MOLECULAR DYNAMICS SIMULATIONS OF COLLAGEN MODEL PEPTIDES: IMPLICATION FOR COLLAGEN DISEASES AND RECOGNITION

By Iwen Fu

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Computational Biology and Molecular Biophysics
Written under the direction of
Dr. David A. Case and Dr. Jean Baum
and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey
January 2015

ABSTRACT OF THE DISSERTATION

MOLECULAR DYNAMICS SIMULATIONS OF COLLAGEN MODEL PEPTIDES: IMPLICATION FOR COLLAGEN DISEASES AND RECOGNITION

BY IWEN FU

DISSERTATION DIRECTOR:

DR. DAVID A. CASE

Understanding the structure and dynamics of the collagen triple helix is critical to understanding the effect of mutations that arise in connective tissue and to understanding its interactions with receptors. Defects in triple helix domain of collagen have been associated with a number of human collagen diseases. Osteogenesis Imperfecta (OI) characterized by brittle bones affects roughly one in 10,000 individuals and results from mutations in type I collagen. The severity of the disease varies widely, ranging from mild to lethal cases. The molecular basis of the disease is still not understood. Here we integrate computational approaches and NMR experiments to provide insight into the effects of variations in amino acid sequence on the conformation and dynamics of the collagen triple helix using collagen model peptides (CMPs), and to provide a molecular interpretation of several types of NMR experiments. Theoretical calculations of quantum mechanical NMR chemical shifts on the MD selected snapshots represent the first example of a reasonable fit of experimental data and highlight the sensitivity of the NH chemical shift to the complex dynamics of the H-bonds.

The degree of inter-chain hydrogen bonding and the dynamics with different G-X-Y triplets are considered important for collagen recognition and binding sites containing disease-related mutation. A analysis of dynamics of triple helix have been studied using ^{15}N relaxation measurements and model-free approach. This work is one of the first detailed MD studies of NMR relaxation data for proteins with high shape anisotropy. We interpret ^{15}N relaxation measurements arising from the a variety of types of motions in the triple helical peptides.

Both molecular dynamics simulations and analyses of NMR data have pointed to significant amounts of bending and twisting motions existing in the triple helical peptide, and we compare collagen-like peptides to other deformable rod-like molecules such as DNA duplexes. A Weakly Bending Rod (WBR) model is introduced to mimic the elastic properties of these flexible rod-like CMPs and provide insight into the interpretation of the relaxation data and order parameter profiles.

ACKNOWLEDGEMENT AND DEDICATION

First and foremost, I would like to thank my advisor Dr. David A. Case for all of the mentorship and advice he provided me in the past six years. He responded to my questions quickly and was practical in his answers. I will always remember that he went through every single slide for my qualifier exams and PhD defense for hours. I am grateful for his patience and attention to my long and rambling presentations with my limited English ability. I truly thank him for his insight, moral support and for seeking out opportunities to make me a well-rounded as a scientist with complete focus. I would also like to thank Dr. Jean Baum, who is ready with helpful comments and criticisms. Their enthusiasm and devotion to science encouraged me to face and overcome many difficulties in my research.

I want to acknowledge the members of Dr. Case's group for their guide, suggestion and support over the years. I want to thank Dr. Jianxi Xiao for starting initial project of this thesis with me. I want to thank Dr. Sishi Tang, Dr. Tyler, Dr. Juncho Xia and Pawel Janowski for sharing their knowledge, experimence and friendship.

I would like to thank Gautam Singh for staying with me through all difficult times in the process of pursuing PhD degree. I truly thank him for all the understanding, patience, love, kindness, caring and moral support.

Finally, I would like to thank my parents and my whole family, to whom I also dedicate this thesis, for loving me, supporting me, and believing in me.

Contents

List of Tables

List of Figures

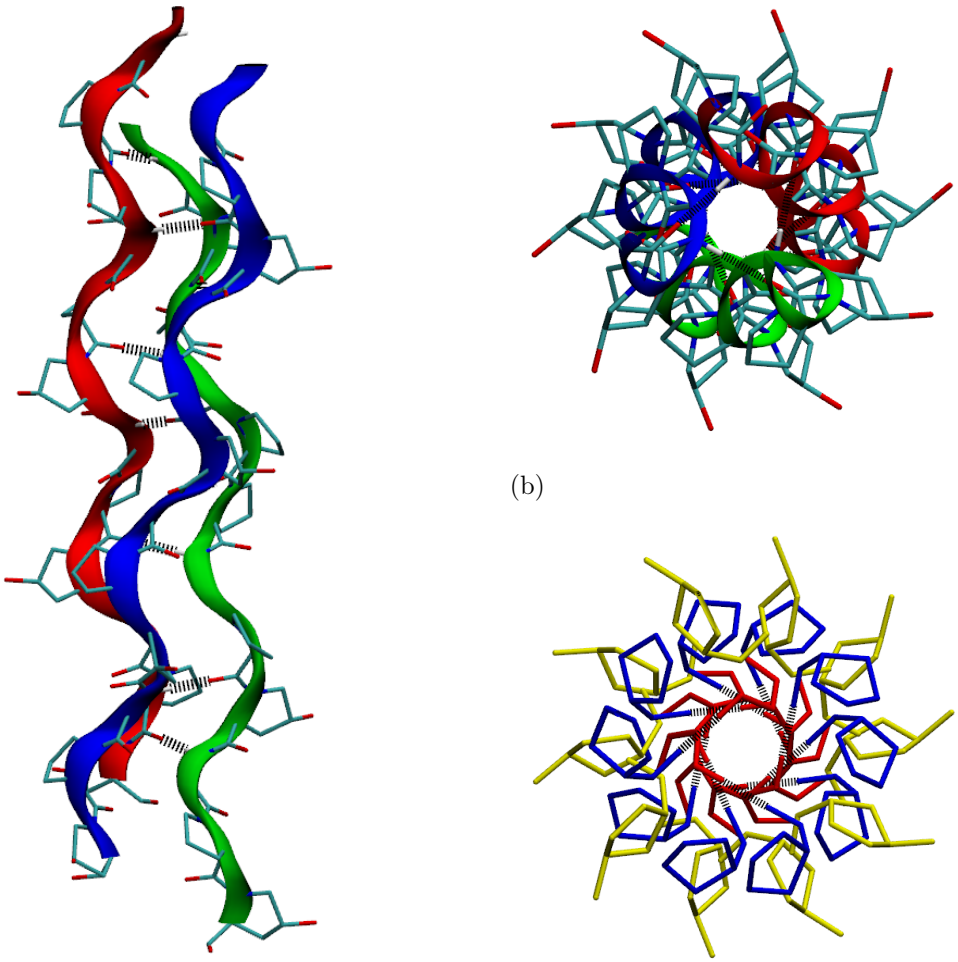
1 Introduction

1.1 Features of Collagen

Collagen, the most abundant protein found in vertebrates, is the main component of bone, skin, tendon, ligament, cornea, and other connective tissues. In humans, collagen makes up one-third of the total protein mass and is the most prevalent component of the extracellular collagen matrix. The defining structural feature of collagen is its right-handed superhelix, so-called the collagen triple helix, consisting of three parallel left-handed polyproline-like II helices, called α chain, with one-residue staggering (Figure 1.1 (a)). Some collagen molecules contain three identical α chains, whereas some other types of collagen have two or three different chains. In fibrillar collagens, the presence of glycine (Gly, G) in every third position is required because the close packing of triple helix only allows this smallest residue to fit into the interior of the helix without distortion (Figure 1.1 (b) and (c)). Thus the primary sequence of collagen content is characterized by the repeats of the Gly-X-Y triplet, where X and Y can be any amino acid but are often occupied by proline (Pro, P) and hydroxyproline (Hyp, O) at the X and Y position, respectively. Because both Pro and Hyp residues are rigid, their limited rotation of the polypeptide prevents the structure from adopting other conformations but the polyproline- II-like (PPII-like) structure. To be able to form close packing of the triple helix, the rings of Pro and Hyp must be pointing outward. Because of this, they are solvent-accessible and believed to play important roles in interactions with other molecules, such as the extracellular matrix protein. Three helices can be further stabilized by inter-chain hydrogen bonds (Figure 1.1 (c)). The individual triple helices, at approximately 300 nm long and 1.5 nm in diameter for each, are organized into fibrils which have high tensile strength and flexibility. These fibrils can be further assembled and cross-linked to support stress efficiently in tissues [5–7].

1.2 Collagen Types, Associated Diseases, and Molecular Recognition

To date, 28 different types of collagens have been identified [8]. The most common types are Type I, II, III belonging to fibrillar collagen, and type IV to non-fibrillar types and basement membrane collagen. For example, type I collagen, the major structural protein of bone, tendon, skin and cornea, is a heterotrimer consisting of two α1-chains and one α2-chain encoded by COL1A1 and COL1A2, respectively. This type has approximately 1000 amino acid residues with length of about 300 nm. Type II collagen, a homotrimer,



(a) (b) (c)

Figure 1.1: *Ball and stick diagrams of the triple helical structure in two projections: (a) side view showing the three helices wrapping around each other with one-residue staggering , and (b) and (c) look down the helical axis showing the cross section of the triple helical structure for fiber collagen model with one inter-chain hydrogen bond per triplet. The sequence shown is (Gly-Pro-Hyp)$_4$ and the backbone of each chain is colored with ribbon drawn. Inter-chain hydrogen bonds are represented as black dashed lines. In (a) and (b), the various atoms are colored as C-cyan, N-blue, O-red, and W(water)-yellow. In (c), The Gly, Pro and Hyp residues are colored in red, blue and yellow, respectively. To form the close packing of the triple helix, the Gly residues are closely packed in the center of the triple helix, while the rings of Pro and Hyp are pointing outward.*

makes up 50% of all cartilage collagen fibrils. Type III collagen, composed of three α1-chains, is the second most abundant collagen in human tissues after type I and is primarily found in tissues exhibiting elastic properties, such as skin, lung, and blood vessel walls. All the fibrillar collagens maintain a precise Gly-X-Y repeat sequence throughout their 1000 residues. Defects in the fibrillar collagen regular repeating sequences have been related to different diseases associated with connective tissues [8–12]. For example, Osteogenesis Imperfecta (OI), characterized by fragile bones, is caused by a mutation in Type I collagen. Osteoarthritis diseases, known as skeletal disorder, are caused by a mutation in type II collagen. Ehlers-Danos Syndrome type IV (EDS IV) affects blood vessels, joints, and internal organs as a result of mutations in type III collagen. In contrast to fibrillar collagens, non-fibrillar collagens do not obey the requirement of the presence of Gly as every 3rd residue. However, the replacement of one Gly residue in the Gly-X-Y repeat still can lead to a clinical phenotype. For example, Alport syndrome results in progressive kidney disease due to mutations in type IV collagen. An understanding of the basis for collagen stability can lead to new biomaterials and effective therapies for these connective tissue diseases.

The most common mutations in the collagen are single base mutation that replaces a Gly residue with larger residues causing structural distortion, which might prevent or delay the folding of triple helices as well as decrease the stability of the collagen molecules to different extent. It has been determined that mutation of a glycine to alanine is 19% lethal, serine is 30% lethal, cysteine is 39% lethal, and Aspartic acid is 68% lethal [13]. The existence of a connection between certain collagen mutations and the diseases described above has been established. For example, the most severe form of EDS IV is correlated with point mutations that substitute a Gly residue near the C-terminal end of the triple helix, including G1012R, G1018V, and G1021E [14]. However, the mechanism by which mutations types cause the human phenotype is still poorly understood.

As the major structural protein in extracellular matrix, collagen also plays an important role in cellular processes at many aspects, e.g. cell development, cell adhesion, cell migration and wound healing, through interactions with other matrix molecules [15]. Collagen can specifically bind to other molecules such as integrins or matrix metalloproteinases (MMPs). Despite the relative uniform rod-like conformation and the repeating Gly-X-Y sequence [16], collagen undergoes conformational change during the binding process. The best example of the collagen binding is the triple helical peptide $(GPO)_2GFOGER\ (GPO)_3$ complexed with the I-domain of integrin $\alpha 2\beta 1$ with bended backbone, as evidenced by the comparison of the structure of this peptide which has been solved in both isolated state and complex with integrin [17, 18].

The recognition and cleavage of collagen by MMPs are crucial in the process of collagen degradation, which has been implicated in a variety of different disease processes. Stability

and geometry of the triple helix makes collagen resistant to most proteases. Accelerated breakdown of collagen may result in many diseases such as arthritis and tumor cell invasion. Type III collagen and its cleavage are an important step in the pathogenesis of atherosclerotic plaque rapture, acute myocardial infarction and death [15, 19]. These MMPs can bind human type III collagen at unique cleavage sites characterized by a Gly-Leu or Gly-Ile bond which is preceded by the region rich in imino acid residues Pro/Hyp at X/Y positions and is followed by an imino acid-poor region [19, 20]. Although there are numerous potential cleavage sites presented in the collagens, they are not usually cleaved by collagenases, indicating that this selective recognition must be related to the local conformation of the extended sequence, particularly, the imino-acid deficient section on the C-terminal side of the cleavage sites. The underlying structural basis for this selective recognition in the uniform rod-like triple helical collagen is not well understood.

1.3 Conformational and Dynamical Studies on Collagen Model Peptides

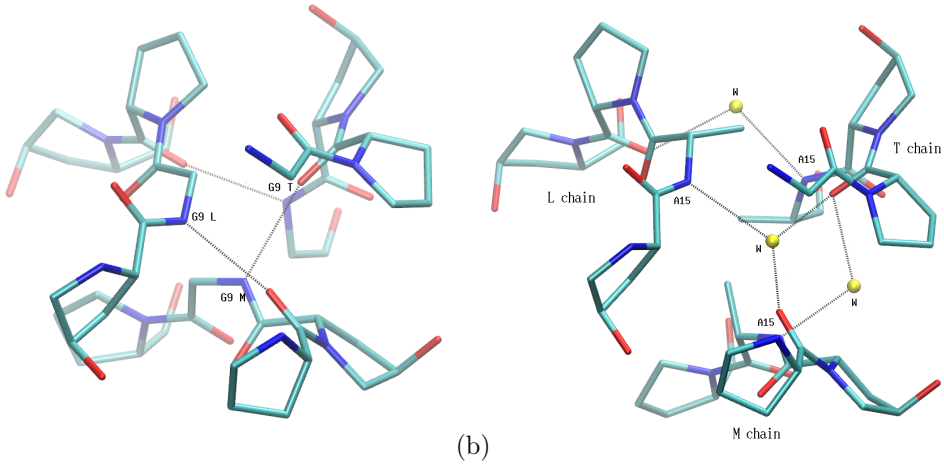
Due to large size, insolubility, repetitive sequence, and complex hierarchical structures of native collagens, it has been difficult to perform the atomic-level analyses on the full-length collagen to interpret the structural basis in stability and folding. This inability to crystallize full-length collagen and its unsuitability for multidimensional NMR techniques have been compensated by the studies on the collagen model peptides. Collagens are usually modeled by short triple helical peptides, which satisfy $(\text{Gly-X-Y})_n$ trimer repeats and high content of imino acids, forming the stable rod-like triple helices in solution. They are helpful to understand the factors that stabilize the triple helix and the folding of triple helix as well as its destruction and reformation upon sequence. In 1973, Prockop and coworkers first synthesized the small mimics of the collagen triple helix [21], revealing that thermal stability of a triple helix is enhanced by Hyp residues. In 1994, Berman, Brodsky, and coworkers revealed the first high-resolution structure of a triple-helical collagen model peptide (CMP) with a Gly mutation into Ala using X-ray diffraction analysis [22]. The sequence of each strand in this model peptide is $(\text{Pro-Hyp-Gly})_4$-Pro-Hyp-Ala-$(\text{Pro-Hyp-Gly})_5$ denoted as Gly→Ala. Subsequently, high-resolution crystal structures of CMPs have been determined including several $(\text{Gly-Pro-Hyp})_n$ and $(\text{Gly-Pro-Pro})_n$ peptide structures as well as biologically relevant $(\text{Pro-Hyp-Gly})_3$-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-Pro-Hyp-Gly-$(\text{Pro-Hyp-Gly})_3$ denoted as T3-785 peptide, the charged peptide $(\text{Pro-Hyp-Gly})_4$-Glu-Lys-Gly-$(\text{Pro-Hyp-Gly})_5$ [22–29]. Many studies, e.g. x-ray diffraction, CD spectroscopy, and NMR spectroscopy, have used these designed CMPs as a guide to gain a better insight into the forces which are responsible for the conformational stability of the triple helix as well

as the mechanism by which enzymes and other molecules recognize it [7, 16, 22, 24, 30–39].

1.3.1 Glycine Mutations

There are extensive studies of the CMPs containing a replacement of Gly with other amino acid [7, 13, 40–51]. These lead to the conclusion that the destabilization caused by these Gly mutations depends on the identity and location of the mutation as well as the degree of flexibility in the surrounding peptide [37]. It has been shown that a substantial loss in triple-helix stability with a degree of destabilization depends on the identity of the residue replacing Gly: Ala $<$ Ser $<$ Cys $<$ Arg $<$ Val$<$Gly,Asp$<$ Trp [40]. Although Gly mutated into Ala (the simplest side chain next to Gly) residue peptide is relatively highly under-represented in OI disease, the Gly^{928} to Ala substitution in the $\alpha 1$ chain of human type I collagen does causes lethal OI [52]. These mutation sites must have some unusual features due to local sequence environment or location at a biological important site.

The Gly to Ala mutation, introduced first into the peptide $(\text{Pro-Hyp-Gly})_{10}$ and still forming a very stable triple helix, has been shown to exhibit a reduced thermostability ($T_m = 29^0C$ melting temperature of CMPs used in NMR titration) and a slower overall folding rate relative to $(\text{Pro-Hyp-Gly})_{10}$ ($T_m = 60^0C$) [49]. Bella *et al.* [22] confirmed that when introducing glycine into alanine, the peptide adopts a triple helical structure with a slight untwisting at the substitution sites and the direct inter-chain hydrogen bonds are replaced with water bridges between the backbone groups (Figure 1.2).



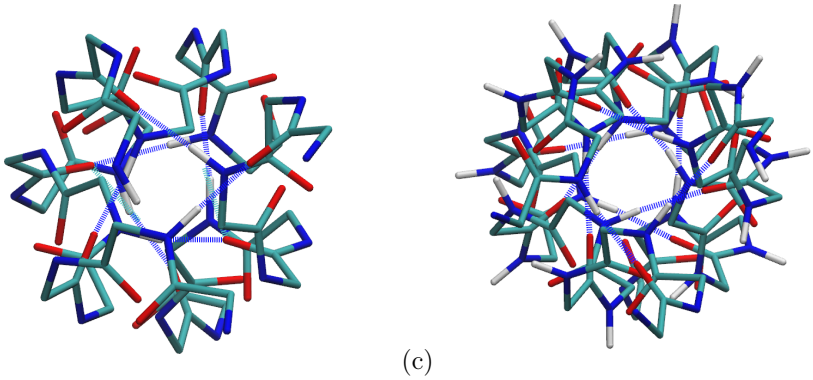
(a) (b)

Figure 1.2: *Top view down of the representative triple helical conformation of the X-ray structure of mutation peptide with PDB 1CAG from N terminus, showing details of (a) the direct inter-chain hydrogen bond between amide group of Gly and carbonyl oxygen of residue in X position in the neighboring chain in the regular collagen region with Gly-Pro-Hyp sequence and (b) the water bridge hydrogen bonds connecting amide group of Ala (A15) and carbonyl oxygen of residue in X position, in the substitution region with Ala-Pro-Hyp sequence. The various atoms are colored as C-cyan, N-blue, O-red, and W(as water molecule)-yellow. Hydrogen is not shown. The chain in leading, middle, and trailing are denoted as L, M, and T respectively.*

1.3.2 Helical Conformation

High resolution structures of the CMPs have also provided new insight into the sequence-dependent helical parameters [32, 35, 36, 45, 53, 54]. Almost all CMPs adopt a 7/2 helical symmetry (tightly packed triple helical segment) due to the geometrical restriction of the proline ring, whereas the peptide containing non-imino acid residues adopts an alternative conformation (relative loose structure than a 7/2) close to 10/3 super-helical properties, i.e. the integrin-binding protein complexed with integrin and the T3-785 peptide (Figure 1.3) [16, 28, 35, 53, 55]. It has been proposed that this looser conformation of an imino-acid-deficient region may unwind more easily and afford the extra contacts by polar or charges residues, which likely plays an important role in fibrillar assembly and interactions with other extracellular components, e.g. MMPs. With X-ray analysis, however, only a static picture of the conformation differences between a tight triple-helical structure and a loose conformational state around the cleavage state and not the full dynamics picture of the flexibility change upon the loose/tight conformation state.

(a)



(b) (c)

Figure 1.3: *(a) Schematic representation of amino acid composition of the T3-785 (PDB 1BKV). Top view of two sticks representative triple helical conformation of the X-ray structure of the peptide with N terminus in the front, showing details helical conformation. (b) Residues labeled as green in (a) located at the C-terminal zone matches the 7-fold idealized model. (c) The central zone with sequences shown as red bold in (a) adopts the helix close to the 10-fold idealized model. In (b) and (c), the various atoms are colored as C-cyan, N-blue, O-red, and inter-chain hydrogen bonds-dashed blue. Side chains are not shown in the plot.*

1.3.3 Dynamics

The biological function of collagen is most likely determined not only by its sequence-dependent characteristics and static conformation but also by its dynamics since some small number of specific residues at some point must be accessible to the macromolecules in order to provide sequence-specific recognition and high affinity binding [20, 56, 57]. The dynamics of collagen-like peptides have been studied experimentally for over 40 years [28, 58–64]. It has been suggested that regions with a relatively low imino acid content in collagen may be important for recognition, due to the difference in triple helical conformation and local molecular mobility from the imino-acid rich (POG) regions. Lazarev and colleagues [56, 57] dealt with collagens with different amino acid composition, and very different rigidity and stability using infrared spectroscopy and hydrogen-exchange method. They have found that the low-frequency large-amplitude backbone dynamics belonging to cooperative interactions, depends directly on the protein crystallinity and stability, while the vibrational high-frequency dynamics on the faster time scale in the collagens remains very close to each other.

1.4 NMR Spectroscopy

CMPs are well suited for NMR studies because of the high solubility of both the monomer and trimer forms. NMR studies on the Gly→Ala sequence has been reported in the literature [65] by Liu et al. In their studies, ^{15}N labeling of the Ala15 and Gly24 (as an internal reference) in Gly→Ala is compared with Gly15 (as an external reference) of $(POG)_{10}$ to elucidate how a mutation affects the structure, dynamics and hydration of the triple helix in solution. They have reported that *(i)* the hydrogen exchange rates for three A15 are different from one another (^{1}A15=4.7, ^{2}A15=12.1, and ^{3}A15=11.9 $10^{-4}min^{-1}$) and approximately 2 orders faster than the G24 ($< 0.1\ 10^{-4}min^{-1}$) as control within the same peptide. The Gly15 ($<0.04\ 10^{-4}min^{-1}$) in the $(POG)_{10}$ is 10 times faster slower than G24 in the Gly→Ala peptide, which may be consistent with the lower overall stability of the peptide relative to $(POG)_{10}$; (*ii*) one of three A15 has a chemical shift temperature gradient that is more positive than other two, suggesting the formation of a hydrogen bond in one chain, whereas the other two A15 residues are involved in either weak hydrogen bonding or exposed to solvent; (*iii*) the HSQC NMR data on the Gly→Ala peptide showed three well-resolved and distinct trimer peaks for three Ala15 downfield shifted in the proton dimension relative to monomer A15. It was suggested that these three Ala residues are in different environment because of one-residue staggering of the chains, while three G24 in the $(Pro\text{-}Hyp\text{-}Gly)_n$ repeats zone are in a similar environment as evidenced by only one trimer peak; (iv) furthermore, one of the three A15 is shifted down field (∼9.5 ppm) relative to other two (∼ 8.9 and 9.2 ppm, respectively). Such a significant difference in chemical shift caused by their relatively small differences between their chemical environment somehow is not expected [2, 51, 64, 66]. Due to lack of stagger information from the labeling of the isolated Ala residues, chain assignment could not be obtained, therefore, which A15 among these three residues has such unusual ^{1}H chemical shift is still not clear.

NMR experiments, a powerful tool for studying molecular motions at a wide range of time scales, have been applied to CMPs for many years. Fan *et al.* have investigated the backbone dynamics of the T3-785 and $(POG)_{10}$ peptides using hydrogen-exchange and NMR measurements of ^{15}N relaxation [64]. Hydrogen exchange studies have shown that the Gly15 NH and Leu16 NH in triplet Gly15-Leu16-Ala17 in the T3-785 exchange very slowly and slowly (average rate for three residues are 0.5 for Gly15 and 0.8 for Leu16), respectively, and the Ala 17 NH (∼ 9.6) exchanged very rapidly which is expected for a residues in the Y position of the $(Gly\text{-}X\text{-}Y)_n$ sequence. In contrast to hydrogen-exchange data, the order parameters S^2 derived from the ^{15}N relaxation data are similar for the central Gly15 residue in both the imino-acid poor (in T3-785) and rich region (in $(POG)_{10}$). Since ^{15}N relaxation occurs on a much faster time scale than hydrogen exchange, these results suggest that a Gly

in the imino-acid rich environment is dynamically different than one in the imino-acid poor environment in terms of the slower motion rather than the faster motion (ps).

1.5 MD Simulations

CMPs allow for accurate and tractable simulation because of their small size and regular structure. Molecular dynamics simulations have been extensively employed to study the factors that destabilize/stabilize the collagen triple helix and the folding upon the amino acid sequences along the collagen. The use of MD simulations to interpret NMR data is an alternative to standard analytical models [67–70], since they provide a high-resolution view of the motions of biological macromolecules and produce continuous trajectories with the potential to connect static structural snapshots, and give very detailed (if imperfect) description of protein structures and dynamics in solution.

Klein *et al.* showed that the destabilization caused by introducing a Ala into the central Gly position of the short homotrimer collagen peptide are contributed by unfavorable steric and electrostatic interactions [71, 72]. Their simulations observed the breaks of inter-chain hydrogen bonds and the distortion of PPII-like helical conformation near the substitution sites. Continuing their modeling of CMPs, Mooney *et al.* [43] looked into the structural consequences of the OI-associated mutations in collagen and observed a decrease in inter-chain hydrogen bonds and an increase in backbone-bound water molecules but the total backbone hydrogen bonds remain. However, an MD study of the nonequivalent environment of the three chains due to mutations as evident from hydrogen exchange and chemical shift analysis seen in solution NMR has not been investigated.

The dynamics of collagen model peptides (CMPs) has been studied experimentally for over 40 years [28, 58–64]. Until now, there are only a few examples of MD studies of the interpretation of ^{15}N relaxation data for global proteins with intermediate anisotropy [67, 73–79]. Collagen peptide with very large anisotropy have not been explored yet. Furthermore, the triple-helical model peptide is not a rigid rod in solution: both MD simulations and analyses of NMR relaxation point to significant amounts of bending and twisting motion. These aspects will be explored in details in the following chapters.

1.6 Overview of the Dissertation

The use of MD simulations is an alternative to standard analytical models to assist the sequence-specific assignment and interpret the NMR data for the triple helical collagen peptide. This thesis concentrates on characterizing the structural and dynamical properties of triple helical peptides including $(POG)_4(POA)(POG)_5$ (denoted as Gly→Ala) and the

T3-785 peptide, which have been investigated extensively by NMR techniques. Other two model peptides: $(POG)_{10}$, and $(PPG)_{10}$ are included as external references.

Chapter 2 focuses on the effect of amino acid sequence environment by a glycine mutation site leading to collagen disease and interpretation of NMR data on the mutation peptides. We specifically look into the formation and dynamics of hydration-induced hydrogen bonds around the mutation sites and draw the connection between the peaks of Ala residues in HSQC spectrum and the type of the Ala backbone hydrogen bonds.

Chapter 3 explores the types of motions, including the global tumbling and internal motions as well as the correlated motions, existing in the triple helical peptide in solution. We describe explicit solvent MD simulations for four collagen-like peptides, including a 0.5 μs long simulation of the T3-785 peptide, and 100 ns simulations of other three model peptides: Gly→Ala , $(POG)_{10}$, and $(PPG)_{10}$ to explore potential sequence-dependent motions. We evaluate overall tumbling of the triple helix with high anisotropy and the internal motions monitored by NMR. The back calculation of relaxation parameters becomes possible, thus, MD simulation facilitates the tests of the assumptions underlying the model-free approach. The triple-helical collagen-like peptide is not a rigid rod in solution: both molecular dynamics simulations and analyses of NMR relaxation point to significant amounts of bending and twisting motion [64, 80]. Other motional model instead of model free approach is proposed to study NMR relaxation analysis.

In **Chapter 4,** we suggest a motional model that should correctly deal with both high anisotropies and (weakly) flexible geometries. We describe the weakly bending rod (WBR) model and its application to the interpretation of NMR relaxation data. Excellent fits are possible, suggesting that global bending and twisting modes dominate the deviations from rigid-rod behavior.

2 Structure and dynamics in disease-related glycine mutations

Abstract. The results of molecular dynamics simulations of the collagen-like peptide, $(POG)_4$-(POA)-$(POG)_5$ (referred to as the Gly→Ala peptide hereafter) and $(POG)_{10}$ are presented. To characterize the structure and dynamics of collagen containing mutations associated with *osteogenesis imperfecta*, we examined rmsd fluctuations, backbone bending, predominant triple helical structures, hydrogen bonds and hydration pattern. Particularly, we look into the formation and dynamics of hydration-induced hydrogen bonds around the mutation sites. Our results reveal that *(i)* a Gly mutation causes unwinding of the triple helix in the mutated zone, imposing more structural and dynamical changes towards the C-terminal end of the triple helix; *(ii)* a disruption of inter-chain hydrogen bond is observed at the Gly18, which is one triplet C-terminal to the mutation site but not at the Gly12, which is one triplet N-terminal to the mutation sites*; (iii)* at the mutation sites, the three Ala residues in each chain behave differently from one another in terms of inter-chain hydrogen bonding and PPII conformation. These changes provide insight into the observed NMR chemical shift and hydrogen-exchange rates in the triple helix. A comparison between calculated and available experimental findings is presented. The structural ensemble at the mutation site, as inferred from the simulation and NMR data, is different from that seen in a crystal structure of the same peptide. The greatest difference is, in the solution simulation, there is no evidence of penetration of water inside the triple helix. In contrast to the study, the penetration of water inside the triple helix is observed in the crystal simulation of the Gly→Ala. Our computational results are in agreement with NMR experimental data found in the literature and the data found in the crystal, indicating the MD simulations can provide a realistic description of the collagen in different environments.

2.1 Introduction

Collagen, the most abundant protein found in the vertebrates, is the main component of a number of tissues, such as tendons, ligaments, skin, cornea, and bone. In humans, collagen comprises one-third of the total proteins and is the most prevalent component of the extracellular matrix [7, 22, 23, 34, 38, 53, 81]. The triple helix is a motif found in fibril-forming collagen and is identified as the ligand-binding regions of the other receptors. Usually, triple helices are rod-like domains, but kinks or other perturbations have been reported at the regions where the Gly-X-Y repeating triplet is interrupted [22, 28, 40, 43, 48, 82, 83]. Defects in the triple helix domain of collagen have been associated with a number of human collagen diseases [8–12]. *Osteogenesis imperfecta* (OI), often called brittle bone disease, results from mutations in type I collagen, the major collagen in bone, tendon

and skin. The most frequent mutations in type I collagen are single base substitutions that convert the required Gly residue to a larger residue in the collagen triple-helix. Such mutations in collagen prevent or delay the folding of the triple helix beyond the mutation point or produce a disruption in triple helical conformation or inter-chain hydrogen bond pattern [8, 22, 40]. Because the triple helix of most collagen is propagated from the C terminus to the N terminus, a Gly mutation nearer to the C-terminal end of the triple-helical domain may produce a more severe phenotype than a similar mutation nearer to the N-terminal end [42]. The correlation between OI severity and mutation location is still unclear.

Studies of the replacement of Gly with other amino acids have been done using CD, NMR and MD simulations [7, 13, 40–51]. These studies all pointed to a conclusion that a degree of destabilization caused by a Gly mutation depends on the identity and location of the mutant as well as the degree of flexibility in the surrounding [37]. For example, a substantial loss in triple-helical stability with a degree of destabilization depends on the identity of the residue replacing Gly: Ala < Ser < Cys < Arg < Val < Gly, Asp < Trp [40]. Although Gly mutated into Ala peptide is relatively highly under-represented in OI disease, the Gly^{928} to Ala substitution in the $\alpha 1$ chain of human type I collagen does causes lethal OI [52], indicating some unusual features due to local sequence environment or location at a biological important site.

A Gly mutated to Ala was introduced first into the peptide $(POG)_{10}$, referred as Gly→Ala, which has been shown to form a very stable triple helix but with a significant decrease in thermal stability ($T_m = 29^oC$ melting temperature of mini-collagen used in NMR titration) relative to $(POG)_{10}$ ($T_m = 60^oC$) [49]. Bella *et al.* [22] first confirmed that this peptide in crystal adopts a triple helical structure with an untwisting and the local disruption of direct inter-chain hydrogen bonds, which are replaced with water bridges between the backbone groups at the substitution sites.

However, solution NMR data showed some observations on the same peptide in solution deviated from crystal environment [65]. Liu *et al.* presented the solution NMR studies on the Gly→Ala peptide in which ^{15}N labeling of the A15 (referred as Ala residue at the position 15, which is the middle of the chain) and G24 (Gly residue at the position 24, which is treated as an internal reference) are compared with the G15 (as an external reference) of $(POG)_{10}$. They reported that (*i*) the hydrogen exchange rate for G15 ($<0.04\ 10^{-4} min^{-1}$) in the $(POG)_{10}$ is 10 times slower than G24 in the Gly→Ala peptide, an observation consistent with the lower overall stability of a Gly→Ala peptide relative to $(POG)_{10}$; (*ii*) the hydrogen exchange rates for three A15 are different from one another (^{1}A15=4.7,^{2}A15=12.1, and ^{3}A15=$11.9\ 10^{-4} min^{-1}$) and approximately 2 orders faster than the G24 ($< 0.1\ 10^{-4} min^{-1}$) within the same peptide; (*iii*) one of three A15 has a chemical shift temperature gradient

that is more positive than other two, suggesting the primary formation of a hydrogen bond in one chain, whereas the other two A15 residues are involved in either weak hydrogen bonding or exposed to solvent; *(iv)* the HSQC NMR data on the Gly→Ala peptide showed three well-resolved and distinct trimer peaks for three A15 residues downfield shifted in the proton dimension relative to monomer A15. Such clear and distinct three peaks in the triple helix where the Ala residues in the leading and trailing have the identical chemical environments are not expected; (*v*) the HSQC spectrum further exhibits that one of the three A15 is shifted down field (∼9.5 ppm) relative to other two (∼ 8.9 and 9.2 ppm). Such a significant difference in chemical shift caused by their relatively small differences between their chemical environments is not clear [2, 51, 64, 66].

MD simulations have the ability to complement the experimental observations. Prior computational studies have investigated some factors, which might influence the triple helix conformation of the mutant peptides. Klein *et al.* showed that the destabilization caused by a substitution of Gly to Ala is contributed by unfavorable steric and electrostatic interactions [71, 72]. Their simulations observed the breaks of the inter-chain hydrogen bonds and a distortion of PPII-like helical conformation near the substitution sites. Continuing their modeling of CMPs, Mooney *et al.* [43] looked into the structural consequences of the OI-associated mutations in collagen. They observed a decrease in inter-chain hydrogen bonds and an increase in backbone-bound water molecules, suggesting that mutant peptides seemed to compensate for mutation-induced loss of stability with a majority of highly exchangeable nonspecific solvent backbone hydrogen bonds that were present less than 10% of the time. However, an MD study that discusses the nonequivalent environment of the chains as evidenced by hydrogen exchange and unusual downfield shift seen in solution NMR has not been investigated.

Several aspects suggested in the NMR studies on the Gly→Ala are being addressed in this work from point of view of MD simulation: *(i)* Are the fast hydrogen exchange rate for Ala residues caused by the absence of inter-chain hydrogen bonds? Or as a result of the dynamics of the bound water nearby? *(ii)* If so, is the nonequivalence in hydrogen exchange rates for three Ala residues as a result of different degrees of freedom of bound waters or different types of water hydrogen bonds? *(iii)* What causes three trimer peaks of Ala residues in the HSQC data where the one has unusual downfield shift of the resonances relative to the other? *(iv)* Can the three trimer peaks reflect the different type of interactions or the strength of interactions with waters or main-chain carbonyl groups? *(v)* It is known whether the Gly→Ala peptide shows a local "under-twist" at the site of the interruption and a slight "over-twist" at the two regions flanking the substitution in a static conformation. However, does this under-twist exhibit more backbone dynamics and greater access to other receptors? *(vi)* Can the asymmetric distortion be seen in the highly symmetric sequences on the both

sides of the mutation sites in terms of hydrogen bond pattern, PPII-like conformation, or internal triple helical parameters?

This chapter is organized as follows. Section 2.2 presents the structural changes upon mutation in terms of rmsd analysis, the global motions, and diameter of rod-like triple helix as well as PPII-like conformation. The local helical twist angle against triplet (cross three chains) are analyzed. In Section 2.3 we provide a qualitative molecular interpretation of the unusual trends that are observed in NMR data. We end with summary of main points and relate to what seen in the NMR data in Section 2.4, followed by the MD system setup and methods in Section 2.5.

2.2 Collagen-like Peptide Structural Stability and Dynamics

2.2.1 Structural Perturbations upon Mutation

The values of Root-Mean Square Deviation (RMSD) of the Cα atoms from the x-ray structures of Gly→Ala and $(POG)_{10}$ peptides are shown in Fig. 2.1. The $(POG)_{10}$ has an average value at 1.29 Å, confirming that the CMP retains the triple helical conformation (Fig. 2.1 (b). Although the rmsd of Gly→Ala peptide has a larger average value of 1.95Å, this mutated peptide still assumes a stable triple helical conformation throughout the solution simulation, indicating that the force field and simulation setup were able to resume the initial conformation in the crystal structure.

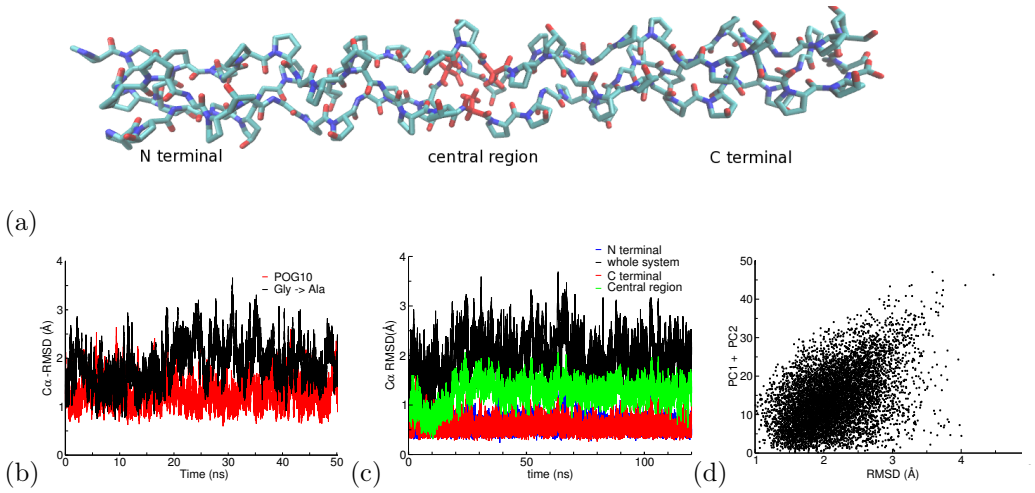


Figure 2.1: *The structure and dynamics of the mutated peptide Gly→Ala in solution simulation. (a) View of the crystal structure (PDB: 1CAG) as initial structure for our solution simulation. This model peptide is composed of three chains with sequence [$(GPO)_4APO(GPO)_5$]. The mutant sites (a glycine mutated into a alanine) were highlighted in red in the central region of the helix. (b) The root-mean square deviation (rmsd) of the Cα atoms from the X-ray structures of Gly→Ala (black) and $(POG)_{10}$ (red). (c) In Gly→Ala, the Cα atom RMSD of the trajectory structure superimposed to the x-ray structure for the whole peptide (black), the central (green) region as well as the C(red)- and N (blue)-terminal regions. (d) Correlation plot between the norm of the sum of the two first PCA (PC1 +PC2) eigenvectors and RMSD values for the Gly→Ala.*

In Gly→Ala, the larger value of rmsd seems to be dominated by the central regions containing the substitution site. In Fig. 2.1 (c), the RMSD values of the N and C zones are relatively low, indicating that both terminal regions maintain stability and remain close to the crystal structure. The central zone exhibits the largest deviation from the starting X-ray model, implying the central region, which reveals either global bending or local bulge fluctuations, dominates the dynamics of the triple helical peptide. The plot of the correlation between the RMSD values and the PCA eigenvalues of the bending motions (the details see the PCA discussion below) suggests that the structural deviations of the triple helical peptide in solution relative to the crystal structure are largely due to a low-frequency backbone bending (Fig. 2.1 (d)) .

