Structural Analysis of Type III Collagen Using Two Dimensional Nuclear Magnetic Resonance

Tamara Todorovic Meister and Melinda Harrison*

Cabrini College, 610 King of Prussia Road, Radnor, PA 19087, USA

Type III collagen belongs to one of the twenty-eight known collagen types. It is found throughout the body, including being a major component of blood vessels, internal organs, and skin. A collagen molecule comprises of over a thousand amino acids, which are all arranged in three left-handed peptide chains folded together into a characteristic right-handed triple helix. Type III collagen is known to play a role in wound healing and tissue repair, and is the cause for hereditary Ehlers-Danlos syndrome (EDS) type IV which often leads to ruptures of large arteries and internal organs and, in severe cases, death. However, due to the large size of its molecule, not much is known about how type III collagen's structure influences the functions it has in the body. Two-dimensional nuclear magnetic resonance (2D NMR) is a relatively new analytical technique that has become increasingly more utilized in the study of biological molecules, since it has the ability to detect relationships of atoms both through their bonds and through space.   Still, because of the length and repetitive nature of its sequence, the best way to study the collagen molecule by NMR is with model peptide sequences that mimic a particular collagen segment of interest. This review summarizes the research conducted up to this point on collagens using nuclear magnetic resonance.

*Corresponding author: Melinda Harrison, Cabrini College, 610 King of Prussia Road, Radnor, PA 19087, USA. E-mail: Melinda.a.harrison@cabrini.edu.

Introduction

Collagen is a fibrous protein that makes up close to a third of all protein in the body, and is found in connective tissues like bone, cartilage, tendons, skin, blood vessels, etc. [1]. It is a rigid, non-stretchable protein that enables running and jumping activities, and whose strength originates from the unique structure of peptides that organize in a triple helical conformation.

The amino acid sequence of a basic structural unit of collagen, tropocollagen, includes approximately one thousand amino acids in each polypeptide chain, with the length of three hundred nanometers, diameter of 1-2 nanometers, and weight around 285kDa [2]. There are twenty eight known types of collagen that are found in human tissues, with the most abundant being type I collagen heterotrimer, located mostly in bones and tendons [3]. Also

common are type II and III collagen homotrimers, consisting of three identical chains and existing in cartilage (type II) and blood vessels (type III) among other tissues.

The uniqueness in the structure of collagen comes from the repeating $(X-Y-G)_n$ sequence, where G is glycine and X and Y are often proline and proline-modified amino acids, such as 4-hydroxyproline or 3-hydroxyproline for example [2]. Although proline and hydroxyproline make up nearly thirty percent of collagen residues, the modification of proline to hydroxyproline does not occur until after the polypeptides are synthesized. The replacement of proline's gamma or delta-carbon's hydrogen by a hydroxyl group occurs with the help of propyl hydroxylase, an iron (II) ion activated enzyme that speeds up the reaction of proline residues with oxygen, α-ketoglutarate, and ascorbic acid (vitamin C). As a result of the reaction, products formed include hydroxyproline residues, carbon dioxide, succinate, and dehydroascorbate [4]. The three left-handed peptide chains fold in such a way to form a right-handed triple helix instead of the usual alpha helices or beta sheets. This is due to the high concentration of glycine/proline/ and hydroxyproline repeating sequence in the peptides, with proline rings causing highly rigid angles incapable of substantial rotation (see Figure 1). The spacing within the backbone of the triple helix is tight and, glycine is found in every third residue so that it always occupies the central interior position [5].Once formed, the tropocollagen molecules, specifically collagen types I, II, III, V and XI, further assemble into fibrils, which are staggered arrangement of triple helices that create a banded pattern with gaps between adjacent molecules. These gaps play an important role in covalent attachment of sugars, as well as present a possible nucleation site for bone formation [1].

Nuclear magnetic resonance is a common technique used to a obtain structural information of molecules of all sizes, which has rapidly improved in resolution and range of applications from the time it was first introduced

in the 1970s. Since then, it has become increasingly utilized in the study of biomolecules at the atomic level, especially in determining ways in which molecular structure influences different processes in the body. Previous work performed by a team of researchers has already shown a correlation between type III collagen and wound healing with a decreased scar formation [6]. As an accidental side-effect during that study, it was observed that the genetically modified type III collagen deficient mice were spontaneously developing more tumors than their wild type counterparts. Therefore, based on the hypothesis that certain tumors can be treated as "wounds that do not heal", the research is now being performed to determine if there is a documented connection between the two, and if so what are the mechanisms involved [6]. However, currently, there is no single, efficient way to test for the presence of type III collagen in the tissues of interest, nor is there enough knowledge about its structure to fully understand the mechanisms of how type III collagen influences wound healing or tumor growth on a molecular level.
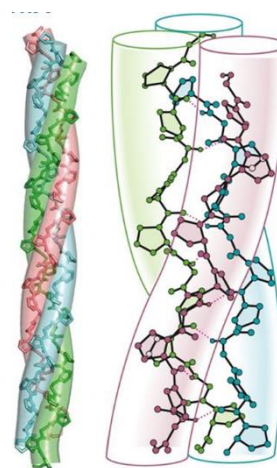

Figure 1. Glycine/proline/hydroxyproline section of collagen. The figure illustrates the three left-handed polypeptides that come together to form a right-handed triple helix (photo taken from Garret & Grisham, 2012, p. 158).

