szyperski, T. (2004). GFT NMR experiments for polypeptide backbone and ^{13}Cbeta chemical shift assignment. J Biomol NMR *28*, 117-130.

Kleinman, H.K., McGoodwin, E.B., Martin, G.R., Klebe, R.J., Fietzek, P.P., and Woolley, D.E. (1978). Localization of the binding site for cell attachment in the alpha1(I) chain of collagen. J Biol Chem *253*, 5642-5646.

Knight, C.G., Morton, L.F., Peachey, A.R., Tuckwell, D.S., Farndale, R.W., and Barnes, M.J. (2000). The collagen-binding A-domains of integrins alpha(1)beta(1) and

alpha(2)beta(1) recognize the same specific amino acid sequence, GFOGER, in native triple-helical collagens. J Biol Chem *275*, 35-40.

Korzhnev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M., and Kay, L.E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. Nature *430*, 586-590.

Kramer, R.Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H.M. (1999). Sequence dependent conformational variations of collagen triple-helical structure. Nat Struct Biol *6*, 454-457.

Kramer, R.Z., Venugopal, M.G., Bella, J., Mayville, P., Brodsky, B., and Berman, H.M. (2000). Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. J Mol Biol *301*, 1191-1205.

Kramer, R.Z., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A., and Berman, H.M. (1998). X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). J Mol Biol *280*, 623-638.

Kuivaniemi, H., Tromp, G., and Prockop, D. (1997). Mutations in Fibrillar Collagens (Types I,II,II, and IX), Fibril-Associated Collagen (Type IX), and Network-Forming Collagen (Type X) Cause a Spectrum of Diseases of Bone, Cartilage, and Blood Vessels. Human Mutation *9*, 300-315.

Lauer-Fields, J.L., and Fields, G.B. (2002). Triple-helical peptide analysis of collagenolytic protease activity. Biol Chem *383*, 1095-1105.

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

Lemmink, H.H., Schroder, C.H., Monnens, L.A., and Smeets, H.J. (1997). The clinical spectrum of type IV collagen mutations. Hum Mutat *9*, 477-499.

Li, M.H., Fan, P., Brodsky, B., and Baum, J. (1993). Two-dimensional NMR assignments and conformation of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and a designed collagen triple-helical peptide. Biochemistry *32*, 7377-7387.

Li, Y., Brodsky, B., and Baum, J. (2007). NMR Shows Hydrophobic Interactions Replace Glycine Packing in the Triple Helix at a Natural Break in the $(Gly\text{-}X\text{-}Y)_n$ Repeat. J Biol Chem *282*, 22699-22706.

Li, Y., Kim, S., Brodsky, B., and Baum, J. (2005). Identification of partially disordered peptide intermediates through residue-specific NMR diffusion measurements. J Am Chem Soc *127*, 10490-10491.

Liu, X. (1997). Real-time NMR investigations of triple helical peptide folding and collagen folding diseases. Ph D Dissertation, 101.

Liu, X.Y., Kim, S., Dai, Q.H., Brodsky, B., and Baum, J. (1998). Nuclear magnetic resonance shows asymmetric loss of triple helix in peptides modeling a collagen mutation in brittle bone disease. Biochemistry *37*, 15528-15533.

Liu, X.Y., Siegel, D.L., Fan, P., Brodsky, B., and Baum, J. (1996). Direct NMR measurement of the folding kinetics of a trimeric peptide. Biochemistry *35*, 4306-4313.

Long, C.G., Braswell, E., Zhu, D., Apigo, J., Baum, J., and Brodsky, B. (1993). Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. Biochemistry *32*, 11688-11695.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins *50*, 437-450.

Lucas, L.H., Yan, J., Larive, C.K., Zartler, E.R., and Shapiro, M.J. (2003). Transferred nuclear overhauser effect in nuclear magnetic resonance diffusion measurements of ligand-protein binding. Anal Chem *75*, 627-634.

Mansfield, S.L., Jayawickrama, D.A., Timmons, J.S., and Larive, C.K. (1998). Measurement of peptide aggregation with pulsed-field gradient nuclear magnetic resonance spectroscopy. Biochim Biophys Acta *1382*, 257-265.