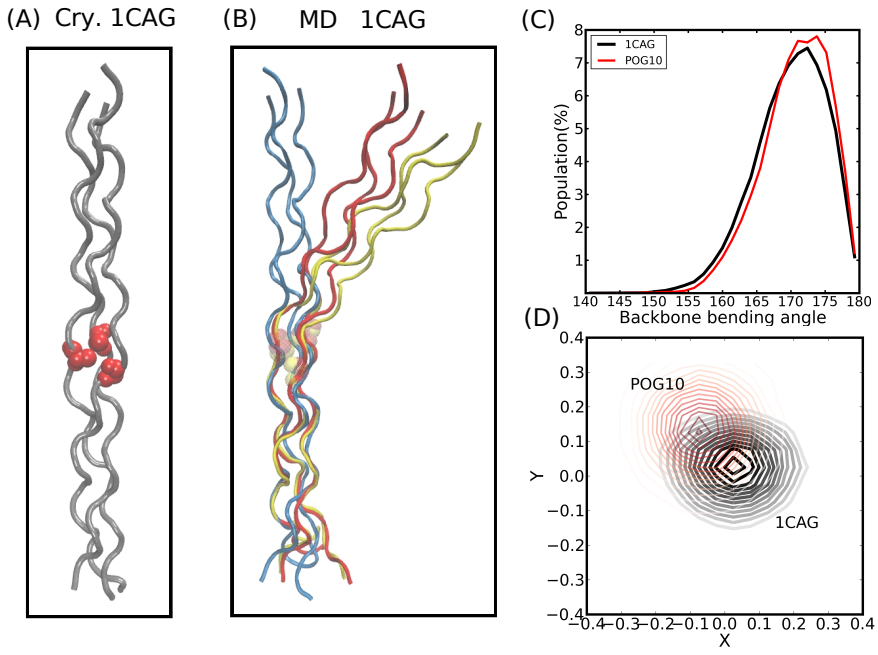


Figure 2.2: *Structure and dynamics of the Gly mutated peptide: Gly→Ala (A) Cartoon representation of crystal structure (PDB 1cag) with Ala residues shown as highlight red VDW balls. (B) Cartoon representation of few snapshots obtained from the 150 ns MD simulation of Gly→Ala in water. (C) Histogram of the backbone bending angle (in degree) of Gly→Ala and $(POG)_{10}$ of MD simulation. (D) Projection of principal axis of backbone of the C terminal region on the plane with normal vector obtained from the principal axis of N terminal zone. All the structures are superimposed with N terminal backbone atoms. In (C) and (D), Gly→Ala and $(POG)_{10}$ are colored as black and red lines, respectively.*

The high resolution crystal structure of Gly→Ala appears to be overall straight with standard triple helical structure at both ends and a localized conformational deformation around Ala residues. (Fig. 2.2 (A)). Our structural ensemble observed that the triple helical peptides are flexible molecules capable of changing from a linear to a curved structure (Fig. 2.2 (B) and (C)).

	%variance accounted		
Mode	POG10	Gly→Ala	description
1	33.1	33.1	global bending
2	28.2	23.4	global bending
3	8.9	11.5	helical twist
4	5.2	5.5	double bending
5	4.4	4.6	double bending
cumulative	79.7	78.2	

Table 2.1: *Percentage of variance and physical description of the five most important Principal Components obtained from diagonalization of the Cα covariance matrix for the triple helical peptide $(POG)_{10}$ and Gly→Ala trajectories*

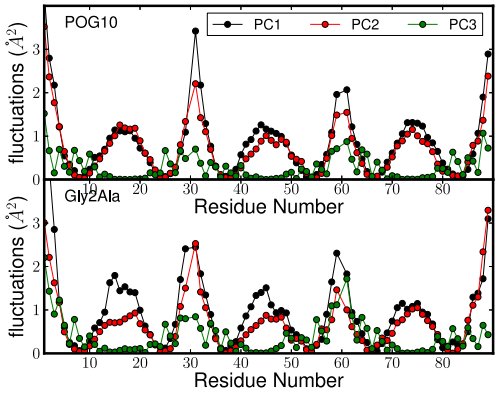


Figure 2.3: *The atomic fluctuation of essential motions of Cα atoms of POG10 and Gly→Ala projecting on the three most significant eigenvectors: PC1, PC2, and PC3. Residue number 1-29, 31-59, and 61-89 belong to leading, middle, and trailing chain, respectively.*

The flexibility of the triple helical peptides upon mutations is characterized by performing a principal component analyses (PCA) on the backbone Cα atoms of $(POG)_{10}$ and Gly→Ala (Table 2.1). The results of PCA on both peptides share some similarities including that (*i*) the first five PCAs account for up to ∼80% of the structural variance of the Cα atoms. Over half of this structural variability is mainly described by the PC1 and PC2 modes, which correspond to orthogonal global bending-like oscillatory modes of the backbone; (*ii*) the PC3 mode describes the twisting-like motion; (iii) the next higher modes represent the higher harmonics of motions. In overall, visualization of the projection of the PCA modes onto the MD trajectories shows that the essential dynamics of the two peptides is very similar (Fig. 2.3).

However, there are some slight differences induced by mutations in the projection of the PCs modes onto the MD trajectories. *First*, the atomic fluctuations of PC1 of each chain

are slightly different from one another, with the leading chain having the largest fluctuations while the lowest one revealing in the trailing chain. *Second*, although both peptides have similar distribution of backbone bending (Fig. 2.2 (C)), in Gly→Ala the total variance accounted for PC2 mode is significantly higher than one in PC1 mode particularly in the leading chain, suggesting the characteristics of the anisotropic backbone bending as seen in Fig. 2.2 (D). This asymmetric atomic fluctuation of each chain in essential motions may be the causality of the disruption and inequivalence of inter-chain hydrogen bonding at the substitution sties. It is suggested in the study [84] of DNA bending showing that DNA bends and becomes very flexible once a single base pair has been disrupted.

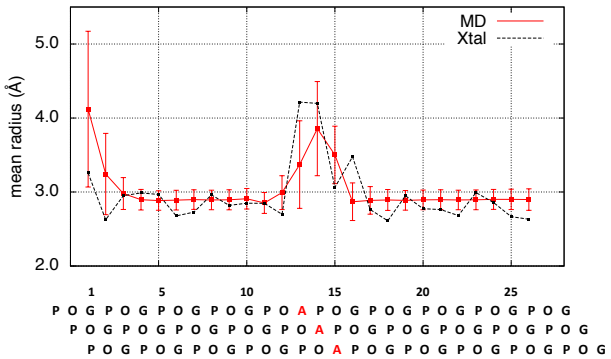


Figure 2.4: *Evolution of the average helical radius (as ●) at each Gly position to the center of triplet along the amino acid composition in the Gly→Ala of the simulation. Helical radius from x-ray crystal structure is marked as ●. Schematic representation of amino acid composition of the Gly→Ala is labeled in x-axis.*

The structural variations upon mutation are characterized by the helical radius mean and its deviation. The average helical radius measured at the Gly position ranges from 2.88 Å to 3.85 Å (Fig. 2.4), indicating that the triplet containing mutation sites exhibits the unwinding of triple helix. Interesting, this unwinding region shows a larger deviation around its mean value, suggesting increased flexibility due to the mutations.

2.2.2 Internal Helical Parameters

To investigate the interruption in the triple helix associated to the mutations, we computed the local helical conformations, which are characterized by the local step parameters: shift, slide, rise, roll, tilt, and twist between two adjacent triplets using 3DNA algorithm [85–87].

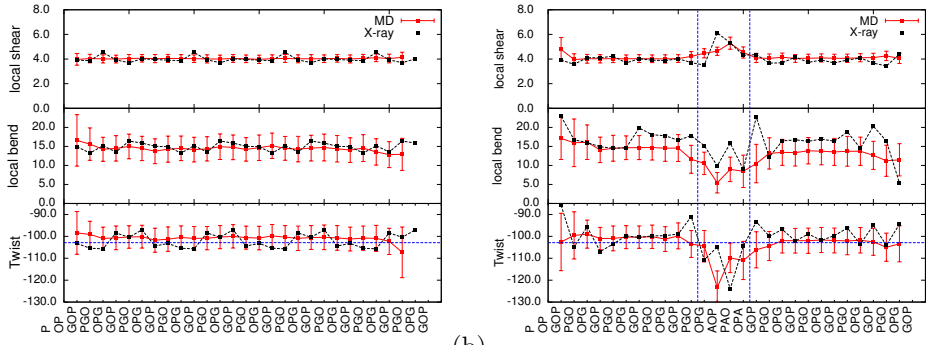


Figure 2.5: *Average step parameters: local shear ($\sqrt{Shift^2 + Slide^2}$), local bend ($\sqrt{Tilt + Roll^2}$), and twist are calculated for the MD trajectory of (a) $(POG)_{10}$ and (b) Gly→Ala peptides. Schematic representation of the triplet (cross all three chains) sequence along the peptide is labeled in x-axis.*

In Fig. 2.5, the local shear, local bend and the twist values of Gly→Ala and $(POG)_{10}$ along the triplet (cross all three chains) sequences are compared: the Gly→Ala peptide in solution adopts the "under-twisted" and "bulge of the triplets" located in the central region of the helix (triplets from 12 to 16), as evidenced by the twist values (ranging from θ = -103 ± 6.1 to -123 ± 7.6 in degrees) and local shear (ranging from d = 4.2 ± 0.4 to 5.2 ± 0.5 Å) larger than those in the simulation of $(POG)_{10}$ (average value of θ = -100.3 ± 5.3 for the triplets at the central region of the helix). These results confirm the deformation induced around the substitution sites and show the consistency with the studies of other groups [7, 32, 53]. Furthermore, the local step parameters between two triplets containing the Ala residue exhibit greater deviations than those in host regions within the same peptide or in $(POG)_{10}$, suggesting increased flexibility. Such deformation induced by the mutations may cause the loss of direct inter-chain hydrogen bonds, and hence offer less resistance to twist and local shear than other regular collagen zones.

2.2.3 Predominate Structure

The cluster analyses were conducted on both peptides to reveal the predominated structure in solution simulation. The representative structures of these five sets are shown in Fig. 2.6 and each set has at least 1.2 Å rmsd from the other. Table 2.2 points that the $(POG)_{10}$ peptide has high popularity in adopting an uniform straight triple helix in solution and the representative structures of set1, set2 and set3 seem to be close to each other in terms of the helical conformation and the backbone bending (Fig. 2.6(a)). In contrast to $(POG)_{10}$, the representative structure of each setin Gly→Ala peptide is easily distinguished from each

other by visualization, and each set has an rmsd value of 1.44 Å ∼3.63 Å from the other (Fig. 2.6(b)). The predominant structure (set1) in Gly→Ala has a slightly bent backbone and bulge region in the mutation zone. All the sets (except set 4 $<$ 0.4 %) have similar N-terminal triple helical conformations, in particular in terms of Rise and Twist. Whereas the helical structure in the C-terminal region has slightly more freedom to adopt different conformations or fluctuates more between over-twisted and under-twisted helix compared to the N terminal region (Data not shown). The under-twisted conformations in substitution site and the C-terminal regions in Gly→Ala show more fluctuations, which are reflected in the larger values in the standard deviations of the local step parameters, indicating, again the increased flexibility as well as asymmetric distortion between N and C termini.

This asymmetric distortion between N and C termini surrounding the mutation sites has been observed experimentally. Sagaya and coworkers [88] found that bacterial collagen-binding domain binds (CBD) to either the Gly to Ala substitution sites or the C-terminus of mini-collagens using 1H -15N HSQC-NMR. Bella et al. [32] revealed that the local helical conformation in the substitution site and the C-terminal toward the end region of the Gly→Ala crystal structure is under-twisted and the end N-terminus is over-twisted.

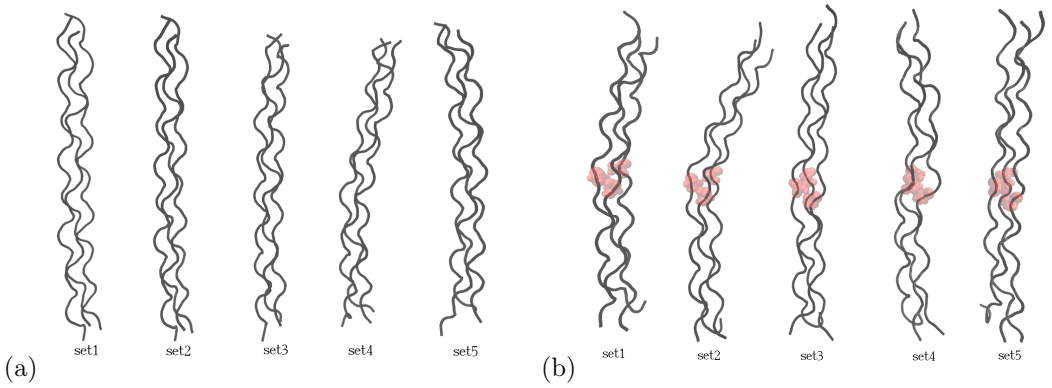


Figure 2.6: *Representative structures of five clusters obtained from cluster analysis on the (a) $(POG)_{10}$ and (b) Gly→Ala peptides.*

	set1	set2	set3	set4	set5
POG10	25.6	**32.6**	12.7	27.9	1.2
Gly-Ala	**70.5**	4.2	23.3	0.4	1.6

Table 2.2: *Calculated occupancy of collagen-like peptides in MD simulation using cluster analysis.*

2.2.4 Hydrogen Bond and Hydration around the Mutation Sites

2.2.4.1 Inter-chain Hydrogen Bonds

Formation of the direct inter-chain hydrogen bonds NH(Gly)$\cdots$ OC(X) between the amide proton of Gly residue in one chain and the carbonyl groups of the X position in the adjacent chain, is well-known for the stability of the CMPs. Bella *et al.* [22] confirmed that the direct inter-chain hydrogen bonds of all three Ala residues are replaced with water bridges between the backbone groups when introducing glycine into alanine. Fig. 2.7 shows that, in the substitution zone of the crystal Gly→Ala peptide, the amide groups of all three A15 residues and one ^{L}Gly18 (in the leading chain, L) lose the direct hydrogen bonding due to increased distance of carbonyl and amide groups, but gain the hydrogen bonds with interstitial waters bridging carbonyl groups of Pro. In contrast to crystal environment, the solution NMR studies on the same peptide showed that one of three A15 has a chemical shift temperature gradient that is more positive than other two, suggesting the formation of a hydrogen bond in one chain, whereas the other two A15 residues are involved in either weak hydrogen bonding or exposed to solvent [65].

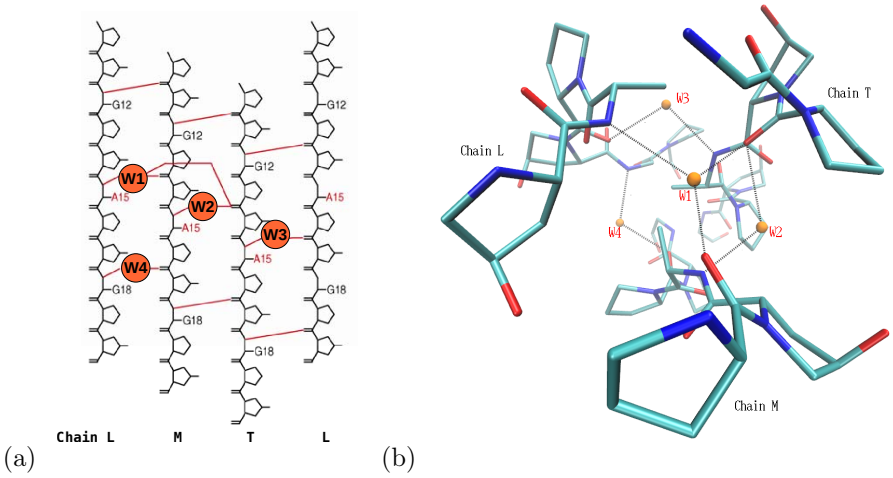


Figure 2.7: *X-ray crystal structure of Gly→Ala. (a) Schematic representation of hydrogen bonding topology around the substitution region. (b) Top view down of the same structure from N terminus. Water bridge hydrogen bonds (W1-W4) connecting amide group of Ala (A15) and carbonyl oxygen of residue in X position in the substitution region with Hyp-Ala-Pro-Hyp-Pro sequence each chain. The various atoms are colored as C-cyan, N-blue, O-red, and W1 to W4 (as bridge water)-orange. Hydrogen is not shown. The chain in leading, middle, and trailing are denoted as L, M, and T, respectively.*

In our MD simulation of the Gly→Ala, the results of the inter-chain hydrogen bond (the criteria shown in Eq. 2.3) analyses are depicted in Table 2.3. In the collagen zone, all the

H-bonds have an average value of 1.99 $\sim$ 2.01Å with maximum standard deviation of 0.17Å, and an average H-donor-acceptor angle less than 23.7 degree with a maximum standard deviation of 10.01 degrees, indicating all inter-chain hydrogen bonds are maintained well throughout the simulation. In the substitution sites, significant decreasing occupancies of inter-chain hydrogen bonds are found at ^{M}A15 and ^{T}A15 as expected, but the inter-chain hydrogen bond of ^{L}A15 still maintains up to 70% of the simulation time. The average lifetime of inter-chain hydrogen bonds decreases to 10 ps in the mutation zone from 30 ps in the collagen zone, implying the accordance with an increase in flexibility of backbone hydrogen bonds and suggesting these short lived interaction may facilitate the hydrogen exchange process for the A15 amide proton seen in NMR data.

MD	Occupied(%)			Average Lifetime(ps)			N-H$\cdots$O=C(Å)		
Chain	L	M	T	L	M	T	L	M	T
G12	91.7	90.5	87.3	26.2(29.2)	22.0(25)	20.0(28.4)	1.98	1.88	1.98
A15	72.0	16.9	30.6	24.4(28.6)	13.4(24)	9.2(11.8)	2.01	2.06	2.19
G18	65.6	74.2	90.3	24.6(36.0)	18.2(26)	24.0(27.0)	2.06	2.05	1.99
G24	92.2	92.1	92.0	27.0(30.2)	26.8(30)	26.8(29.2)	1.99	1.99	1.99

Table 2.3: *Main chain hydrogen bond (N-H$\cdots$O=C) parameters, occupancies, length as well as lifetime means and its standard deviation (given in parentheses), for G12, A15, G18, and G24 on chains L, M and T, respectively.*

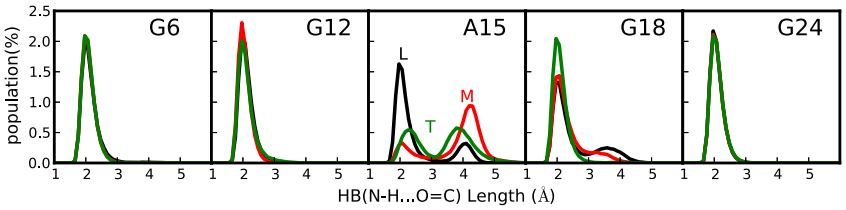


Figure 2.8: *Distribution of the distance of NH(Gly/Ala)-O=C(Pro) inter-chain hydrogen bonds of G6, G12, A15, G18, and G24 residues in the MD simulation of the Gly→Ala. Chain L, M and T are colored as green, red, and black, respectively.*

Now, we look into the distribution of the distance between the amide groups of G/A and the carbonyl groups of Pro at the X position in the adjacent chain from the MD simulation. Fig. 2.8 shows that all the Gly residues consistently form the inter-chain hydrogen bond with only one peak locating at the distance within 3.3 Å, but Ala residues reveal two distinct conformations, which correspond to the direct inter-chain forming ($<$ 3.3 Å) and breaking ($>$ 3.3 Å). It is obvious that three A15 have completely different behaviors in terms of this distribution: the ^{L}A15 has primarily inter-chain hydrogen bond, and ^{T}A15 has about half chance in forming direct hydrogen bond, whereas ^{M}A15 shows the least favor in forming the

inter-chain hydrogen bond. Such inequality seems to agree with the observations in HSQC spectrum and chemical shift temperature gradients data (which will be discussed in details in Sec. 2.3). Additionally, two states conformations of the inter-chain hydrogen bond are slightly observed at $^{\mathrm{L}}$G18 and $^{\mathrm{M}}$G18, which are a triplet C-terminal to the substitution sites. Such small percentage of disruption of inter-chain hydrogen bond in Gly18 is related to the study of folding pathway of collagen-like peptide in the experimental peptide studies [42, 51], in which the residues C-terminal to the substitution site show signs of increased mobility and unwinding.

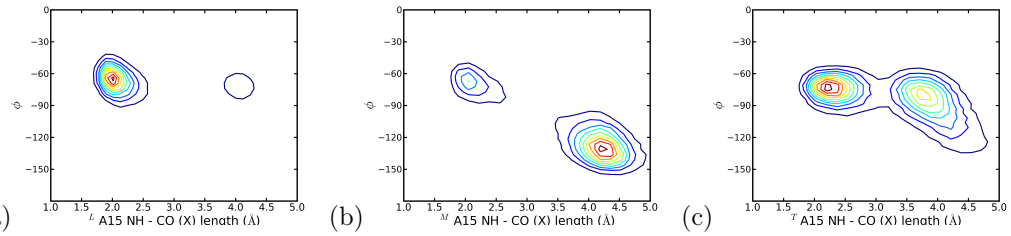


Figure 2.9: *Correlation between direct inter-chain hydrogen bond length and the backbone torsion angle ϕ of (a) ^{L}A15, (b) ^{M}A15 and (c) ^{T}A15 residues.*

Fig. 2.9 shows a strong correlation between the direct inter-chain hydrogen bond length and the ϕ angle for Ala residue, particularly for $^{\mathrm{M}}$Ala15. Similar correlation is seen in other two A15 residues, the main distribution of Φ angle still remains within the range of the PPII conformation. The presence of inter-chain hydrogen bond in $^{\mathrm{L}}$A15 reduces the diameter of the local helix which might cause more distortion of the PPII conformation for the $^{\mathrm{M}}$A15 than what observed in crystal structure (Table 2.6, discussed in Sec. 2.3).

2.2.4.2 Water Bridging Hydrogen Bonds

The existence of the interstitial waters and the dynamics of solvent/solute hydrogen bond are supported in Table 2.4 and Fig. 2.10, indicating that the $^{\mathrm{L}}$Ala15 maintains primarily stable direct hydrogen bonds similar to Gly in collagen zone (i.e. G24) while $^{\mathrm{M}}$Ala15 and $^{\mathrm{T}}$Ala15 exchange between two states: inter-chain hydrogen bonds and solvent/solute hydrogen bonds.

		A15 NH-Wat		G18 NH-Wat		
Chain	L	M	T	L	M	T
Occupied		73.5	57.9	15.0	10.9	
Distance	×	1.99(0.16)	2.03(0.17)	2.03(0.17)	2.0(0.16)	×
Angle		161(8.8)	162(10.1)	161.8(9.6)	161.2(9.5)	
Ave. Lifetime		61.6(42.4)	35.4(22.4)	31.6(17.8)	68(42.8)	

Table 2.4: *Water Hydrogen Bond parameters of A15 and G18. Standard deviations are given in parentheses. Hydrogen bond analysis on the 20ns trajectory. Distance cutoff 3.5 Å and the angle cutoff 135°. (X) denotes no hydrogen bonding with solvent.*

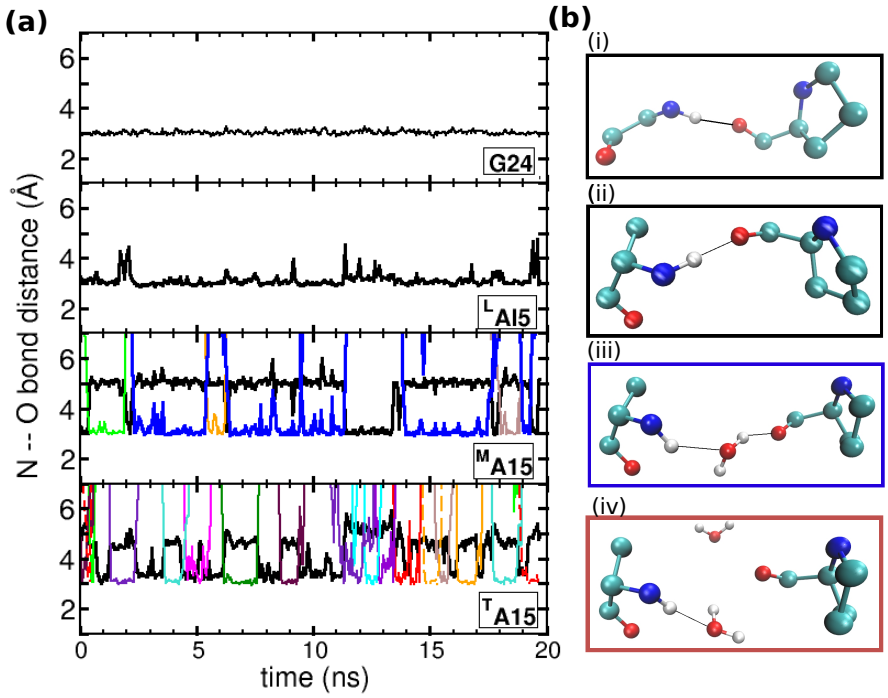


Figure 2.10: *The N-O length between the amide group of G24/A15 and the acceptor throughout the simulation of Gly→Ala peptide as well as types of hydrogen bonds of G24/A15 residues including the direct inter-chain NH(G/A)-OC(P) hydrogen bond, interstitial hydrogen bond NH(A)-Wat-OC(P), and the backbone-bounded water hydrogen bond NH(A)-Wat. The direct inter-chain hydrogen bond, N(G/A)–OC(P), between amide group of G24/A15 and the carbonyl group of Pro is labeled as black line, which is illustrated in (i/ii) black box. The interstitial water hydrogen bond, N(A)-W-OC(P), between amide group of A15 and the carbonyl group of Pro through a water molecule is colored with blue line and its representation is represented in (iii) blue box. Backbone-bound water hydrogen bond, N(A)–OH(W) is marked with other colors and its representation is depicted in (iv) red box.*

Distinct differences between ^{M}Ala15 and ^{T}Ala15 residues are observed when they are bound to water molecules. For example, interstitial water bridges are observed some time for ^{M}Ala15 residue, but not really seen in ^{T}Ala15. Fig. 2.11 indicates that water molecules generally exchange hydrogen bonds quickly around the amide group of ^{M}Ala15 and ^{T}Ala15, and ^{T}Ala15 has slightly more exchangeable nonspecific hydrogen bonds with water molecules than ^{M}Ala15. These results point to that the reorientation dynamics of amide protons and the mobility of water molecules at the substitution sites is fairly different for each chain.

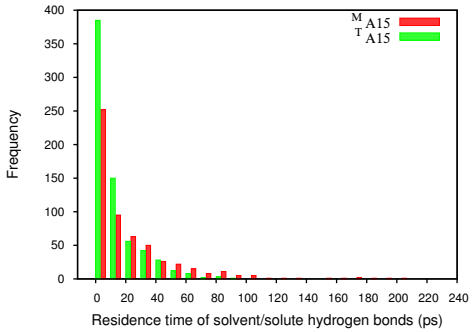
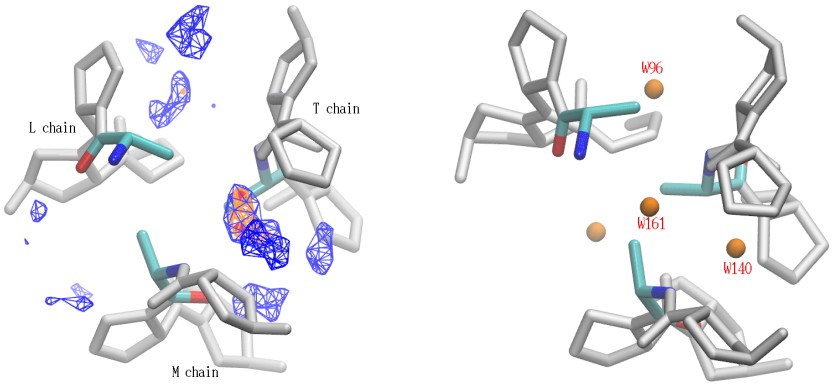


Figure 2.11: *The distribution of the residence time of solvent/solute hydrogen bond for ^{M}Ala15 and ^{T}Ala15 residues. The Y-axis is the frequency and the X-axis is the residence time that the solvent/solute hydrogen bond persists. Every bin denotes every 10 ps.*



(a) (b)

Figure 2.12: *Comparison of the MD simulated (b) and the experimental (a) solvent/solute hydrogen bonds in the substitution sites of the Gly→Ala. (a) Crystallographic interstitial water mediated hydrogen bonds are marked as orange ball. The three Ala residues are represented as Licorice in color and other residues in gray. (b) The simulated water density is shown as contour lines, which are color code by the total occupancy level of the volume. Volumes were occupied > 75%(red) > 50%(orange) > 25% (blue). The simulated water density well reproduces the crystal location of the interstitial waters 140 and 96, but not water 161.*

Water occupancy density around the peptide surface is conducted. Fig. 2.12 shows the comparison of the experimental and simulated water distribution around the Ala residues in Gly→Ala. The greatest difference between MD simulated and crystal peptide is that, in the simulation, there is no evidence of any penetration of water inside the center of triple helix (Fig. 2.12(a)), while in crystal environment, one of the interstitial water bridges sits inside the center of the triple helix (W161 next to ^{L}A15 in Fig. 2.12 (b)). Although the simulated water density is well reproduced the interstitial waters in crystal (W140 and W96 which are next to ^{M}A15 and ^{T}A15, respectively in Fig. 2.12 (b)), as mentioned in previous section, there are two distinct solvent/solute hydrogen bonds behaviors for ^{M}Ala15 and ^{T}Ala15 in the simulation; ^{M}Ala15 exchanges between interstitial water and solvent hydrogen bonds and ^{T}Ala15 exchanges between inter-chain and solvent hydrogen bonds.

2.2.4.3 Hydration Patterns

The mutation-induced disruption is also seen in the extended and repetitive water network, which can be characterized by the radial distribution function (RDF) of the residue from MD trajectories of Gly→Ala.

Hydration around the amide/hydroxyl groups The RDFs of water with respect to amide groups of Gly/Ala are shown in Fig. 2.13. The presence of the mutation increases the interaction of A15 and G18 amide proton with water. Interesting, the ^{M}A15 (red line in Fig. 2.13) has two distinct peaks in RDFs compare to ^{T}A15 (black line): the first peak that represents the stronger interaction with solvent might be referred as the interstitial water hydrogen bond; the second peak might indicate the water hydrogen bond. The analyses of RDFs are fairly in agreement with the results of water hydrogen bonds analysis on amide group of Ala residues in previous section. Hydroxyl groups of Hyp are fully hydrated and hydration pattern is not affected upon mutation (Data not shown).

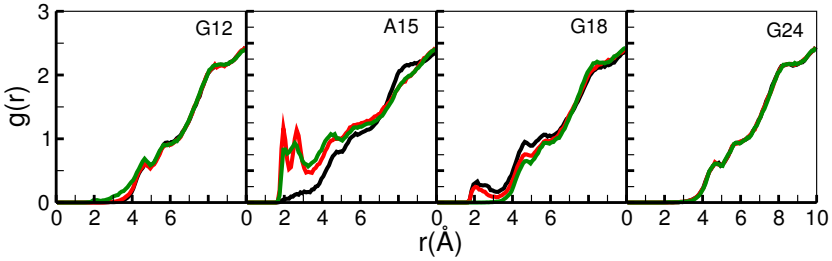


Figure 2.13: *Radial distribution function (RDF) of amide groups of G12, A15, G18, and G24. The chain L, M, and T are colored in black, red, and green, respectively.*

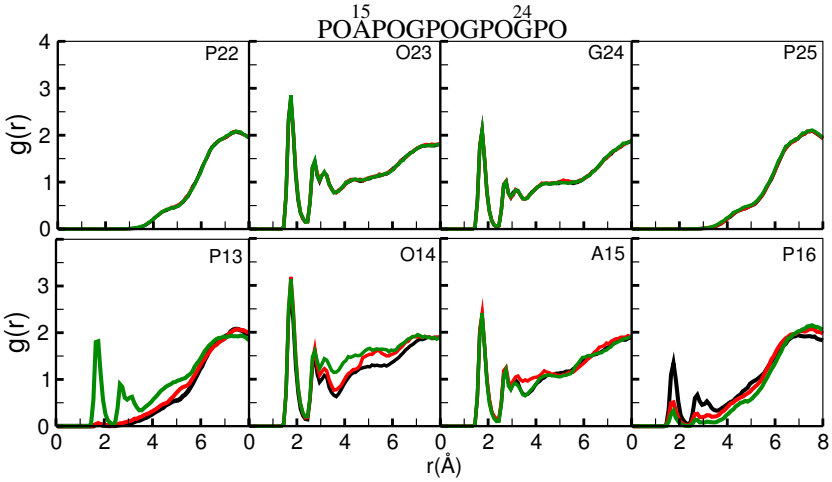


Figure 2.14: *Radial distribution function (RDF) of the carbonyl groups of the residues P13, O14, A15, P16 surrounding the substitution sites as well as the P22, O23, G24 and P25 at the collagen zone along each chain. The chain L, M, and T are colored in black, red, and green, respectively.*

Hydration around the carbonyl groups In the regular collagen zone, of the three carbonyl groups of the triplet (Gly-Pro-Hyp), the carbonyl groups of Gly and Hyp interact directly with water molecules, while carbonyl groups of Pro in Xaa positions do not participate in water hydrogen bond as it is involved in the direct inter-chain hydrogen bonds. Fig. 2.14 clearly shows that hydration pattern of the carbonyl groups of Ala and Hyp is not affected by the mutation (as comparing the residues in the mutated sites and the regular collagen zone). However, mutations do disturb the RDFs of carbonyl groups, it is particularly noticeable in ^{T}P13 and ^{L}Pro16 because both of them are the residues close to the mutation sites (see Fig. 2.16(a)).

2.3 Analysis of NMR Measurements

2.3.1 Dihedral Angles and Coupling Constants

Table 2.5 shows the mean of the average backbone dihedral angles ϕ and ψ and the corresponding standard deviations surrounding the mutations sites (residues Hyp14, A15, P16) from the solution simulation. The mean values of dihedral angle for the nearby residues Hyp15 or Pro16 are similar for each chain. While it is not the case for three A15 mutation sites where the average values of the ϕ angle range from -69^{o} to -114^{o}.

Residue\Chain	L	M	T	L	M	T
Dihedral Angle		ϕ			ψ	
G24	-71.3(9.6)	-71.5(9.6)	-71.2(10.5)	169.5(8.1)	169.5(8.2)	169.3(8.3)
Hyp14	-65.9(10.4)	-69.6(10.4)	-63.5(11.0)	152.4(11.0)	161.7(13.1)	144.1(15.6)
A15	-69.8(14.3)	-114.0(26.9)	-85.1(19.5)	152.1(11.0)	157.1(11.6)	149.2(11.2)
P16	-69.1(10.9)	-69.0(10.3)	-70.0(10.5)	150.6(12.0)	153.9(11.0)	158.4(9.9)

Table 2.5: *Average values of the dihedral angle (ϕ ,ψ) mean and standard deviation given in parentheses for the residues Hyp14, A15, P16 and G24 in each chain over the MD trajectory of the Gly→Ala.*

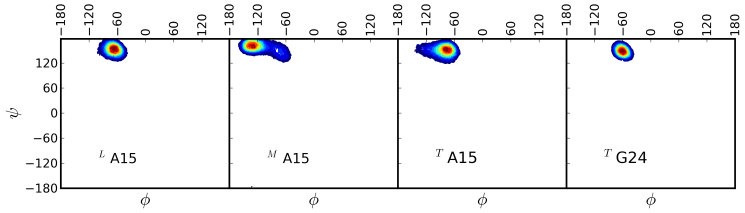


Figure 2.15: *Ramachandran plot (ϕ, ψ) of three Ala and one G24 residues. Warmer color represents higher population.*

In Fig. 2.15, the structural ensemble from the simulation exhibits that three A15 have similar distribution around the mean value of the ψ angle but different degree of dispersions in ϕ angle. The ϕ value of $^{\mathrm{L}}$A15 is almost identical to one in G24, and $^{\mathrm{T}}$A15 is also close to $^{\mathrm{L}}$A15, while $^{\mathrm{M}}$A15 deviates from one of G24 the most. In other words, the leading and trailing chains remains/closes in the PPII helical conformation, whereas the M chain deviates the most from the PPII helical conformation. It is worthy to note that $^{\mathrm{M}}$A15 has the most standard deviation up to $\sim 27^{\mathrm{o}}$ is because of the possibility of forming inter-chain hydrogen bond, which requires the PPII helical conformation.

Comparison of MD results and the experimental data is displayed in Table 2.6. MD results are consistent with experimental data $^{3}\mathrm{J}_{\mathrm{NHHA}}$ coupling constants on the Gly→Ala, in which $^{\mathrm{M}}$A15 has a very high J-coupling relative to other two A15 and Gly24 residues [51]. This significant structural change at the substitution sites is also seen in the crystal structure [22]. For most angles, the crystal structure values are within 10 degrees of the simulation mean values, except $^{\mathrm{T}}$A15 in which the value differs by around 24 degrees (which is outside the standard deviation of the mean values) between the simulation and the crystal structure, indicating that how the local conformation of solution collagen peptide differs from model collagen peptide in crystal.

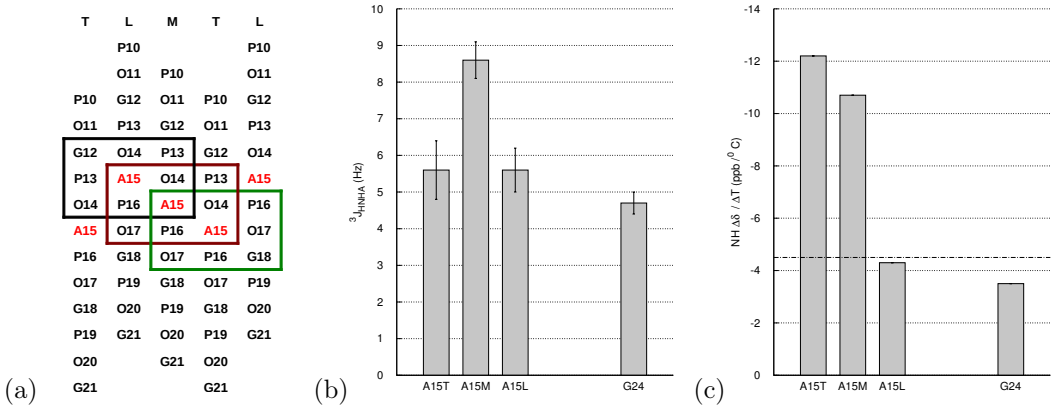


Figure 2.16: *(a) Schematic representation of the nearest neighbor residues for the Ala15 in the triple helix. The chain leading, middle and trailing are denoted with L, M, T, respectively. Boxes refer to the residue environment of the A15. (b)Experimental $^3J_{NHHA}$ coupling constants measurement on peptide Gly→Ala. (c) Amide NH temperature gradients ($\Delta\delta/\Delta T$) plot in Gly→Ala peptide. The dashed horizontal line corresponds to $\Delta\delta/\Delta T = -4.5$ ppb/^oC which provided the cutoff line of hydrogen bonding. G24 and ^{L}A15 appear to have similar less negative NH temperature gradients that are consistent with hydrogen bonding seen in solution simulation.*

	^{3}J$_{\mathrm{NHHA}}$		ϕ		
Residue	Solution	MD	Xtal	Simulation	$\phi_{calc} - \phi_{xtal}$
G24[a]	4.7 (0.3)	5.6 (1.42)	-77.4	-71.3(9.6)	+6.1
^{L}A15	5.6 (0.6)	5.3 (1.6)	-81	-69.8 (14.3)	+10.2
^{M}A15	8.6 (0.5)	8.5(1.8)	-104	-114.1(26.9)	-10.1
^{T}A15	5.6 (0.8)	7.0 (1.8)	-61	-85.1(19.5)	-24.1

a. Average ϕ angle of G24 in three chains was used.

Table 2.6: *$^3J_{NHHA}$ coupling constant and of Gly24 and three Ala15 in the Gly→Ala peptide from experimental measurement and MD simulation. Average angle of G24 over three chains. $^3J_{NHHA}$ coupling constants we estimated from the parameterized Karplus equation: $^3J_{HNHA} = 6.51cos^2(\phi - 60) - 1.76cos(\phi - 60) + 1.6$ [1].*

2.3.2 Amide Chemical Shift Temperature Dependence

Yingjie et al. [51] reported the data of the amide temperature gradient and ^{3}J$_{\mathrm{NHHA}}$ coupling constants on peptide Gly→Ala to characterize the amides of G24 and A15. Since the three chains of the CMP are staggered by one residue, three A15 are not all in an identical environment (boxes shown in Fig. 2.16 (a)). Both ^{L}A15 and ^{T}A15 have identical chemical environments that are distinct from ^{M}A15. This seems to be consistent with the measurement of ^{3}J$_{\mathrm{NHHA}}$ coupling constants; one of the Ala trimer peaks has a very high J-coupling relative to the other two A15, and they are indistinguishable in nmr (Fig. 2.16 (b)).

However, non-equivalent chemical shift temperature gradients for three A15 residues are observed; G24 and one of the three A15 appear to have similar less negative gradient, indicating the forming of the direct inter-chain hydrogen bonds, while other two A15 residues with more negative gradients are likely involved in either weak hydrogen bonding or exposed to solvent (Fig. 2.16 (c)). Such a significant difference in chemical shift temperature gradients (^{L}A15= -4.3, ^{T}A15= -12.2 $ppb/^{o}C$) between ^{L}A15 and ^{T}A15 is unexpected.

With the aid of MD results reported in Sec. 2.2.4, it can nicely explain what observed in NMR studies. MD simulation suggests that when introducing an Ala into a Gly in the middle of each chain $(POG)_{10}$ peptide, the mutations do affect the local structural conformation, particularly at ^{M}A15 with two A15 nearby. ^{L}A15 and ^{T}A15 keep close to standard PPII conformation due to the Gly residue as an neighbor. That is why ^{L}A15 and ^{T}A15 with identical sequence environment have similar $^{3}J_{NHHA}$ coupling constants, which are only slightly away from one of Gly24. However, the solvent environments are completely different between ^{L}A15 and ^{T}A15; ^{L}A15 has primarily inter-chain hydrogen bond, while ^{T}A15 has highly exchangeable nonspecific hydrogen bonds with solvent. This asymmetric effect reflected in the solvent environment may induce the asymmetric variations between ^{L}A15 and ^{T}A15, further extend to one triplet N and C terminal to the ^{M}A15 site. It has been suggested that the degree of flexibility due to mutations depends on the region along the triple helix: mutated sites > C-terminal region > N-terminal region.