Type III collagen is one of the most abundant collagens in tissues, only second to type I collagen, and can be found in blood vessels,

internal organs, skin, and other elastic tissues [1]. While type III collagen often coexists with type I collagen in many tissues, it is almost exclusively found in large blood vessels such as the aorta but is not found in bone where type I collagen is abundant [7]. Although the role of type III collagen in the body has not yet been fully understood, there have been studies that confirmed its functions in wound healing and tissue repair. Additionally, type III collagen mutations are known to be the cause of hereditary Ehlers-Danlos syndrome type IV, which leads to ruptures of large arteries, other blood vessels and major internal organs [1, 6, 7].

In particular, the study performed by Volk et al. compared wound healing abilities and scar deposition of both young and aged type III collagen deficient mice (Col3+/-) with wild type mice (Col3+/+) in a blind comparison study, with the intention of determining the role of type III collagen in tissue repair [6]. What they found was that the aged Col3+/- mice (56 to 65 weeks old) had wounds that closed more rapidly, but the closure was due to wound contraction and increased scar formation, and not due to healthy tissue regeneration. They also noticed a greater number of myofibroblast cells in granulation tissue, both in vitro and in vivo, in the collagen III deficient mice, as well as more scaring after twenty one days from wounding as a result of increased numbers of myofibroblasts [6]. The study concluded that type III collagen plays a role in wound healing by regulating myofibroblast differentiation, but also conceded that the detailed mechanisms of this modulation are not known, and that a need exists for additional studies on the molecular level that would further explain these findings [6].

Similarly, another study found that mutations in the COL3A1 gene that encodes the sequence for type III procollagen are the cause of Ehlers-Danlos syndrome (EDS) type IV, an inherited connective tissue disorder that results in thinner skin, impaired wound healing, more prominent bruising, and premature death often caused by blood vessel ruptures [8]. This study investigated skin biopsies of individuals with EDS syndrome, to determine the effects of eighteen different COL3A1 mutations on skin and collagen fibril structure. Researchers were able to conclude that different mutations of this gene have a different effect on the location of the sequence change, and that the locations of the sequence change influence the severity of the disease [8]. For example, a glycine substitution near the carboxyl terminal of the chain leads to the most severe type of Ehlers-Danlos syndrome, with highest intracellular accumulation of abnormal procollagen in the rough endoplasmic reticulum (RER) and smaller than normal fibril diameters. On the other hand, a substitution near the amino terminal had variable fibril size and did not show as much retention in the RER. Further research on the molecular mechanisms of how the sequence change leads to the seriousness of EDS symptoms will need to be studied [8].

Structure of Type III Collagen

The structure of type III collagen mostly consists of a homotrimer triple helix, meaning that the three peptides making up the triple helix have an identical amino acid sequence. Due to its size, the determination of structure of full length type III collagen molecules has been a challenging task, and many different techniques, including X-ray crystallography and nuclear magnetic resonance, were used in order to fully understand the molecular basis of its function [9]. It is known that type III collagen is comprised of three alpha1 chains, with a characteristic region at the C terminal that stabilizes the helix. This region, called the disulfide or cysteine knot, is formed by three disulfide bonds between the chains, but so far these bond connectivity's have not been conclusively determined by NMR or X-ray crystallography due to difficulties in replicating the cysteine knot in model peptides [1]. It is also known that the alpha 1 chain of type III collagen is fifteen amino acids longer than alpha 1 or alpha 2 chains of type I collagen. As a result this causes different hydropathy plots of collagen III versus collagen I and therefore

different structure and function. In addition, it has been established that collagen III contains multiple charged residues in the X and Y positions of its sequence, but their specific roles are yet to be determined [7].

The long sequence of over a thousand amino acids in the collagen molecule, as well as the fairly low thermal stability of some of its parts, has presented a problem in structure analysis for both X-ray crystallography and NMR studies. Greater flexibility of longer segments of the triple helix will often obstruct crystallization, and similarly, multiple competing resonances in an NMR spectrum will cause overlapping peaks and poor resolution [1, 5]. Therefore, the study of collagen structure has, so far, been mainly focused on examining individual segments of the helix, with artificially synthesized peptides that mimic the sequences found in natural collagen. However, when synthesizing these peptide sequences, usually repeating sections of highly stable proline/hydroxyproline/glycinesegments, or a cysteine knot, are added at the ends of the peptide of interesting order to stabilize the helix [1, 10]. These sections provide additional sites of hydrogen bonding which hold the three triple helix chains tightly together.
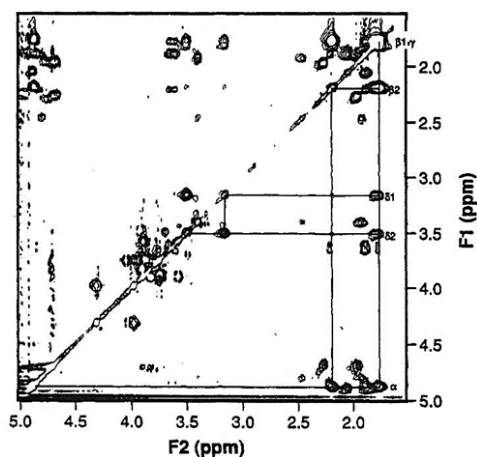