Marini, J.C., Forlino, A., Cabral, W.A., Barnes, A.M., San Antonio, J.D., Milgrom, S., Hyland, J.C., Korkko, J., Prockop, D.J., De Paepe, A.*, et al.* (2007). Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. Hum Mutat *28*, 209-221.

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

Mayo, K.H. (1996). NMR and x-ray studies of collagen model peptides. Biopolymers *40*, 359-370.

Merutka, G., Dyson, H.J., and Wright, P.E. (1995). 'Random coil' ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. J Biomol NMR *5*, 14-24.

Messerle, B.A., Wider, G., Otting, G., Weber, C., and 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.

Miles, A.J., Knutson, J.R., Skubitz, A.P., Furcht, L.T., McCarthy, J.B., and Fields, G.B. (1995). A peptide model of basement membrane collagen alpha 1 (IV) 531-543 binds the alpha 3 beta 1 integrin. J Biol Chem *270*, 29047-29050.

Miles, A.J., Skubitz, A.P., Furcht, L.T., and Fields, G.B. (1994). Promotion of cell adhesion by single-stranded and triple-helical peptide models of basement membrane collagen alpha 1(IV)531-543. Evidence for conformationally dependent and conformationally independent type IV collagen cell adhesion sites. J Biol Chem *269*, 30939-30945.

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

Mohs, A., Li, Y., Doss-Pepe, E., Baum, J., and Brodsky, B. (2005). Stability junction at a common mutation site in the collagenous domain of the mannose binding lectin. Biochemistry *44*, 1793-1799.

Mohs, A., Popiel, M., Li, Y., Baum, J., and Brodsky, B. (2006). Conformational features of a natural break in the type IV collagen Gly-X-Y repeat. J Biol Chem *281*, 17197-17202.

Mooney, S.D., Huang, C.C., Kollman, P.A., and Klein, T.E. (2001). Computed free energy differences between point mutations in a collagen-like peptide. Biopolymers *58*, 347-353.

Mooney, S.D., and Klein, T.E. (2002). Structural models of osteogenesis imperfecta-associated variants in the COL1A1 gene. Mol Cell Proteomics *1*, 868-875.

Myllyharju, J., and Kivirikko, K.I. (2001). Collagens and collagen-related diseases. Ann Med *33*, 7-21.

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

Nagarajan, V., Kamitori, S., and Okuyama, K. (1999). Structure analysis of a collagen-model peptide with a (Pro-Hyp-Gly) sequence repeat. J Biochem (Tokyo) *125*, 310-318.

Okuyama, K., Hongo, C., Fukushima, R., Wu, G., Narita, H., Noguchi, K., Tanaka, Y., and Nishino, N. (2004). Crystal structures of collagen model peptides with Pro-Hyp-Gly repeating sequence at 1.26 Å resolution: implications for proline ring puckering. Biopolymers *76*, 367-377.

Ottl, J., Battistuta, R., Pieper, M., Tschesche, H., Bode, W., Kuhn, K., and 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*, 31-36.

Palmer, A.G. (2004). NMR characterization of the dynamics of biomacromolecules. Chem Rev *104*, 3623-3640.

Palmer, A.G., 3rd (2001). Nmr probes of molecular dynamics: overview and comparison with other techniques. Annu Rev Biophys Biomol Struct *30*, 129-155.

Palmer, A.G., 3rd, and Massi, F. (2006). Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem Rev *106*, 1700-1719.

Palmer, A.G., Grey, M.J., and Wang, C.Y. (2005). Solution NMR spin relaxation methods for characterizing chemical exchange in high-molecular-weight systems. Method Enzymol *394*, 430-465.

Persikov, A.V., Pillitteri, R.J., Amin, P., Schwarze, U., Byers, P.H., and Brodsky, B. (2004). Stability related bias in residues replacing glycines within the collagen triple helix (Gly-Xaa-Yaa) in inherited connective tissue disorders. Hum Mutat *24*, 330-337.

Persikov, A.V., Ramshaw, J.A., and Brodsky, B. (2005). Prediction of collagen stability from amino acid sequence. J Biol Chem *280*, 19343-19349.

Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci U S A *94*, 12366-12371.

Ponder, J.W., and Case, D.A. (2003). Force fields for protein simulations. Adv Protein Chem *66*, 27-85.

Powers, R., Gronenborn, A.M., Clore, G.M., and Bax, A. (1991). Three-dimensional triple-resonance NMR of 13C/15N-enriched proteins using constant-time evolution. J Magn Reson *94*, 209-213.