2.3.3 Chemical Shift Calculation

Protein chemical shifts, which can be precisely measured by NMR and highly accessible, provide an information of local changes in the surrounding chemical environment and the dihedral angle constraint of individual atoms. Among the shifts, the $^{1}H_{N}$ chemical shifts are very sensitive to the local H-bond geometry.

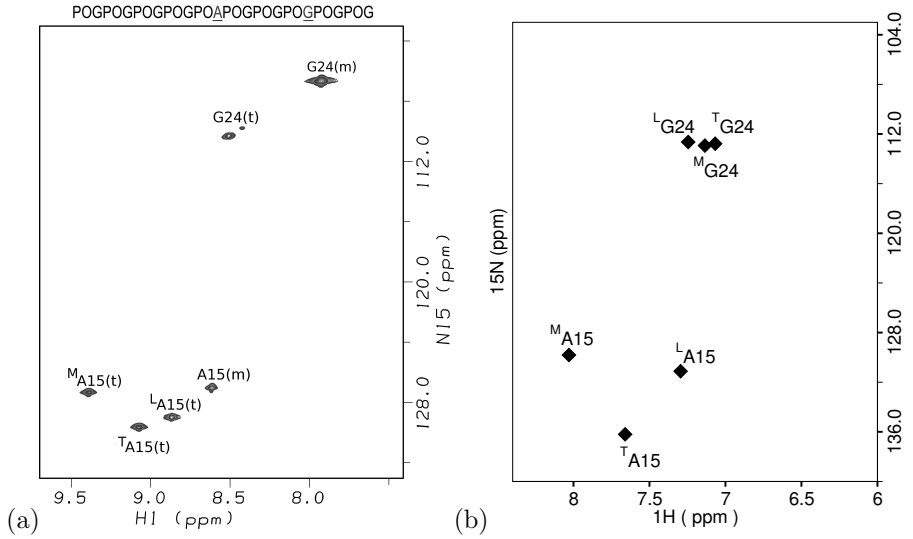


Figure 2.17: *(a) HSQC spectrum of Gly→Ala measured at 15 C. Sequence is shown above the spectrum with ^{15}N labeled residues underlined. The peaks corresponding to the monomer and trimer states are denoted with m and t, respectively, in the parenthesis after the residue. The leading, middle, and trailing chains are denoted with a superscript L, M, and T, respectively, located in the left side of residue. A15(t) trimer are downfield shifted in proton dimension relative to A15(m) monomer, particularly one chain shifted to 9.4 ppm. (b) The QMMM calculation of shifts on the four different sets of MD trajectories of Gly→Ala in the trimer state. In proton dimension, averaging over four sets gives the values of A15 are ^{L}A15=7.29 , ^{M}A15=7.66 and ^{T}A15=8.03 (ppm).*

The HSQC spectrum (Fig. 2.17 (a)) of the Gly→Ala clearly showed three trimer peaks for the three Ala residues, indicating the three distinct local environment of the A15 from one another [48, 65, 89]. One of these three trimers even shifted down field by ∼0.7 ppm in ^{1}H dimension relative to the Ala in the monomer state. NMR measurements have also showed intriguing yet consistent trends in other collagen-like peptides that contain (GXY)n interruptions or other substitutions at Gly positions; in all cases the amide protons in one of the chains is significantly downfield shifted relative to the other two once the mutation or interruption takes place.

This significant downfield shift is unexpected when there is slightly difference between their sequential environments. However, this unusual pattern has been difficult to interpret. It has been found good correlations between hydrogen bond coupling constant and $^{1}H_{N}$ chemical shift both experimentally and computationally, suggesting that this unusual shift might be related to the interaction with the nearby waters.

To interpret this unusual downfield shift due to the mutation, automated fragment quantum mechanics/molecular mechanics (AF-QMMM) approach [90, 91] was applied to compute

the shifts for the NH atoms of the Gly and Ala residues on the selected MD snapshots of Gly→Ala. Calculated ^{1}H chemical shifts (Fig. 2.17) are compared with the calculation of water hydrogen bonds analysis on the Gly and Ala residues: *(i)* $^{\mathrm{L}}$A15 with primarily direct inter-chain hydrogen bond has similar ^{1}H shift as the Gly24 does; *(ii)* interacting of amide proton with water molecules causes a downfield shifts in ^{1}H dimension by ∼0.36 ppm relative to $^{\mathrm{L}}$A15 as seen in $^{\mathrm{T}}$A15 which has highly exchangeable nonspecific hydrogen bonds with solvent; *(iii)* interstitial water hydrogen bonds seems to push the shift even downfield by ∼0.7 ppm relative to $^{\mathrm{L}}$A15. Berglund et al. [92] has suggested that increasing hydrogen bond strength causes a downfield ^{1}H chemical shift. The calculation of the radial distribution function (Fig. 2.13) also shows that water population is slightly higher around $^{\mathrm{M}}$A15 than one in $^{\mathrm{T}}$A15. Our calculated shifts of amide proton are highly related to corresponding hydrogen bonds, suggesting that, if forming hydrogen bond, the strength of the interaction of amide proton of Ala residue depends on the type of hydrogen bonds: interstitial water > water > backbone peptide. Although the calculated $^1\mathrm{H_N}$ chemical shifts show upfield shifts as compared to experimental shifts values (∼1.4 ppm), the relative shifts between residues computationally are in agreement with ones seen experimentally. Qualitatively and quantitatively, the data are beginning to interpret the unusual trends observed in NMR data on the mutation collagen peptides.

2.4 Conclusions

Presentation in this paper is the detailed MD studies of structural and hydrogen binding properties of a CMP where a mutant replacing one Gly to other large residue. Our findings reveal that major changes in the conformation of the mutated Gly peptide occur in solution where the deviations from the starting structure are larger. Replacing Gly increases the backbone flexibility of the triple helical collagen as evidenced by the variations in radius of the triple helix and the backbone bending. The overall dynamics of the peptide is characterized by performing PCA of the $C\alpha$ covariance matrices, showing the majority of the structural variability of the triple helices corresponding to the orthogonal bending and twist motions of the whole backbone. Mutation induces the unequal orthogonal bending motions of backbone, resulting the backbone bending with directional preference. The results of clustering analysis reveal that the predominant conformation of the mutated peptide adopts slightly a curved structure while POG10 adopts a structure similar to an idealized helix. The calculations of the helical parameters reveal that mutated peptide is unwound in the central region helix, and asymmetric distortion at N-and C-terminals where the higher degree of freedom of adopting other conformation in the C-terminal zone compare to N-terminal zone. Although mutant collagen is partially unwounded through its interaction with the solvents

or bridge waters, it remains its triple-helical character.

Calculation of structural conformation and hydrogen bonds of triple helical peptide show how dynamics, solvated collagen model peptide differ from model collagen peptide in crystal structures. This is not surprising due to the difference between the solution and crystallized environment, in which the collagen are closely packed.

The results of this work closely parallel what are observed in NMR data on the Gly→Ala in solution, which in turn support its ability to complement the experimental observations. *(i)* The NMR hydrogen exchange rates observed are not quit the same for all three A15 in the triple helix and one of them has an exchange rate much slower than other two. MD studies suggest this observation may be the result of the inequality of direct inter-chain hydrogen bond of three A15 residues and the ^{L}A15 with primarily direct inter-chain hydrogen bond is the one having the slowest exchange rate compared to other two; *(ii)* the hydrogen exchange rates for A15 are 2 orders faster than the G24 in the same peptide. This is highly supported by the observation that ^{M}A15 and ^{T}A15 residues have highly exchangeable nonspecific hydrogen bonds with solvent relative to ^{L}Ala15 (which is more buried in the POG region) or G24; *(iii)* the HSQC spectrum clearly exhibits that three trimer peaks for three A15 residues, specially one is shifted down field more than other two. This observation is best explained by MD studies showing that the solvent environments are responsible for unusual trends: the ^{M}A15 with most distortion PPII conformation and forming frequently the interstitial water hydrogen bonds shifts down-field the most; the shift of the^{L}A15 having primarily inter-chain hydrogen bond is close to that of G24; while ^{T}A15 has the highly exchangeable nonspecific hydrogen bonds with solvent relative to ^{M}A15. Back-calculated shifts of amide proton suggests that the strength of the interaction of amide proton of Ala residue depends on the type of hydrogen bonds: interstitial water $>$ water molecules $>$ backbone peptide.

In our solution simulation, high correlation between inter-chain hydrogen bond length and the backbone dihedral angle ϕ indicates the distortion of PPII conformation interrupts the inter-chain hydrogen bonds and induces the interactions with waters. This observation explains why G24 and one of A15 residues (^{L}A15) appear to have similar less negative NH temperature gradients. Furthermore, although amide groups in the middle and trailing chains have similar populations for two distinct conformations of amide groups hydrogen bonds, one in the middle chain with the most distortion in ϕ angle interacts with interstitial waters, while one in the trailing chain with less distorted PPII conformation exchanges faster between carbonyl group of Pro and waters (no interstitial water bridge was found in the simulation). The distribution of inter-chain hydrogen bond and the dihedral angle of ϕ further support the inequality of three Ala15 seen in NMR experiments.

Tracking the interactions between A15 amide and the waters has shown that the lifetime

of waters and interstitial waters are in the range of picoseconds timescale, supporting the unusually fast hydrogen exchange rates in NMR. In the simulation, the water occupying the water bridge hydrogen bonds are not tightly bound and the solvent molecules seem to have a fairly degree of mobility. These results are also seen in Klein's study [71], suggesting that water is not an essential contribution to the stability of collagen [93]. It is worthy to note that the average lifetime of interstitial water molecules is still much longer than the ones of solvents, suggesting having interaction with interstitial waters pushes the chemical shift down field more relative to the inter-chain hydrogen bonds by about 0.7 ppm in the proton dimension as reflected in the shift calculation using QMMM. So far, our computational results are in agreement with NMR experimental data found in the literature [48, 51, 89, 94], indicating the MD simulations can actually provide a realistic description of the collagen in solution environment and further complement the experimental observation.

2.5 Material and Methods

2.5.1 System Description and Setup

The initial coordinates of the collagen-like peptides, $(POG)_4(POA)(POG)_5$, $(POG)_{10}$, and $(PPG)_{10}$ were taken from the published structures in PDB 1CAG [22], 1V7H [24], and 1K6F [23], respectively. MD simulations were performed with the Amber ff99SB [95] protein force field using the AMBER10 programs [96]. In the simulation, the protein is placed in a truncated octahedral periodic box of the explicit SPC water model with about 27,700 water molecules. The distance from the surface of the box to the closet atom of the solute is set to 15 Å. The system is heated from 0 to 293 K over 1ns and followed by first 3 ns simulation with restraints then 1 ns without restraints in the NPT ensemble. Finally, MD production simulation is performed in the NVE ensemble. In MD simulations, SHAKE algorithm is used to constrain all the bonds involving hydrogen atoms and the time step is 1 fs. The potential energy function represents the contribution from bond, dihedral angles, van der Waals, and electrostatic terms. The particle-mesh Ewald method was used to calculate the electrostatics interaction. A non-bonded cutoff of 8 Å was used during the MD simulation. The simulations were run for 150 ns for Gly→Ala peptide, and 100 ns for $(POG)_{10}$ and $(PPG)_{10}$ peptides. All trajectories were saved every 2ps for future analysis. Post processing and analysis were carried out using the *cpptraj* analysis tool. VMD was used for visualization of trajectories.

2.5.2 Root Mean Square Deviation

The $C\alpha$ atoms RMSD of the peptide were calculated with respect to its x-ray structure as a function of time using Eq. 2.1

$$RMSD(t) = [\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|r_{ij}(t) - r_{ij}(0)|^2] \tag{2.1}$$

where r_{ij}, the distance between atoms at the time t, is compared with the distance between the same atoms at the t=0 and N is the total number of backbone atoms considered.

2.5.3 Helical Radius

Helical radius is defined for each residue i as the distance from its Cα atoms to the centroid defined by the Cα atoms residues i, i-1 and i-2 on chains leading(L), middle (M), and trailing (T) respectively [97].

2.5.4 Cluster Analysis

The structures are grouped into five sets using cluster analysis based on the rmsd between each frames. The difference in rmsd values from one group to another is at least 1.4 Å. The clustering algorithm is implemented in AMBER package.

2.5.5 J Coupling Constants

The ^{3}J$_{\mathrm{HNHA}}$ coupling constant is a valuable NMR restraint to distinguish $\alpha-$helical from $\beta-$sheet secondary structure in proteins, e.g. 4.8 Hz for α helix and 8.5 Hz for β sheet. Empirical ^{3}J$_{\mathrm{HNHA}}$ coupling values and angular backbone constraints can be inter-converted following well-know Karplus relationship of the form [98]

$$^3J_{HN,HA} = 7.9cos^2(\phi - 60) - 1.04cos(\phi - 60) + 0.65 \tag{2.2}$$

where ϕ is the backbone torsion angle N-H-C-HA.

2.5.6 Hydrogen Bond Analysis

The hydrogen bonds between the chains in collagen-like peptide were calculated using geometrical criteria given in Eq. 2.3

$$H_i = \left\{ \begin{array}{c} 1, (r\,(N_i - H_i \cdots C = O) < 3.5\mathring{A})\; and\; (135^o \leq \alpha\,(O - H \cdots N) \leq 180^o)) \\ 0, otherwise \end{array} \right\} \tag{2.3}$$

where r is the distance between donor and acceptor and α is the angle between the hydrogen donor and acceptor. Also the occupancy of the H-bonds were calculated using the following equation:

$$H - bond = \frac{number\,of\,time\,the\,occurrence\,of\,particular\,hydrogen\,bond}{total\,number\,of\,frames} \times 100 \qquad (2.4)$$

2.5.7 Rigid-Backbone Bending Angle

We calculated the bending angle based on local triads, which were constructed joining the C_α atoms of Gly, Xaa, Yaa of adjacent chains along the length of the peptide (will put a Fig. for it). The C_α atom of Gly of each triangle was chosen as the local coordinate origin and $\hat{e}_2$ were defined as the unit vector along the line pointing from the origin to C_α atom of Xaa position and Yaa position, respectively. Setting $\hat{e}_3$ perpendicular to the triad plane $\hat{e}_3 = \hat{e}_1 \times \hat{e}_2$ as the normal vector of the each triad. There are 28 triads in this triplet helical peptide. The first 9 triads from both end terminal regions are taken separately to get the average normal vectors. The angle between these two average vectors is the angle we define for the backbone bending.

2.5.8 Rigid Body Parameters Calculation

The local step parameters are the internal or local coordinate representations of helices, e.g. DNA that specify how a set of directors attached to the ith base pair are rotated (Tilt, Roll, Twist) and translated (Shift, Slide, Rise) to a set of directors associated with the (i+1)th base pair. The local bend is ($\sqrt{Roll^2 + Tilt^2}$) and local shear is ($\sqrt{Shift^2 + Slide^2}$). Gly→Ala and other collagen-like peptides of studied here contain three polypeptide chains (labeled as Leading (L), Middle (M), and Trailing (T)) with 30 amino acids in each chain. The three Cα atoms from Gly^L_{i+2}, Y^M_{i+1}, and X^T_i residues, lying approximately in a perpendicular plane to the triple helical axis (see Fig.2.18(a)), form a triad (Fig. 2.18(b)). RMSD two consecutive, (i) and (i+1) triads, can give triad local step parameters which specify the configurations of spatially adjacent triads: two bending angles called Tilt(τ) and Roll (ρ), and the rotation angle Twist(Ω), and two in-plane translation termed Shift (D_x) and Slide (D_y), and the vertical displacement Rise(D_z) (See Fig. 2.18 (c)). The calculations of rigid-body parameters follow the mathematical expressions by El Hassan's algorithm [99] and the 3DNA code [86]. Firstly, from the position and orientation of the (i) and (i+1) triplet triads, to get the direction and the magnitude (Γ) of the hinge, we take the cross product of the z-axes of the two successive triads. Secondly, in order to align these two z-axes of these two triads, we rotate the two triads about the hinge axis through the angles of equal magnitude

but opposite signs until z axes are parallel; the current z axis is the z axis of the middle frame located between these two consecutive triads. Thirdly, we have hinge axis lying on the x-y plane of middle frame and inclined at ϕ to the y axis. Roll and Tilt follow from Γ and ϕ. Along the current z axis, the relative rotation of the two triads along the current y (or x) axis of each triad gives the helical twist (Ω). Finally, the displacement,D_x, D_y, and D_z now can be obtained by projecting the vector of origins of the two original triads into components along the axes of the middle frame.

Beside this set of six parameters mentioned above, the local geometry of each triad step can also be described by another set of six helical parameters: x-displacement, y-displacement, helical rise, inclination, tip, and helical twist. There two sets of parameters should be directly convertible, as demonstrated by Calladine and Drew in their 1984 JMB B-to-A transition paper [85, 100].

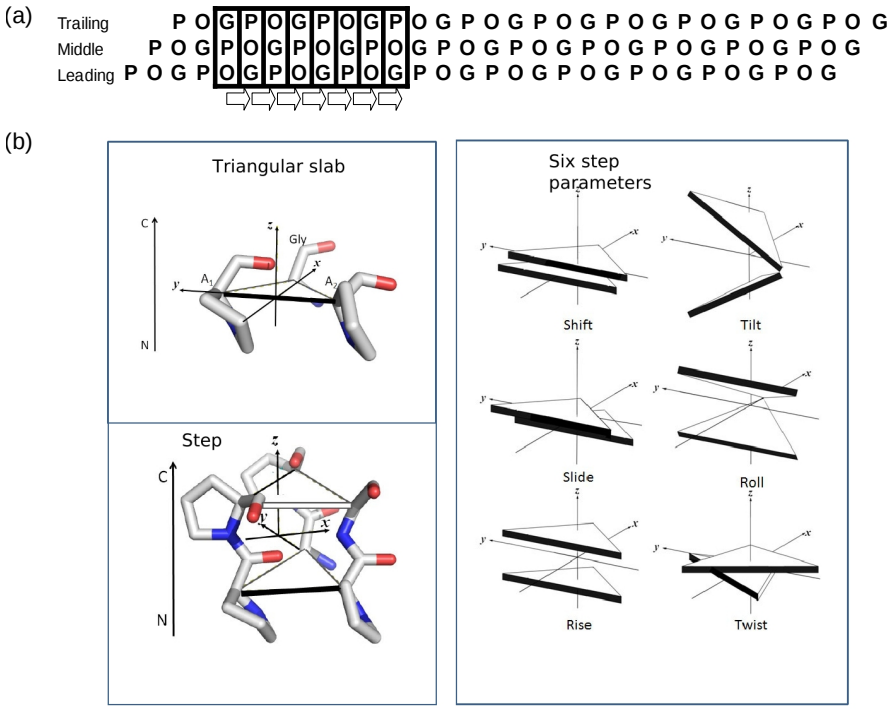


Figure 2.18: *(a) Schematic representation of amino acid composition, base (black box) and bases step parameters (arrow) of POG10 peptide. (b) Pictorial definition of rigid-body parameters describing the geometry of a base and the sequential base planes. Base identification: triangular slab consist of three CA atoms lying approximately in a perpendicular plane to the triple helix axis. Base frame definition: one side of the triangle excluding Gly is defined as Y-axis from Yaa to Xaa positions is shown as heavy black line; Z axis is perpendicular to Y axis and pointing to the side with Gly; the origin is the midpoint of the Xaa and Yaa. Definition of six step parameters characterizing the relative displacements: Shift, Slide, and Rise along the X-, Y- and Z-axis, respectively, and orientation: Tilt, Roll, and Twist about the X-, Y- and Z-axis of the sequential triangular slabs. Image illustrates positive values of the designed parameters.*

2.5.9 Chemical Shift Calculation

Proton and nitrogen chemical shifts for the NH atoms of the Gly and Ala residues were estimated using the "automated fragmentation" model originally introduced by He et al. [90]. In this model, a "quantum region" is defined around the residue whose shift is being calculated, that includes that residue and its neighbors out to a cutoff of about 3.5 Å. Water molecules within a similar distance are also included. Protein residues beyond the cutoff are included a MM (molecular mechanics) atoms, and the effects of waters beyond the cutoff are treated with a continuum dielectric (Poisson) model. Details are given elsewhere

[90, 91, 101, 102]. For these calculations, we used the OLYP density functional [103] with a TZVP basis set, [104] using the *demon-2k* (version 3) program [105]. For each calculation, we averaged 400 equally-spaced snapshots with an interval of 50 ps from the last 20 ns of the MD simulation trajectory. The computed shieldings were compared to those for amide nitrogens and hydrogens in ubiquitin (using the same computational model), to obtain the reported chemical shifts.

2.5.10 Step Parameters(Tilt, Roll, Twist, Shift, Slide, Rise)

The helical structure of collagen-like peptides can be described with step helical parameters including translation displacements: Shift, Slide, Rise, and rotational displacements: Tilt, Roll, and Twist computed using the methodology and algorithms in 3DNA [85, 100]. The translational parameters are simply the projection of the vector link from triad O1 to triad O2 onto the x-, y-, and z-axes of the middle frame which are the average between those of R1 in the triad 1 and R2 in the triad 2. The origin of the middle frames is the geometric average of O1 and O2. These parameters describe the relative rotations and translations along a length of triple helical collagen peptides. Examples are shown in Fig. 2.19.

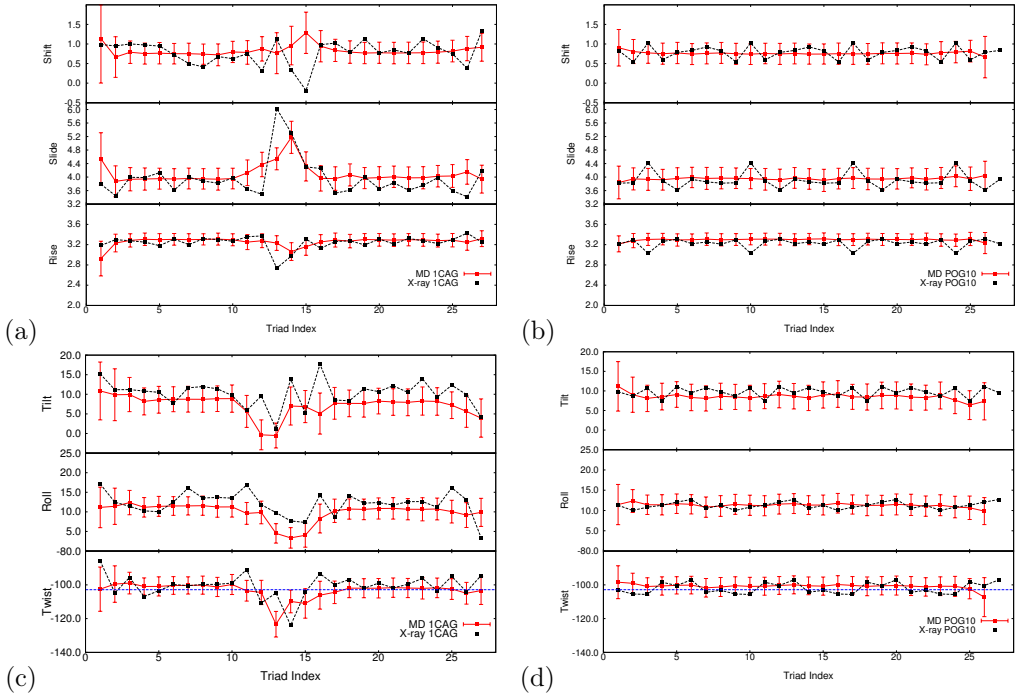


Figure 2.19: *Evolution of the average step parameters along the triplets of the collagen-like peptide, Gly to Ala in (a) and (c), and POG10 in (b) and (d). The rigid-body parameters characterizing the relative position and orientation of the sequential triangular bases steps along the triple helices of the crystal structures are marked as black.*

2.5.11 Helical parameters

Helical parameters for collagen peptides can be defined in ways analogous to those used for DNA: [85, 100] 3DNA

1. Define a local helical axis, we need two triads reference frames(R1=[x1,y1,z1] with origin o1, and R2=[x2,y2,z2] with origin o2). A single-helical axis (h) which is calculated using (x2-x1) $\times$ (y2-y1) brings triad1 to coincide with triad 2 by helical twist angle.

2. Local helical frame of triad 1 H1. TipInclination angle Ψ is computed between h and z1. Hinge axis is computed by a cross product from h to z1. Rotate R1 through the hinge axis by negative Tipinclination angle will give us the local helical reference frame H1 (H1=rotation_matrix($-\Psi$,Hingeaxis)$\cdot$R1)

3. Local helical frame of triad 2 H2 is constructed the same way as H1. TipInclination angle Ψ is computed between h and z2. Hinge axis is computed by a cross product from h to z2. Rotate R2 through the hinge axis by negative Tipinclination angle will

give us the local helical reference frame H2 (H2=rotation_matrix($-\Psi$,Hingeaxis)·R2)

4. Middle helical frame Hm is the average of H1 and H2.

5. Helical twist is the angle between H1(y) and H2(y) or (H1(x) and H2(x)) with reference to the local helical axis h for sign determination

6. Helical rise is the projection of the vector from O1 to O2 onto the local helical axis h. Rise=$\overline{O1O2} \cdot h$

7. Tip and Inclination calculation extracted from TipInclination angle Ψ. Phase angle ψ is defined from hinge axis to positive tip axis which is the 2nd column of H1 with reference the local helical axis h for sign control (product of h and (hingeaxis $\times$ H2(y))). Tip is defined as $\Psi cos(\psi)$

8. Inclination is defined as $\Psi sin(\psi)$

9. Origin of the local helical frames H1, H2, and Hm

   - Vector AB=(o2-o1) - helical_rise*h
   - M=rotation_matrix(h,angle) where angle= 90 - 0.5 * helical_twist (degrees)
   - Vector AD=M*AB
   - H1(o)=o1+|AD|*AD
   - H2(o)=H1(o)+helical_rise*h=H1(0)+$\overline{O1O2}$
   - Hm(0)=(H1(0)+H2(0))/2

10. X-displacement is defined as the projection of the vector from o1 to H1(o) onto the x- and y-axis of H1. X_disp=(o1-H1(o1)*H1(x))

11. Y-displacement is defined as the projection of the vector from o1 to H1(o) onto the x- and y-axis of H1. Y_disp=(o1-H1(o1)*H1(y))

3 Theory of rotational diffusion on deformable macromolecules: collagen model peptides

Abstract. The triple-helical collagen-like peptide is not a rigid rod in solution: both molecular dynamics simulations and analyses of NMR relaxation point to significant amounts of bending and twisting motion. Here, the weakly bending rod (WBR) model (which has been used to analyze the dynamics of duplex DNA) is applied to study the internal motions of collagen-like peptides. This model assumes the existence of both twisting and bending degrees of freedom, and can be used to estimate the effect of collective deformational modes on NMR relaxation rates and position-dependent order parameters S^2. We parameterize the WBR model with results from MD simulations. Each triplet (Gly-Xaa-Yaa) cross all three chains is treated as local cylindrically symmetric disk and its mean squared angular displacements $< \triangle_{i,x}^2 >$ and $< \triangle_{i,z}^2 >$ of the ith bond vector rotating about the instantaneous transverse (x) and symmetry (z) axes, respectively, are extracted from an MD trajectory. The decay of the corresponding correlation functions is unusual with exponentials in t^β(where $0 < \beta < 1$) due to twisting and bending. Results from MD simulation reveal that the angular motions can be quite accurately fit by the WBR model, allowing for local twisting and bending relaxations of the triplets and the overall rotational motion of the triple helix, and that the positional-dependence of the oscillation amplitude along the collagen-like peptide sequence is in good agreement with the predictions of the weakly bending rod theory. We use Lipari-Szabo-style order parameters (described below) as a point of contact, since they can be readily determined from both molecular dynamics simulations and from the WBR model. Excellent fits are possible, suggesting that global bending and twisting modes dominate the deviations from rigid-rod behavior. The understanding gained from this model is expected to contribute toward the more detailed interpretation of experimental order parameter profiles.

3.1 Introduction

Collagen, as the major structural protein in extracellular matrix, plays an important role in cellular processes at many aspects, e.g. cell development, cell adhesion, cell migration and wound healing, through interactions with other matrix molecules [15]. Collagen exhibits various functions through specific interactions with other macromolecules. NMR studies on mini-collagen have revealed that the collagen-binding domain of *Clostridium histolyticum* secretes collagenases that bind unidirectionally to the undertwisted regions, where the induced binding affinity might be due to the ability of the replaced residues to flex. This flexibility might be an intrinsic property of triple-helical collagen and play an important role in its functionality, particularly in the area of molecular recognition [44]. Another great example

is from the aspect of the mechanical property upon sequence content to study how a single point mutation alters and affects the susceptibility of an entire bone to brittle fracture are studied [106]. These functions are most likely characterized by its dynamics, therefore, it is increasingly important to explore the dynamics which might exist in collagen.

NMR measurement offers a powerful tool for studying molecular motions at a wide range of time scales, and has been applied to collagen-like peptides for many years [2, 51, 64, 66]. As described below, solution NMR studies the dynamics at ns-ps time scales by recording of spin relaxation rates R1, R2 and the steady-state heteronuclear nuclear Overhauser effect (NOE), generally for isotopically labeled residues. A popular interpretation of NMR relaxation data, first developed by Lipari and Szabo [76, 107], characterizes the relaxation in terms of overall rotational diffusion (which is a global property) and local internal motions that are described by the order parameters and time constants. The main assumptions of simple and extended model-free approach are that *(i)* the relevant internuclear vectors are assumed to move in isotropic deflection potentials fixed at their orientations in the molecules, in other words, the local amplitude fluctuation of the vector is independent of its orientation with respect to the molecular frame; *(ii)* assumption of the separation of timescales and decoupling of motions for global diffusion and local internal motions; and *(iii)* the molecule undergoes global rotational diffusion as a rigid body. This is a powerful model, but suffers from two limitations for the purposes to the following:

1. In the model-free analysis, the S^2 order parameter does not provide a complete description of the underlying re-orientational motional distribution. Each of the local parameters is independent of the others, so that no information about the correlated modes between the vectors which are not next to each other is obtained in this standard model-free description. Stated in other words, it can be difficult to interpret a set of order parameters in molecular terms. This is both a strength and a limitation: because no "model" of internal motions is assumed, there is also no model at the end to interpret.

2. The simple model-free approach is strictly only to spherical molecules or those with low anisotropies up to $D_{||}/D_{\perp} \sim 2$. In the case of the collagen triple helix (or of nucleic acid helices) with very high anisotropy, the failure of assumption of (*iii*) invalidates the decorrelation or factorization approximation of internal motion from global motion even with the assumption of rigid rod model. Furthermore, triple helical collagen-like peptides are not rigid cylinders, but have small amplitude vibrations and collective internal motions [57]. These collective modes might not be governed by an isotropic potential, but would show substantial anisotropy, such that the internal motion is described multiple exponential functions [77], whose relative weightings depend on the

orientation of the dipole-dipole interactions with respect to the symmetry axis of the diffusion tensor.

Here we explore a motional model, which is weakly bending rod (WBR) model, that should correctly deal with both high anisotropies and (weakly) flexible geometries. The outline of the chapter is as follow. In Section 3.2, we begin with the model-dependent WBR theory and its application to the interpretation of NMR relaxation data. We shall see that the WBR model predicts the dependence of the order parameters on the label position and the orientation of the vector relative to the symmetric axis of the molecular frame, which is not seen in the model-free approach. In Section 3.3, we fit the adjustable parameters in this model (principally the bending and torsional rigidities) to the results of molecular dynamics simulations on collagen peptides. We use both Lipari-Szabo-style order parameters and mean-squared angular displacements (described below) as a point of contact, since they can be readily determined from both molecular dynamics simulations and the WBR model. Excellent fits are possible, suggesting that global bending and twisting modes dominate the deviations from rigid-rod behavior. This has implications for the interpretation of experimental relaxation rates in these peptides, although more complete isotopic labeling will probably be required to test these ideas; such experiments are being planned. We end with some conclusions of main points and an outlook for future research, followed by Appendices that collect some of the math needed for the WBR model and that describe computer programs that implement this model.

3.2 Application of the Weakly Bending Rod Model to NMR Data

3.2.1 Introduction

The weakly bending rod (WBR) model, developed by Song and Schurr [108], has been successfully applied in both electron paramagnetic resonance (EPR) and NMR measurements on short DNA fragments to characterize the motions on time scales between end-to-end tumbling and rapid internal motion [109–113]. In this model, the deformable filament is represented by a series of spherical beads or cylindrical discs, which are coupled to their nearest neighbors by a harmonic potential characterized by torsion and bending elastic force constants, which are related to the mean-square amplitudes of the oscillation of the ith base pair in DNA dynamics. In the following, we give a more complete account of the application of this WBR model to describe the effects of deformable motions on NMR relaxation data, which can be used to ascertain the twisting and bending rigidity, which acts to straighten the chain

Our objective is to evaluate the relaxation rates, R1, R2, and NOE for the case when

the vectors of interest in a molecule that may be deformable, but exhibits mean local cylindrical symmetry [111]. Of special interest here are the angular displacement motions of the N-H bond vectors, which have local angular motions taking place in the frame of the molecular subunit, for example, triplet (the detailed definition in Sec. 3.3). Each subunit participates in all the collective twisting and bending deformations and uniform (rigid-rod) rotation around the symmetry and transverse axes. The angular displacement correlation functions arising from these deformable motions can be expressed in terms of familiar internal parameters, such as S^2 and τ_{int}, that we borrowed from the LS model-free formula, as suggested by Robbison *et al.* [113]. This WBR-derived order parameter profile thus provides an explanation of the dependence of order parameters on label position.

Xiao *et al.* [2, 64] have measured the ^{15}N relaxation rates (data shown in Chapter 3) of a collagen-like peptide with sequence $(POG)_3$-ITG-ARG-LAG-$(POG)_4$, denoted as T3-785 [28]. In analyzing their data, they employed rigid cylinder models that do not admit twisting and bending deformations, and show the divergence of the values of the R2/R1 of Gly NH vectors from the experimental data. They concluded that the Gly amide proton in the triple helical collagen peptide needs to be repositioned relative to the backbone. However, as discussed in Chapter 3, our MD simulations don't appear to support the details of this model, which might be insufficient to explain the experimental data. In order to assess the relative contributions of collective and localized internal motions to the nmr relaxation, it is necessary to analyze the data using a model that actually allows for internal twist and flexure. Here, we consider the internal dynamics of this model peptide and others, including $[(\text{Pro-Hyp-Gly})_{10}]_3$ denoted as POG10 [24], and $[(\text{Pro-Pro-Gly})_{10}]_3$ denoted as PPG10 [23], in terms of a WBR model.

3.2.2 Theory of NMR relaxation

The NMR dipolar relaxation experimental technique employed to study rotational dynamics of macromolecules actually monitor the orientation of a vector attached to the macromolecules in solution with an isotropic equilibrium state. According to Bloch, Wangsness and Redfield, spin relaxation parameters (thermal motions), i.e. the ^{15}N -^{1}H nuclear Overhauser effect (NOE), the spin-lattice relaxation time ($T1 = 1/R1$), and the spin-spin relaxation time ($T2 = 1/R2$), can be expressed as the function of the spectral density $J(\omega)$, defined as the Fourier transformation of the time correlation function:

$$R1 = \frac{d^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \tag{3.1}$$

$$R2 = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{6}[3J(\omega_N) + 4J(0)] \tag{3.2}$$

$$NOE = 1 + (\frac{d^2}{4})(\frac{\gamma_H}{\gamma_N})[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]/R1 \tag{3.3}$$

where $c = \omega_N \delta_N / \sqrt{3}$, and $d = \mu_0 h \gamma_H \gamma_N / (8\pi^2 < r_{NH}^3 >)$. Here μ_0 is the permeability of vacuum, h is Planck's constant, ω_H and ω_N are the ^1H and ^{15}N Larmor frequencies of the experimental magnetic field, γ_H and γ_N are their respective gyromagnetic ratios, δ_N is the chemical shift anisotropy of the ^{15}N nucleus, and $<r_{NH}^3>$ is the average N-H bond length. $J(\omega)$ is the spectral density function with the expression as following

$$J(\omega) = \int_{-\infty}^{\infty} C(t) e^{-i\omega t} dt, \tag{3.4}$$

where $C(t)$ is the time autocorrelation function of the orientation of the internuclear vector in the lab frame. The time autocorrelation function, describing the angular reorientation of the dipole-dipole and ^{15}N CSA interaction, which contribute to ^{15}N relaxation, is thus investigated directly. Methods for calculating the required correlation functions are presented in Section 3.2.3, next.

3.2.3 Time Correlation Functions

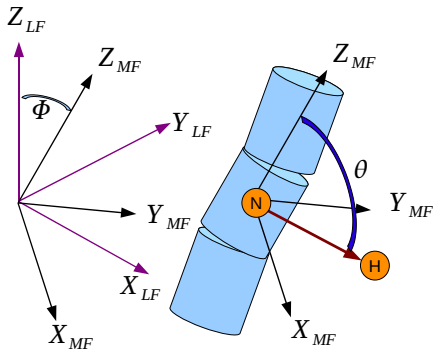


Figure 3.1: *Schematic representation of a vector with internal rational angles attached to a tumbling cylindrical rod. The laboratory coordinate system (LF), molecular tumbling system (MF) and a vector (N-H) attached to the molecule.* $\Phi(t) = [\alpha(t), \beta(t), \gamma(t)]$ *is the Euler rotation that carries a coordinate frame from the lab frame (LF) to the molecular frame (MF), and* θ *is the orientation of the NH bond vector relative to the symmetric axis (which is* Z_{MF}*) of the molecular frame.*

The triple helix is assumed to be of cylindrical symmetry, such that D_z and $D_x = D_y$ denote the diffusion constants relative to the symmetry axis of the molecule. The autocorrelation function of a particular internuclear vector can be written as

$$C(t) = \frac{4\pi}{5} \sum_{m=-2}^{2} < Y^*_{2m}[\Omega_L(0)] Y_{2m}[\Omega_L(t)] >= P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) \tag{3.5}$$

where $\Omega_L(t)$ and $Y_{2m}[\Omega_L(t)]$ denote the polar coordinates and the spherical harmonics function of a vector at time t in the laboratory frame. The second equality is obtained by using the addition theorem for spherical harmonics where $\hat{\mu}(t)$ is instantaneous unit vector in the lab frame.

If Φ is the Euler rotation that carries a coordinate system from coincidence with the lab frame to coincidence with the molecular frame, then $D^2_{pm}(\Phi)$,the Wigner rotation functions, can be used to transform these two frames. Thus

$$Y_{2m}(\Omega_L(t)) = \sum_{p=-2}^{2} D^2_{pm}(\Phi) Y_{2p}[\Omega_M(t)] \tag{3.6}$$

where $\Omega_M(t)$ denotes the instantaneous orientation of a vector in the molecular frame M (Fig.3.1). Using this expression with Eq. 3.5 and assuming the motion of a vector in the molecular frame is now statistically independent of the motion in the lab frame yield

$$C(t) = \frac{4\pi}{5} \sum_{m} \sum_{p} \sum_{q} < Y^*_{2p}[\Omega_M(0)] Y_{2q}[\Omega_M(t)] >< D^{2*}_{pm}(\Phi_0) D^2_{qm}(\Phi) > \tag{3.7}$$

3.2.3.1 Correlation Function for Subunit Motions in a Molecule

For any given subunit (shown as blue cylinder in Fig.3.1), the $< D^{2*}_{pm}(\Phi_0) D^2_{qm}(\Phi) >$ correlation function for subunit motion in molecules with mean local cylindrical symmetry is given by

$$< D^2_{pn}[\Phi_0] D^{2*}_{qn}[\Phi] >= \delta_{pq} \, (\frac{1}{5}) exp[-\frac{1}{2}(6 - n^2) < \Delta^2_x(t) >] \, exp[-\frac{1}{2} n^2 < \Delta^2_z(t) >] \tag{3.8}$$

where $< \Delta^2_z(t) >$the mean squared displacement about the z-axis (body-fixed symmetry) corresponding to twisting of the molecule, whereas $< \Delta^2_x(t) >$the mean squared displacement about the x-axis (transverse axis) corresponding to the bending of the molecule. Again, the mean local cylinder allows the simplifying assumption of$< \Delta^2_x(t) >=< \Delta^2_y(t) >$. Insert Eq. 3.8 into Eq. 3.7, the autocorrelation function of the orientation of a vector can be simplified

as

$$C(t) = \frac{4\pi}{5} \sum_{n=-2}^{2} < Y_{2n}^{*}(\Omega_R(0)) Y_{2n}(\Omega_R(t)) > exp\{-\frac{1}{2}[6-n^2] < \Delta_x^2(t) >\} exp\{-\frac{1}{2} n^2 < \Delta_z^2(t) >\}$$

$$= \sum_{n=0}^{2} I_n(t) B_n(t) F_n(t) \tag{3.9}$$

where the $I_n(t)$ internal correlation functions are defined as

$$I_n(t) = \frac{4\pi}{5} < Y_{2n}^{*}(\Omega_M(0)) Y_{2n}(\Omega_M(t)) > . \tag{3.10}$$

the $B_n(t)$ tumbling correlation functions are defined as

$$B_n(t) = exp\{-\frac{1}{2}[6 - n^2] < \Delta_x^2(t) >\} \tag{3.11}$$

the $F_n(t)$ twisting correlation functions are defined as

$$F_n(t) = exp\{-\frac{1}{2} n^2 < \Delta_z^2(t) >\} \tag{3.12}$$

An important feature of Eq. 3.9 is that the rotational motions of a bead/disc (which is subunit mentioned above) can be expressed simply in terms of the mean-squared angular displacement about its body-fixed symmetry (z) and transverse (x) axes. These displacements include all modes of motion: the length-independent mode (the motion of the vector independent of the molecule at which it is); the internal collective modes (dynamics at one point of the molecule as a consequence of motions from the rest of the molecules); and the uniform modes (overall tumbling of the molecule as a rigid body). For any given "subunit" (which will be a triad or triplet in our case), the mean-squared angular displacements due to the deformable motions about its body-fixed transverse (x) and symmetry (z) axes are given by

$$\frac{1}{2} < \Delta_x^2(t) >= D_{\perp} t + \frac{1}{2} < [\theta_i(t) - \theta_i(0)]^2 > \tag{3.13}$$

$$\frac{1}{2} < \Delta_z^2(t) >= D_{||} t + \frac{1}{2} < [\phi_i(t) - \phi_i(0)]^2 > \tag{3.14}$$

respectively. Therefore, the correlation function can be re-written as

$$C(t) = \sum_{n=0}^{2} I_n(t) C_n^O(t) C_n^I(t) \tag{3.15}$$

where

$$C_n^O(t) = exp[-(6-n^2)D_\perp t]exp(-n^2 D_{||} t) \tag{3.16}$$

is the relaxation attributed from the uniform motion, and

$$C_n^I(t) = exp\{-\frac{(6-n^2)}{2} < [\theta_i(t) - \theta_i(0)]^2 >]\}exp\{-\frac{n^2}{2} < [\phi_i(t) - \phi_i(0)]^2 >]\} \tag{3.17}$$

is the relaxation attributed from the deformable modes, which build to it equilibrium as time goes to infinity.

3.2.3.2 Evaluation of the $I_n(t)$ Correlation Function for Rapid Internal Motions

In the special case of *"rapid isotropic wobble"* of a vector relative to the molecule, the local internal motions relax so rapidly that the collective and the tumbling motions remain nearly stationary during its relaxation, hence, the derived amplitude of isotropic motion should be nearly independent of the assumed equilibrium orientation of the vector. Their only effect is to reduce the total amplitude of the slower relaxation due to bending and twisting motions. The effect of this rapid motion reduces the all three internal decay functions $I_n(t)$ in Eq. 3.9 by the same factor, named "amplitude reduced factor" (ARF) is given by [114]

$$ARF =< P_2(cos\beta) >^2 \cong 1 - 6\sigma^2 \tag{3.18}$$

where β the deflection of a vector away from its equilibrium orientation. The second equality in Eq. 3.18 is the condition for a small amplitude of motion and σ^2 is the root mean squared angle of rotation of the vector around a single transverse axis. This amplitude reduced factor is independent of the orientation of the bond vector and the position where the vector is attached. For instance, ARF=0.9 if $\sigma \cong 7^o$ and ARF=0.73 if $\sigma \cong 12^o$. Therefore, after the complete relaxation of this isotropic wobble, usually less than few hundreds picosecond, the $I_n(\infty)$ internal correlation function are given by

$$I_0(\infty) = ARF[(1/4)(3cos^2\theta - 1)^2] \tag{3.19}$$

$$I_1(\infty) = ARF[3cos^2\theta sin^2\theta] \tag{3.20}$$

$$I_2(\infty) = ARF[(3/4)sin^4\theta] \tag{3.21}$$

where θ the angle of a vector relative to symmetry axis of the molecule. Both ARF and θ will be the adjustable parameters into the fitting protocol. Hereafter, we name $A_0(\theta) = (1/4)(3cos^2\theta - 1)^2$, $A_1(\theta) = 3cos^2\theta sin^2\theta$, and $A_2(\theta) = (3/4)sin^4\theta$.

If macromolecule is a completely **"rigid"** rod, then $ARF = 1$ and $\frac{1}{2} < \Delta_x^2(t) >= D_\perp t$, and $\frac{1}{2} < \Delta_z^2(t) >= D_{||} t$, indicating the rotation of a subunit attributed only from global twisting/tumbling. Thus, the $C(t)$ autocorrelation function in Eq. 3.9 is reduced to Woessner rigid-body diffusion model

$$C^{Rigid}(t) = \sum_{n=0}^{2} A_n(\theta) exp[-(6 - n^2)D_\perp t] exp(-n^2 D_{||} t) \tag{3.22}$$

For the ***flexible*** filaments, assuming the the vector *is under "rapid isotropic wobble"* relative to its molecule with deformable motions, the correlation functions in Eq. 3.9 is expressed as

$$C(t) = ARF \sum_{n=0}^{2} A_n(\theta) B_n(t) F_n(t) \tag{3.23}$$

If removing the overall rotational motions away (e.g superimpose the trajectory into reference structure), then the conventional $C_{int}(t)$ internal correlation function can be expressed by

$$C_{int}(t) = ARF \sum_{n=0}^{2} A_n(\theta) C_n^I(t) \tag{3.24}$$

As we can see clearly, the $C_{int}(t)$ internal correlation function depends on the orientation of the dipole-dipole interactions with respect to the symmetry axis of the diffusion tensor, indicating the dynamics due to deformable motions is governed by anisotropic deflection potentials. Eq. 3.23, the expression for the correlation function, will be used in the rest of the analysis in this paper.

3.2.3.3 The Collective Bending and Twisting Correlation Functions within WBR Model

In the WBR model, the $F_{n,WBR}(t)$ twisting correlation function is given by (see Appendix 3.5.1)

$$F_{n,WBR}(t) = exp\{-n^2[D_{||} t + \frac{1}{2} < [\phi_i(t) - \phi_i(0)]^2 >]\}$$

$$= exp[-n^2 D_{||} t] exp[-n^2 \frac{k_B T}{\alpha} Q \Lambda^{-1} (1 - e^{-\alpha \Lambda t/\gamma_t}) Q^T]$$

$$= F^O_{n,WBR}(t)F^I_{n,WBR}(t) \tag{3.25}$$

where $F^O_n(t)$ is uniform rigid-body rotating about the symmetry axis, and $F^I_n(t)$ is collective torsional motion from deformable modes. $F^I_n(t)$ depends upon the torsional elastic constant α, twisting friction factor γ_t and the N of subunits. The relaxation time for the lth normal twisting mode is

$$\tau(twist)_l = \gamma_t/\alpha\lambda_l. \tag{3.26}$$

where λ_l is the lth eigenvalue in the matrix Λ, in which the longest relaxation time is given by

$$\tau_1 = \gamma_t/(4\alpha sin^2(\pi/2(N+1))). \tag{3.27}$$

Similarly , the $B_{n,WBR}(t)$ tumbling correlation function is given by (see Appendix 3.5.2)

$$B_{n,WBR}(t) = exp[-(6-n^2)D_\perp t]exp[-(6-n^2)\frac{k_BT}{\kappa}(\delta Q_b)\Lambda_b^{-1}(1-e^{-\kappa_b\Lambda_b t/h_b^2\gamma_b})(\delta Q_b)^T]$$

$$= B^O_{n,WBR}(t)B^I_{n,WBR}(t) \tag{3.28}$$

where $B^O_{n,WBR}(t)$ is uniform rigid-body rotation about its transverse axis and $B^I_{n,WBR}(t)$ is internal bending correlation function contributed from deformable modes. The relaxation time for the lth normal bending mode is

$$\tau(bend)_l = \gamma_b h_b^2/\kappa_b\lambda_{bl}. \tag{3.29}$$

where λ_{bl} is the lth eigenvalue in the matrix Λ_b.

We borrow the idea from the work by Robbison *et al.* [113], who have adopted the simplicity of the Lipari-Szabo model in terms of S^2 and internal relaxation times τ_{int} to characterize the internal motions. For each term ($n = 0, 1, 2$) in Eq 3.24 , the WBR internal correlation function from the internal modes are compared to stretched exponentials:

$$C^I_{n,WBR}(t) = B^I_{n,WBR}(t)F^I_{n,WBR}(t) = [S_n^2 + (1-S_n^2)exp(-(\frac{t}{t_n})^{\beta n})] \tag{3.30}$$

where S_n^2 is the order parameter attributed from combination of twisting and bending modes with value of n

$$S_n^2 = exp(-n^2(\frac{k_BT}{\kappa}Q_b\Lambda_b^{-1}Q_b^t))exp(-n^2(\frac{k_BT}{\kappa}Q_b\Lambda_b^{-1}Q_b^t)).$$

Therefore, the deformable order parameters S_D^2 for ith rod unit (triplet) computed when $t \to \infty$ is given by,

$$S_{D,i}^2 = \sum_{n=0}^{2} A_{i,n}(\theta)S_{i,n}^2 \tag{3.31}$$

and finally the WBR-derived order parameter is

$$S_i^2 = ARF_i \sum_{n=0}^{2} A_{i,n}(\theta)S_{i,n}^2 \tag{3.32}$$

which is determined by the local rapid motion, the orientation in the rod unit frame, and the position along the sequence. The deformable-derived order parameters, are different from the conventional order parameter due to dynamics, both local and collective.

3.2.4 Spectral Density Function from Deformable Macromolecules

Using Eq. 3.22,3.24,3.25, and 3.28, total correlation function of a given spin vector in Eq. 3.23 can be used to compute the spectral density function $J(\omega)$ as following

$$J(\omega) = \int_{-\infty}^{\infty} C(t)cos(\omega t)dt$$

$$= ARF \sum_{n=0}^{2} 2A_n(\theta) \int_0^{\infty} B_n^O(t)F_n^O(t)B_n^I(t)F_n^I(t)cos(\omega)t$$

$$= ARF \sum_{n=0}^{2} A_n(\theta)2 \int e^{-t/t_n^o}[S_n^2 + (1-S_n^2)e^{-(t/\tau_n)^{\beta n}}]cos(\omega t)dt \tag{3.33}$$

where $n = 0, 1, 2$. Once we obtain the distribution function P_i shown in Eq 3.63 for the stretched exponential function, $C_{kww}(t)$, spectral density $J(\omega)$, can be computed as following [113],

$$J(\omega) = ARF\{\sum_{n=0}^{2} A_n(\theta)[S_n^2 \frac{2\tau_n^o}{1+(\omega\tau_n^o)^2} + (1-S_n^2)\sum_{i=0}^{N} P_i \frac{2\tau_{n'}u_i}{1+(\omega\tau_{n'}u_i)^2}]\} \tag{3.34}$$

where $(\tau_n^o)^{-1} = [(6-n^2)D_\perp + n^2 D_{||})]$ the overall correlation time and $\tau_{n'}^{-1} = (\tau_n^o)^{-1} + \tau_n^{-1}$ the effective internal correlation times.

In the rigid model, the spectral density function is expressed as following,

$$J(\omega) = S^2 \sum_{n=0}^{2} A_n(\theta) \frac{2\tau_n^o}{1 + (\omega\tau_n^o)^2} + (1 - S^2) \frac{2\tau'}{(1 + (\omega\tau')^2} \tag{3.35}$$

where $1/\tau^{'} = 1/\tau_M + 1/\tau_I$, τ_M and τ_I denoting the effective relation times from overall tumbling of the molecule and internal motion, respectively.

3.2.5 Application to Collagen Model Peptides

To determine whether the backbone motion of the collagen-like peptide in the MD simulations can be interpret by the WBR model or furthermore there is the sequence dependence of collagen-like peptide flexibility, we apply the theory developed by Schurr and co-workers for the elastic motion of the bending duplex DNA. Following their study, some assumptions are also made in collagen-like peptides: (1) collagen-like peptide is assumed to behave like a flexible rod that has mean-local cylindrical symmetry; (2) each triplet is coupled to its neighbors by a harmonic twisting potential with the α force constant between consecutive triplets; (3) each triplet is coupled to its neighbors by a harmonic bending potential and the force constant κ between triplets, which is assumed to be the same for the rotation about x and y axes; (4) we assume that there is a single bending and a single twisting force constant between consecutive triplets, and that each acts to straight the chain; (5) although the collagen-like peptide is not as rigid but still stiff enough that the local twisting and bending of motions are uncoupled; (6) the uniform rotational motions are the rotation around the symmetric axis (which is the helical axis of the triple helix) and about the y/x axis.

3.2.5.1 Parameter Fitting

The parameters used in WBR model are:

1. the ARF is the amplitude reduction factor due to rapid isotropic wobble of the bond vector in the local frame;

2. the torsion elastic constant α, which governs the dynamics of twisting deformations;

3. the bending elastic constant κ, (or equivalently the dynamics persistence length $P_d = h\kappa/k_BT$, where h is the rise between triplets and $k_B$is Boltzman's constant), which governs the dynamics of the bending deformations;

4. the hydrodynamics radius R_H related to the friction coefficient $\gamma \sim R_H^2$;

5. the diffusion coefficient for the symmetry rotation $D_{||}$ and combined with (4), the friction factor γ per triplet for rotation the symmetry axis is obtained, $\gamma = k_BT/[D_{||}(N + 1)]$;

6. the diffusion coefficient for the transverse axis rotation (usually called end-to-end tumbling) $D_{\perp}$;

7. the orientation θ of bond vector in the local frame (triplet);

8. other general constants are N (number of subunit, triplet), h (length of subunit, rise between two triplets), ζ (solvent viscosity), and T (at room temperature).

The parameters are determined through a fitting procedure against atomistic MD simulation results. Some general criteria apply to decide which parameters to fix and which to adjust. Usually, a parameter should be fixed when it is known to be the level of precision from "independent" experiments or simulations or theory. For example, in the case of MD simulation, h, L, and R_H are well-known and are not expected to vary significantly among different collagen-like peptides, e.g. from PPG10 peptide to POG10 peptide. Diffusion coefficients should be not varied among the different samples due to the similar size of R_H, h and N. The value of the equilibrium angle θ of the vector attached to the collagen model peptides can be computed directly from MD simulation. The mean and standard deviations of the orientations of the NH vectors in the collagen-like peptides have been discussed in previous chapter (see Chapter 3, Table 4.6). In the case of POG10 peptide, since the mean value and the corresponding standard deviation are similar through the peptide. Thus, the values of θ and ARF can be set to be the same for each disc. Therefore, in the fitting procedure, the only two adjustable parameters left are the force constants, α and κ which may be varied with different sequence compositions. We fit the specific properties (i.e. order parameters S^2, mean-squared angular displacements $< \Delta_x(t)^2 >$), which can be achieved by both MD trajectories and the WBR model.

The constants used for the calculation in the rest of paper are: N=27 (number of subunits in the collagen-like peptide), R_H =7 Å, h=3.2 Å, T=293 K, and ζ=0.01 poise, and $\gamma = 4\pi\zeta {R_H}^2 h$ dyne cm s. With those input parameters, the uniform diffusion constants about the symmetry and transverse axes computed from theory by Tirado and Garcia (TG) [115] are 0.0611 and 0.005 ($\times 10^9/s$), respectively. The complete set of parameters for the WBR model are summarized in Table 3.1.

Parameters	Numerical values
Equilibrium disk length h [Å]	3.2
Equilibrium polar angle θ [degree]	90
Number of discs	27
solvent viscosity ζ [poise]	0.01
temperature [K]	293
Hydrodynamics radius R_H[Å]	7

Table 3.1: *Summary of parameters of the weakly bending rod model, derived from atomistic MD simulations*

3.2.5.2 Model Fit: MD Simulation v.s. WBR model

In this section, we briefly examine whether the convectional order parameters S^2 derived from the MD trajectory of collagen model peptides POG_{10} and PPG_{10} can be fitted with WBR model. As these two peptides with identical triplets (containing Gly-Pro-Hyp and Gly-Pro-Pro for POG_{10} and PPG_{10}, respectively) have uniform height and radius throughout the triple helix, we assume that each triplet is connected to its neighbors on both sides through bending and twisting springs with the same elastic force constant κ and α, respectively, which acts to straighten the chain.

The plot of the order parameters S^2 versus the position of glycine of NH vectors (Fig. 3.2) shows that plateaus of the internal correlation functions of NH vectors from MD trajectories are captured well with WBR model, indicating a decrease in S^2 toward the termini as a consequence of rod flexure and twisting. The best fitted parameters are listed in Table 3.2. However, the best fitted values are not really consistent with what observed in MD trajectories. For example, infinity value in bending force constant represents the absence of bending modes in the collagen-like peptides, but this motion has been seen clearly in PCA results. The main reason for this infinity of the bending force constant is the fitting procedure with two variables having no constraints when having order parameter S^2 as a target.

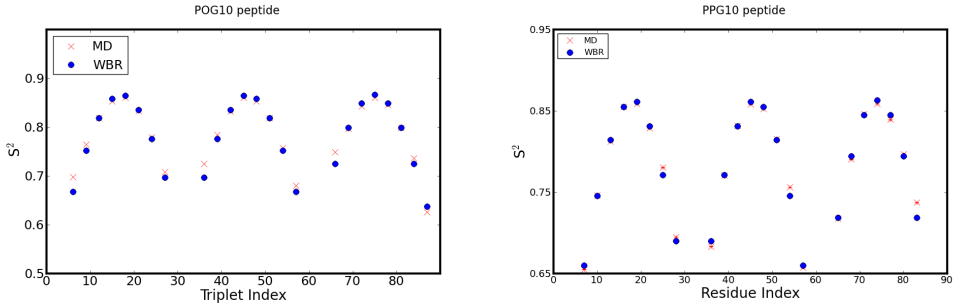


Figure 3.2: *The order parameter S^2 as a function of position of vectors along the collagen-like peptides: POG10(left) and PPG10 (right). MD-derived S^2 denoted as Red ($\times$) and WBR-derived S^2 denoted as filled blue ($\bullet$). The fitted force constants are shown in Table 3.2. All the vectors are assumed to have the same orientation with $\theta = 90^0$ relative to the symmetric axis in the molecular frame.*

	α (K_BT)	κ (K_BT)
POG10	46	$\sim 10^8$
PPG10	44	$\sim 10^8$

Table 3.2: *Summary of the force constants κ and α, computed using the best-fit parameters for the S^2 of NH vectors of the collagen-like peptides.*

Using order parameters as a point of contact may not give the appropriate twisting and bending force constants, which can match nicely the collective motions existing in the collagen-like peptide from MD simulation, but gives a general idea for the order parameters profiles having highest values in the middle of the helix and gradually decreased values toward the end. (In next section, we will use other property as a target to find more rational force constants) As discussed in previous chapter, the analyses on the orientation of the NH vector have shown that the NH vector at the end does not deviate that much from the one in the middle, suggesting that the local rapid internal motion might not the one dominating the lower value of order parameter at the termini than the one in the middle by ~0.2. Excellent fits with WBR model shown here suggest that it is the global bending or twisting modes dominates the deviations from rigid-rod behavior.

3.3 Using the WBR Model to Analyze MD Simulations of Collagen Model Peptides

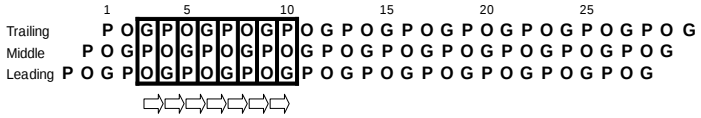
It has been shown in Sec. 3.2 how the time correlation function of a vector attached to a subunit of the filament could be expressed simply in term of the mean squared angular

displacements. The objective in this section is to compute the mean-squared angular displacements $< \Delta_x(t)^2 >$ in Eq. 3.13 and $< \Delta_z(t)^2 >$ in Eq. 3.14 separately as a point of contact, since they can be readily determined from both molecular dynamics simulations and the WBR model. As discussed in earlier section, excellent fits in order parameters suggest the collective bending or twisting modes dominating the deviations from rigid-rod behavior. WBR model also assumes the rod flexes and twists independently, therefore, one expects to fit $< \Delta_x(t)^2 >$ to obtain the bending restore force constant, and fit $< \Delta_z(t)^2 >$ to compute the twisting force constant separately. Once these force constants are given, all other properties, i.e. time correlation functions, order parameters, relaxation rates are readily computed within WBR model (Sec. 3.2.3, 3.2.4) and the impact of the rod flexure can be explored.

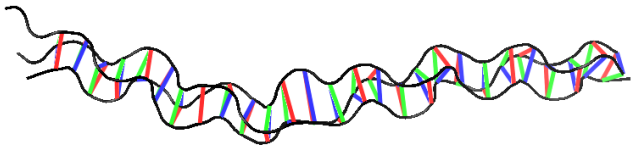
An attempt here is to demonstrate how to compute the values of $< \Delta_x(t)^2 >$ and $< \Delta_z(t)^2 >$ for each subunit (triplet) around its transverse (x) axes and instantaneous symmetry (z), respectively, from MD snapshots. First, we define a subunit and its corresponding rotation axes in collagen-like peptide. The connected subunit and the body-fixed coordinate axes of a given subunit are shown in Fig. 3.3. An instantaneous coordinate rod frame is defined in the molecule (each snapshot) as follows. The z-axis of the whole rod is taken parallel to the instantaneous end-to-end vector and is chosen pass through the center of mass of the molecule. The x and y axes are taken to be any two orthogonal axes in a plan perpendicular to z axis.

From the point of view of WBR model, the collagen-like peptide containing only (Gly-Pro-Hyp) or (Gly-Pro-Pro) repeated triplets is assumed to be composed of N "identical" triplet with unit height h ∼3.2 Å and hydrodynamics radius R_H. Each triplet is connected to its neighbors on both sides through bending and twisting springs with elastic force constant κ and α, respectively, which acts to straighten the chain.

(a)



(b)



(c)

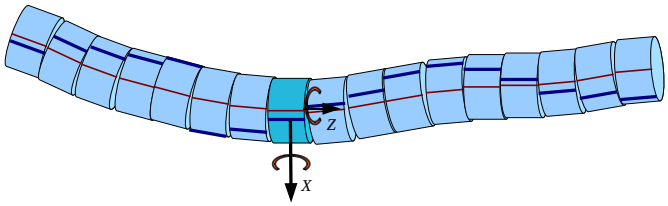


Figure 3.3: *The mean local symmetry top model for torsion and bending motions in collagen-like peptide. (a) Schematic representation of amino acid composition, subunit (black box) and subunits step parameters (arrow) of POG10 peptide. (b) The collagen-like peptide is considered consist of a set of triplets (subunits) with different orientations in torsion and bending angles relative to each other. One triplet, triangle plane, is defined with three atoms Gly, Pro, and Hyp from each strand in the same helical axis level. Lines between Gly and Pro, Pro and Hyp, Hyp and Gly are marked as red, blue and green, respectively. (c) The peptide can be modeled as a string of discs connected by Hooke twisting and bending springs. The mean squared angular displacement* $< \Delta_j(t)^2 >$, $j = x$ *or* z. *rotation about the instantaneous* $\hat{z}$ *and* $\hat{x}$ *axes of a subunit disk. Cylindrical symmetry here in the sense that* $< \Delta_x(t)^2 > = < \Delta_y(t)^2 >$ *for any pair of transverse* $\hat{x}$ *and* $\hat{y}$ *axes. Twisting modes:* $< \Delta_z(t)^2 >$*is the mean squared angular displacement about body-fixed z axis. Bending modes:* $< \Delta_x(t)^2 >$ *is the mean squared angular displacement about body-fixed x axis.*

3.3.1 Mean-Squared Angular Displacements

In MD, the instantaneous twisting angle $\phi(t)$ between any two consecutive triplets of collagen-like peptide for each snapshot is calculated using 3DNA algorithm [86], with which the "six step parameters": *shift, slide, rise, tilt, roll* and *twist* between the any successive triplets are computed (more detailed information can be found in Methods section in Chapter 3). The instantaneous bending angle $\theta(t)$ is defined as the angle between the local z-axis

(normal vector of each triplet) and the end-to-end axis. Inserting the twisting and bending angles into Eq. 3.13 and 3.14, respectively, yields the values of$< \Delta_x(t)^2 >$and $< \Delta_z(t)^2 >$of each triplet, which are displayed in Fig. 3.4.

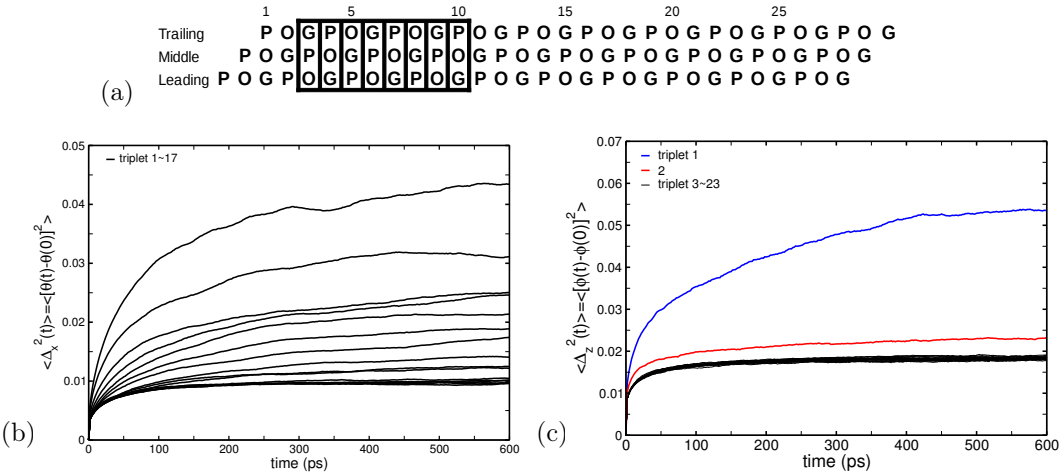


Figure 3.4: *(a) Schematic representation of amino acid composition, base (black box) and bases step parameters (arrow) of POG10. The mean-squared angular displacements: (b) bending,* $< \Delta_x^2(t) >=< [\theta(t) - \theta(0)]^2 >$*, where* $\theta(t)$ *is the angle between the normal axis of the triplet to the end-to-end axis; (c) twisting,* $< \Delta_z^2(t) >=< [\phi(t) - \phi(0)]^2 >$*, where* $\phi(t)$ *is the angle between two successive triads at time t and is calculated using 3DNA. Number n triplet represents the angles between the nth triad and the* $(n+1)$*th triad counting from N terminus of the peptide POG10.*

As all conformers are superimposed to the reference frame, indicating that no uniform motion contribute to these displacements, so that the slopes $\frac{d<\Delta_x(t)^2>}{dt}$ and $\frac{d<\Delta_z(t)^2>}{dt}$ become zero after all the internal modes have completely relaxed. For example, in Fig. 3.4 (a), the middle triplet (13∼15) in peptide POG10 has a slope of $\frac{d<\Delta_x(t)^2>}{dt}$ that reaches zero after ∼100 ps, whereas the triplet toward the terminal take longer time to relax completely.

With the definition for the bending angle, Fig. 3.4 (a) shows that the value of $< \Delta_x(t)^2 >$ of a triplet decreases gradually when the triplet towards to the center of the helix, consist with the results from WBR theory. However, it is not the case for the value of $< \Delta_z(t)^2 >$. Fig. 3.4 (b) shows that expect the two triplet at the termini, all the triplets have more and less similar curves in $< \Delta_z(t)^2 >$, indicating that this angular motion behaves the "local rapid-like" motion instead of the collective-like motion. It might be the definition for the angle used here, therefore, an appropriate twisting angle representing the normal mode is necessary to characterize the twisting elasticity of the rod. We will not investigate the calculation of twisting angle in this dissertation but in the future research. In the following, we focus on the rod flexure and characterize the rigidity of the rod in terms of the bending

force constant.

3.3.2 Model Fit

In this fit, the comparison of the values of the $< \Delta_x^2(t) >$ for the central triplet with those computed from normal mode theory yields the bending force constant $\kappa \sim 460$ (K_BT). The agreement is fairly good for the central triplet, but a discrepancy emerges for the triplet toward the end, as is evident from Fig. 3.5, indicating that the bending angle of a triplet at the end experiences some kind of motions other than the collective bending modes. Overall, we conclude that the WBR theory works well for the backbone global bending motions of collagen-like peptide in MD simulation, especially for the triplets away from the termini.

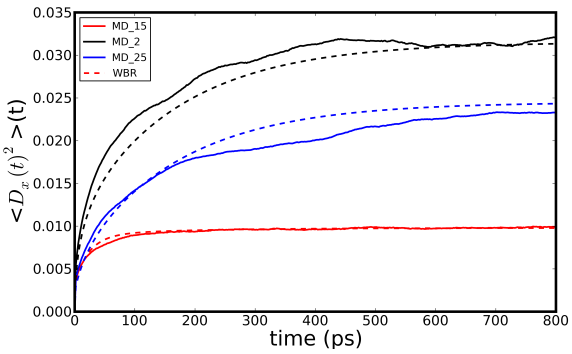


Figure 3.5: *Comparison of the MD simulations (solid lines) with the normal mode theory (dashed lines) in terms of the value of* $< \Delta_x^2(t) >$ *for the central triplet 15 and the two (2,25) triplets at the end. The fitted bending force constant* $\kappa \sim 460$ (K_BT)

We have addressed the problem of getting the values of $< \Delta_z^2(t) >$ for the triplets that are dominated from rod twists from MD simulation. But with the fitted bending force constant κ, it is possible to obtain the twisting force constant α. Excellent fits to S^2 gives the twisting force constant $\alpha \sim 52$ (K_BT), suggesting the significant contribution of rod twists to the order parameters (the difference between red and blue circles) as is evident from Fig. 3.6, if the model admits both bending and torsion motions. In conclusion, neglecting the influences of collective twisting and bending motions leads to a very large estimate for the amplitudes of the local librational modes of the NH bond vectors. It has been suggested in the literature by Schurr et al. [110], in which collective modes are taken into account when reevaluating experimental NMR relaxation data, resulting in a significantly smaller amplitude for the local motions of the bases.

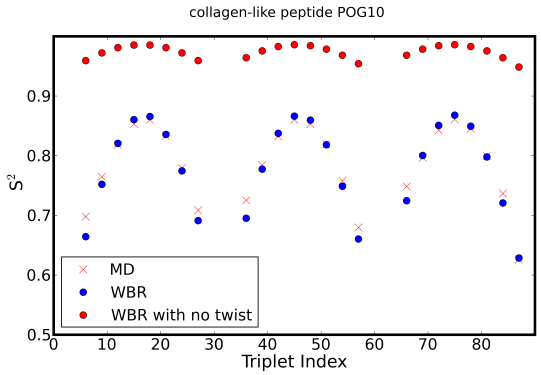


Figure 3.6: *The order parameter S^2 as a function of the NH vectors of the Gly residues in POG10. MD-derived S^2 denoted as Red (×) and WBR-derived S^2denoted as filled dot. The WBR model that admit only bend with force constant $\kappa \sim 460$ (K_BT) gives the values of S^2 denoted as filled dot (●). With a given bending force constant $\kappa \sim 460$ (K_BT), the best fitted twisting force constants is $\alpha \sim 52$ (K_BT) shown as filled dot (●). All the vectors are assumed to have the same orientation with $\theta = 90^0$ relative to symmetric axis in the molecular frame.*

Additionally, this deformable-induced order parameters shows the orientational-dependence if the rigidities between torsion and bending motions are different (Fig. 3.7). This dependence is more significant for the rod toward the end (blue dot in Fig. 3.7). This difference can be understood as follows. Internal torsion modes do not affect the vector pointing to the symmetric axis but become pronounced for vector orienting away from this main axis, while bending modes affect the most for the vector aligns along the rod. Only when the torsion and bending have similar rigidities, this deformable order parameter becomes less sensitive to its orientation (diamond in Fig. 3.7). This relation is not possible with the model-free formula because it assumes that the local amplitude fluctuation of the vector is independent of its orientation with respect to the molecular frame.

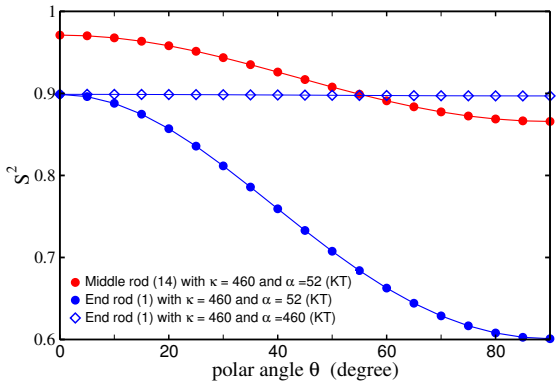


Figure 3.7: *The WBR-derived order parameter S^2 as a function of the orientation of a NH vector in the molecular frame. A vector at the middle rod labeled as red and at the end rod shown as blue. The WBR model that admit both bend with force constant $\kappa \sim 460$ (K_BT) and torsional force constant $\alpha \sim 52$ (K_BT) denoted as filled dot (●, ●), and rigidities with $\kappa \sim 460$ (K_BT) and $\alpha \sim 460$ (K_BT) denoted as triangle (◇).*

As mentioned earlier, using order parameters S^2 as a contact between MD and WBR model requires the information of the local internal motion of a vector attached to the triplet i.e. equilibrium polar angle and ARF, since the order parameter is the product of the vector orienting in the molecular frame. An advantage of using the mean-squared angular displacement of a triplet as a target is that the knowledge of the local internal motions is not needed. As we have assumed that the value of the mean-squared angular displacement of a triplet is purely contributed from the collective modes of bending or twisting. However, this assumption could be incorrect, as the collagen bends in solution simulation, the triplets do not actually move in straight lines perpendicular to the end-to-end vector as assumed in the WBR theory.

The point in this work is not that bending force constant is actually 460 (K_BT) or twisting force constant is 52 (K_BT), but that, under some simple assumptions, to access the possible and relative contributions of collective and localized internal motions to the MD data i.e. order parameters, which might not require that much larger amplitude of the internal angular motions as predicted in rigid body model. We show below that how this WBR model that admit twist and bend deviates the NMR data from rigid rod behavior.

3.3.3 Relaxation Data: WBR model v.s. Rigid-Rod Model

With the given information (i.e. MD-derived order parameter S^2 and geometry structure for POG10) from MD simulation, we look forward to computing the nor relaxation data using two motional models: rigid-rod and the WBR rod model. Calculated values of R1, R2

and NOE for the rigid rod and the present flexible model with the force constants, which are obtained in previous section, are plotted as a function of the NH vector of Gly residues for POG10 peptide in Fig. 3.8.

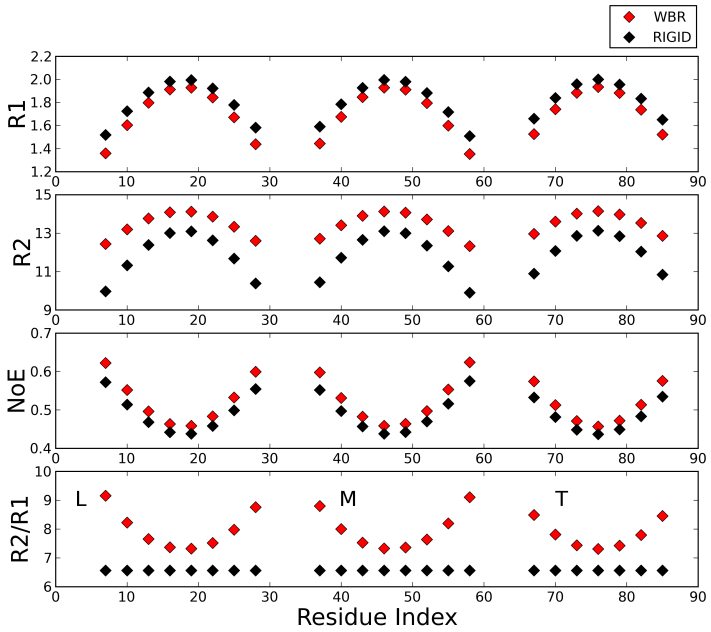


Figure 3.8: *Relaxation data (R1,R2,NOE, and R2/R1) for each NH vector of glycine residues in POG10 peptide using WBR model (red ◇) and Rigid model (blue ◇). The calculations of relaxation data excludes the contributions from internal relaxation times. The polar angle of each NH vector is set to* $\theta = 90^0$. *In the WBR model, the torsion force constant* $\alpha = 50\ K_BT$ *and the bending force constant* $\kappa \sim 460\ K_BT$.

In the rigid rod model, only the large amplitudes of the local internal motions are considered into the decrease in order parameters and the expression of the spectral density function in Eq. 3.35 is applied to get the nmr relaxation data. This model assumes the isotropic internal motion, in other words, the local amplitude fluctuation of the vector is independent of its orientation in the molecular frame. As a result, taking the ratio of R2 and R1 should eliminate the contribution from internal motions and only be related to the orientation of that vector in molecular frame. It is expected to get a constant value of R2/R1 for the NH vectors in Fig. 3.8 since all the vectors are assumed to be perpendicular to symmetric axis of the collagen rod.

When account is taken of the collective twisting and bending deformation, there are some deviations from rigid-rod model as evident in Fig. 3.8. The internal twisting and bending motions shift the values of R1 and R2 compared to the rigid rod but keep the similar trend

along the collagen-like peptide. The most striking deviation is the positional-dependence of the values of R2/R1 for the vector with same orientations due to the different twisting and flexure rigidities along the rod. A vector toward the end is more sensitive to the deformable motion than the one in the middle, where local amplitude fluctuations induced by the collective modes is governed by a less anisotropic deflection potential.

In conclusion, if the twisting and bending rigidities are smaller, their amplitude fluctuation as isotropic internal motion will be reduce, hence, the positional-dependence of the ratio of R2/R1 become pronounced.

3.4 Conclusion

NMR relaxation has been recognized as having the greatest potential to provide the most comprehensive picture of motional amplitudes and frequencies in a protein through the interpretation by the model-free approach. However, this simple approach is only valid for isotropic overall tumbling and when internal motions occur in a highly symmetric fashion. When it comes to the collagen-like peptide, which is better approximated as a cylinder and with highly anisotropic overall tumbling, this widely used Lipari-Szabo model free approach should not be used.

The primary focus of this work concerns the modeling of the torsional and bending motions of linear triple helical collagen-like peptide with 30 residues on each strand to explain the possible internal motions causing the order parameter profile. Both torsional and bending motions are explained by analytic theories with torsion or bending elastic constant which is uniform along the helix (with host sequence POG or PPG repeated triplets). This present flexure model provides some features that model-free approach does not offer. Firstly, WBR model predicts orientational-dependence of order parameters if the torsion and bending rigidities are small and different from each other. This is not possible with the model-free formula because it assumes that internal motion of a given vector undergoes in isotropic deflection potential. Secondly, WBR model shows the internal motion is in dependence on the position along the triple helical peptide. The order parameters S^2 decreases as the position of the bond vector away from the middle to the terminus of the peptide. Therefore, the site specific label can be a good test of the internal dynamics.

In the next step, we turn to a more general formulation of the WBR model that considers overall tumbling as well as internal motions. Here instead of having one unit for each peptide "step" of the triple helix, the model is more coarse-grained, using eight spheres to model the thirty-residue per chain peptides considered here, and we call this the "eight-bead" model. This provides a model for the coupling between overall (tumbling) and internal (twisting and bending) motions. We consider a range of parameters that could describe many sorts of

rod-like molecules. Our preliminary explorations suggest that, in most cases, such motional coupling has only a small effect on correlation times of the orientation of a vector (data not shown), but further exploration of this model seems warranted.

Another on-going investigation in this project is the concern of the dependence of collagen-like peptide flexibility on sequence composition. MD simulations have clearly indicated that the bending behavior of T3-785 peptide with amino-acid rich residues is much different than other peptides (in Chapter 3). It is not clear at this point how severe this dependence is, but elucidating the effects of the varying amino acid sequences on this behavior will be necessary to understand better how the chemical structure and overall mechanical deformation motions are related in details.

3.5 Appendix

Here the theory of the weakly bending model developed by Schurr and coworkers is briefly presented and the reader can refer the papers for more details of derivations [108, 112, 113]. The deformable macromolecule is regarded as a linear but it is not complete rigid and straight. The model assumes that the rod is consist of N identical spherical beads whose centers are connected by N-1 virtual bond vector of fixed length h as shown h_i of fixed length. Between each bead of successive bond vectors are the Hooke spring and bending spring that act to straighten the chain. The z-axis of the rod is chosen to be parallel to the instantaneous end-to-end vector passing through the center of mass. The x and y axes are initially taken to be any two orthogonal axes in a plane perpendicular to z. Input parameters of WBR model are

1. Coupling constant $\alpha$and κ. for constructing the matrix A.

2. Hydrodynamics interaction H.

   a) Since the hydrodynamics interactions are only know for spherical beads, when we are dealing with a series of cylindrical rod, it must be substituted by an equivalent sphere. The optimum sphere size is chosen to ensure that relaxational times for the effective DNA model reproduce the translational and rotational dynamics for the uniform modes of the rigid object subject to the constraint $hN = h_{bead}N_{bead}$ with N_{bead} has to be integer. For example, according to Song's work [108], it has been shown that the optimum bead size has radius $R = 15.9$ or $h_b = 31.8$ for DNA chain.

   b) Size of beads (radius R) for computing the friction factor γ_b or γ_{rb}.

   - The translational friction factor $\gamma_b = 6\pi\eta R = 3\pi\eta h_b$

- Rotational friction factor, $\gamma_{rb} = 8\pi\eta R^3 = \pi\eta h_b^3 = \frac{1}{3}\gamma_b h_b^2$ where η is defined as the force per unit area needed to maintain unit velocity gradient between two parallel surfaces moving relative to one another in the fluid. The cgs units of η are , or poise.

a) scale factor S

- $\Lambda_b(N)S^{0.33} = \Lambda_b(N_b)$

3.5.1 Twisting

The objective in this section is to seek the mean-squared angular amplitudes as a function of time. We start with the diffusion equation for twisting motion of the N discs, which is most easily represented by the Langevin equation than bending motion.