Privalov, P.L. (1982). Stability of proteins. Proteins which do not present a single cooperative system. Adv Protein Chem *35*, 1-104.

Qing, G., Ma, L.C., Khorchid, A., Swapna, G.V., Mal, T.K., Takayama, M.M., Xia, B., Phadtare, S., Ke, H., Acton, T.*, et al.* (2004). Cold-shock induced high-yield protein production in Escherichia coli. Nat Biotechnol *22*, 877-882.

Raghunath, M., Bruckner, P., and 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. In Treatise on Collagen, G.N. Ramachandran, ed. (New York, Academic Press), pp. 103-184.

Ramachandran, G.N., and Kartha, G. (1955). Structure of collagen. Nature *176*, 593-595.
Raynal, N., Hamaia, S.W., Siljander, P.R., Maddox, B., Peachey, A.R., Fernandez, R., Foley, L.J., Slatter, D.A., Jarvis, G.E., and Farndale, R.W. (2006). Use of Synthetic Peptides to Locate Novel Integrin {alpha}2beta1-binding Motifs in Human Collagen III. J Biol Chem *281*, 3821-3831.

Renner, C., Sacca, B., and Moroder, L. (2004). Synthetic heterotrimeric collagen peptides as mimics of cell adhesion sites of the basement membrane. Biopolymers *76*, 34-47.
Rich, A., and Crick, F.H. (1961). The molecular structure of collagen. J Mol Biol *3*, 483-506.

Rosenberg, K., Olsson, H., Morgelin, M., and Heinegard, D. (1998). Cartilage oligomeric matrix protein shows high affinity zinc-dependent interaction with triple helical collagen. J Biol Chem *273*, 20397-20403.

Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and Prockop, D.J. (1973). Synthesis of $(Pro\text{-}Hyp\text{-}Gly)_n$ of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxypyroline. Biochim Biophys Acta *303*, 198-202.

Schittny, J.C., and Yurchenco, P.D. (1989). Basement membranes: molecular organization and function in development and disease. Curr Opin Cell Biol *1*, 983-988.

Siljander, P.R., Hamaia, S., Peachey, A.R., Slatter, D.A., Smethurst, P.A., Ouwehand, W.H., Knight, C.G., and Farndale, R.W. (2004). Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. J Biol Chem *279*, 47763-47772.

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

Summerfield, J.A., Sumiya, M., Levin, M., and Turner, M.W. (1997). Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series. BMJ (Clinical research ed *314*, 1229-1232.

Super, M., Thiel, S., Lu, J., Levinsky, R.J., and Turner, M.W. (1989). Association of low levels of mannan-binding protein with a common defect of opsonisation. Lancet *2*, 1236-1239.

Sutcliffe, M.J., and Dobson, C.M. (1991). Relaxation data in NMR structure determination: model calculations for the lysozyme-Gd3+ complex. Proteins *10*, 117-129.

Tillett, M.L., Lian, L.Y., and Norwood, T.J. (1998). Practical aspects of the measurement of the diffusion of proteins in aqueous solution. J Magn Reson *133*, 379-384.

Tjandra, N., and Bax, A. (1997a). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science *278*, 1111-1114.

Tjandra, N., and Bax, A. (1997b). Solution NMR measurement of amide proton chemical shift anisotropy in N^{15}-enriched proteins. Correlation with hydrogen bond length. J Am Chem Soc *119*, 8076-8082.

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

Vuister, G.W., Yamazaki, T., Torchia, D.A., and Bax, A. (1993). Measurement of two- and three-bond ^{13}C-^{1}H J couplings to the C delta carbons of leucine residues in staphylococcal nuclease. J Biomol NMR *3*, 297-306.

Wagner, G. (1993). Prospects for NMR of large proteins. J Biomol NMR *3*, 375-385.
Wagner, G., Pardi, A., and Wuthrich, K. (1983). Hydrogen bond length and proton NMR chemical shifts in proteins. J Am Chem Soc *105*, 5948-5949.

Wallis, R., Shaw, J.M., Uitdehaag, J., Chen, C.B., Torgersen, D., and Drickamer, K. (2004). Localization of the serine protease-binding sites in the collagen-like domain of mannose-binding protein: indirect effects of naturally occurring mutations on protease binding and activation. J Biol Chem *279*, 14065-14073.