Single oscillator

To be less complicated, we first consider an isolated oscillator in a harmonics potential $U = \frac{1}{2}\alpha\phi^2$. The Langevin equation for the twist angle $\phi(t)$ for the one oscillator is expressed in terms of few variables:

$$I\ddot{\phi}(t) + \gamma_t\dot{\phi}(t) = F(\phi) + R(t) \tag{3.36}$$

where $\gamma_t\dot{\phi}$ represents drag force with friction factor γ_t; I is the rotational inertia; $F(\phi) = -\frac{dU}{d\phi} = -\alpha\phi$ the torque arising from the potential U; and $R(t)$ is the torque from the Brownian fluctuations of the solvent with the properties of $< R(t) >= 0$, and $< \phi(t^{'})R(t) >= 0$, and $< R(t^{'})R(t)) = 2(k_BT/\gamma_t)\,\delta(t - t^{'}) >$. In the overdamped regime or when twisting potential U is large compared to inertial term, i.e. $F(\phi) >> I\ddot{\phi}$, then the Langevin equation is approximately $\gamma_t\dot{\phi} + \alpha\phi = R(t)$. With the properties of $R(t)$, we can write down the Langevin equation for the average value of ϕ and $\phi(t)\phi(0)$ separately,

$$\gamma_t < \dot{\phi}(t) > +\alpha < \phi(t) >= 0 \tag{3.37}$$

$$\gamma_t\frac{d < \phi(0)\phi(t) >}{dt} + \alpha < \phi(0)\phi(t) >= 0 \tag{3.38}$$

and the solutions to these two equations are

$$< \phi(t) >=< \phi(0) > e^{-\alpha t/\gamma_t} \tag{3.39}$$

$$< \phi(0)\phi(t) >=< \phi(0)^2 > e^{-\alpha t/\gamma_t} \tag{3.40}$$

Since the distribution is independent of time (as the diffusive process is assumed to be a stationary process) and is always at equilibrium state, the correlation function amplitude is $< \phi(0)^2 >=< \phi(t)^2 >= k_BT/\alpha$. With these properties of the results, the mean-squared angular displacement $< [\phi(t) - \phi(0)]^2 >$ can be further expressed as

$$< (\phi(t)-\phi(0))(\phi(t)-\phi(0)) >= 2(< \phi(0)^2 > - < \phi(t)\phi(0) >) = \frac{2k_BT}{\alpha}(1-e^{-\alpha t/\gamma_t}) \quad (3.41)$$

N coupled torsional oscillators

Similar to above expressions for a single oscillator, the Langevin equations for N coupled torsional oscillators (i.e. discs) in the WBR can be written together in terms of a matrix equation:

$$I\ddot{\vec{\phi}} + \gamma_t\dot{\vec{\phi}} + \alpha A\vec{\phi} = \vec{R}(t) \quad (3.42)$$

where a vector of angular displacements $\vec{\phi}$ for N oscillators and N by N matrix A, which are

$$\vec{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ \phi_N \end{pmatrix} \quad (3.43)$$

$$A = \begin{pmatrix} 1 & -1 & 0 & . & . & . & 0 \\ -1 & 2 & -1 & & & & 0 \\ 0 & -1 & 2 & & & & . \\ . & 0 & -1 & 2 & -1 & . & . \\ . & . & & & . & & . \\ . & . & & & -1 & 2 & -1 \\ 0 & . & . & . & 0 & -1 & 1 \end{pmatrix} \quad (3.44)$$

These N coupled Langevin equations (as the form in A matrix) can be decoupled by diagonalizing the matrix A, which yields N eigenvalue Λ, by operating transformation matrix Q to A: $Q^tAQ = \Lambda$. Therefore, the uncoupled normal twisting Langevin equations can be expressed as

$$I\ddot{\overrightarrow{\rho}} + \gamma_t\dot{\overrightarrow{\rho}} + \alpha\Lambda\overrightarrow{\rho} = Q^tR(t) \quad (3.45)$$

The lowest eigenvalue of Λ, $\lambda = 0$, corresponding to the uniform motion, and the associated

column vector from Q will be removed in the calculation of internal motions. The resulting matrix $Q\Lambda^{-1}Q^T$ is so-called the pseudo-inverse of A (due to removing the uniform mode).

There are couple things about these uncoupled Langevin equations. (i) Since molecular inertial motions in liquids normally has damping times of the order of picoseconds or less, contributions to the equations of motion involving inertial terms might be ignored; (ii) since the torsional oscillations of small elements probably do not perturb the fluid enough to cause any significant effect, the hydrodynamics interaction through twisting is also omitted (this term will be included in the case of bending). Same procedures used for a single oscillator are used here, therefore, the twist autocorrelation functions for the normal modes can be expressed as

$$< \overrightarrow{\rho}(0)^2 >=< \overrightarrow{\rho}(t)^2 >= \frac{k_BT}{\alpha}\Lambda^{-1} \tag{3.46}$$

$$< \overrightarrow{\rho}(t)\overrightarrow{\rho}^t(0) >=< \overrightarrow{\rho}(0)^2 > exp(-\frac{\alpha\Lambda}{\gamma_t}t) = \frac{k_BT}{\alpha}\Lambda^{-1}exp(-\frac{\alpha\Lambda}{\gamma_t}t) \tag{3.47}$$

By substituting $\overrightarrow{\rho} = Q^t\overrightarrow{\phi}$ back to those equations above, the self and cross correlations of the mean-squared angular amplitudes for the twisting motions is

$$< \overrightarrow{\phi}(t)\overrightarrow{\phi}^t(0) >= \frac{k_BT}{\alpha}Q\Lambda^{-1}exp(-\frac{\alpha\Lambda}{\gamma_t}t)Q^t \tag{3.48}$$

, and the equilibrium average in a compact matrix form is:

$$< \overrightarrow{\phi}(0)\overrightarrow{\phi}^t(0) >= \frac{k_BT}{\alpha}Q\Lambda^{-1}Q^t = \frac{k_BT}{\alpha}A^{-1} \tag{3.49}$$

The mean-squared angular displacement due to twisting motions is

$$\begin{aligned} < (\overrightarrow{\phi}(t) - \overrightarrow{\phi}(0))(\overrightarrow{\phi}(t) - \overrightarrow{\phi}(0))^t >= 2(< \overrightarrow{\phi}(0)\overrightarrow{\phi}(0)^t > - < \overrightarrow{\phi}(t)\overrightarrow{\phi}(0)^t >) \\ = 2\frac{k_BT}{\alpha}Q\Lambda^{-1}(1 - e^{-\alpha\Lambda t/\gamma_t})Q^t \end{aligned} \tag{3.50}$$

Apparently, the mean-squared angular displacement is the sum of the contribution of each normal mode. Besides these deformational normal modes, there is also uniform axil spinning mode in which the contribution to mean- squared angular displacement is just $D_{||} = D_1/N$ where $D_1 \equiv k_BT/\gamma_t$ is the diffusion for each identical rod.

3.5.2 Bending

The similar procedure is used fTwisting and Bending or constructing bending correlation functions. The bending potential energy in the x and y direction for the N coupled oscillators

can be written in terms of the displacement coordinates,

$$U = U_x + U_y = 1/2\ \kappa\vec{\eta}A\vec{\eta} + 1/2\ \kappa\vec{\theta}A\vec{\theta} \tag{3.51}$$

where η_i and θ_i are the displacement coordinate angles between the i and $i+1$ bead (or disk) in the x and y direction, respectively. Note that there are only $N-1$ such angles between the N oscillators unlike the problem in twisting where every oscillator possesses a twist angle. In the matrix form, we can relate η and x, θ and y by

$$\eta = \frac{1}{h_b}\delta\ x \tag{3.52}$$

and

$$\theta = \frac{1}{h_b}\delta\ y \tag{3.53}$$

where

$$\delta = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \tag{3.54}$$

if it is in the case of $N = 3$. δ is the $(N-1)\times N$ finite difference matrix and h_b is the diameter of the effective sphere, or the distance between the centers of the contiguous beads. The bending potential energy in the x and y direction may be re-written in terms of the displacement coordinates $U = U_x + U_y = 1/2\ \frac{\kappa}{h_b^2}\vec{x}\delta^t A\delta\vec{x} + 1/2\ \frac{\kappa}{h_b^2}\vec{y}\delta^t A\delta\vec{y}$. Since collagen-like peptide can be assumed to be molecule with symmetry top, $U_x = U_y$. We just focus on beads bending in the y direction. The bending force is $\overrightarrow{F_b}(y) = -\frac{dU_y}{dy} = -\frac{\kappa}{h_b^2}\delta^t A\delta\vec{y}$ and the associated Langevin equation containing both the hydrodynamics interactions and the bending force ($\overrightarrow{F_b}$)is

$$m\ddot{y} + \gamma_b\dot{y} + \frac{\kappa}{h_b^2}H\delta^t A\delta\vec{y} = R(t) \tag{3.55}$$

where m is the mass of each bead or disk, γ is the translation friction factor, H is the hydrodynamics interaction tensor and is the sum of the identity matrix and the Rotne-Prager tensor: $H = 1 + T$. T is real, symmetry Toeplitz matrix that is accurate for the equilibrium position of the rod and hence to all thermally accessible states of deformation in the WBR limit. The elements of T can be expressed as

$$T_{lm} = \frac{3a}{4R_{lm}}\{I + \hat{R}_{lm}\cdot\hat{R}_{lm} + \frac{2a^2}{R_{lm}^2}(\frac{1}{3} - \hat{R}_{lm}\cdot\hat{R}_{lm})\} \tag{3.56}$$

where $R_{lm}$is the center-to-center vector from the $l^{th}$to the m^{th}bead, I is the unit tensor, a is the hydrodynamics radius of a bead. Note that the definition of the hydrodynamics interactions only works for the case of spherical beads. We also assume that the weakly bending rod is effectively straight so that $R_{lm} = |l - m|h$ where h is the distance between center of two contiguous beads. Then, the elements of the hydrodynamics interaction matrix for the y motion will be given by

$$H_{lm} = \delta_{lm} + (1 - \delta_{lm})\frac{3}{8|l = m|}\{1 + \frac{1}{6|l - m|^2}\} \tag{3.57}$$

The generic element of the hydrodynamics interaction matrix H_{lm} describes how the motion of the lth residue produces instantaneous waves in the solvent, which will perturb the velocity of the fluid surrounding a generic mth residue when $l \neq m$. The perturbation of the velocity decreases as the inverse of the inter-residue distance R_{lm}; in this way the hydrodynamics interaction is a long-term scale dynamics of the protein.

Transformation matrix Q_b is chosen so that the matrix $H\delta^t A\delta$ is diagonalized and diagonal eigenvalue Λ_b can be found which is $\{Q_b^{-1}HDQ_b = \Lambda_b\}_{n,n}$, where $D = \delta^t A\delta$. The matrix HD has two eigenvector with zero eigenvalue and corresponds to the translation and rotation modes of mode, respectively. After moving these two modes from Λ_b and Q_b, the mean-square angular displacement of internal motion can be expressed as

$$\begin{aligned} < (\overrightarrow{\eta}(t) - \overrightarrow{\eta}(0))(\overrightarrow{\eta}(t) - \overrightarrow{\eta}(0))^T >= 2(< \overrightarrow{\eta}(0)\overrightarrow{\eta}(0)^T > - < \overrightarrow{\eta}(t)\overrightarrow{\eta}(0)^T >) \\ = 2\frac{k_B T}{\kappa_b}(\delta Q_b)\Lambda_b^{-1}(1 - e^{-\kappa_b \Lambda_b t/h_b^2 \gamma_b})(\delta Q_b)^T \end{aligned} \tag{3.58}$$

where $\Lambda_b$is $(N-2)\times(N-2)$ and $Q_b$is $N \times (N-2)$.

Unlike the case for twisting motions, the single disc diffusion coefficient $D^B_{disc} = kT/\gamma^B$ is not identical to that of the uniform mode D^B_1 because of the hydrodynamics interaction involved. The relationship between the entires in $D^B_\perp$ is not so simply related to D^B_1 as is true with twisting, but the magnitude of ratio of $D^B_\perp$ over D^B_1 is $\sim 1/N^2$. The diffusion matrix for the entire system can be written as

$$D^B_\perp = D^B_1 H^{\frac{1}{2}}(1 - A^{-1}A)H^{\frac{1}{2}} \tag{3.59}$$

3.5.3 Hydrodynamics

1. Viscosity of water η is defined as the force per unit area needed to maintain unit velocity gradient between two parallel surfaces moving relative to one another in the fluid. The cgs units of η is dyne-sec/cm^2 or poise in honor of Poiseuille. The viscosity of water is a strongly decreasing function of temperature. Hydrodynamics results are

	Cylinder
R_e	$(3/2p^2)^{1/3}(L/2)$
F_t	$\frac{(2p^2/3)^{1/3}}{lnp+\gamma}$, $\gamma = 0.312 + \frac{0.565}{p} + \frac{0.1}{p^2}$
$F_r(a)$	$0.64(1 + \frac{0.677}{p} - \frac{0.183}{p^2})$
$F_r(b)$	$\frac{2p^2}{9(lnp+\delta)}, \delta = -0.662 + \frac{0.917}{p} - \frac{0.05}{p^2}$

Table 3.3: *The ratio of translational and rotational frictional coefficients of cylindrical rods to spheres of the same volume.* $F_t = f_t/6\pi\eta R_e$ *and* $F_r = f_r/8\pi\eta R_e^3$

commonly adjusted to 20^0C, at which $\eta \sim 0.01002$poise. or 1.002 centipoise (cp).

2. Frictional coefficients of sphere. It has translational frictional coefficient $f_t = 6\pi\eta R$, where R is often called the Stokes radius, and rotational frictional coefficient $f_r = 8\pi\eta R^3$. Note that translational friction varies only as the first power of the molecular size, therefore, it is much less sensitive to molecular dimensions than rotational friction, which varies as molecular volume changes.

3. Frictional coefficients of ellipsoid and cylinders: the ellipsoid model is widely used because its properties can be calculated exactly. The are often used to model cylindrical rods, giving reasonably accurate results if length and volume of the two objects are equated ($2a = L, b_{elli} = \sqrt{3/2}b_{cyl}$). The translational and rotational friction coefficients for cylinders are listed in Table 3.3. The equivalent radii $R_e$of the spheres of equal volume are given in the first row of Table 3.3. For cylinders, $p = L/2b$, where $2b$ is the diameter.

3.5.4 Stretched Exponential Function

Multi-exponential methods indeed provide better fits of the twisting and the bending correlation functions. However, they are incapable of providing insight into the basis of the physical relaxation phenomena. The intention of the research is to find a function to be able to relate the relaxation behavior to physical quantities and with relatively small number of degrees of freedom. Stretched exponential function seems to be a good candidate in this case. Stretched exponential relaxation $C(t) = exp[-(t/\tau)^\beta]$ is ubiquitous in many research areas, where τ is a characteristic relaxation time and the stretching exponent β is the range $0 < \beta < 1$. As β decreases from unity, more and more pronounced positive curvature is evident at small t. Here the correlation functions from rotational motions are found fit well in time space using the Kohlraush -Williams-Watts(KWW) or stretched exponential functions.

$$C_{kww}(t) = exp[-(t/\tau_w)^{\beta_w}] \tag{3.60}$$

There is no analytic expression for the Fourier transform of stretched exponential function but a numerical transform can always carried out. A natural used interpretation of the stretched exponential function decay is in terms of a set of pure exponential decays, with a particular probability distribution $\rho(r)$ of independently relaxing with relaxation times τ where $u \equiv \tau/\tau_n$. Therefore, the stretched exponential function can be rewritten in the form of

$$C_{kww}(t) = \int_0^\infty dr\rho(u,\beta_w)exp[-\frac{t}{\tau_w u}] \cong \sum_{i=1}^{N} P_i exp[-(-\frac{t}{\tau_w u_i})])] \tag{3.61}$$

where the time constant for each exponential in the discrete summation term is $u_i = \tau_w/\tau$. Since $\rho(u,\beta_w)$ (or P_i) is a probability density, one has to meet the normalization condition.$\int_0^\infty \rho(u,\beta_w) = 1$ (or $\sum_{i=1}^{N} P_i = 1$). $\rho(u,\beta_w)$ can be computed from the series expansion:

$$\rho(u,\beta_w) = -\frac{1}{\pi u}\sum_{k=0}^{\infty}\frac{\Gamma(\beta_w k+1)}{\Gamma(k+1)}sin(\pi\beta_w k)(-1)^k u^{\beta_w k} \tag{3.62}$$

$$P_i = -\frac{1}{\pi}\sum_{k=0}^{\infty}\frac{\Gamma(\beta_w k+1)}{\Gamma(k+1)}sin(\pi\beta_w k)(-1)^k u^{\beta_w k} \tag{3.63}$$

3.5.5 Determination of Diffusion Coefficient of a Rigid Circular Cylinder

The hydrodynamics properties which depend on the particle's size and shape are the rotational coefficients ,$D_{||}$ and $D_\perp$, for spinning and tumbling, respectively. They are the primary rotational properties, the quantities that are experimentally observed are the rotational relaxation times. They can be also computed using different models and theory. Several dynamics models have been developed to compute the overall rotation diffusion tensor, i.e. from the NMR relaxation rates depending on the orientation of N-H vectors with respect to the principal axes frame of diffusion tensor [75, 116–118]. A number of theoretical models have also been developed for predicting the diffusion properties of protein from its three-dimensional structure [119, 120]. There are also attempts to estimate diffusion properties using molecular dynamics simulation [79, 121].

1. One of the popular theories by Broersma in 1960 is valid well for the long cylinder with stick boundary condition. The Broersma theory made some corrections to mimic the slip boundary condition, which is valid for the ratio greater than about 4. The hydrodynamics properties of a rod with length L, diameter d and the aspect ratio

$p = L/d$ can be expressed in the following forms:

$$D_{||} = \frac{k_B T}{\pi \eta d^2 L} \tag{3.64}$$

$$D_{\perp} = 3\frac{k_B T}{\pi \eta L^3}\{ln(p) - 1.57 + 7[\frac{1}{ln(p)} - 0.28]^2\} \tag{3.65}$$

where k_B is Boltzmann constant, T is the absolute temperature, η is the viscosity of the solvent.

2. Follow the theory of Tirado and Garcia (TG) [115], the diffusion coefficients for uniform rigid body rotation around two axes are

   a) the rotational diffusion constant about the transverse axis

   $$D_{\perp} = \xi 3\frac{k_B T(lnP + C_r^{\perp})}{\pi \eta L^3} \tag{3.66}$$

   where ξ is a scale factor that takes into account the slight reduction in end-to-end distance due to spontaneous curvature appropriate for an equilibrium persistence length;

   b) the rotational diffusion constant about the symmetry axis

   $$D_{||} = \frac{4k_B T p^2}{A_0 \pi \eta L^3 (1 + C_r^{||})} \tag{3.67}$$

   where $A_0 = 3.84$. Therefore, the friction per triplet for rotation about the symmetry axis is

   $$\gamma = f_{||}/N + 1 = A_o(1 + C_r^{||})\pi \eta R_H^2 h \tag{3.68}$$

In the above-mentioning equations, the $C_r^{\perp}$ and $C_r^{||}$, the so-called end-plate corrections, were obtained as numerical constants. The numerical results for the end-effect terms are expressed as

$$C_r^{\perp} = -0.662 + 0.917/p - 0.05/p^2 \tag{3.69}$$

$$C_r^{||} = 0.677/p - 0.183/p^2 \tag{3.70}$$

3. from MD simulation [79],

4. from NMR relaxation data [122],

5. from combination of exact and approximation methods [118].

3.5.6 Correlation Function for a Subunit in the Molecule with Mean Local Cylindrical Symmetry

A diffusion equation for the probability density $P(\Phi, t)$ of finding a non-spherical macromolecule pointing direction ϕ at time t is as following

$$\frac{\partial P(\Phi,t)}{\partial t} = -\sum \hat{I}_i^2 \cdot D_{i,j} \cdot \hat{I}_j^2 P(\Phi,t) \tag{3.71}$$

where $\hat{I}$ is an angular momentum operator, and Φ is the Euler angle specifying the orientation of a given molecule. The conditional probability density $G(\Phi t|\Phi_0)$, Green's function which is the solution of the diffusion equation Eq. 3.71, represents that the macromolecule with orientation at Φ_0 at time $t = 0$ will move to Φ at time t. This $G(\Phi t|\Phi_0)$ is used to evaluate correlation function

$$< D_{pn}^2[\Phi_0] D_{qn}^{2*}[\Phi] >= \int d\Phi \int d\Phi_0 P(\Phi_o) \times G(\Phi t|\Phi_0) D_{pn}^2(\Phi_o) D_{qn}^2(\Phi)$$

$$= \delta_{pq} < D_{pn}^2(\Phi_o) D_{qn}^2(\Phi) > \tag{3.72}$$

where $D_{pn}^2(\Phi)$the rotation function in the Wigner-Edmonds convention is just the rotation that carries a coordinate frame from coincidence with the laboratory frame to coincidence with the principal-axis frame of the molecule. The rotations of any particular subunit are characterized by the time-dependent mean-squared angular displacements around its body-fixed axes $< \Delta_x^2(t) >, < \Delta_y^2(t) > and < \Delta_z^2(t) >$ along x, y, and z axes, respectively. If and only if the macromolecule exhibits the mean local cylindrical symmetry,$< \Delta_z^2(t) >$contains the effects of collective twisting deformations and uniform rigid-body rotation around the symmetry axis, while $< \Delta_x^2(t) >=< \Delta_y^2(t) >$contains the effects of collective bending deformations and uniform rigid-body rotations around a subunit transverse axis. For more details of derivation of diffusion equation is presented in the paper [108]. The correlation function of a subunit motion in the molecule in the molecule with mean local symmetry top can be written as

$$< D_{pn}^2[\Phi_0] D_{qn}^{2*}[\Phi] >= \delta_{pq} \, (\frac{1}{5}) exp[-\frac{1}{2}(6 - n^2) < \Delta_x^2(t) >] \, exp[-\frac{1}{2} n^2 < \Delta_z^2(t) >]. \tag{3.73}$$

4 Dynamics of the collagen triple helix: Implications for molecular recognition

Abstract. The dynamics of inter-chain hydrogen bonding is important for collagen recognition sites, including those containing disease-related mutations or excessive collagen degradation. Here we study the dynamics of NH bond vectors in collagen model peptides (CMPs) using molecular dynamics simulations. Both overall rotations and internal motions appearing in molecular dynamics simulations (MD) using explicit solvent, allow the back-calculation of NMR spectra and compare to the experimental data. The MD results show *(i)* the rotational diffusion constants are larger than one in experiment, and it is likely that the details of hydration layer around the CMPs are inaccurate in this protein-water potential; *(ii)* simulated CMPs exhibit significant amounts of bending and twisting motions, and their dependence upon the amino acid composition, challenging the assumption of decoupling between overall and internal motions that are generally made to interpret NMR relaxation data. To test this separation, we performed the isotropically reorientational eigenmode dynamics (iRED) analysis, by which the overall rotational diffusion constants are in a good agreement with those obtained from an approach that uses backbone superposition to define overall tumbling. Computed diffusion coefficients are rescaled to fit experimentally measured ^{15}N relaxation data for one of the CMPs, (T3-785).

4.1 Introduction

The triple helix is the main structural component in fibrillar collagens, the most well characterized form of collagen, as well as an important structural motif found in several membrane proteins. For instance, the members of fibril-forming family include type I collagen, a heterotrimer, found in tendon and bone, and type III collagen, a homotrimer that found together with type I collagen in skin, blood vessels and other tissues. The strict amino acid sequence with glycine residue in every third position (Gly-X-Y)$_n$ is the main feature of triple helical conformation of three peptide chains which are supercoiled around a helical axis and stabilized by a set of inter-chain hydrogen bonds. Beside the sequence-dependent characteristics and its static conformation, the biological function of collagen is also likely determined by its dynamics, since individual residues at some point must be accessible to the macromolecules in order to provide sequence-specific recognition and high affinity binding [20, 56, 57]. It has been suggested that regions with a relatively low imino-acid content in collagen (i.e., fewer proline and hydroxyproline residues) may be involved in collagen recognition sites, due to the difference in triple helical conformation and local molecular mobility compared the imino-acid-rich (POG) regions. The great example is that matrix

metalloproteinase 1, 3, and 8 bind human type III collagen at unique cleavage sites characterized by a Gly-Leu or Gly-Ile bond which is preceded by the region rich in imino acids Pro/Hyp at the X/Y positions, and is followed by an imino-poor region [19, 20]. Although there are numerous potential cleavage sites presented in the collagens, they are not usually cleaved by collagenases, indicating that this selective recognition must be related to the local conformation of the extended sequence, particularly, the imino-acid deficient segment on the C-terminal side of the cleavage sites. This change seems to be minor, but it has been suggest that even small changes in conformation and mobility of the triple helix might be sufficient to make the triple helix accessible to collagenases without additional enzymatic contributions to unwinding [16, 20, 28, 80, 83, 88, 123].

4.1.1 Experimental Background

The dynamics of collagen model peptides (CMPs) has been studied experimentally for over 40 years [28, 58–64]. Lazarev and colleague [56, 57] dealt with collagens with different amino acid composition, and hence different rigidity and stability, using infrared spectroscopy and hydrogen-exchange methods. They have found that the large amplitude of backbone dynamics with low frequency belonging to cooperative interactions are related to the protein crystallinity and stability, whereas the high frequency dynamics on faster time scale remains similar among these collagens. Brodsky and colleagues investigated the sequence-dependent stability for various CMPs designed by host-guest approach [7]. Data from circular dichroism suggest that the melting temperatures of these CMPs are related to their imino acid content: the presence of imino acids at the X and Y positions confers maximum stability to collagen, while non-imino acids at the Xaa and Yaa position result in variations in global and local stability. For example, a CMP T3-785 (PDB entry 1BKV [28]), with sequence $(POG)_3$-ITG-ARG-LAG-POG-$(POG)_3$, was designed to mimic a fragment with an imino-acid poor region containing 12 residues occurring one triplet C-terminal to the unique collagenase cleavage sites Gly-Ile or Gly-Leu in type III human collagen. The study of circular dichroism showed that T3-785 forms a stable triple-helix with a melting temperature $T_m = 18^oC$, far below the $T_m = 60^oC$ for $(POG)_{10}$, consistent with the notion that imino-acid poor regions of collagen have decreased stability [16].

To quantify the degree of thermal fluctuations within a crystal environment, Kramer *et al* [28] reported that the B factors associated with the central imino acid-poor region are considerably smaller than those associated with the imino acid-rich N- and C-terminal regions in the T3-785 peptide, indicating less mobility in the imino acid-poor region. This is seen in contrast to the thermal stability measurements, and suggests that further analysis of the relation between stability and flexibility is warranted.

4.1.2 NMR Spectroscopy

In this work, we are particularly interested in NMR spectroscopy, as it offers a powerful method for studying molecular motions at a wide range of time scales, and has been applied to CMPs for many years [2, 51, 64, 66]. Fan *et al.* [2, 64] investigated the backbone dynamics of the specific residues in peptide T3-785 and $(POG)_{10}$ on the different time scales. In the literature, they summarized with some main points: (i) the effect of chemical shift on the relaxation process is ignored due to the very slow rate; (ii) the contribution of the correlation time of internal motion to the relaxation processes is also neglected on the basis of the observed large value of ^{1}H-^{15}N NOE; (iii) under the conditions (i) and (ii), the R2/R1 ratio is only a function of the overall motional parameters; (iv) the motional properties on the picosecond time scale from ^{15}N relaxation data are rather uniform for the central Gly residues in both the imino-acid poor and rich regions, suggesting that the backbone motion on the fast time scale is not sensitive to the sequence composition; (v) in contrast to the relaxation data, hydrogen exchange data on T3-785 showed that the Gly15 NH and Leu16 NH exchange very slowly and slowly, respectively, while the Ala 17 NH exchanged very rapidly; (vi) hydrogen-exchange data further indicates that the central Gly NH group (in the sequence Gly-Leu-Ala in the middle of T3-785) exchanges at a faster rate than one at the center of $(POG)_{10}$ peptide, suggesting that imino acid-poor regions indeed increase mobility in solution. These results point to the dynamics of Gly residues within imino-rich or poor regions that differs at slower time scales.

While NMR spectroscopic experiments have implied that collagen may be flexible in some areas [28, 58–64], only limited information on the nature of conformational changes is provided. The use of MD simulations to interpret NMR relaxation data is an alternative to standard analytical models [67–70], since they provide a high resolution view of the motions of biological macromolecules and produce continuous trajectories with the potential to connect static structural snapshots, and give very detailed (if imperfect) description of protein structures and dynamics in solution. Because the whole timescale of both rotational and internal motions is now accessible by molecular dynamics simulations using explicit solvent, the back calculation of relaxation parameters becomes possible.

MD simulations also provide an ability to explore correlated dynamics, which are often responsible for the structural and biological activity of proteins and can be separated from the less-relevant local motions of molecules with high-frequencies. Two methods used here for analyzing collective dynamics are principal component analysis (PCA) [124] and isotropically reorientational eigenmode dynamics (iRED) analysis [67]. These techniques, which are described in detail below, diagonalize covariance matrices in Cartesian or angular variables, respectively, and construct collective modes based on the resulting eigenvectors.

4.1.3 Molecular Dynamics Simulations

All MD simulations were conducted with the X-ray structure of each CMP of interest (Table 4.1) as a starting structure. Crystallographic water molecules were not included in the starting model to avoid any bias on the MD hydration sites. The peptides were immersed in a truncated octahedral box of SPC/E water model, originally introduced by Berendsen et al. [125], with at least 15 Å between any peptide atom and the edge of the water box. The Amber99SB force field was applied to model all the systems. The parameters for hydroxyproline are presented in Appendix [71, 126]. Details about the starting structures and the number of water molecules are given in Table 4.1. All simulations were carried out with the SHAKE algorithm enabled for hydrogen atoms and with Particle-Mesh Ewald approach with a 8 Å cut-off for non-bounded interactions. After minimization, relaxation and equilibrium, the production run was using NVE and the time step of integration was 1.0 fs and the Cartesian coordinates were saved every 2ps.

Table 4.1: *Details of MD Simulations*

peptide	PDB	No. amino acid	water model	No. waters	ns
POG10	1V7H [24]	90	SPC/E	28989	86
PPG10	1K6F [23]	87	SPC/E	30991	97
GlyAla	1CAG [22]	88	SPC/E	27662	150
	sequence in chain: $(POG)_4POA(POG)_5$				
T3-785	1BKV [28]	89	SPC/E	30008	500
	sequence in chain: $(POG)_3ITGARGLAG(POG)_4$				

4.1.4 Chapter Outline

The objective in this paper is to explore the types of motions, including the overall tumbling and internal motions as well as the correlated motions. In Section 4.1.3, we start with explicit solvent MD simulations for four CMPs, including a 0.5 μs long simulation of the T3-785 peptide, and 100 ns simulations of Gly→Ala, $(POG)_{10}$, and $(PPG)_{10}$ to explore potential sequence-dependent motions. During these simulation times, the CMPs are stable and exhibit fluctuations about their equilibrium configuration. The outline of this Chapter is as follows. Section 4.2 we consider overall tumbling of triple helix with the high anisotropy. Section 4.3 looks at NMR relaxation behavior and compare with experimental data. Internal motions monitored by NMR are considered in Section 4.4 and the internal motions of the side chains explored by iRED method are discussed in Section 4.5. The following Section 4.6, we consider a correlated dynamics existing in the triple-helical peptide. We end with the conclusion of main points in Section 4.7. In the Methods section, we describe the methodology we used to analyze simulation data.

4.2 Overall Rotational Diffusion

The overall tumbling motion (or rotational Brownian motion) of the triple-helical collagen peptide is the dominant contribution to the NMR spin-relaxation in solution, and it is considered first, since the analysis of internal motions in NMR depends upon an accurate estimates of overall diffusion. Since molecular motions are rather complex, the nmr relaxation data are almost universally interpreted using some simple models: the overall rotational dynamics of spherical molecule can be expressed in terms of a single scalar quantity, the overall correlation time τ_M, or the equivalent diffusion constant $D = 1/(6\tau_M)$; for non-spherical proteins, D becomes a rotational diffusion tensor and anisotropy [4]. A number of theoretical models have also been developed for predicting the diffusion properties of protein from its three-dimensional structure [4, 119, 120], or from molecular dynamics simulations [79, 121]. Here we consider what can be learned about collagen peptides via both experimental and computational analyses of rotational diffusion.

Several protocols have been developed to derive the overall rotational diffusion anisotropy from NMR spin-relaxation rates [75, 116–118, 127]. Torchia with colleagues [128] suggested that a rotational diffusion anisotropic model is suitable for describing the overall motion of the collagen with high anisotropy. Xiao et al. [2] collected a set of relaxation rates of the labeled Gly-Leu-Ala-Gly residues in the central region of T3-785 peptide and they applied Lipari-Szabo model-free formalism [76, 107] to obtain the diffusion tensor. Table 4.2 lists some results of rotational diffusion analysis on the triple helical peptides: the effective diffusion time and the anisotropy derived from NMR relaxation data on T3-785 reported in the literature [2] and different computational approaches, which are discussed in the following section.

	exp[a]		MD (RotDif)		Ellipsoid model		Hydronmr[d]	
	τ_M	$D_{\|\|}/D_\perp$	τ_c	$D_{\|\|}/D_\perp$	$I_x : I_y : I_z$ [b]	$D_{\|\|}/D_\perp$ [c]	τ_M	$D_{\|\|}/D_\perp$
T3-785	6.9	13.1	3.5	12.4	36.34 : 36.29 :1.00	12.68	6.8	8.93
PPG10			3.1	13.3	43.56: 43.55: 1.00	14.42	6.8	8.83
GlyAla			4.1	10.2	36.25:36.20: 1.00	12.66	6.9	8.54

Table 4.2: *Diffusion tensors of collagen-like peptides.* [a]*: NMR-derived data from Ref. [2].* [b] *: The ratio of principal components of inertia tensor of heavy atoms of peptides from crystal structures.* [c] *: Calculation is done using Torchia's [3] empirical formula* $D_{||}/D_\perp \sim (I_\perp/I_{||})^{1/\sqrt{2}}$*, where* $I_\perp/I_{||}$ *is the ratio of the principal components of the inertia tensor and* $I_x = I_y = I_\perp$ *and* $I_z = I_{||}$ *in the case of prolate ellipsoid.* [d] *: HYDROPRO calculation [4] using the X-ray coordinates. For the three peptides listed, values of* a *(see the text) were 2.9, 2.4, and 2.3, respectively;* σ_i *was 0.7 and* σ_x *was 1.65 for all three peptides.*

4.2.1 Hydrodynamic Models

We begin with an overview of "classical" models for the rigid bodies undergoing Brownian motion in a continuum solvent. We know from rigid-body hydrodynamics that rotational diffusion is quite sensitive to the size and shape of macromolecules in solution. X-ray crystallography has indicated that the shape of triple helical peptides can be approximated as ellipsoids or cylinders. The simplest calculations of rotational anisotropy are made by modeling a triple helical peptide from crystal structure as a prolate ellipsoid (Table 4.2). This estimation is done using moment of inertia derived from atomic structure and empirical relationship, $D_{||}/D_{\perp} \sim (I_{\perp}/I_{||})^{1/\sqrt{2}}$ [3, 115]. The ratios of inertia moments (e.g. T3-785 peptide) around the principal axes of the inertia tensors ($I_{xx} : I_{yy} : I_{zz}$) of the triple helix are 36.3: 36.3: 1.00, indicating axially symmetric diffusion tensors. Using the empirical relationship, the diffusion anisotropy value of helix is approximately ∼12.7. This ellipsoid approximation does not include the effects of bound waters, and the calculated inertial axes from one static structure do not necessarily have to agree with the principal axes of the actual rotational diffusion tensor.

Shell models, developed by Garcia de la Torre *et al.* [4], provide a more elaborate description of the protein surface than what is afforded by ellipsoid approximation. Each non-hydrogen atom is replaced by a spherical element with a radius a, which is called atomic element radius (AER). In the case of T3-785 peptide, with the value of AER $\sim$ 1.4 for each heavy atom, hydrodynamics shell model yields a reasonable estimate of the diffusion anisotropy ($D_{||}/D_{\perp}$∼10.71) but significantly overestimates the rate of molecular rotational motion (τ_M ∼4.8 ns) compared with experimental value, τ_M ∼6.92 ns [2]. With value of AER between 2.3 Å∼ 2.5 Å, shell model yields an effective overall relaxation time close to the experimental data, but has the values of $D_{||}/D_{\perp}$ much lower than anisotropy predicted from ellipsoid model or from experiment (See Table 4.2). The difference between AER and the typical van der Waals radii of atoms would amount to the thickness of the hydration shell. Therefore, the increased bound waters tend to decreases the value of anisotropy $D_{||}/D_{\perp}$ and slow down the tumbling of the molecule. The need for an unknown (and seemingly arbitrary) value of the AER parameter has seriously limited the application of hydrodynamic theory to the analysis of NMR relaxation data.

4.2.2 MD Simulation Results

The procedure of evaluating rotational diffusion tensor from MD trajectory follows the paper by Wong and Case [79]. First, 500 vectors randomly and uniformly distributed on the unit semi-sphere are selected, as shown in Fig. 4.1 and treated as initial pointing direction for each vector. Overall (rigid-body) rotational motion monitored by computing the rotation

matrices $R(t)$, which is generated using an rms superposition of $C\alpha$ atoms at time t to the average structure. Subsequently, the pointing direction $n_i(t)$ of the ith vector at time t is generated by operating the rotation matric $R(t)$ to the ith vector at their initial direction $n(0)$. Thus, the $C_{rot,i}(t)$ autocorrelation function of the rotational motion for that vector is determined by

$$C_{rot,i}(\tau) =< P_2[n(\tau) \cdot n(0)] >$$

where $P_2(x)$ is the Legendre polynomial.

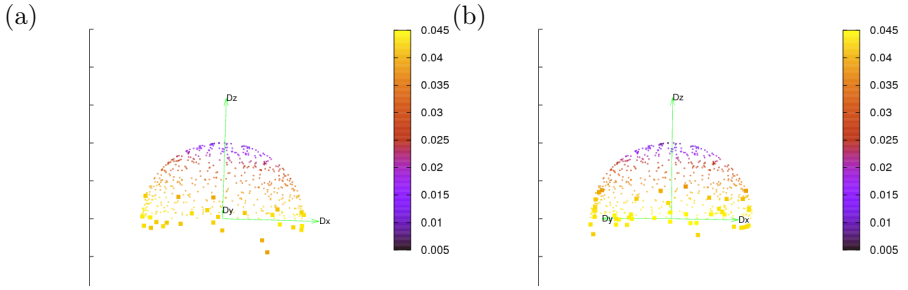


Figure 4.1: *Local diffusion constants calculated with RotDif for (a) Gly→Ala and (b) T3-785 peptides. Orientations of the random 500 vectors (cross points) and N-H vectors (filled squares) from origin and its corresponding local diffusion constant d_{obs} in the range of 0.005 to 0.045 ($10^9 s^{-1}$). The green lines represent the principal axes of the derived diffusion tensor.*

Following, we consider the rotational diffusion time of the ith vector, it is done by integral the correlation function of its overall tumbling motion,

$$\tau_{MD,i} = \int_0^\infty C_{rot,i}(t)dt \tag{4.1}$$

To estimate the full integral, Wong *et al.* used the Romberg integrator to numerically integrate the correlation function with a "single" decay time (see Theory in paper [79]). Subsequently, the Woessner rigid body diffusion model, τ_{RB} ,(shown in Eq 4.2) is used to optimize the correspondence between back-calculated and the MD-derived local effective relaxation times, τ_{MD}, to determine $D_\perp$and $D_{||}$.

$$\tau_{RB,i} = \frac{0.25(3cos^2\theta_i - 1)^2}{6D_\perp} + \frac{3cos^2\theta_i sin^2\theta_i}{5D_\perp + D_{||}} + \frac{0.75sin^4\theta_i}{2D_\perp + 4D_{||}} \tag{4.2}$$

where we integral the rigid body correlation function

$$C_{RB,i}(t) = \frac{1}{4}(3cos^2\theta - 1)^2 e^{-6D_\perp t} + 3cos^2\theta sin^2\theta e^{-(5D_\perp + D_{||})t} + \frac{3}{4}sin^4\theta e^{-(2D_\perp + 4D_{||})t}. \tag{4.3}$$

Fig. 4.2 shows the scatter plots of the local correlation time, τ_{MD}, of any given vector from the MD trajectory and Woessner rigid- body diffusion model correlation time τ_{RB} in Eq 4.2 for two peptides Gly→Ala and T3-785. As expected for a non-spherical protein, the local diffusion constant profile, d_{MD}, which is $1/6\tau_{MD}$, shows the dependence on the orientation of the vector in the molecular reference frame. The fits of the MD to a Woessner rigid body diffusion model are not quite accurate for these rod-like peptides as the deviations of τ_{MD} take place for the vectors with the orientation between 70° to 90°. These deviations can be seen easily in the correlation between τ_{MD} and τ_{RB}.

In the rotational "rigid rod" model, when the anisotropy is larger than some specific value, $D_{||}/D_{\perp} > 4$, a slight peak at 90 degrees occurs in the plot of the correlation time τ_{RB} against the orientation of bond vector. That is, when the anisotropy changes from 4 to 15, the global minima of the plot (θ,τ_{RB}) shifts from 90° to 59.2°. However, this shift is not really seen in our calculation. The possible reason for this is as following. In our calculation, only a single exponential decay is used to describe the correlation function of each random vector; this is usually valid for the system with small anisotropy, but it fails in the limit of the anisotropy larger than 4, as it cannot capture the slow decay mode which dominates the decay after long time. Therefore, at least two exponential terms should be used to fit the correlation function of individual vector from MD simulation. Furthermore, Fig. 4.2 exhibits that rigid body model does not match the values of diffusion constant of the vectors, which are aligning along the symmetry axis of the diffusion frame. It is because of minimization of χ^2 of the least-squared fitting of all random vectors that are dominated by the vectors pointing away from the long axis and only few vectors align along the long axis. Re-doing the analysis along these lines is planned.

The values of the anisotropy are similar with those from ellipsoid model and experiment (Table 4.2). However, the rotational diffusion constant in the simulation of T3-785 peptide is about twice smaller (τ_c ∼3.4ns) than experimental value (τ_c ∼7ns [2]). This result is in line with earlier results on globular proteins [79], which also found current simulation protocols to give rotational diffusion that is too fast.

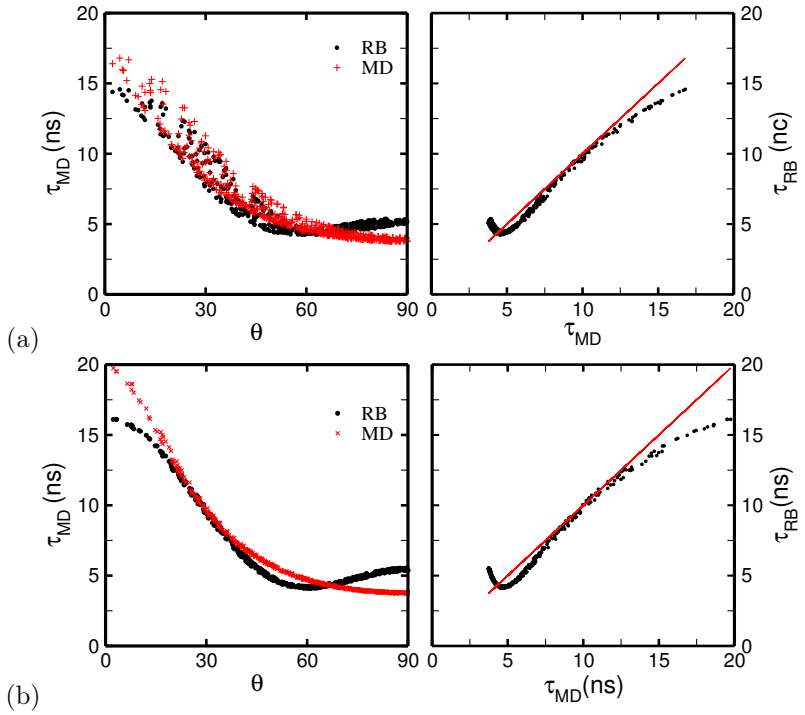
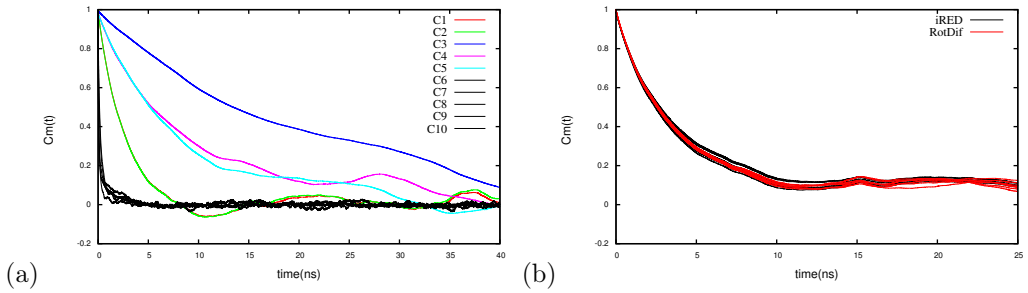


Figure 4.2: *Comparison of τ_{MD} values derived from the trajectory with those τ_{RB} from "rigid-body" diffusion theory for Gly→Ala peptide in panel (a), and T3-785 in panel (b). Left: Distribution of local effective τ values against the orientation of vectors with respect to the principal axis of diffusion tensor. Right: Scatter plot of τ_{MD} and τ_{RB} .*

4.2.3 iRED Approach

NMR relaxation data are not the only experiments capable of generating dynamics information such as global tumbling and order parameters. The iRED calculation is an alternative method for analyzing the molecule's motion that involves the angular fluctuations of spin interactions; it can describe both the overall reorientation of the molecule, along with collective motions of internal fluctuations of chemical groups. In the RotDif analysis discussed above, the triple helical peptide seems to tumble fast, which might be caused by an insufficient sampling or the validation of separation of overall and internal motions, or weak interaction with bound waters. Here, we perform iRED method on the peptide T3-785 with 500 ns to characterize its rotational diffusion constant and compare it with previous MD result (RotDif).



(a) (b)

Figure 4.3: *iRED analysis on T3-785 peptide. (a) Time-Correlation functions $C_m(t)$, of the ten largest modes of the matrix M averaged over a 500ns trajectory from iRED analysis, computed according to Eq 4.17. The C1∼C5 represent the overall correlation function modes and C6∼C10 represent the most collective internal modes. (b) Comparison of overall correlation functions of all the Gly NH vectors from iRED analysis (black lines), with those from the RotDif calculation (red lines).*

Fig. 4.3(a) displays the time correlation functions $C_m(t)$ of each eigen modes which are computed using Eq 4.17; *(i)* the internal correlation functions ($C_6 \sim C_{10}$), converging to zero during the first nanosecond, is well described with monoexponential decay; *(ii)* the first five overall correlation functions ($C_1 \sim C_5$), which are distinct from those of internal modes, are grouped into three categories, indicating that this triple helical collagen peptide tumbles in a nearly axially symmetric fashion.

The correlation functions of the overall reorientational modes $m = 1, .., 5$ do not reach a stable plateau at a value well above zero after 5 ns simulation, suggesting that the reliable estimates τ_m cannot be made directly from the iRED original output file (see Eq. A19 in Appendix 2 in ref. [67]). The first 5 ns of the correlation functions seem to be fit well to a monoexponential decay. These improving τ_ms' for the global modes are listed in Table 4.3 and the improving overall correlation function can be built and compared with the results of RotDif analysis. Fig. 4.3(b) shows a good agreement in overall correlation functions of Gly NH vectors between iRED and RotDif analysis, but the discrepancy becomes significant as the time increases.

It is possible to estimate the principal axis values of the rotational diffusion tensor using Woessner's equations (see Appendix). For example, from the $\tau_1 \sim \tau_5$ values (the 2nd column in Table 4.3), the following diffusion tensor principal axes values are obtained using Eq.4.22 (Table 4.4). A measured for the anisotropy of the diffusion tensor is $2D_z/(D_x + D_y)$=11.4, which compares well with the corresponding value of 13.1 found by Xiao *et al.* [2]. Compared to RotDif calculation, iRED analysis gives a slightly better agreement in diffusion time with the value (τ_M = 6.92 ns) reported in the literature [2], but still is about 1.5 times faster (τ_M ∼4.5ns) in simulation. To further improve agreement between iRED results and experimental relaxation data, the correlation times of the reorientational eigenmodes were possibly adjusted by minimized the χ^2 in Eq. 4.6 (more details see the text in Sub.

eigenmode (ns)	original[a]	5ns fitted[b]	LeastSq[c]
τ_1	2.0	2.51	3.02
τ_2	1.90	2.50	3.65
τ_3	9.60	19.6	31.3
τ_4	6.23	7.30	16.3
τ_5	6.09	7.32	13.4

Table 4.3: *Correlation times of overall modes from iRED analysis on the peptide T3-785.* [a]*: original data (see text);* [b]*: Fitted the row correlation function of the largest five modes with monoexponential functions using only first 5ns long period of time.* [c]*: Adjusting the correlation times of the five largest modes to reproduce the experimental relaxation data (see Sec. 4.3.3).*

4.3.3). Since the adjusted correlation times given in the 3rd column in Table 4.4 do not follow the axially symmetry diffusion motion, which gives $\tau_1 = \tau_2$ and τ_4=τ_5, the fully asymmetric diffusion Woessner equations in Eq. 4.21 are used to obtain the principal axis values: $D_x = 0.006(10^9/s)$, $D_y = 0.0022(10^9/s)$, and $D_z = 0.0645(10^9/s)$. This result is not surprising since in iRED analysis, these modes depend on the associated eigenmode directions, which do no necessarily coincide with the diffusion tensor principal axes. The corresponding anisotropy is $2D_z/(D_x + D_y)$=15.7, which is slightly larger than the value of 13.1. This new set of adjusted τ_m yields correlation time ($\tau_M = 6.86$ ns), which is in a good agreement with the value reported by Xiao *et al.* [2].

T3-785 peptide	$D_\perp(10^9/s)$[b]	$D_{\|\|}(10^9/s)$	$D_r = \frac{D_{\|\|}}{D_\perp}$	D_{iso}	τ_M(ns)[c]
RotDif	0.0099	0.1229	12.4	0.0476	3.5
iRED(5ns)	0.0084	0.0957	11.4	0.0375	4.44
[a]L-A(r1r2_diffusion) [2]	0.0048	0.0624	13.1	0.0240	6.92
iRED (LeastSq)	0.0041	0.0645	15.7	0.0243	6.86

Table 4.4: *Calculation of rotation diffusion coefficients derived with model-free analysis based on the experimental NMR relaxation data for T3-785 with the results of the iRED and RotDif analysis of our MD trajectories.*
[a.] *L-A (r1r2_diffusion): the derivation of the diffusion tensor using the fitting program "r2r1_diffusion" to the experimental relaxation data of Leu16 and Ala17 only.*
[b.] $D_\perp = (D_x + D_y)/2$
[c.] $\tau_M = 1/(6D_{iso}) = 1/(4D_\perp + 2D_{||})$ *the average diffusion time*

The NMR-derived diffusion constants are significantly smaller than those calculated with iRED or RotDif analysis, indicating the use of the SPC/E water model here underestimates of the overall correlation timescale by speeding up the rotational diffusion compared to the experimental value. This discrepancy between MD simulation and experiment is expected, as it is observed in other study in which the SPC/E water model under-predicts the viscosity by around 19% than the value determined experimentally at the room temperature [69, 121, 129]. Furthermore, Maragakis *et al.* [69] have accessed to μs long timescale simulation of

ubiquitin using SPC water model and still observed diffusion constant about twice times than the experimental values. It suggests the fast tumbling of molecule in SPC/E water model is not from the insufficient sampling over time. It is likely because there is no sustained hydration layer over period time long enough to affect the hydration dynamics properties of the protein or the interactions between bound waters and protein seem not to be strong enough to slow down the tumbling of the protein in the simulation. Therefore, finding optimized potential which gives a better value of global diffusion constant seems to be a prerequisite for reproducing NMR relaxation data directly from MD simulation.

4.3 NMR Relaxation Data

NMR ^{15}N relaxation parameters (Appendix 4.8.2), which are related to the spectral density functions, are evaluated from the MD trajectory in three ways. In method *(1)*, we follow Lipari-Szabo (or model-free) approach in which the assumption of separability of the overall and internal motions is made. It gives

$$C_{tot}(t) = C_{rot}(t)C_{int}(t) \tag{4.4}$$

where the internal correlation function defined in Eq.4.7 and overall correlation function in Eq. 4.3 from RotDif analysis. In methods *(2)* and *(3)*, superposition of all conformers to the average structure is not needed. In method *(2)*, S^2 and overall tumbling time τ_M are obtained from the iRED (see Sec. 4.8.4) where the conformers do not need to be superimposed and the orientation of any given vector is not considered. In method *(3)*, the raw data of $C(t)$ is extracted directly from the raw MD trajectory and the parameters are obtained using the first 10 ns to fit the data by a least-squares fit with a function of three exponential terms

$$C_{best}(t) = A_1 exp(-t/\tau_1) + A_2 exp(-t/\tau_2) + (1 - A_1 - A_2) exp(-t/\tau_3) \tag{4.5}$$

Here there is no need to label S^2 or τ_M in above equation. The purpose of this method is to find the best function $C_{best}(t)$ being able to describe the total correlation function directly from MD trajectory without any intermediate assumptions.

It is worth emphasizing that methods *(1)* and *(2)* for computing spectral density function use the effective correlation time, τ_M for each individual bond vector, therefore, Eq. 4.24 that is attributed from overall tumbling, is applied to compute the spectral density function.

4.3.1 Experimental Data

Experimental relaxation data depend strongly on both global and internal reorientational motions: (i) the value of R1 mainly comes from $J(\omega_N)$ (see Eq. 4.8) and is more sensitive to the variation of the correlation times τ of either overall or internal motions; (ii) the value of R2 is mainly dominated by how fast the molecule tumbles in solution (represented by the $J(0)$ term); (iii) if the tumbling of the molecule is described well by "rigid body" rotational model, then the value of R2/R1 is just a function of diffusion tensor and the orientation of that vector relative to the principal axes of the diffusion tensor. Therefore, it is common to plot the back-calculated value of R_2/R_1 based on the diffusion tensor derived from a set of experimental R2/R1 data, as shown in Fig.4.4, in which blue line denotes the optimized rigid body diffusion model for the relaxation data on the peptide T3-785 (Table 4.5 [2]).

The back-calculated R2/R1 values of the G15 and G18 deviate from the experimental values by about 12%, suggesting a re-positioning of the backbone amide protons of Gly residues (relative to the crystal structure) by adjusting the angle θ to ∼71 degree, out of the $C-N-C\alpha$ plane toward the $N-C\alpha$ bond. Such a modification of the NH position of Gly residues is shown to improve the hydrogen bonds of amide group of Gly residues to the carbonyl groups of Pro residues [2]. However, this conclusion depends strongly on the neglect of internal motions, and might be reconsidered in light of the MD results reported here.

POGPOGPOGITGAR$G^{15}L^{16}A^{17}G^{18}$POGPOG^{24}POGPOG

	R1	error	R2	error	$\frac{R2}{R1}$	error	NOE
1G15	1.99	0.03	11.4	0.2	5.73	0.11	0.63
2G15	2.02	0.02	11.8	0.1	5.84	0.12	0.68
3G15	1.99	0.02	11.8	0.2	5.93	0.12	0.64
1L16	1.9	0.02	11.5	0.2	6.05	0.12	0.68
2L16	1.96	0.03	11.4	0.2	5.82	0.12	0.72
3L16	1.93	0.02	11.5	0.2	5.96	0.12	0.72
1A17	1.82	0.03	11.8	0.3	6.48	0.13	0.6
2A17	1.79	0.02	11.7	0.2	6.54	0.13	0.65
3A17	1.81	0.02	12	0.2	6.63	0.13	0.69
1G18	2.02	0.03	11.6	0.3	5.74	0.11	0.67
2G18	2.02	0.03	11.3	0.2	5.59	0.11	0.71
3G18	2.04	0.03	11.9	0.1	5.83	0.12	0.72
G24	2.04	0.03	11.5	0.2	5.64	0.11	0.59

The value of error was computed using 500, 600, 800 relaxation data.

Table 4.5: *NMR relaxation R1,R2, NOE data obtained using 500, 600, 800 MHz [2].*

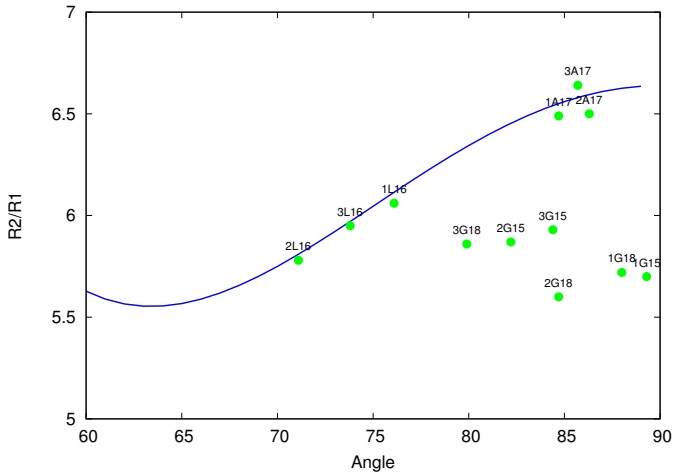


Figure 4.4: *The ratio of R2/R1 against the angle between NH vector and the principal axis of triple helical collagen peptide. Green circles are the NMR experimental data [2]. Blue solid line follows the theoretical anisotropic diffusion model with* $D_{ave} = 0.024 \times 10^9$*(1/s) and* $D_{||}/D_{\perp}$*=13.1. The number 1, 2 and 3 before the residue name denote the leading, middle and trailing chain, respectively.*

4.3.2 MD simulation Results

In our solution simulation of T3-785, the calculation of ^{15}N relaxation data for a set of residues is shown in Fig. 4.5. Three different approaches, methods *(1)*∼*(3)*, give similar overall correlation times, which dominates the value of R2. On the other hand, these methods show somewhat different values in R1, which depends on the parameters characterizing the internal motions. In method *(1)*, extended model-free formula gives S^2 and extra parameter, the internal correlation time τ_{int}, for each NH vector in the peptide. This τ_{int} increases the value of R1, as the 2nd term in Eq. 4.23, explaining why the values of R1 are literally higher than those derived from other two approaches. The higher values of R1 (compared to experiment) is a result of the fast overall tumbling of molecule in simulation.

Our objective here is not to accurately reproduce experimental relaxation measurements reported by Xiao *et al.* [2], as these measurement strongly depend on the global tumbling of the molecule, and it has been demonstrated earlier in Section 4.2 that molecule tumbles fast in simulation than in the experiment. The discrepancy of relaxation data is depicted in Fig. 4.5. Rather, our study here aims to exam if the rigid body model is sufficient to accurately capture the picture of internal motions existing in collagen-like peptide in solution, at the same time, to provide a distribution of relaxation data of the residues in simulation. More details will be discussed next.

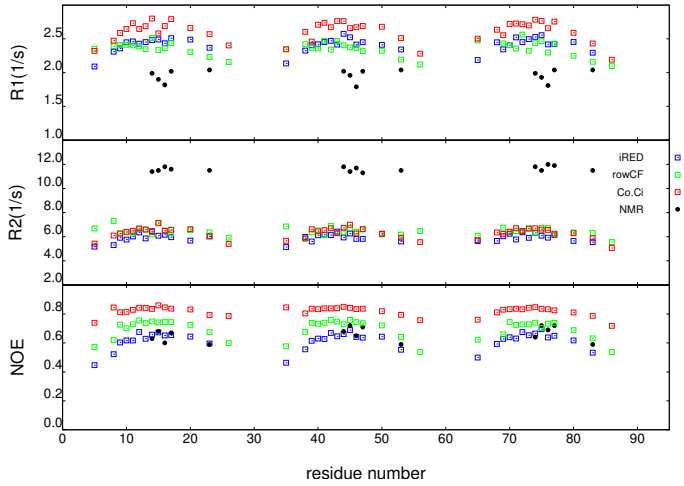


Figure 4.5: *^{15}N relaxation data R_1, R_2, and NOE for labeled residues in peptide T3-785 are computed from MD simulation. Experimental relaxation data (black circles) are collected at 20^0C at spectral frequency 500 MHz.*

4.3.3 Re-scale overall Tumbling in Simulation (iRED analysis)

Our previous calculation of the overall tumbling motion has suggested that the protein-water interaction parameters in current force field may underestimate the strength of the interaction between the protein and the solvent. Here, we examine how the simulated diffusion tensor can be re-scaled to achieve a better fit to the relaxation data, then provide an insight into the interpretation of the internal motions in solution nmr.

According to iRED analysis introduced by Bruschweiler and Promers [67], the agreement between calculated results and experimental data can be effectively improved by adjusting the correlation times of the first largest eigenmode to the experimental data. Treating only 10 correlation times τ_m as variable parameters, we used the χ^2 as a measure of proximity to the experimentally determined relaxation rate constants:

$$\chi^2 = \sum_i^N \frac{(X_i^{calc} - X_i^{exp})^2}{(\sigma_i^{exp})^2} \tag{4.6}$$

where X_i^{calc} are the calculated and X_i^{exp} are the experimental R1, R2, and NOE parameters, and σ_i are the corresponding experimental errors. As described below, the correlation times τ_m can be adjusted to minimize χ^2 using a least-squares fitting procedure.

Fast tumbling of molecule extracted in simulation leads to a χ^2 value of 1.5×10^5. The five largest overall tumbling modes and the five largest internal modes are adjusted through the fitting procedure of minimizing value of χ^2 in Eq. 4.6. This fitting improves χ^2 value by about a factor 35 ($\chi^2 = 4.18 \times 10^2$). The back-calculated relaxation parameters are

displayed as blue circles in Fig. 4.6. For the most residues (except the values of R1 in Ala residues), the data is well reproduced when considering only 10 fit parameters. The adjusted correlation times are given in Table 4.3 together with the original values.

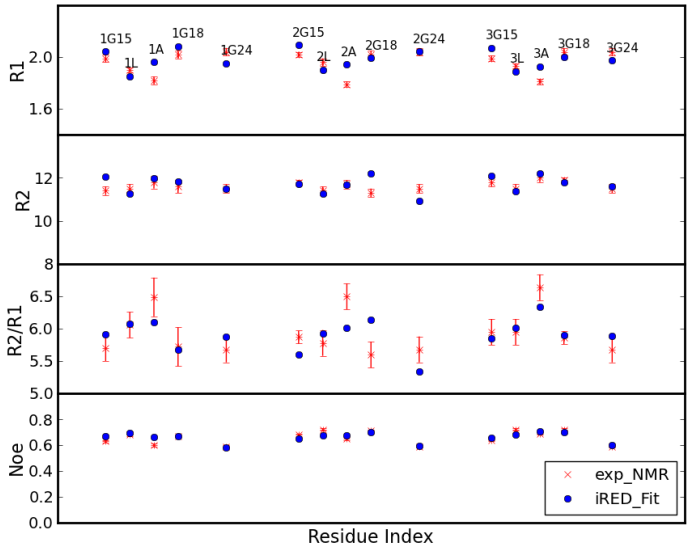


Figure 4.6: *The ^{15}N R1, R2, and NOE relaxation data at 500MHz proton frequency for labeled residues -(G15-L16-A17-G18)- and G24 in T3-785. l, 2, and 3 represent leading, middle and trailing chains, respectively. Relaxation parameters calculated using the iRED analysis after fitting the correlation times of the 10 largest modes are shown as blue circles. Experimental relaxation data are shown as* × *with errors at 20^0C at 500 MHz.*

Re-scaling the overall tumbling times eliminates the possibility of an overall increase in R1 attributed from the fast tumbling in solution compared to experiment (as discussed above). Therefore, the internal modes from IRED analysis clearly give the higher values of R1 than ones from experiment for the labeled residues Ala16 (Fig. 4.6), implying that the force field may overestimate the rigidity of amide N-H vectors at Y position in the imino-acid poor region.

Another possible explanation (borrowed from Schurr *et al. [113, 114]*) suggests that this increase in R_1 values would be interpreted as greater flexibility contributed from the collective bending and twisting motion of backbone. This greater flexibility will give the larger internal relaxation time and explain why the experimental R_1 values of Gly15 NH and Gly 18 NH are slightly higher than the others *e.g.* Leu16 NH and Ala17 NH (Table 4.5). If this statement is true, then it may imply that dynamics of backbone of the collagen-like peptide might be non-rigid body like. It may also suggest that a single reference frame cannot be defined for analysis of the MD trajectory.

4.4 Internal Motions for NH Bond Vectors

Next, we turn to a discussion of internal motions in collagen-like peptides. These are most commonly described in terms of order parameters S^2, which range from one to zero, and measure the spatial restriction of that vector in the molecular frame.

4.4.1 Order Parameters

Here, we calculated S^2 from the MD trajectory in two ways, In method *(1)*, overall motion of the protein is first removed by superposing the backbone heavy atoms of each conformer to a reference frame. Then, $C_I(t)$ is calculated directly from the MD trajectory after superposing, and S^2 is obtained by least-squares fit of the "extended" model-free (EMF) model (more details in Appendix) for the correlation function introduced by Clore, *et al* [107], which is defined as

$$C_I(t) = [S^2 + (1 - S_f^2)e^{-t/\tau_f} + (S_f^2 - S^2)e^{-t/\tau_s}]. \tag{4.7}$$

EMF model allows the internal motion on two different time scales; τ_f and τ_s are the relaxation time for the fast and slow internal motions as well as S_f^2 and S_s^2 are the squared order parameters on the fast and slow time scales, respectively, where $S^2 = S_f^2 S_s^2$. In this model, S^2serves as an approximation to the long-time limit of the internal autocorrelation function $limit_{\tau\to\infty} C_I(\tau)$. In method *(2)*, S^2 is computed from the iRED where the conformers do not need to be superimposed. Additional details on this analysis are provided in Sec. 4.8.4.

Two analysis methods yield S^2 values in the T3-785 that are consistent with each other (Fig. 4.7): high consistency in the center region with higher value of S^2 but divergence a bit at the sites with lower values of S^2 toward the end (Fig. 4.8(a)). Because there is no need to separate internal and overall motions in iRED analysis, high correlation between two methods implies the decorrelation of the bond vector orientation from molecular tumbling. The larger discrepancy between the two methods is observed among the residues toward the terminus, suggesting that the capability to separate internal and overall tumbling motions might be limited.

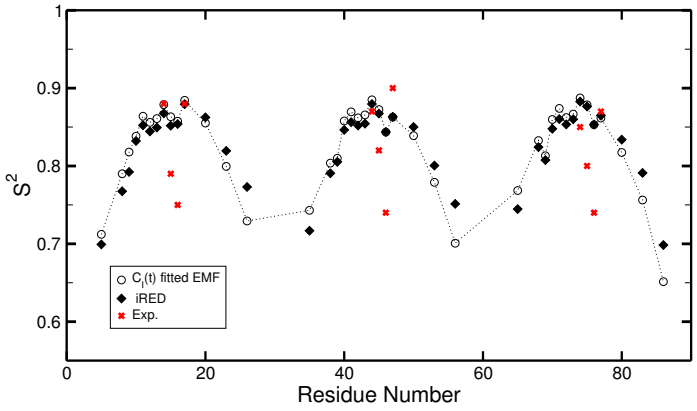


Figure 4.7: *Comparison of the generalized order parameters. NMR-derived S^2 based on model free approach (Gly modified) 500+600+800 NMR on the T3-785 are labeled as red cross. Calculated values were obtained from a 0.5 μs MD simulation using iRED and internal correlation functions fitting as described in the Methods.*

The comparison of the values of S^2 obtained from MD simulation and those obtained from experiment is often presented as a test of MD force fields. However, the limitation of number of experimental observables makes this comparison difficult and incomplete [2, 64]. Qualitatively, the S^2 values obtained from both methods for most Gly residues in the triplet Gly-Leu-Ala-Gly in T3-785 peptide are consistent with experiment. It suggests that the MD simulation using the Amber ff99SB force field with the SPC/E water model can give a good agreement for the Gly residues. The EMF-derived S^2 of Gly residues exhibits a general correlation with the presence of main chain hydrogen bonding of its amide proton (Fig. 4.8 (b)). This hydrogen bond-related correlation was also observed in the ^{15}N relaxation measurement on the globular protein human ubiquitin [77]. These observations lead to a conclusion that triple helical tight interactions provide the dominant restriction to high frequency motion of the main chain with hydrogen bonding interactions.

Unlike Gly residues, the simulated S^2 for Leu and Ala (especially) residues are significantly higher than NMR-derived S^2. These sites are highly exposed to the solvent, it is possible that AMBER ff99SB force fields underestimate the flexibility of amide N-H vectors at the X and Y positions in the imino-acid poor region, or in solution there are regions of conformational space that give additional flexibility but have not been accessed in the simulation, or that the fluctuations that are responsible for the experimental relaxation data possibly more complex than can be described by the model-free formalism. But it is also possible that the method for extracting order parameters from the experimental data is flawed, and we explore this more in the next chapter.

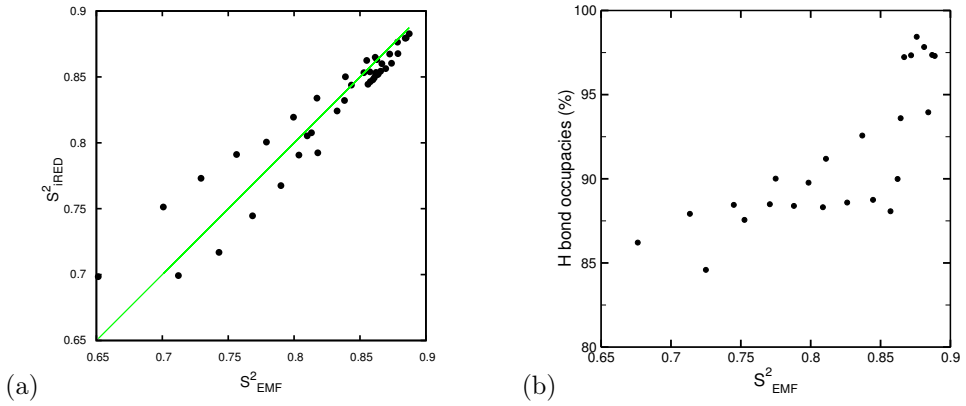


Figure 4.8: *Predictions of the values of S^2 from MD simulation. (a) A correlation plot between iRED-derived S^2 and the EMF-derived S^2 values for the vectors considered and the correlation coefficient r= 0.98. (b) A correlation plot between EMF-derived S^2 and main-chain hydrogen bonding of an amide hydrogen of Gly residues in T3-785.*

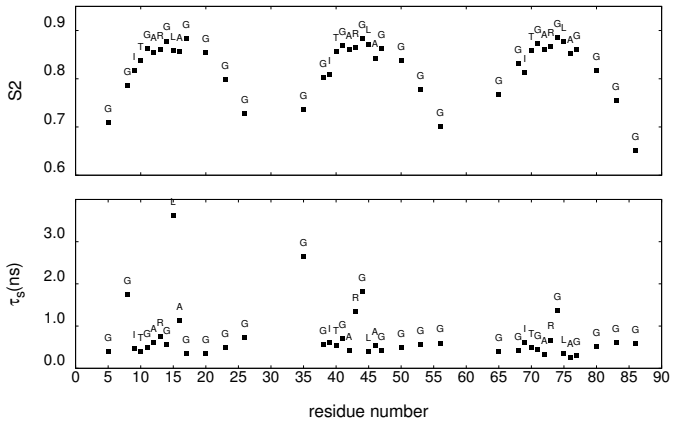


Figure 4.9: *The S^2 and τ_s extracted from the Clore model-free formula $C_I(t) = [S^2 + (1 - S_f^2)e^{-t/\tau_f} + (S_f^2 - S^2)e^{-t/\tau_s}]$ fitted to the correlation function of NH bond vectors in the peptide T3-785 of MD simulation.*

Last but not least in this section, we present internal correlation time τ_s (Fig. 4.9) from the Clore model-free analysis on the peptide T3-785. Most τ_s values are around half nanosecond, while for some of bond vectors, τ_s values are even close to the overall tumbling time ($\sim$ 3.5 to 4.5 ns) in the simulation. The physical nature of these slow fluctuations of the backbone is unknown, however, the nanosecond time scale might be related to the occurrence of large

amplitude or more complicated concerted motions. These ideas will be explored in more details in the next chapter.

4.4.2 Orientational Fluctuations

The predicted S^2 profiles can be reflected in the reorientational fluctuation of bond vectors in molecular frame [130]. The mean value of orientation of individual NH bond vector with respect to the molecular frame and its standard deviation are given in Table 4.6. That the distribution of orientations of N-H bond vectors in simulation follows "Gaussian distribution" indicates that the internal motion, particularly in the backbone, is highly restricted and dominate in the case of harmonic dynamics.

		G6	G9	I10	T11	G12	A13	R14	G/A15	L16	A17	G18	G21	G24
T3-785	L	90(10.6)	86.3(9.1)	69.8(9.1)	77.4(6.8)	84.1(7.8)	70.2(7.5)	80.9(5.9)	84.1(7.5)	70.7(7.2)	82.1(7.0)	84.7(7.4)	89.5(8.9)	90.3(9.9)
	M	90.2(10)	86.0(9.1)	72.1(9.1)	79.1(6.2)	84.6(7.8)	68.4(7.0)	82.2(5.7)	86.5(7.4)	70.4(7.2)	79.5(7.1)	86.0(8.0)	88.1(9.1)	88.2(10)
	T	88.0(9.8)	85.5(8.4)	72.4(9.1)	80.4(6.3)	83.4(7.6)	69.7(6.8)	84.0(5.7)	85.2(7.3)	67.7(7.2)	81.4(6.8)	89.1(8.8)	89.3(9.4)	88.1(10)
POG10	L	88.5(10.4)	89.2(9.6)			89.6(9)			89.2(8.7)			89.3(8.7)	89.5(9.0)	89.5(9.8)
	M	88.7(10)	88.6(9.5)			89.3(8.9)			89.3(8.5)			88.9(8.8)	88.9(9.2)	89.2(10)
	T	89.1(11)	89.2(10)			89.3(9.1)			88.9(8.8)			89.4(8.7)	89.3(8.8)	88.8(9.5)
PPG10	L	91.2(11)	89.9(9.9)			89.0(10)			88.9(8.6)			89.6(8.5)	89.6(8.9)	89.2(9.5)
	M	90(10.6)	89.4(9.4)			89.7(8.8)			89.4(8.5)			89.2(8.6)	89.6(9.1)	90.2(10)
	T	88.8(9.9)	88.3(9.2)			89.1(8.7)			89.8(8.5)			89.1(8.7)	88.8(9.3)	88.9(10.2)
Gly2Ala	L	91.3(11)	89.6(10)			88.8(9.4)			85.6(10.9)			84.3(8.6)	91.2(9.0)	90.1(9.5)
	M	89.4(11)	89.5(9.6)			89.8(8.8)			100.7(10.3)			84.4(9.2)	89.0(9.4)	90.3(10.2)
	T	88.2(10)	88(9.5)			89.4(9.3)			82.2(11.2)			89.(8.7)	86.9(9.6)	87.7(11)

Table 4.6: *Mean*[1] *value of orientation of N-H vectors with respect to principal axis of inertia tensor*[2] *of triple helix. The value in the parenthesis is the standard deviation.*
[1] *Mean,* μ*, of the Gaussian distribution function* $p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-\mu)}{2\sigma^2})$ *and its standard deviation,* σ *shown in bracket.*
[2] *The orientation of inertia axis is aligned to the N terminus of the triple helix.*

The iRED-derived order parameters S^2_{iRED} of the N-H vectors are presented in Fig. 4.10 for four different triple helical collagen peptides. All of the backbone bond N-H vectors exhibit similar S^2 profiles with gradually reduced S^2 values toward both termini of the triple helices (except where the Gly is replaced by Ala residue); a maximum S^2 values of around 0.88 occurs near the middle of the sequence, while a minimum value of around 0.65 is observed at the ends. This fraying effect is accomplished by a relatively low separability index (g$\sim$ 2.5) between overall and internal motions, which suggest that the capability to distinguish between internal and overall motion is limited. The same trend is observed in a study of backbone dynamics of α-Helix in the uncorrelated relation in dihedral angles, in which the distortions is maximized relative to the reference helix and the fraying effect is high [78].

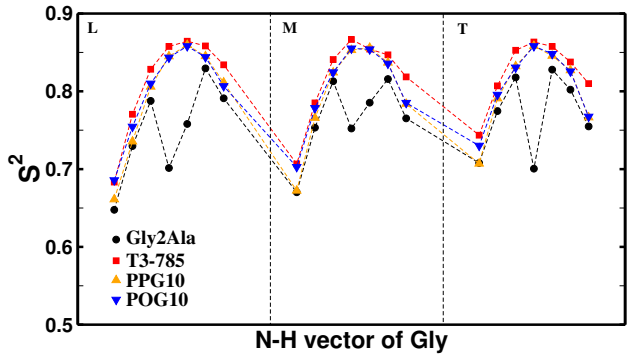


Figure 4.10: *Comparison of the model-free derived order parameters from iRED analysis for all peptides*

It is worthy noting that the values of S^2 of Gly residues in imino-acid poor region is slightly higher than those at the same sites in the imino-acid rich region. This is consistent with the trend of orientation fluctuations of the Gly NH vectors in Table 4.6, in that the fluctuation amplitudes increase near the termini.

4.4.3 Inter-chain Hydrogen Bonding

Here we look carefully into the inter-chain hydrogen bonds to understand the higher order parameters S^2 observed in the simulation than those in experiment in the peptide T3-785.

In crystal, a report of T3-785 structure has been published on the formation of interchain water-mediated hydrogen bonds as well as its hydration pattern [28, 55]. On average, two of the three amide groups in each triplet in the imino acid poor region are involved in hydrogen bonds in T3-785 in crystal (Fig. 4.11); one is the conventional inter-chain hydrogen bond, NH(Gly)$\cdots$OC(X), between the amide proton of Gly residue in one chain and the carbonyl groups of the X position in the adjacent chain; another hydrogen bond; the second type of hydrogen bond runs in the opposite direction from the first type hydrogen bonds through water molecules. On average, these nine water molecules (W1m$\sim$W9 shown in Fig. 4.11) are 2.85 Å away from N atoms and 2.99 Å away from the O atoms, and make an average angle of 107.1°. Despite of the lack of imino acids in the central zone, there is little deviation of these values from the average values, indicating that the three chains in this imino-acid poor region are not pulled apart to any significant extent.

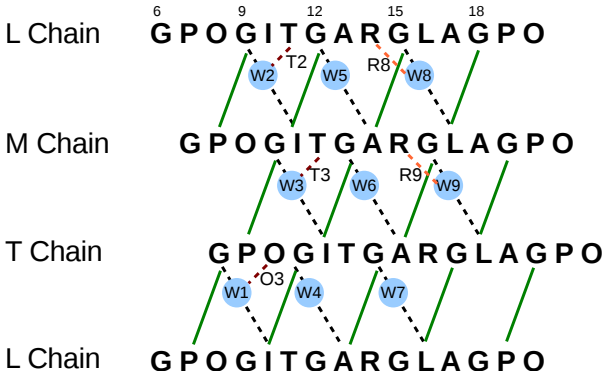


Figure 4.11: *Hydrogen bonding pattern in amino acid-rich region of peptide T3-785 in "crystal structure". The conventional main chain direct hydrogen bond NH(Gly)-O=C(X) is shown as solid green line. All water-mediated hydrogen bonds are marked with dotted lines. There are three types of water-mediated hydrogen bonds shown: (i) water molecules, denoted as blue circles, form hydrogen bonds with the carbonyl oxygen of Gly residues and with backbone nitrogen of residues in X position of an adjacent chain (black dotted lines); (ii) water molecules, W1, W2, W3 form extra hydrogen bonding to side chains of Hyp (T), Thr (L), Thr (M), respectively (red dotted lines). Although Thr in T chain is close to W4, its side chain is not properly oriented for H bond forming in this crystal structure; (iii) water molecules W8 and W9 form extra hydrogen binding to the long side chains of Arg (L) and Arg (M), respectively.*

In solution, NMR hydrogen exchange measurements showed that Gly and Leu have similar exchange rates and are more protected from exchange than Ala, suggesting that the amide groups of the Gly and X positions have similar hydrogen bonding and different from those at the Y position [64]. Both experimental observations are accordance with the model with two hydrogen bonds per tripeptide, which has been proposed for Gly-X-Y sequences where X is not an imino acid by Ramachandran and Chandrasekhar [81].

In our simulation, these two hydrogen bonds are also observed for every triplet (Gly-X-Y) in the imino-acid poor region, and the amide groups in the Y position are essentially external and exposed to the solvent molecules. In Table 4.7, higher occupancies are displayed by sites W1, W8, and W9, in which the water molecules are further stabilized by H-bond interactions withe the side chain of the residues at the Y position. This high occupancy of W8 and W9 might be due to mobility of Arg14 side chain and the IIe. Additionally, the high occupancy of W1 is due to the participation of the hydroxyl group of Hyp residue water mediated interaction with carbonyl group of Gly in the same strand. These water sites locate around the edge of imino-acid rich and poor region might be crucial for the triple helix stability.

The imino acid poor region of the T3-785 molecule is heavily hydrated than the POG regions at the ends of the molecule. Then the C-terminal zone is next well hydrated, and N-terminal zone is least well-hydrated [55]. It has been suggested that the extensive hydration in the central region is partially the result of the additional donor groups of the amide nitrogen atoms in contrast to POG regions in which the only donor groups are the hydroxyl groups of Hyp and water molecules.

	W1[1]	W2	W3	W4	W5	W6	W7	W8	W9	T2[2]
Occupy%	80.3	58.9	39.8	51.4	79.6	67.7	66.5	88.9	83.1	6.7
Ave. time(ps)	45	21	31	32	44	28	28	157	53	12
Max. time	156	68	98	126	152	118	108	476	158	90

Table 4.7: *Water mediated hydrogen bonds (labels are shown in Fig. 4.11) parameters in the imino-acid poor region in T3-785 peptide of MD trajectory. HB length cutoff 3.3 A and HB angle cutoff 120^o.*
[1] *W denotes NH(X)-WAT-OC(Gly) water-mediated hydrogen bonding two different strands: the water molecule bound to the amide group of the amino acid in the X position and the carbonyl groups of Gly in the adjacent strand.*
[2] *T2 denotes hydroxy group of Thr in leading chain hydrogen bonding to the water molecule W2 as shown in Fig. 4.11, which is Oγ1Hγ1-WAT-OC(Gly) hydrogen bond connecting the same strand: the water molecule interacts with the side chain of the Thr in the same chain as the Gly carbonyl which is at the N-terminal side of Thr.*

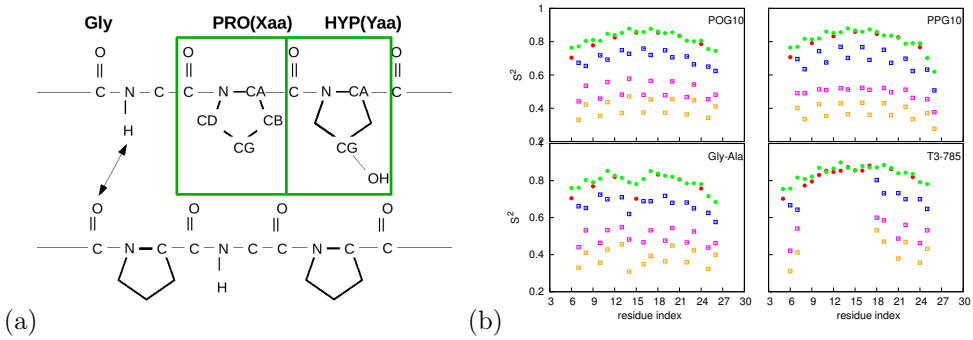
It should be noted that the low B-factors of the amino acid residues do not reflect on the stability of the triple helical conformation, but rather are indicative of the numerous interactions mediated by these residues to symmetric-related molecules. Thus, the presence of non imino acid residues plays an important role in collagen structure by providing a looser helical conformation while also permitting the formation of important hydrogen bonding that stabilizing not only the triple helix but also mediated helical helices contacts that are important in collagen packing [54]. The main reason to have a water molecule bridging N-H(X)$\cdots$O=C (Gly) hydrogen bond is to avoid the distortion of the triple helix [131] and further get a set of interactions almost as strong as direct inter-chain hydrogen bonds.