Weisemann, R., Ruterjans, H., Schwalbe, H., Schleucher, J., Bermel, W., and Griesinger, C. (1994). Determination of H(N),H-Alpha and H(N), C' Coupling-Constants in C^{13},N^{15}-Labeled Proteins. J Biomol NMR *4*, 231-240.

Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A., and Smith, L.J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry *38*, 16424-16431.

Wittekind, M., and Mueller, L. (1993). HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. J Magn Reson B *101*, 201-205.

Word, J.M., Lovell, S.C., Richardson, J.S., and Richardson, D.C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol *285*, 1735-1747.

Wuthrich, K. (2001). The way to NMR structures of proteins. Nat Struct Biol *8*, 923-925.

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

Xu, Y., Gurusiddappa, S., Rich, R.L., Owens, R.T., Keene, D.R., Mayne, R., Hook, A., and Hook, M. (2000). Multiple binding sites in collagen type I for the integrins alpha1beta1 and alpha2beta1. J Biol Chem *275*, 38981-38989.

Xu, Y., Keene, D.R., Bujnicki, J.M., Hook, M., and Lukomski, S. (2002). Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. J Biol Chem *277*, 27312-27318.

Xu, Y.J., Hyde, T., Wang, X., Bhate, M., Brodsky, B., and Baum, J. (2003). NMR and CD spectroscopy show that imino acid restriction of the unfolded state leads to efficient folding. Biochemistry *42*, 8696-8703.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R., and Kay, L.E. (1994). A Suite of Triple-Resonance Nmr Experiments for the Backbone Assignment of N^{15}, C^{13}, H^2 Labeled Proteins with High-Sensitivity. J Am Chem Soc *116*, 11655-11666.

Yang, W., Battineni, M.L., and Brodsky, B. (1997). Amino acid sequence environment modulates the disruption by osteogenesis imperfecta glycine substitutions in collagen-like peptides. Biochemistry *36*, 6930-6935.

Yurchenco, P.D., and Ruben, G.C. (1988). Type IV collagen lateral associations in the EHS tumor matrix. Comparison with amniotic and in vitro networks. Am J Pathol *132*, 278-291.

Zhou, J., Hertz, J.M., Leinonen, A., and Tryggvason, K. (1992). Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. J Biol Chem *267*, 12475-12481.

Zhu, G., and Bax, A. (1992). Two-dimensional linear prediction for signals truncated in both dimensions. J Magn Reson *98*, 192-199.

Curriculum Vita

Yingjie Li

1996-2000	B.S. in Microbiology, Fudan University, Shanghai, P.R.China
2001-2002	Graduate Fellowship. Rutgers/UMDNJ Graduate Programs in Molecular Biosciences, Rutgers University, Piscataway, NJ.
2002-2007	Research Assistant. Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ.
2002-2008	Teaching Assistant. Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ.
2005	Mohs, A., Li, Y., Doss-Pepe, E., Baum, J. and Brodsky, B. "Stability junction at a common mutation site in the collagenous domain of the mannose binding lectin." Biochemistry 2005, 44, 1793.
2005	Li, Y., Kim, S., Brodsky, B. and Baum, J. "Identification of partially disordered peptide intermediates through residue specific NMR diffusion measurements." J. Am. Chem. Soc. 2005,127, 10490.
2006	Mohs, A., Popiel, M., Li, Y., Baum, J. and Brodsky, B. "Conformational features of a natural break in the type IV collagen Gly-X-Y repeat." J. Biol. Chem. 2006, 281, 17197.
2006	Winner of the Molecular Biophysics Minisymposium Poster Contest. Center for Molecular Biophysics and Biophysical Chemistry, Rutgers University, Piscataway, NJ.
2006-2007	Reid Award for senior graduate students who have demonstrated outstanding accomplishments in their thesis research. Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ.
2007	Stanley and Francine Mandeles Graduate Research Award. Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ
2007	Li, Y., Brodsky, B. and Baum, J. "NMR shows hydrophobic interactions replace glycine packing in the triple-helix at a natural break in the $(Gly\text{-}X\text{-}Y)_n$ repeat." J. Biol. Chem. 2007, 282, 22699.
2007	Ph. D. in Biochemistry. Department of Biochemistry, Rutgers University, New Brunswick, NJ.