4.5 Internal Motion for Collagen Side-chains

4.5.1 Mobility of Side-chains upon the Sequence of the Triple Helix

Collagen side chains, except one at Gly residues are accessible to solvent, therefore, side-chain flexibility and mobility are important for understanding the interaction of collagen with receptors, other collagen and other matrix molecules. NMR studies of amino acids incorporated into collagen have showed a high degree of side-chain and backbone mobility in

fibrils [62]. High resolution x-ray structures of model peptides show side-chains with multiple occupancies, indicating that fluctuations of the side-chain conformation can occur within crystals [132]. Studies of intrinsic fluorescence in the collagen triple helix have provided an insight into the mobility and accessibility of individual side chains at the X and Y position [133].



(a) (b)

Figure 4.12: *(a) Structure of collagen motif, Gly-X-Y triplet, where X is often Pro and Y is often Hyp residue. (b) The S^2_{iRED}(Eq. 4.15) for backbone bond vectors and side chain of aromatic residues (Pro/Hyp) for four collagen-like peptides. Filled circles represent backbone N-H (red) and CA-HA (green) bond vectors, respectively. Open squares represent side chain CD-HD (blue), CB-HB (magenta), and CG-HG (orange) bond vectors.*

However, specific models for characterizing the internal motion has not yet been obtained for large side chains and the data analysis is complicated by the fact that a side chain can assume many different conformations; meanwhile, the experimentally determined Lipari-Szabo S^2 provide minimal data about small regions of the collagen-like peptide. Th iRED approach allows the calculation of the S^2_{iRED} for any N-H or C-A bond vector as it requires no needs of superimposing the conformers (Fig. 4.12 (a)). Among the C-H bond vectors in side chain at X and Y position, CB-HB and CG-HG are more accessible to the solvent than GD-HD vector. The iRED analyses on the simulation of the four CMPs show that all the backbone bond vectors exhibit similar S^2_{iRED} profiles with gradually reduced S^2 value toward both termini by around 0.15 compared to ones in the middle (Fig. 4.12 (b)).

The profiles of S^2_{iRED} of these four peptides reveals the extent to which side-chain mobility reflects differences in local flexibility upon the sequence content of the triple helix. For the GPO region in $(POG)_{10}$, Gly→Ala, and T3-785 peptides, the values of S^2_{iRED} of side chain at the X position are lower than those at the Y position, indicating that side chains at the X positions are more mobile than ones at Y sites in the host region. It is consistent with the observation by Kristein et al., [133] who employed Trp fluorescence to compare the X and Y positions of collagen in solution. It has been suggested that this higher accessibility

of side chain at X position may be responsible for interacting with other molecules or self-association [133, 134]. It is possible that this extra mobility allows side chains at X position to fluctuate to fit into the binding to other macromolecules. Replacing Hyp residues with Pro residues in Y position $(PPG)_{10}$ or mutating Gly with Ala (Gly→Ala) apparently interrupts this pattern of mobility of side chain at X and Y sites; e.g. an increase in side chain flexibility at Y position and a decrease in the mobility of side chain of proline at X position.

4.6 Correlated Dynamics

4.6.1 Principal Component Analysis

General speaking, the intramolecular dynamics of protein is very complex compared to overall rotational diffusion, since it consists of motions of different types, at different time-scales, and amplitudes. One strength of MD simulations is its ability to reveal correlated dynamics through principal component analysis (PCA).

The results of PCA on all the four peptides share some similarities including that (*i*) the first five PCAs account for up to ∼80% of the structural variance of the Cα atoms. Over half of this structural variability is mainly described by the PC1 and PC2 modes, which corresponds to orthogonal global bending-like oscillatory modes of the backbone (Fig. 4.13); (*ii*) the PC3 mode describes the twisting-like motion; (iii) whereas the next higher modes represent the higher harmonics of motions. In overall, visualization of the projection of the PCA modes onto the MD trajectories shows that the essential dynamics of these peptides is very similar.

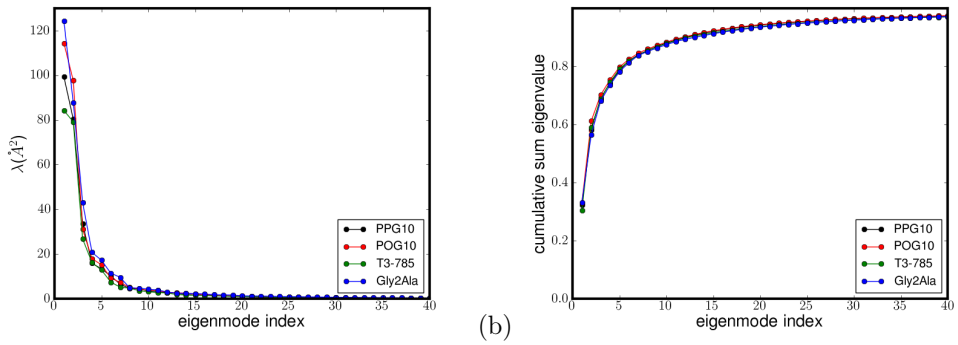


Figure 4.13: *(a) Eigenvalues plotted in descending order of magnitude are obtained from Cα-atoms coordinate covariance matrix in the MD simulation of collagen model peptides. (b) Relative cumulative sum eigenvalues to the total eigenvalues of the motions along the eigenvectors obtained from the Cα atoms coordinates covariance matrix in all peptides.*

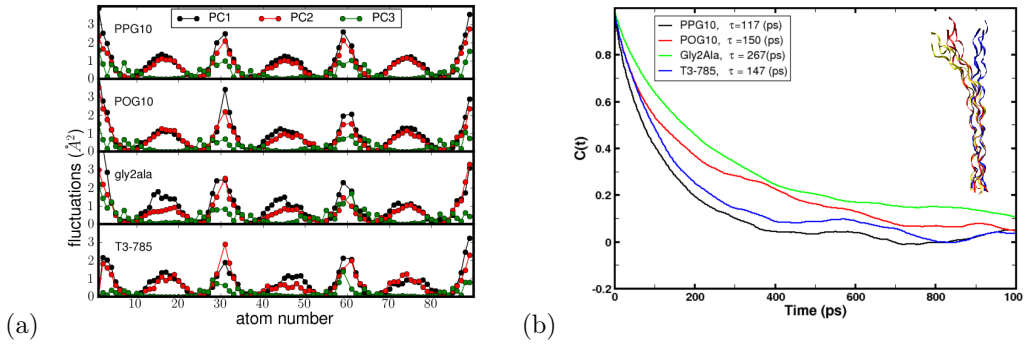


Figure 4.14: *(a) The atomic fluctuation of the essential motions of Cα atoms of the collagen mode peptides along the most three significant eigenvectors: PC1, PC2, and PC3. Residue numbers 1-29, 31-59, and 61-89 belong to leading, middle, and trailing chain, respectively. (b) Time correlation function of backbone bending and its associated correlation time.*

The atomic fluctuations from the three largest PC modes on the four peptides are shown in Fig. 4.14 (a). Although, these fluctuations have some similarities in both ordering and the nature of PCs for all four sequences, it does reveals some slight differences induced by mutations in the projection of the PCs modes onto the MD trajectories. For example, a Gly→Ala has the most flexibility in terms of the backbone bending. The Gly→Ala peptide has the longest bending correlation time among the peptides (Fig. 4.14 (b)).

In addition, the Gly→Ala also exhibits an anisotropic backbone bending, which is reflected in the atomic fluctuations of PC1 of each chain that are slightly different from one another (This chain asymmetry was discussed more fully in Chapter 2, above). While the T3-785 exhibits the least fluctuations in both bending (PC1 and PC2) and twisting-like (PC3) modes. This is consistent with the previous calculations of the S^2 profiles on the same peptides, in which T3-785 has the highest values of S^2_{iRED} than other peptides (Sec. 4.4).

4.6.2 iRED Approach

iRED is an alternative method for detecting correlated molecular motions through an analysis of the angular fluctuations of spin interactions as well as describing both the overall reorientation of the molecule, along with collective motions of internal fluctuations of chemical groups.

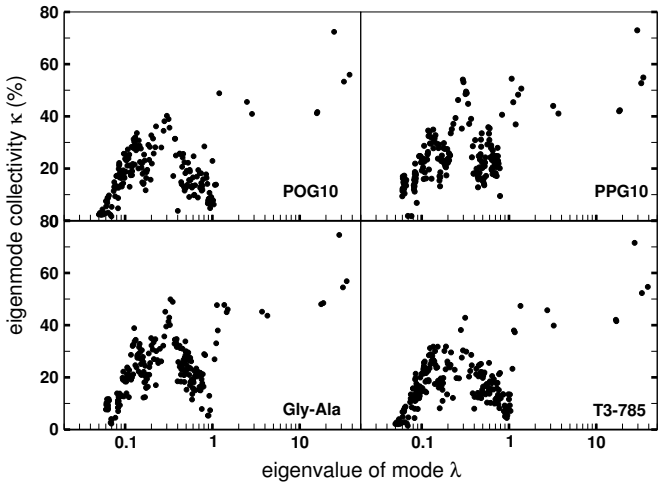


Figure 4.15: *Eigenmode collectivity plots. The collectivity factor, κ (defined in Eq 4.14), as a function of eigenvalueλ of each mode for the peptides POG10, PPG10, Gly-Ala, and T3-785. The degree of separability between global and internal modes can be quantified by a separability index g_2. The value of g_2 is 2.83, 2.76, 2.58 and 2.62 for peptides POG10, PPG10, Gly-Ala, and T3-785, respectively.*

It can give a quantitative measure for the separability of global and internal motions in collagen-like peptides. A good separability is characterized by a large gap in two dimensions: eigenvalues λ and collectivity κ. The plot of the eigenmode collectivity κ versus the eigenvalue λ of each mode for each collagen-like peptide provides a summary of the motions accessible to each peptide (as shown Fig. 4.15). This plot summaries the properties of the dynamics of system and examines if the separability assumption can be applied to the system, if not, the experimental data will have to be analyzed using something other than the Lipari-Szabo model-free formalism.

The results of iRED analysis lead to a variety of conclusions (Fig. 4.15):

1. collagen-like peptide processes overall "axially" anisotropic rotation as the three catalogs out of these five largest amplitudes modes;

2. the overall tumbling of these four peptides are similar as a results of the similar distributions of eigenvalues of the five largest modes;

3. these four collagen-like peptides have similar value of separability gap between 2.58 and 2.83;

4. there are some differences in the first two largest amplitude modes corresponding to the correlated internal motions. For instance, in Gly→Ala peptide, the two largest local modes shift to higher value in the λ dimension compared the ones in POG10 or

PPG10 peptide, while in T3-785, the first two largest internal modes are the lowest ones among four peptides. The decrease in the value of λ in the local modes is reflected in an increase in the value of S^2_{iRED} (Fig. 4.10);

5. although the eigenvalues corresponding to the global tumbling motions can be distinct from those of internal motions in λ dimension by visualization, these values of separability index are close to that of a less ordered system where the overall and internal motions are "not" separable [67];

6. furthermore, for the system in which the internal and overall motions are separable, the collectivities (κ in the Y axis) should decrease with descending λ. However, such a gap is not really observed in the collectivities dimension, indicating that some internal modes should be the collective motions through the whole system instead the high frequently local motion. These observations above implies that the global and internal dynamics of the collagen-like peptides are likely not rigorously separable at some extent, it might lead to the inaccurate NMR-derived order parameters, and the discrepancies of S^2 toward the terminus between iRED method and model-free approach.

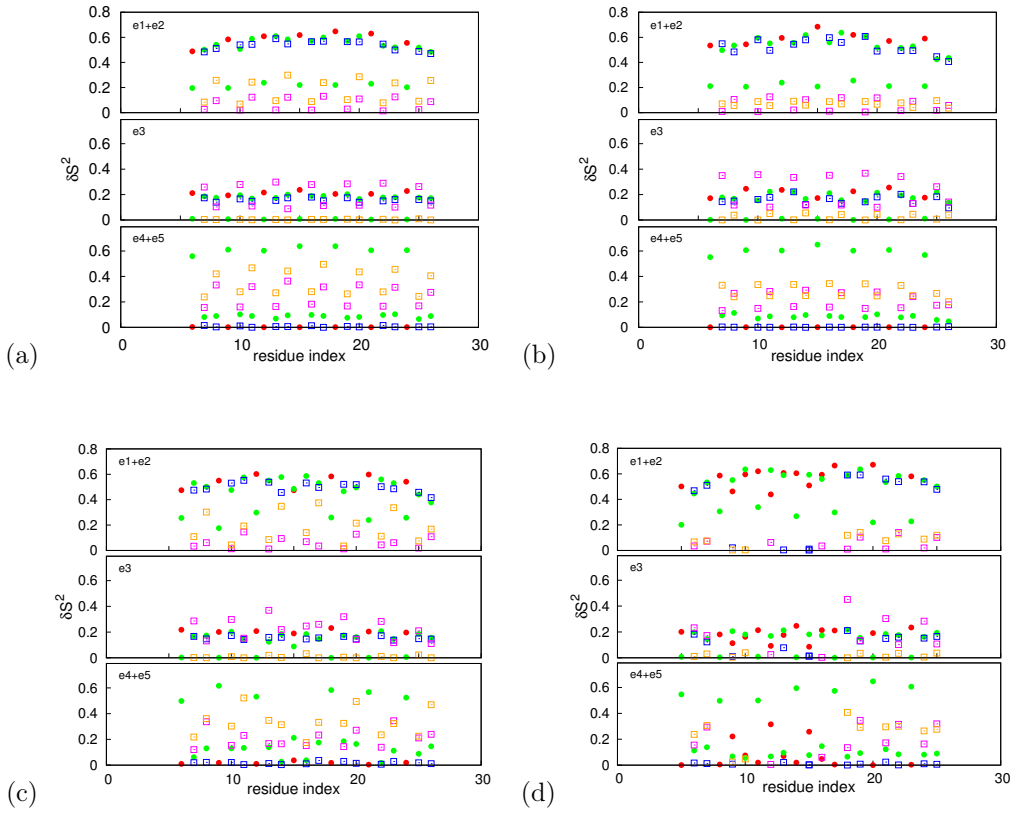


Figure 4.16: *The values of δS^2 of the bond vectors along the triple helix for the reorientational modes $e = 1, ..., 5$ in the peptide POG10 (a), PPG10 (b), Gly-to-Ala (c), and T3-785 peptide (d). Filled Circles represent spin pairs of the backbone: N-H (red) and CA-HA (green) interactions, respectively. Open squares represent spin pairs of side chain: CD-HD (blue), CB-HB (magenta), and CG-HG (orange) interactions.*

Fig. 4.16 depicts how the iRED modes affect the sites collectively. The overall reorientational modes $m = 1, .., 5$ are highly collective affecting a large portion of all bond vectors in a correlated way. The residues that are not affected by the $e1$ mode are significantly affected by the $e2$ eigenmode (data not shown), indicating that they are indeed separate modes with global character since they demonstrate a large collectivity index, as shown in Fig. 4.15.

According to an axially anisotropic rigid-body diffusion theory, the overall correlation time of any spin pair should be only dependent on its orientation regardless the position of the label sites. However, the positional-dependence of mobility is seen by the modes of $e1 + e2$ along the triple helix: the effect of $e1 + e2$ on backbone decreases as the sites move away from the center (excluding the mutated sites), implying the coupling between the sites and

the intrinsic flexibility of the triple helix. The third mode $e3$, the longest decay mode, affects over 70% spin interactions, where the side chain CB-HB at the X position are affected the most. The effect of this mode on the sites at the end is similar to the sites in the center. The second longest decay mode with $e4$ and $e5$ together contributes to backbone CA-HA vector of Gly residues the most throughout entire helix. Overall, these five modes show a very similar sequence profile.

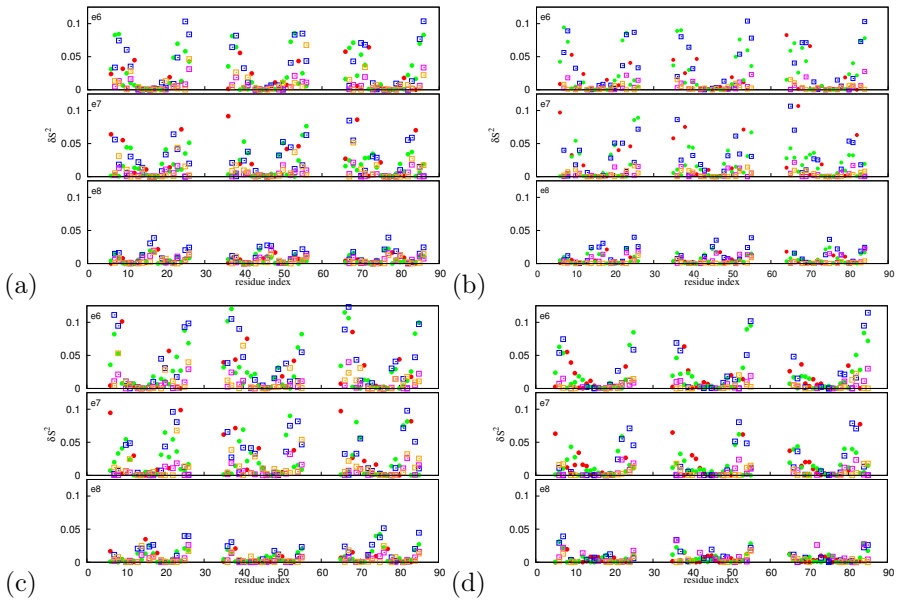


Figure 4.17: *The values of δS^2 of the bond vectors along the triple helix for the modes e6, e7, and e8 in the peptide POG10 (up left), PPG10 (up right), Gly-to-Ala (bottom left), and T3-785 peptide (bottom right). Filled Circles represent spin pairs of the backbone: N-H (red) and CA-HA (green) interactions, respectively. Open squares represent spin pairs of side chain: CD-HD (blue), CB-HB (magenta), and CG-HG (orange) interactions.*

The remaining modes, associated to the fast motional modes in Fig. 4.17, have much smaller and localized contributions to the internal dynamics of peptides. The modes $e6$ and $e7$ mainly affect the terminal end but the sites in the center region of helix, indicating some kind of internal motions, e.g. twisting-like motion. Mode $e8$ mainly affects the spin interactions of residues around the center of triple helix, implying some kind of bending motions which seems to be the least in T3-785 compared to other peptides. The mobilities attributed by these modes are in lines with the observations found in PCA.

4.7 Conclusions

The structural features of collagen-like peptides have been investigated experimentally in studies using x-ray crystallography and solution NMR. The global diffusion properties and internal motion have also been investigated by NMR spectroscopy. The results presented in this chapter can be summarized as follows:

4.7.1 Overall Tumbling Times τ_M

One long 0.5 μs MD simulation of T3-785 peptide solvated with SPC/E waters at 298 K allows us to look into the influence of the flexibility of a molecule on NMR relaxation parameters (R1, R2, and NOE). Various aspects of the triple helical peptide dynamics have been considered. In the simulation at 293 K, the peptide stays in a folded conformation, and the internal dynamics is confined to motions in the side-chains and in the terminal residues.

The rotational correlation time computed with iRED method is approximately 4.5 ns, which is 30% lower than the value indicated by NMR experiments at 293 K ($\tau_M \sim 6.9$ ns). Theoretical calculations based on hydrodynamic modeling shows that the rotational diffusion tensor is affected by the extent of bound water. It was found that the faster overall tumbling from simulations is the consequence of the absence of the hydration layer attached to the peptides, indicating inaccuracy of the strength of protein-water interactions in the simulation.

Many of the common rigid models overestimate the diffusion constant. Previous studies of the viscosity for different water models commonly showed the significant differences between the calculated results and the experimental data [135]. For a SPC/E water model, Balasubramanian *et al.* [136] calculated its viscosity at T=303.15 K about 18% less than the experimental value; Hess [137] computed the viscosity at T=300K, which is about 30% lower than the experimental value; Wu *et al.* simulated its shear flow at T=298.5 K and reported an error about 23% between experimental data and calculated results [138]. As would be expected, the use of the SPC/E water model here underestimated the overall tumbling times by speeding up the rotational diffusion [135, 139]. Further improvement of current water models for use in simulation setups to treat accurately the strength of protein-water interaction which is important in the study of the NMR relaxation parameters and the hydration layer attached to the peptides.

4.7.2 Order Parameters

The order parameter represents the information about the reorientational fluctuation of inter-nuclear vector at any site of protein. The order parameters S^2 computed in this work

for triple helical peptide are basically *(1)* the plateau value of local or internal autocorrelation function of a given N-H bond vector after superposing the ensembles to the reference frame; and *(2)* iRED analysis without the requirement of superimposition. Two of methods provide highly consistent results for the triple helical peptides of studies. The variation of S^2 along the triple helical peptide reflects contribution from long-range motion onto the local environment. The fraying effect is observed in the order parameter profiles toward the termini of the triple helices and the fluctuation amplitudes of the NH bond orientation with respect to the molecular frame increase from middle toward the termini. Although the origin of this effect is nonlocal, it is seen locally in S^2.

The present paper further discusses hydrogen bond factor on the local order parameters. The MD simulation of T3-785 peptide show a second inter-chain backbone hydrogen bond mediated by a water. NMR hydrogen exchange studies and X-ray structure support its presence as well. The studies of sequence-dependent variation in triple-helical parameters, the collagen regions with rich in imino acids, e.g. the (Pro-Hyp-Gly) repeat units, adopt tighter helix, close to 7/2 symmetry (3.5 residues/turn), while regions with deficiency in imino acids are likely to have a looser helix. This looser conformation might fit the second type water-mediated inter-chain hydrogen bond into the helix.

4.7.3 Correlated Motions Accessible to Collagen Model Peptides

MD simulations have pointed to significant amounts of bending and twisting motions existing in the triple helical collagen peptides. PCA has showed that over half of the backbone structural variability are mainly described by the orthogonal global bending-like oscillatory modes of the backbone as well as the twisting-like motion.

Similar to PCA, the iRED analysis have exhibited that some fast motional modes mainly affect the terminal sites, indicating some kind of twisting-like motion, and other modes affect the spin interactions of residues around the center of triple helix, implying some kind of bending motions. Both PCA and iRED methods suggest that the imino-acid poor region in T3-785 has the smallest-amplitude low frequency motions in terms of bending and twisting compare to other peptides. These concerted motions imply the undetermined reference frame, which will result in systematic errors to the model-free derived S^2 value.

4.7.4 Extra Information by iRED Approach

The iRED method basically provides a general framework for characterizing the macromolecular dynamics on the timescales accessed by MD simulation without any assumptions. Therefore, the results of this method have been all over the places in this chapter. Most results of the iRED method have been compared with others to test the validation of the

other approaches. However, it also provides some information that other methods have not explored yet:

1. NMR relaxation measurement of proteins provided the information of motional amplitudes and times scales under the standard procedures, such as model-free approach and analytical model, for data interpretation. The iRED analysis of rank 2 provides such information depicting correlated dynamics of different NH vectors together with mode-specific correlation times;

2. unlike model-free approach, the iRED analysis is not restricted to globular systems for which overall and internal motions are well separable. Isotropic RED is based on a principal component analysis of irreducible lattice functions. The overall tumbling observed during the finite MD trajectory is displayed as isotropic ensemble. In this study on the collagen-like peptide, the satisfactory agreement between experimental and theory is achieved by adjusting the correlation times of 10 largest amplitudes modes;

3. an axially anisotropic rigid-body rotational model states that the overall correlation time of the spin-pair is orientational-dependent. However, iRED analysis reveals the positional-dependence of mobilities by the first two largest overall amplitude modes for the spin-pair with similar orientation: the effect of mobilities gradually decreases as the sites move away from the center, implying the coupling between the sites and the intrinsic flexibility of the triple helix;

4. while some specific models for characterizing the internal motion has not yet been obtained for large side chains and the data analysis is complicated by the fact that a side chain can assume many different conformations. The iRED method allows one to calculate the order parameters for the C-H bond vector of side-chain. The results have shown that side chains at the X positions are more mobile than ones at Y sites in the host region. Replacing Hyp residues with Pro residues in Y position $(PPG)_{10}$ or mutating Gly with Ala (Gly→Ala) apparently interrupts this pattern of mobility of side chain at X and Y sites.

MD simulations can be thought of as a computational "experiment"; when there are discrepancies between its results and those of the real experiment, it can often be hard to gain an understanding of the source of the differences. In these circumstances, simpler model calculations can be of use, and we will turn in that direction in the next Chapter.

4.8 Appendix

4.8.1 Parameterization of Hydroxyproline

Atom	Charge	Atom	Charge	Atom	Charge
N	-0.25480	HG2	0.02130	HA	0.0641
CD	0.00192	CB	-0.0070	C	0.5896
HD2	0.03910	HB2	0.02530	O	-0.5748
HD3	0.03910	HB3	0.02530	OD	-0.6761
CG	0.30610	CA	-0.0266	HD	0.41020

Table 4.8: *Point charges for hydroxyproline (Hyp)*

	$\chi 1$	ξ
Hyp	29.73	147.86

Table 4.9: *Dihedral angles*
a. ξ is the dihedral angle N-C^{α}-C^{γ}-O^{α}
b. χ_1 is dihedral angle (N-C^{α}-C^{γ}-C^{β})

4.8.2 General Relations in NMR Relaxation Data

According to the relaxation theory of Bloch, Wangsness, and Redfield [140, 141], the spectral density function determines the relaxation parameters measured in NMR, i.e. spin-lattice relaxation rate ($R1$), spin-spin relaxation rate ($R2$) of a 15N nucleus and the NOE factors, which are contributed by the dipolar-dipolar relaxation and the chemical shift anisotropy relaxation, according to the following

$$R1_{,j} = q_{dd}^2\{3J_j(\omega_N) + J_j(\omega_H - \omega_N) + 6J_j(\omega_H + \omega_N)\} + q_{CSA}^2 J_j(\omega_N) \tag{4.8}$$

$$R2_{,j} = \frac{q_{dd}2}{2}\{4J_j(0) + 3J_j(\omega_N) + J_j(\omega_H - \omega_N) + 6J_j(\omega_H)+$$

$$(6J_j(\omega_H + \omega_N)\} + \frac{q_{CSA}^2}{6}\{4J_j(0) + 3J_j(\omega_N)\} \tag{4.9}$$

$$NOE_j = 1 + \frac{\gamma_H}{\gamma_N}\frac{\Gamma}{R1_{,j}} \tag{4.10}$$

where $q_{dd}^2 = \frac{1}{20}(\frac{\mu_0}{4\pi})^2(\frac{h}{2\pi})^2\gamma_N^2\gamma_H^2 < \frac{1}{r_{NH}^3} >^2$, $q_{CSA}^2 = \frac{1}{15}\omega_N^2(\Delta\sigma)^2$, $\Gamma_j = q_{dd}^2\{6J_j(\omega_H + \omega_N) - J_j(\omega_H - \omega_N\}$ which is the ^{1}H to ^{15}N cross-relaxation rate constant, μ_0 the permeability of vacuum, h Planck's constant, γ_N and γ_H the gyromagnetic ratios of ^{15}N and ^{1}H nuclei, and $\Delta\sigma$ the chemical shielding anisotropy constant, and r_{NH} the N-H distance; ω_N and ω_H

Larmor frequencies (in rad/s) of ^{15}N and ^{1}H nuclei, respectively. We set $\Delta\sigma$ = -176 ppm for the ^{15}N spin , r_{NH}= 1.01 Å, and external magnetic field of 11.74 T. The other values are 1.054× 10E34 for the Plank constant, 4π × 10E7 TmA^{-1} for the permeability of free space μ_0, 2.675 × 10E8 T$^{-1}S^{-1}$ for γ_H and -2.71 × 10E7 T^{-1}S^{-1}for γ_N.

For the rate of transfer of longitudinal magnetization from the ^{15}N to its proton R1 in Eq. 4.8, as ω_H and $\omega_H \pm \omega_N$ are large frequencies compared to ω_N, consequently $J(\omega_H \pm \omega_N)$ and $J(\omega_H)$ are altogether negligible in compare to $J(\omega_N)$, and make no significant contribution to R1. For the rate of transfer of transverse magnetization, only the term with zero angular frequency make a non-vanishing contribution.

These relaxation constants for backbone amide moieties in proteins in solution depends on the autocorrelation functions, $C(t)$, of the N-H bond unit vector orientations $\mu(t)$ in laboratory frame. This autocorrelation function is defined by

$$C(t) =< P_2[\mu(0) \cdot \mu(t)] > \tag{4.11}$$

where $P_2[x] = (3x^2 - 1)/2$ Legendre polynomial and the angle brackets denoted ensemble averaging. The Lipari-Szabo model interprets the explicit expression to $J(\omega)$, which is the Fourier transformation of $C(t)$, in terms of the global tumbling correlation time (τ_M) as well as internal correlation time (τ_e) associated with the motion for each spin pair and an order parameter (S^2), which is a measure of the spatial restriction that the vector experiences in a molecular frame. This expression provides a path for refining collagen structures and the local flexibility from NMR experimental data.

4.8.3 Principal Component Analysis (PCA)

A principal component analysis (PCA) carried out on the molecular dynamics trajectory involved two main steps: *(1)* the construction of the covariance matrix, using positional deviation of the $C\alpha$ atoms from the ensemble which must be superposed to filter internal motion from overall rotational and translational motions. This is accomplished by a least-squares fit each frames to a reference structure (e.g x-ray structure or average structure); *(2)* the diagonalization of this covariance matrix, yielding a set of eigenvectors, which are the indicators the directions of the motions, and the associated eigenvalues, which represent the amplitudes of that motion. In PCA, eigenmodes are usually sorted according to variance rather than frequency. Nevertheless, the largest-amplitude mode of a PCA also represents the slowest dynamical mode. Projecting the trajectories onto individual eigenvectors provides information about the time dependence of protein conformation along these eigenvectors.

4.8.4 Theory of iRED

The iRED is an alternative method for detecting the molecular motion through the angular fluctuations of spin interactions, describing both the overall tumbling of the molecule, along with collective motions of internal fluctuations of chemical groups. The effect of these fluctuations on the relaxation parameters are reflected in variances and covariance of the spherical harmonics $Y_{2m}(\Omega)$ with L=2. "Isotropic" reorientation means that the overall orientation of molecule is isotropically distributed to get the averaged covariance matrix with each element calculated between any two relevant bond vectors (e.g. N-H or C-H vectors) in the same snapshot. Reorientational eigenmodes (eigenvectors) and amplitudes (eigenvalues) are estimated using principal component analysis of the isotropically averaged covariance matrix. Mathematical details are given in following.

The theory of iRED method are summarized briefly in this section. For the further derivation and explanations can be found in appendix of Prompers and Brüschweiler's paper [67]. The iRED analysis was used to detect molecular motions through the angular fluctuations of interaction vectors within the molecules. The isotropically averaged covariance matrix M is computed with elements

$$\overline{M}_{ij} =< P_2(cos(\Omega_i - \Omega_j)) >, \tag{4.12}$$

where P_2 is the Legendre polynomial of rank L=2, $(\Omega_i - \Omega_j)$ is the angular difference between two vectors Ω_i and Ω_j in the same snapshot, and the overbar indicates the isotropic averaging over all snapshots of the trajectory. Unlike RotDif analysis described above, in which each snapshot is rotated uniformly as a rigid body, reliance on the single reference frame is eliminated. In this approach, the conformers do not need to be superimposed because the covariance matrix M constrains only inner products which are rotationally invariant. In this analysis, all N-H, CA-HA, CB-HB, CD-HD, and CG-HG bond vectors (n=210 for peptide POG10, PPG10, and Gly-to-Ala, and n=198 for T3-785 peptide) were used to generate a second order iRED matrix. This set of inter-nuclear interactions provide both backbone and side chain dynamics.

Next, a principle component analysis is performed by solving the eigenvalue problem $M|m >= \lambda_m|m >$, where $=|m >$ describing correlated modulations of different spin interactions, and eigenvalues, λ_m representing the eigenvalues. The values of λ_m are sorted in descending order and correspond to the variance of the amplitude fluctuations of the trajectory along each eigenmode.

4.8.4.1 Separability Index g_2

The distribution of the eigenvalues is used as a direct indicator of whether the assumption of the separability between internal and overall motions is fulfilled. The separability index g_2, providing a measure of the separability of internal and overall modes of motion, is calculated as

$$g_2 = \frac{\Sigma_{i=1}^{n} \lambda_i}{\Sigma_{i=6}^{n} \lambda_i} \tag{4.13}$$

If the molecule is complete rigid, indicating no eigenmode attributed from the internal dynamics, so $g_2 \to \infty$. For a flexible molecule, g_2 will reflect how good the separability is between internal and overall tumbling motions. A large value of g_2 reflects good separability between internal and overall motions.

4.8.4.2 Eigenmode Collectivity κ

Eigenmode collectivity κ , a measure indicating the percentage of the bond vectors collectively affected by this given mode, can be calculated as

$$\kappa_m = \frac{1}{n} \exp\{-\Sigma_{k=1}^{n} | | m >_k |^2 \log | | m >_k |^2\} \tag{4.14}$$

where $|m >_k$ is the kth component of the normalized eigenvector $|m >$. The eigenmode collectivity (κ) plot versus the eigenvalues of the mode (λ) provides the summary of the motions accessible to a system. For an ideally rigid molecule, indicating no internal motion, thus there should be only five points with nonzero eigenvalues and collectivities of the associated eigenmodes should be very high. Thus, the separability gap, defined in Eq. 4.13, between internal and global motions ideally is infinity. At the same time, the gap in collectivity phase, proportional to difference in the percent of sites significantly affected by a given eigenmode between the five largest global modes and the n-5 internal modes, should go infinity as well. For the non-rigid molecule with correlated internal motions, the corresponding modes would appear at lower (κ, λ) plot. This plot can examine whether the separability assumption is suitable for the system, if not, the experimental data will have to be analyzed using something other than the Lipari-Szabo model-free formalism.

4.8.4.3 The Principal Order Parameter

The S_j^2 value for the jth vector is calculated as

$$S_j^2 = 1 - \sum_{m=6}^{n} \delta S_{jm}^2 \tag{4.15}$$

where δS^2_{jm} the principal order parameter components and the sum extends over the eigenvector $|m>$ belonging to the N-5 smallest eigenvalues, which correspond to internal motions. The δS^2_{jm} represents the contribution of a given mode with corresponding eigenvector $|m>$ to the decay of the correlation function for a given interaction j, can be expressed by

$$\delta S^2_{jm} = \lambda_m(|m><m|)_{jj} \tag{4.16}$$

where $\delta S^2_{j,m} \geq 0$, and $\Sigma_m \delta S^2_{j,m} = 1$.

4.8.4.4 Derived Overall Correlation Function

As the time information is lost when constructing the covariance matrix, here we demonstrate how to construct the time-dependent correlation function $C_j(t)$ for the jth vector in iRED analysis. First, timescale information within the trajectory can be reconstructed by calculating correlation function, $C_m(t)$ of each eigenmode m as following

$$C_m(t) = \Sigma^2_{l=-2} < a^*_{m.l}(t+\tau)_{m,l} a(\tau) > \tag{4.17}$$

Second, in Lipari-Szabo model, the $C^{over}_j(t)$ overall correlation function of the jth bond vector is contributed by the five largest eigenmodes

$$C^{over}_j(t) = \sum_{m=1}^{5} \delta S^2_{jm} C_m(t) / (\sum_{m=1}^{5} \delta S^2_{jm}), \tag{4.18}$$

if taking more modes into consideration, this overall correlation function will approach to the total correlation function, which is given by

$$C^{tot}_j(t) = \sum_{m=1}^{n} \delta S^2_{jm} C_m(t) / (\sum_{m=1}^{n} \delta S^2_{jm}) \tag{4.19}$$

where n is equal to the number of the spin vectors considered in the iRED matrix. For example, the overall and total correlation functions of Gly8 NH vector are plotted in Fig. 4.18 in the case of T3-785.

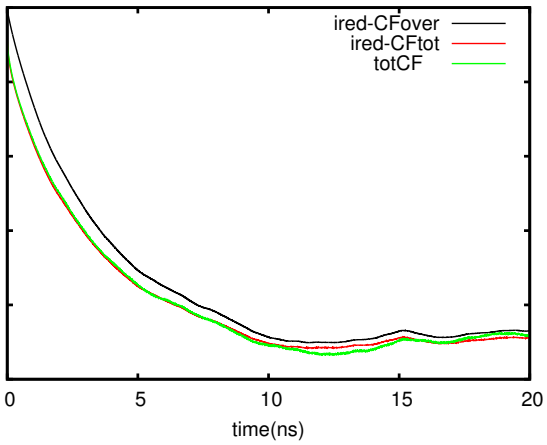


Figure 4.18: *The overall and total correlation functions extracted from iRED method. The overall correlation function $C^{over}_{NH=8}(t)$ extracted from the eigenmodes of the matrix M averaged over 500ns trajectory of peptide T3-785, is computed using Eq. 4.18(black line). The total correlation function (red lines), Eq. 4.19, of the same bond vector, is computed with all the eigenmodes and is expected to approach to the real total correlation function of NH residue 8 extracted directly from row trajectory depicted as green line.*

4.8.4.5 Special Density Functions

With the assumption of monoexponential decaying correlation function of each mode, the correlation time τ_m of each mode can be computed by integration of $C_m(t)$. In the triple helical peptide T3-785, the internal modes were found to decay in a good approximation mono-exponentially and a lifetime τ_m associated with motional mode can be established. Thus, the spectral density function $J(\omega)$ is analytically expressed as

$$J_j(\omega) = \int_{-\infty}^{\infty} C_j^{tot}(t)cos(\omega t)dt = (\sum_{m=1}^{n} \delta S^2_{j,m} \frac{2\tau_m}{1+\omega^2\tau_m^2})/(\sum_{m=1}^{n} \delta S^2_{j,m}) \tag{4.20}$$

4.8.4.6 Convergence Issues

The use of $C_m(t)$ can help in determining the convergence of the trajectory, indicating that a well-behaved correlation function will asymptotically decay stably to zero and can be fitted with monoexponential function. A well-converged simulation, statistical error from non-linearly fitting the correlation function is negligible. In the case of peptide T3-785, the correlation functions corresponding to the overall modes displayed a less ideal decay shape, indicating that these modes do not follow in good approximation a monoexponential

decay as in the case for most of other internal modes. Therefore, an reliable estimates τ_m determining the spectral density function cannot be extracted just from trajectory.

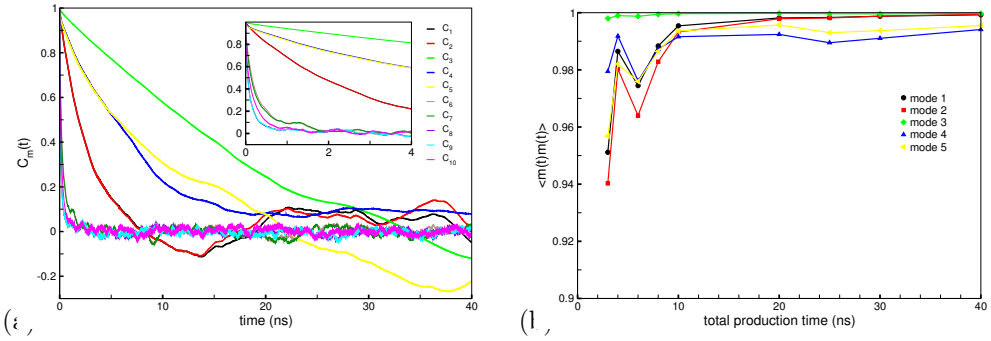


Figure 4.19: *(a) Correlation functions corresponding to the largest ten amplitude modes of the matrix M averaged through 180ns trajectory from iRED analysis. Among the modes representing the overall modes, only $C_4(t)$ decays to a relative stable asymptote. C_3 and C_5 failed to reach an asymptote. Insect plot shows the mode correlation functions in 5 ns delay time. (b) Convergence of iRED eigenvectors. The dot products of the iRED eigenvectors corresponding to the five largest modes at different simulation time to task the convergence of our MD simulation.*

In Fig. 4.19, the correlation functions of the five overall reorientational modes $m = 1, .., 5$ seem to follow the monoexponential function well during the first 5 ns simulation, to obtain the better mode correlation functions following in a good exponential decay, only the 5 ns correlation function is used to fit.

These poorly converged correlation functions indicates a problem with trajectory or its length. To task the convergence of the MD trajectory, we also calculated the dot products of the normalized modes obtained at different trajectory length (Fig. 4.19 (b)). It shows that the covariance matrix is well converged at 10ns timescale for the peptide T3-785, suggesting that the average iRED covariance matrix through 10ns long trajectory is sufficient to describe the motions of the peptide.

4.8.5 Woessner's Equations

Anisotropic diffusion for the fully asymmetric system ($D_x \neq D_y \neq D_z$), the correlation times can be expressed in terms of diffusion constants as following:

$$\tau_1^{-1} = D_x + D_y + 4D_z$$

$$\tau_2^{-1} = 6D + 6\sqrt{D^2 - L^2}$$

$$\tau_3^{-1} = 6D - 6\sqrt{D^2 - L^2}$$

$$\tau_4^{-1} = 4D_x + D_y + D_z$$

$$\tau_5^{-1} = D_x + 4D_y + D_z \tag{4.21}$$

where $D = (D_x + D_y + D_z)/3$ and $L^2 = (D_xD_y + D_yD_z + DD_{xz})/3$. In the case of axial symmetry top ($D_x = D_y = D_\perp$), then the above equations will reduce to

$$\tau_1^{-1} = \tau_2^{-1} = 2D_\perp + 4D_{||}$$

$$\tau_3^{-1} = 6D_\perp$$

$$\tau_4^{-1} = \tau_5^{-1} = 5D_\perp + D_{||} \tag{4.22}$$

4.8.6 Lipari-Szabo Approach - Spectral Density Function

This section, we discuss the effects on calculation of the spectral density functions, which is the Fourier transform of the total correlation function of the vector described in the extended model-free (EMF) approach:

$$J(\omega) = S^2 J_{over}(\omega) + (S_f^2 - S^2) J_{int}(\omega) \tag{4.23}$$

where $J_{int}(\omega)$ and $J_{over}(\omega)$ are the spectral density functions attributed from internal motions and overall tumbling, respectively. In Eq. 4.23, $J_{int}(\omega)$ is expressed as $2\tau'/(1+(\omega\tau')^2)$ where $1/\tau^{'} = 1/\tau_M + 1/\tau_I$, τ_M and τ_I denoting the relation times from overall tumbling of the molecule and internal motion, respectively.

4.8.6.1 Effect of the Iso- and Aniso calculation of Overall Tumbling Time on $J_{over}(\omega)$

For the simple isotropic motion of a spherical molecule, the overall tumbling motion gives the Lipari-Szabo spectral density function $J(\omega)$

$$J_{rot}^{iso}(\omega) = \frac{2\tau_M}{1 + (\omega\tau_M)^2} \tag{4.24}$$

where τ_M is the relaxation time for uniform (rigid-body) motion.

For the rigid cylindrical rod such as the collagen-like peptide of studied here, the corresponding spectral density function attributed from overall tumbling "should be" expressed as

$$J_{rot}^{aniso}(\omega) = \sum_{k=0}^{2} \frac{2A_k\tau_k}{1+(\omega\tau_k)^2} \tag{4.25}$$

where ω is the angular resonance frequency, and terms $A_0 = \frac{1}{4}(3cos^2\theta - 1)^2$, $A_1 = 3cos^2\theta sin^2\theta$, and $A_2 = \frac{3}{4}sin^4\theta$ (θ is the angle between the time-averaged N-H bond orientation and the long axis of the diffusion tensor); time constants $\tau_0 = 1/(6D_\perp)$, $\tau_1 = 1/(5D_\perp + D_{||})$, and $\tau_2 = 1/(2D_\perp + 4D_{||})$. The effective relaxation time for that vector can be expressed by

$$\tau_M = A_0\tau_o + A_1\tau_1 + A_2\tau_2 \tag{4.26}$$

In the case of a molecule with low anisotropy r (i.e. $r \sim 1$), three relaxation times are not far from one another, indicating the formula of Eq. 4.24 is still suitable for computing spectra density function at frequency ω. However, it is not the case for the molecule with high anisotropy, e.g. the collagen model peptide with $10 < r < 15$. The explanation is following.

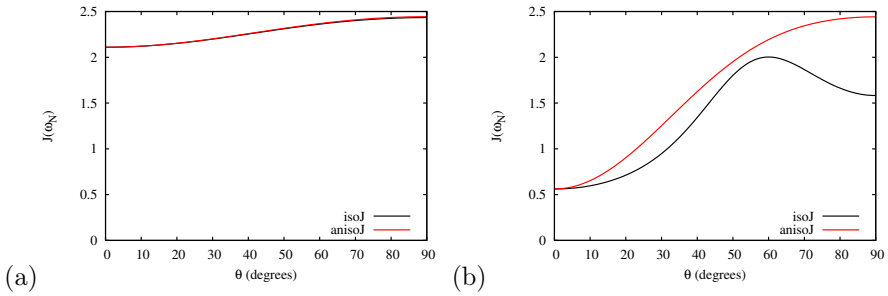


Figure 4.20: *Spectral density function using Eq. 4.24 (black line) and Eq. 4.25 (red lines) against the orientation of the NH vectors relative to the molecular frame with (a) low anisotropy $r = 1.5$ and (b) high anisotropy $r = 13$. Here the order parameter $S^2 = 1$ is assumed.*

Fig. 4.20 (a) depicts that when the anisotropy is low (i.e. $r = 1.5$), the discrepancy between Eq. 4.24 and Eq. 4.25 in calculation of $J(\omega)$ is negligible. While the anisotropy is high e.g. 13, the discrepancy becomes significant, particularly for the vectors perpendicular to the symmetry axis of the diffusion frame as shown in Fig. 4.20 (b). Therefore, in the case of the triple helical collagen peptide with most NH vectors pointing away from symmetric axis of diffusion frame, knowledge of the orientation of the NH vector relative to the symmetric axis of diffusion frame is required to accurately estimate the relaxation data.

4.8.6.2 Effects of Internal Parameters

Due to the mathematics form of the spectral density function, it is clear that it reaches its maximal value as the correlation time approaches the reciprocal of the spectrometer frequency. For relaxation of the nitrogen by the neighboring proton R1 in Eq. 4.8, as ω_H and $\omega_H \pm \omega_N$ are large frequencies compared to ω_N, consequently $J(\omega_H \pm \omega_N)$ and $J(\omega_H)$ are altogether negligible in compare to $J(\omega_N)$, and make no significant contribution to R1. Hence, an increase in $J(\omega)$ will increase R1 rates. As the total R1 includes the contribution of the pure dipolar and pure chemical shift anisotropy terms, to make less complicated, we will focus on the R1 contributed from the dipolar relaxation, $R1^D$ the first term in Eq. 4.8.

The spectral density function $J(\omega)$ is a complicated mixture of order parameter S^2 and the correlation times. Here we are interested in looking at how the internal parameters S^2 and τ_I affects the $J(\omega_N)$ at $f = 500$ MHz so as to $R1^D$(Fig. 4.21),

1. Fig. 4.21 (a) demonstrates that when the internal correlation time, τ_I , for the local angular motion is sufficiently small (i.e. $\omega\tau \ll 1$), a reduction in $J(\omega_N)$ is strongly caused by the decreasing S^2. However, when the τ_I approaches its reciprocal of the spectrometer frequency, the internal motions can pronouncedly increase the spectral density functions as the order parameter is reduced. When the τ_I becomes comparable to the correlation times of the uniform (rigid-body) rotations, $J(\omega_N)$ reaches its maxima value and becomes a lot less sensitive to the internal motions. Therefore, the value of the internal correlation time is extremely important in determining whether the spectral densities increase or decrease as the order parameter changes.

2. For the higher spectrometer frequency in , e.g. $\omega_H - \omega_N \sim 2.82 \times 10^9 (1/s)$ in $J(\omega_N - \omega_H)$ in Fig. 4.21 (b), when τ_I is approaching its reciprocal of the spectrometer frequency, $J(\omega)$ decreases as the S^2 is increasing. Other than that, $J(\omega_N - \omega_H)$ becomes insensitive to the S2.

3. The part of R1 in Eq. 4.8 contributed from dipole-dipole interactions, $R1^D$ which is proportional $3J(\omega_N) + 7J(\omega_H)$, is sensitive to internal correlation time τ_I or order parameter S^2 or both under some circumstances. The calculated values of $3J(\omega_N) + 7J(\omega_H)$ are presented in Fig. 4.21(c). When the relaxation time for the local angular motion is sufficiently small (i.e. $\omega\tau \ll 1$), $R1^D$ becomes independent of τ, but declines with decreasing S^2. When τ becomes comparable to the correlation times for the uniform (rigid-body) rotations, $R1^D$ becomes insensitive to the S^2. Apparently, $J(\omega)$ for low frequency dominates the variation of R1 more than the one with higher frequency.

4. Slow down the molecule tumbling yields the lower value of $J(\omega)$ in average as shown in Fig. 4.21 (d). That point τ_x where all curves with the same overall tumbling

time cross, only exist when $\tau_{over} > 1/\omega_N$. The internal correlation timeτ_I below and beyond τ_x shows how varied S^2 affects $J(\omega_N)$ differently. For example, the increased S^2 reduces the value of $J(\omega_N)$ when $\tau_I > \tau_x$. The value of τ_x reduces to the region of time scale less than 1 ns when the molecule tumbles very slowly, e.g. 12 ns. We need to be careful about the interpretation of R1 rates if the real internal correlation time is close to the value of τ_x, Since higher value in R1 does not commit to be more rigid in backbone dynamics based on the circumstances of (2). Therefore, knowledge of internal correlation times for each inter-nuclear bond vector of studied is crucial when the molecules tumble slowly.

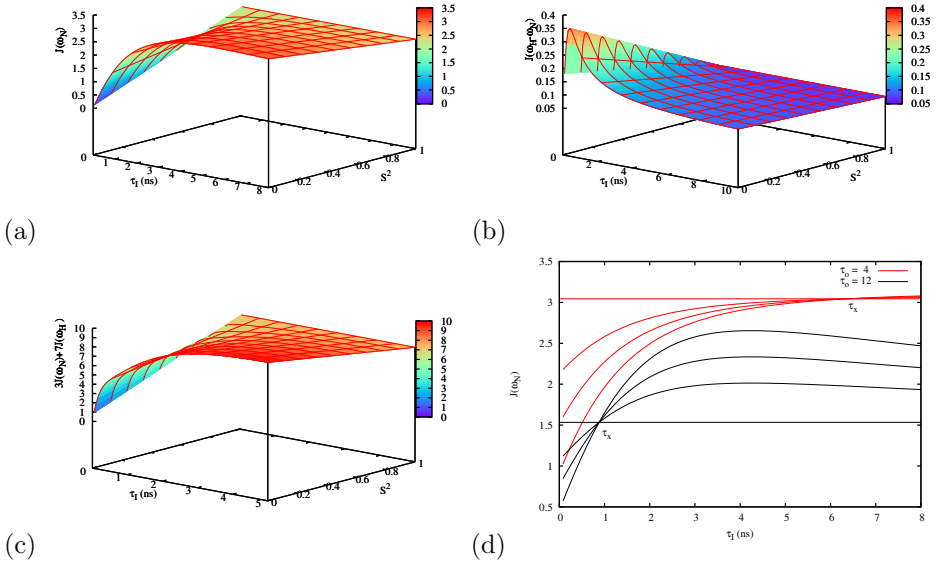


Figure 4.21: *Model-free spectral density function $J(\omega)$ for simple isotropic motion in Eq. 4.23 over a range of internal correlation time τ_I versus a spectrometer frequency, $f = 500MHz$, which gives (a)$\omega_N \sim 0.319 \times 10^9 1/s$, (b) $\omega_H - \omega_N \sim 2.82 \times 10^9$, (c) $\omega_N \sim 0.319 \times 10^9$ with different overall correlation time τ_o. (c) The value of spectral density function from dipole dipole interaction term is approximately proportional to R_1^D in Eq. 4.8. The results are shown in (a), (b), and (c) using overall correlation time for the overall molecules used for rigid cylinder model are $\tau_{over} \sim 6.9$ ns. (d) the distribution of $J(\omega_N)$ at different overall tumbling times.*

Bibliography

[1] Vuister, Geerten W.; Wang, Andy C.; Bax, Ad. Measurement of three-bond nitrogen-carbon j couplings in proteins uniformly enriched in nitrogen-15 and carbon-13. *Journal of the American Chemical Society*, **1993**, *115*(12), 5334–5335.

[2] Xiao, J.; Baum, J. Structural insights from (15)n relaxation data for an anisotropic collagen peptide. *J Am Chem Soc*, **2009**, *131*(51), 18194–5.

[3] CopiÃ©, V.; Tomita, Y.; Akiyama, S. K.; Aota, S.; Yamada, K. M.; Venable, R. M.; Pastor, R. W.; Krueger, S.; Torchia, D. A. Solution structure and dynamics of linked cell attachment modules of mouse fibronectin containing the rgd and synergy regions: comparison with the human fibronectin crystal structure. *J Mol Biol*, **1998**, *277*(3), 663–82.

[4] GarcÃa de la Torre, J.; Huertas, M. L.; Carrasco, B. Hydronmr: prediction of nmr relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson*, **2000**, *147*(1), 138–46.

[5] Jenkins, C. L.; Raines, R. T. Insights on the conformational stability of collagen. *Nat Prod Rep*, **2002**, *19*(1), 49–59.

[6] Bhattacharjee, A.; Bansal, M. Collagen structure: the madras triple helix and the current scenario. *IUBMB Life*, **2005**, *57*(3), 161–72.

[7] Brodsky, B.; Persikov, A. V. Molecular structure of the collagen triple helix. *Adv Protein Chem*, **2005**, *70*, 301–39.

[8] Myllyharju, Johanna; Kivirikko, Kari I. Collagens and collagen-related diseases. *Annals of Medicine*, **2001**, *33*(1), 7–21.

[9] Tromp, G; Kuivaniemi, H; Stolle, C; Pope, F M; Prockop, D J. Single base mutation in the type iii procollagen gene that converts the codon for glycine 883 to aspartate in a mild variant of ehlers-danlos syndrome iv. *Journal of Biological Chemistry*, **1989**, *264*(32), 19313–19317.

[10] Kuivaniemi, Helena; Tromp, Gerard; Prockop, Darwin J. Mutations in fibrillar collagens (types i, ii, iii, and xi), fibril-associated collagen (type ix), and network-forming collagen (type x) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Human Mutation*, **1997**, *9*(4), 300–315.

[11] Lemmink, Henny H.; SchrÃ¶der, Cornelis H.; Monnens, Leo A. H.; Smeets, Huber J. M. The clinical spectrum of type iv collagen mutations. *Human Mutation*, **1997**, *9*(6), 477–499.

[12] Marini, Joan C.; Forlino, Antonella; Cabral, Wayne A.; Barnes, Aileen M.; San Antonio, James D.; Milgrom, Sarah; Hyland, James C.; Korkko, Jarmo; Prockop, Darwin J.; De Paepe, Anne; Coucke, Paul; Symoens, Sofie; Glorieux, Francis H.; Roughley, Peter J.; Lund, Alan M.; Kuurila-Svahn, Kaija; Hartikka, Heini; Cohn, Daniel H.; Krakow, Deborah; Mottes, Monica; Schwarze, Ulrike; Chen, Diana; Yang, Kathleen; Kuslich, Christine; Troendle, James; Dalgleish, Raymond; Byers, Peter H. 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. *Human Mutation*, **2007**, *28*(3), 209–221.

[13] Bodian, Dale L.; Madhan, Balaraman; Brodsky, Barbara; Klein, Teri E. Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry*, **2008**, *47*(19).

[14] Srnith, Lynne T.; Schwarze, Ulrike; Goldstein, Jayne; Byers, Peter H. Mutations in the col3a1 gene result in the ehlers-danlos syndrome type iv and alterations in the size and distribution of the major collagen fibrils of the dermis. *J Investig Dermatol*, **1997**, *108*(3), 241–247.

[15] Myllyharju, J.; Kivirikko, K. I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet*, **2004**, *20*(1), 33–43.

[16] Fields, G. B. A model for interstitial collagen catabolism by mammalian collagenases. *Journal of Theoretical Biology*, **1991**, *153*(4), 585–602. Times Cited: 104 107.

[17] Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J.; Liddington, R. C. Structural basis of collagen recognition by integrin alpha2beta1. *Cell*, **2000**, *101*(1), 47–56.

[18] Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J. Structure of the integrin alpha2beta1-binding collagen peptide. *J Mol Biol*, **2004**, *335*(4), 1019–28.

[19] Miller, E. J.; Harris, E. D.; Chung, E.; Finch, J. E.; McCroskery, P. A.; Butler, W. T. Cleavage of type ii and iii collagens with mammalian collagenase: site of cleavage and primary structure at the nh2-terminal portion of the smaller fragment released from both collagens. *Biochemistry*, **1976**, *15*(4), 787–92.

[20] Yu, Z.; Visse, R.; Inouye, M.; Nagase, H.; Brodsky, B. Defining requirements for collagenase cleavage in collagen type iii using a bacterial collagen system. *J Biol Chem*, **2012**, *287*(27), 22988–97.

[21] Berg, Richard A.; Prockop, Darwin J. The thermal transition of a non-hydroxylated form of collagen. evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochemical and Biophysical Research Communications*, **1973**, *52*(1), 115–120.

[22] Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. Crystal-structure and molecular-structure of a

collagen-like peptide at 1.9-angstrom resolution (gly to ala peptide). *Science*, **1994**, *266*(5182), 75–81.

[23] Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal structure of the collagen triple helix model (pro-pro-gly)(10) (3). *Protein Science*, **2002**, *11*(2), 262–270.

[24] Okuyama, K.; Hongo, C.; Fukushima, R.; Wu, G.; Narita, H.; Noguchi, K.; Tanaka, Y.; Nishino, N. Crystal structures of collagen model peptides with pro-hyp-gly repeating sequence at 1.26 a resolution: implications for proline ring puckering. *Biopolymers*, **2004**, *76*(5), 367–77.

[25] Nagarajan, V.; Kamitori, S.; Okuyama, K. Crystal structure analysis of collagen model peptide (pro-pro-gly)(10). *Journal of Biochemistry*, **1998**, *124*(6), 1117–1123.

[26] Nagarajan, V.; Kamitori, S.; Okuyama, K. Structure analysis of a collagen-model peptide with a (pro-hyp-gly) sequence repeat. *Journal of Biochemistry*, **1999**, *125*(2), 310–318.

[27] Kramer, R. Z.; Vitagliano, L.; Bella, J.; Berisio, R.; Mazzarella, L.; Brodsky, B.; Zagari, A.; Berman, H. M. X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (pro-pro-gly). *Journal of Molecular Biology*, **1998**, *280*(4), 623–638.

[28] Kramer, R. Z.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M. Sequence dependent conformational variations of collagen triple-helical structure. *Nat Struct Biol*, **1999**, *6*(5), 454–7.

[29] Kawahara, Kazuki; Nishi, Yoshinori; Nakamura, Shota; Uchiyama, Susumu; Nishiuchi, Yuji; Nakazawa, Takashi; Ohkubo, Tadayasu; Kobayashi, Yuji. Effect of hydration on the stability of the collagen-like triple-helical structure of [4(r)-hydroxyprolyl-4(r)-hydroxyprolylglycine]10. *Biochemistry*, **2005**, *44*(48), 15812–15822.

[30] Goodman, M.; Bhumralkar; Jefferson, E. A.; Kwak, J.; Locardi, E. Collagen mimetics. *Biopolymers*, **1998**, *47*(2), 127–42.

[31] Baum, J.; Brodsky, B. Folding of peptide models of collagen and misfolding in disease. *Curr Opin Struct Biol*, **1999**, *9*(1), 122–8.

[32] Bella, Jordi. A new method for describing the helical conformation of collagen: Dependence of the triple helical twist on amino acid sequence. *Journal of Structural Biology*, **2010**, *170*(2), 377–391.

[33] Brodsky, Barbara; Baum, Jean. Structural biology: Modelling collagen diseases. *Nature*, **2008**, *453*(7198), 998–999. 10.1038/453998a.

[34] Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M. Crystal and molecular structure of a collagen-like polypeptide (pro-pro-gly)10. *Journal of molecular biology*, **1981**, *152*(2).

[35] Okuyama, K. Revisiting the molecular structure of collagen. *Connective Tissue Research*, **2008**, *49*(5), 299–310. Cited By (since 1996):31 Export Date: 25 March 2014 Source: Scopus.

[36] Okuyama, K.; Miyama, K.; Mizuno, K.; Bachinger, H. P. Crystal structure of (gly-pro-hyp)9: Implications for the collagen molecular model. *Biopolymers*, **2012**, *97*(8), 607–616. ISI Document Delivery No.: 944GV Times Cited: 0 Cited Reference Count: 38 Okuyama, Kenji Miyama, Keita Mizuno, Kazunori Baechinger, Hans Peter Japan Society for the Promotion of Science; Shriners Hospital Contract grant sponsors: Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science, Shriners Hospital WILEY-BLACKWELL MALDEN.

[37] Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. Amino acid propensities for the collagen triple-helix. *Biochemistry*, **2000**, *39*(48), 14960–7.

[38] Persikov, A. V.; Ramshaw, J. A.; Brodsky, B. Prediction of collagen stability from amino acid sequence. *J Biol Chem*, **2005**, *280*(19), 19343–9.

[39] Ramshaw, J. A.; Shah, N. K.; Brodsky, B. Gly-x-y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J Struct Biol*, **1998**, *122*(1-2), 86–91.

[40] Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A.; Brodsky, B. Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc Natl Acad Sci U S A*, **2000**, *97*(8), 4273–8.

[41] Radmer, R. J.; Klein, T. E. Severity of osteogenesis imperfecta and structure of a collagen-like peptide modeling a lethal mutation site. *Biochemistry*, **2004**, *43*(18), 5314–23.

[42] Bhate, Manjiri; Wang, Xin; Baum, Jean; Brodsky, Barbara. Folding and conformational consequences of glycine to alanine replacements at different positions in a collagen model peptide. *Biochemistry*, **2002**, *41*(20), 6539–6547.

[43] Mooney, S. D.; Klein, T. E. Structural models of osteogenesis imperfecta-associated variants in the col1a1 gene. *Mol Cell Proteomics*, **2002**, *1*(11), 868–75.

[44] Bauer, R.; Wilson, J. J.; Philominathan, S. T.; Davis, D.; Matsushita, O.; Sakon, J. Structural comparison of colh and colg collagen-binding domains from clostridium histolyticum. *J Bacteriol*, **2013**, *195*(2), 318–27.

[45] Philominathan, S. T.; Koide, T.; Matsushita, O.; Sakon, J. Bacterial collagen-binding domain targets undertwisted regions of collagen. *Protein Sci*, **2012**, *21*(10), 1554–65.

[46] Cheng, H. M.; Rashid, S.; Yu, Z. X.; Yoshizumi, A.; Hwang, E. L.; Brodsky, B. Location of glycine

mutations within a bacterial collagen protein affects degree of disruption of triple-helix folding and conformation. *Journal of Biological Chemistry*, **2011**, *286*(3), 2041–2046.

[47] Hyde, T. J.; Bryan, M. A.; Brodsky, B.; Baum, J. Sequence dependence of renucleation after a gly mutation in model collagen peptides. *J Biol Chem*, **2006**, *281*(48), 36937–43.

[48] Liu, X.; Kim, S.; Dai, Q. H.; Brodsky, B.; Baum, J. Nuclear magnetic resonance shows asymmetric loss of triple helix in peptides modeling a collagen mutation in brittle bone disease. *Biochemistry*, **1998**, *37*(44), 15528–33.

[49] Long, C. G.; Li, M. H.; Baum, J.; Brodsky, B. Nuclear magnetic resonance and circular dichroism studies of a triple-helical peptide with a glycine substitution. *J Mol Biol*, **1992**, *225*(1), 1–4.

[50] Philominathan, S. T.; Koide, T.; Hamada, K.; Yasui, H.; Seifert, S.; Matsushita, O.; Sakon, J. Unidirectional binding of clostridial collagenase to triple helical substrates. *J Biol Chem*, **2009**, *284*(16), 10868–76.

[51] Li, Yingjie; Brodsky, Barbara; Baum, Jean. Nmr conformational and dynamic consequences of a gly to ser substitution in an osteogenesis imperfecta collagen model peptide. *Journal of Biological Chemistry*, **2009**, *284*(31).

[52] Lamande, S R; Dahl, H H; Cole, W G; Bateman, J F. Characterization of point mutations in the collagen col1a1 and col1a2 genes causing lethal perinatal osteogenesis imperfecta. *Journal of Biological Chemistry*, **1989**, *264*(27), 15809–15812.

[53] Okuyama, K.; Wu, G.; Jiravanichanun, N.; Hongo, C.; Noguchi, K. Helical twists of collagen model peptides. *Biopolymers*, **2006**, *84*(4), 421–32.

[54] Boudko, Sergei P.; Engel, Juergen; Okuyama, Kenji; Mizuno, Kazunori; Baechinger, Hans Peter; Schumacher, Maria A. Crystal structure of human type iii collagen gly(991)-gly(1032) cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *Journal of Biological Chemistry*, **2008**, *283*(47).

[55] Kramer, Rachel Z.; Bella, Jordi; Brodsky, Barbara; Berman, Helen M. The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *Journal of Molecular Biology*, **2001**, *311*(1), 131–147.

[56] Lazarev, I.A; Esipova, N. G. [spectral study of structural ordering of collagen-type proteins]. *Dokl Akad Nauk SSSR*, **1973**, *209*(2), 478–81.

[57] Lazarev, Y. A.; Lazareva, A. V.; Komarov, V. M. Backbone dynamics of triple-helical collagen-like structure. *J Biol Phys*, **1999**, *24*(2-4), 217–22.

[58] Jelinski, L. W.; Torchia, D. A. Investigation of labeled amino acid side-chain motion in collagen using 13c nuclear magnetic resonance. *J Mol Biol*, **1980**, *138*(2), 255–72.

[59] Jelinski, L. W.; Sullivan, C. E.; Torchia, D. A. 2h nmr study of molecular motion in collagen fibrils. *Nature*, **1980**, *284*(5756), 531–4.

[60] Torchia, D. A.; Lyerla, J. R.; Quattrone, A. J. Molecular dynamics and structure of the random coil and helical states of the collagen peptide, alpha 1-cb2, as determined by 13c magnetic resonance. *Biochemistry*, **1975**, *14*(5), 887–900.

[61] Torchia, D. A.; VanderHart, D. L. 13c magnetic resonance evidence for anisotropic molecular motion in collagen fibrils. *J Mol Biol*, **1976**, *104*(1), 315–21.

[62] Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. Characterization of leucine side-chain reorientation in collagen-fibrils by solid-state 2h nmr. *Proc Natl Acad Sci U S A*, **1982**, *79*(2), 386–9.

[63] Torchia, D. A.; Batchelder, L. S.; Fleming, W. W.; Jelinski, L. W.; Sarkar, S. K.; Sullivan, C. E. Mobility and function in elastin and collagen. *Ciba Found Symp*, **1983**, *93*, 98–115.

[64] Fan, P.; Li, M. H.; Brodsky, B.; Baum, J. Backbone dynamics of (pro-hyp-gly)10 and a designed collagen-like triple-helical peptide by 15n nmr relaxation and hydrogen-exchange measurements. *Biochemistry*, **1993**, *32*(48), 13299–309.

[65] Baum, Jean; Brodsky, Barbara. Real-time nmr investigations of triple-helix folding and collagen folding diseases. *Folding and Design*, **1997**, *2*(4), R53–R60.

[66] Li, Ming Hua; Fan, Pei; Brodsky, Barbara; Baum, Jean. Two-dimensional nmr assignments and conformation of (pro-hyp-gly)10 and a designed collagen triple-helical peptide. *Biochemistry*, **1993**, *32*(29), 7377–7387.

[67] Prompers, J. J.; BrÃŒschweiler, R. General framework for studying the dynamics of folded and nonfolded proteins by nmr relaxation spectroscopy and md simulation. *J Am Chem Soc*, **2002**, *124*(16), 4522–34.

[68] Markwick, P. R.; Bouvignies, G.; Blackledge, M. Exploring multiple timescale motions in protein gb3 using accelerated molecular dynamics and nmr spectroscopy. *J Am Chem Soc*, **2007**, *129*(15), 4724–30.

[69] Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen,

M.; Xu, H.; Trbovic, N.; Friesner, R. A.; Iii, A. G.; Shaw, D. E. Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J Phys Chem B*, **2008**, *112*(19), 6155–8.

[70] Markwick, P. R.; McCammon, J. A. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys Chem Chem Phys*, **2011**, *13*(45), 20053–65.

[71] Klein, Teri E.; Huang, Conrad C. Computational investigations of structural changes resulting from point mutations in a collagen-like peptide. *Biopolymers*, **1999**, *49*(2), 167–183.

[72] Mooney, Sean D.; Huang, Conrad C.; Kollman, Peter A.; Klein, Teri E. Computed free energy differences between point mutations in a collagen-like peptide. *Biopolymers*, **2001**, *58*(3), 347–353.

[73] Ryabov, Y. E.; Geraghty, C.; Varshney, A.; Fushman, D. An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation. *J Am Chem Soc*, **2006**, *128*(48), 15432–44.

[74] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *Journal of the American Chemical Society*, **1996**, *118*(9).

[75] Lee, L. K.; Rance, M.; Chazin, W. J.; Palmer, A. G. Rotational diffusion anisotropy of proteins from simultaneous analysis of 15n and 13c alpha nuclear spin relaxation. *J Biomol NMR*, **1997**, *9*(3), 287–98.

[76] Lipari, Giovanni; Szabo, Attila. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *Journal of the American Chemical Society*, **1982**, *104*(17), 4546–4559.

[77] Schneider, D. M.; Dellwo, M. J.; Wand, A. J. Fast internal main-chain dynamics of human ubiquitin. *Biochemistry*, **1992**, *31*(14), 3645–52.

[78] Johnson, E.; Showalter, S. A.; BrÃŒschweiler, R. A multifaceted approach to the interpretation of nmr order parameters: a case study of a dynamic alpha-helix. *J Phys Chem B*, **2008**, *112*(19), 6203–10.

[79] Wong, V.; Case, D. A. Evaluating rotational diffusion from protein md simulations. *J Phys Chem B*, **2008**, *112*(19), 6013–24.

[80] La Penna, G.; Mormino, M.; Pioli, F.; Perico, A.; Fioravanti, R.; Gruschus, J. M.; Ferretti, J. A. Smoluchowski dynamics of the vnd/nk-2 homeodomain from drosophila melanogaster: first-order mode-coupling approximation. *Biopolymers*, **1999**, *49*(3), 235–54.

[81] Ramachandran, G. N.; Kartha, G. Structure of collagen. *Nature*, **1955**, *176*(4482).

[82] Dombrowski, Kenneth E.; Vogel, Bruce E.; Prockop, Darwin J. Mutations that alter the primary structure of type i procollagen have long-range effects on its cleavage by procollagen n-proteinase. *Biochemistry*, **1989**, *28*(17), 7107–7112.

[83] Nerenberg, P. S.; Salsas-Escat, R.; Stultz, C. M. Do collagenases unwind triple-helical collagen before peptide bond hydrolysis? reinterpreting experimental observations with mathematical models. *Proteins*, **2008**, *70*(4), 1154–61.

[84] Ramstein, J; Lavery, R. Energetic coupling between dna bending and base pair opening. *Proceedings of the National Academy of Sciences*, **1988**, *85*(19), 7231–7235.

[85] Lu, XiangJun; Olson, Wilma K. 3dna: a software package for the analysis, rebuilding and visualization of three dimensional nucleic acid structures. *Nucleic Acids Research*, **2003**, *31*(17), 5108–5121.

[86] Lu, X. J.; Olson, W. K. 3dna: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc*, **2008**, *3*(7), 1213–27.

[87] Zheng, G.; Lu, X. J.; Olson, W. K. Web 3dna–a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res*, **2009**, *37*(Web Server issue), W240–6.

[88] Philominathan, Sagaya Theresa Leena; Koide, Takaki; Matsushita, Osamu; Sakon, Joshua. Bacterial collagen-binding domain targets undertwisted regions of collagen. *Protein Science*, **2012**, *21*(10), 1554–1565.

[89] Liu, Xiaoyan; Siegel, Donald L.; Fan, Pei; Brodsky, Barbara; Baum, Jean. Direct nmr measurement of the folding kinetics of a trimeric peptide. *Biochemistry*, **1996**, *35*(14), 4306–4313.

[90] He, X.; Wang, B.; Merz, K.M. Protein NMR Chemical Shift Calculations Based on the Automated Fragmentation QM/MM Approach. *J. Phys. Chem. B*, **2009**, *113*, 10380–10388.

[91] Zhu, T.; Zhang, J.Z.H.; He, X. Automated Fragmentation QM/MM Calculation of Amide Proton Chemical Shifts in Proteins with Explicit Solvent Model. *J. Chem. Theory Comput.*, **2013**, *9*, 2104–2114.

[92] Berglund, B.; Carson, Douglas G.; Vaughan; W., Robert. The chemical shift tensor for a strongly

hydrogen bonded proton: The carboxylic proton in khch2(coo)2. *JCP*, **1980**, *72*(824).

[93] Engel, J.; Prockop, D. J. Does bound water contribute to the stability of collagen? *Matrix Biol*, **1998**, *17*(8-9), 679–80.

[94] Li, Yingjie; Brodsky, Barbara; Baum, Jean. Nmr shows hydrophobic interactions replace glycine packing in the triple helix at a natural break in the (gly-x-y)n repeat. *Journal of Biological Chemistry*, **2007**, *282*(31), 22699–22706. 10.1074/jbc.M702910200.

[95] Hornak, V. et al. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, **2006**, *65*, 712–725.

[96] Case, D.A.; Darden, T.A.; T.E. Cheatham, III; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R.C.; Zhang, W.; Merz, K.M.; Wang, B.; S.Hayik; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K.F.; Paesani, F.; J.Vanicek; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; C. Tan, J. Mongan; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman; P.A. Amber 10. **2008**.

[97] Bodian, D. L.; Radmer, R. J.; Holbert, S.; Klein, T. E. Molecular dynamics simulations of the full triple helical region of collagen type i provide an atomic scale view of the protein's regional heterogeneity. *Pac Symp Biocomput*, **2011**, pp 193–204.

[98] Karplus, Martin. Contact electron spin coupling of nuclear magnetic moments. *The Journal of Chemical Physics*, **1959**, *30*(1), 11–15.

[99] el Hassan, M. A.; Calladine, C. R. The assessment of the geometry of dinucleotide steps in double-helical dna; a new local calculation scheme. *J Mol Biol*, **1995**, *251*(5), 648–64.

[100] Lu, XiangJun; Olson, Wilma K. 3dna: a versatile, integrated software system for the analysis, rebuilding and visualization of three dimensional nucleic acid structures. *Nat. Protocols*, **2008**, *3*(7), 1213–1227. 10.1038/nprot.2008.104.

[101] Tang, S.; Case, D.A. Calculation of chemical shift anisotropy in proteins. *J. Biomol. NMR*, **2011**, *51*, 303–312.

[102] Case, D.A. Chemical shifts in biomolecules. *Curr. Opin. Struct. Biol.*, **2013**, *23*, 172–176.

[103] Wilson, P.; Amos, R.; Handy, N. Toward coupled-cluster accuracy in the prediction of nuclear shielding constants: a simple and efficient DFT approach. *Chem. Phys. Lett.*, **1999**, *312*, 475–484.

[104] Godbout, N.; Salahub, D.R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *Can. J. Chem.*, **1992**, *70*, 560–571.

[105] Koster, A.M.; Geudtner, G.; Calaminici, P.; Casida, M.E.; Dominguez, V.D.; Flores-Moreno, R.; Gamboa, G.U.; Goursot, A.; Heine, T.; Ipatov, A.; Janetzko, F.; del Campo, J.M.; Reveles, J.U.; Vela, A.; Zuniga-Gutierrez, B.; Salahub, D.R. *deMon2k, Version 3*. The deMon developers, 2011.

[106] Gautieri, A.; Vesentini, S.; Redaelli, A.; Buehler, M. J. Single molecule effects of osteogenesis imperfecta mutations in tropocollagen protein domains. *Protein Sci*, **2009**, *18*(1), 161–8.

[107] Clore, G. M.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M. Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15n-1h nmr spectroscopy. *Biochemistry*, **1990**, *29*(32), 7387–401.

[108] Song, Lu; Schurr, J. Michael. Dynamic bending rigidity of dna. *Biopolymers*, **1990**, *30*(3-4), 229–237.

[109] Langowski, J.; Fujimoto, B. S.; Wemmer, D. E.; Benight, A. S.; Drobny, G.; Shibata, J. H.; Schurr, J. M. Deformational dynamics and nmr relaxation of supercoiled dnas. *Biopolymers*, **1985**, *24*(6), 1023–56.

[110] Schurr, J. M.; Fujimoto, B. S. The amplitude of local angular motions of intercalated dyes and bases in dna. *Biopolymers*, **1988**, *27*(10), 1543–69.

[111] Schurr, J. M.; Fujimoto, B. S.; Nuutero, S. Theory of relaxation of quadrupolar nuclei in deformable molecules in isotropic solutions. application to dna. *Journal of Magnetic Resonance, Series A*, **1994**, *106*(1), 1–22.

[112] Okonogi, T. M.; Reese, A. W.; Alley, S. C.; Hopkins, P. B.; Robinson, B. H. Flexibility of duplex dna on the submicrosecond timescale. *Biophysical Journal*, **1999**, *77*(6), 3256–3276.

[113] Smith, Alyssa L.; Cekan, Pavol; Rangel, David P.; Sigurdsson, Snorri Th; Mailer, Colin; Robinson, Bruce H. Theory for spin lattice relaxation of spin probes on weakly deformable dna. *The Journal of Physical Chemistry B*, **2008**, *112*(30), 9219–9236.

[114] Naimushin, A. N.; Fujimoto, B. S.; Schurr, J. M. Dynamic bending rigidity of a 200-bp dna in 4 mm ionic strength: a transient polarization grating study. *Biophys J*, **2000**, *78*(3), 1498–518.

[115] Tirado, MarÃa M.; de la Torre, JosÃ© GarcÃa. Rotational dynamics of rigid, symmetric top macromolecules. application to circular cylinders. *The Journal of Chemical Physics*, **1980**, *73*(4), 1986–1993.

[116] N, Tjandra; S.E., Feller; R.W., Pastor; A.; Bax. Rotational diffusion anisotropy of human ubiquitin from 15n nmr relaxation. *J. Am. Chem. Soc.*, **1995**, *117*, 12562–12566.

[117] Fushman, D.; Cowburn, D. The effect of noncollinearity of 15n-1h dipolar and 15n csa tensors and rotational anisotropy on 15n relaxation, csa/dipolar cross correlation, and trosy. *J Biomol NMR*, **1999**, *13*(2), 139–47.

[118] Ghose, R.; Fushman, D.; Cowburn, D. Determination of the rotational diffusion tensor of macromolecules in solution from nmr relaxation data with a combination of exact and approximate methods–application to the determination of interdomain orientation in multidomain proteins. *J Magn Reson*, **2001**, *149*(2), 204–17.

[119] Bloomfield V.A., D.M. Crothers; Tinoco, I. The quantised hall effect. *Journal of Physical Chemistry of Nucleic Acids*, **1974**.

[120] Garcia de la Torre, J. G.; Bloomfield, V. A. Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. *Q Rev Biophys*, **1981**, *14*(1), 81–139.

[121] Smith, Paul E.; Pettitt, B. Montgomery. Modeling solvent in biomolecular systems. *The Journal of Physical Chemistry*, **1994**, *98*(39), 9700–9711.

[122] Tillett, Marcus L.; Blackledge, Martin J.; Derrick, Jeremy P.; Lian, Lu-Yun; Norwood, Timothy J. Overall rotational diffusion and internal mobility in domain ii of protein g from streptococcus determined from 15n relaxation data. *Protein Science*, **2000**, *9*(6), 1210–1216.

[123] Stultz, C. M. Localized unfolding of collagen explains collagenase cleavage near imino-poor sites. *J Mol Biol*, **2002**, *319*(5), 997–1003.

[124] Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins*, **1993**, *17*(4), 412–25.

[125] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, **1987**, *91*(24), 6269–6271.

[126] Park, Sanghyun; Radmer, Randall J.; Klein, Teri E.; Pande, Vijay S. A new set of molecular mechanics parameters for hydroxyproline and its use in molecular dynamics simulations of collagen-like peptides. *Journal of Computational Chemistry*, **2005**, *26*(15), 1612–1616.

[127] Walker, O.; Varadan, R.; Fushman, D. Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from (15)n relaxation data using computer program rotdif. *J Magn Reson*, **2004**, *168*(2), 336–45.

[128] Tjandra, N.; Garrett, D. S.; Gronenborn, A. M.; Bax, A.; Clore, G. M. Defining long range order in nmr structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat Struct Biol*, **1997**, *4*(6), 443–9.

[129] Smith, P. E.; van Gunsteren, W. F. Translational and rotational diffusion of proteins. *J Mol Biol*, **1994**, *236*(2), 629–36.

[130] Brueschweiler, Rafael; Wright, Peter E. Nmr order parameters of biomolecules: A new analytical representation and application to the gaussian axial fluctuation model. *Journal of the American Chemical Society*, **1994**, *116*(18), 8426–8427.

[131] Fraser. *Conformation in fibrous proteins and related synthetic polypeptides*. academic press, new york, 1973. QP 551.F7 R.

[132] Kramer, R. Z.; Venugopal, M. G.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M. Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *J Mol Biol*, **2000**, *301*(5), 1191–205.

[133] Simon-Lukasik, Kristine V.; Persikov, Anton V.; Brodsky, Barbara; Ramshaw, John A. M.; Laws, William R.; Alexander Ross, J. B.; Ludescher, Richard D. Fluorescence determination of tryptophan side-chain accessibility and dynamics in triple-helical collagen-like peptides. *Biophysical journal*, **2003**, *84*(1), 501–508.

[134] Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. Characterization of leucine side-chain reorientation in collagen-fibrils by solid-state 2h nmr. *Proceedings of the National Academy of Sciences*, **1982**, *79*(2), 386–389.

[135] Markesteijn, A. P.; Hartkamp, R.; Luding, S.; Westerweel, J. A comparison of the value of viscosity for several water models using poiseuille flow in a nano-channel. *J Chem Phys*, **2012**, *136*(13), 134104.

[136] Balasubramanian, S.; Mundy, Christopher J.; Klein, Michael L. Shear viscosity of polar fluids: Molecular dynamics calculations of water. *The Journal of Chemical Physics*, **1996**, *105*(24), 11190–11195.

[137] Hess, Berk. Determining the shear viscosity of model liquids from molecular dynamics simulations. *The Journal of Chemical Physics*, **2002**, *116*(1), 209–217.

[138] Wu, Yujie; Tepper, Harald L.; Voth, Gregory A. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of Chemical Physics*, **2006**, *124*(2), –.

[139] Sgourakis, Nikolaos G.; Day, Ryan; McCallum, Scott A.; Garcia, Angel E. Pressure effects on the ensemble dynamics of ubiquitin inspected with molecular dynamics simulations and isotropic reorientational eigenmode dynamics. *Biophysical journal*, **2008**, *95*(8), 3943–3955.

[140] Bloch, F. The principle of nuclear induction. *Science (New York, N.Y.)*, **1953**, *118*(3068), 425–30.
[141] Redfield, A. G. Nuclear spin thermodynamics in the rotating frame. *Science (New York, N.Y.)*, **1969**, *164*(3883), 1015–23.