Figure 2. Total correlation spectra of the (POG)$_{10}$ chain. The spectra show a different chemical shift for hydrogens belonging to proline, hydroxyproline and glycine due to the downfield pull of nitrogen and hydroxyl group. Also, at leasttwo peaks are seen for each residue from the proline's beta and gamma carbon hydrogens and one peak from the alpha carbon hydrogens (adapted from Li *et al* 1993).

Nuclear Magnetic Resonance Background and Technique

Richard R. Ernst, a Swiss professor and scientist, won a Nobel Prize in Chemistry in 1991 for performing the first experiment using two dimensional nuclear magnetic resonance (2D NMR) [11]. NMR is a spectroscopy technique that studies ways in which atoms respond to radiofrequency electromagnetic radiation. In one dimensional nuclear magnetic resonance (1D NMR), the superconducting magnets direct a single pulse to the sample, with having a single time variable t before the signal is detected. In two dimensional NMR, there is a second pulse following the first one, with a t1 time interval from the first pulse. Therefore, another time variable is created and a second Fourier Transform operation is needed, in order to give rise to two frequency axes which convey the correlation of atoms between them. Also, with both 1D and 2 D NMR, the intensities of signals can be determined from the spectra, in 1D NMR the intensity of the peak height is used and the intensity of 2D NMR resonances is calculated from the peaks' volumes [12].

Two-dimensional NMR spectra can be divided into two main types, depending on the nature of relationships they analyze. When NMR active nuclei are irradiated with external magnetic radiation, the energy gained becomes dispersed to the neighboring nuclei both through the electron cloud of a bond, and directly through space between the dipoles. The main types of through bond correlation (see Figure 2) consist of homonuclear spectra. Homonuclear spectra look at the relationship of the same types of atoms in a molecule and include correlation spectroscopy (^{1}H-^{1}H COSY), total correlation spectroscopy (^{1}H-^{1}H TOCSY), and heteronuclear spectra, like heteronuclear single quantum correlation (HSQC), heteronuclear multiple quantum correlation (HMQC), and heteronuclear multiple bond correlation (HMBC) [12]. Heteronuclear spectra investigate the association of two different types of atoms, such as hydrogen and nitrogen (^{1}H-^{15}N) or hydrogen

and carbon (^{1}H-^{13}C). Nuclear Overhauser effect spectroscopy (NOESY) and rotational Overhauser effect spectroscopy (ROESY), on the other hand, portray the direct distances of nuclei in space, and are becoming an increasingly useful tool in biochemical studies where there is a particular interest in the knowledge of spatial proximity of atoms in different parts of a large biomolecule [12].

Studying the Folding Pathway of Collagen using NMR

The first application of two dimensional nuclear magnetic resonance in the analysis of a collagen-like homotrimer triple helix was performed by Li et al. in 1993 [5]. They were interested in comparing the structure and function of a highly stable section of a homotrimer, having a repeating sequence of proline(P)/hydroxyproline(O)/glycine(G), with the less stable T3-785 peptide sequence found at the trypsin and collagenase cleavage sites of native type III collagen [5]. That particular sequence of type III collagen also contains the site of a single glycine to serine substitution responsible for Ehlers-Danlos syndrome, and is a particularly cyclic amino acid poor region, meaning it is not comprised of many proline or proline-like residues that contain a ring structure [5].
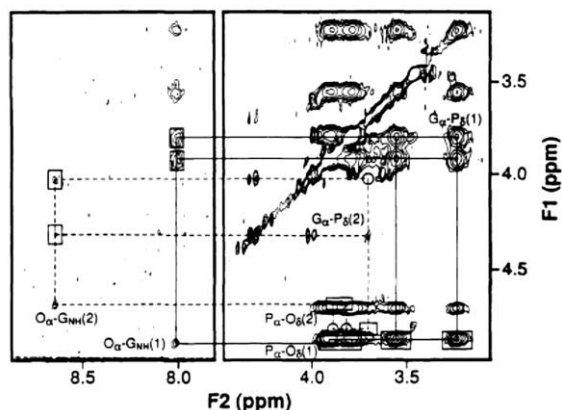

Figure 3. NOESY spectra of the (POG)$_{10}$ chain. The amide region and aliphatic region of the spectra show correlations within the tripeptide. The squares around the peaks show positions in the TOCSY spectra, with the solid and dashed lines connecting tripeptides one (more orderly triple helix) and two (less orderly triple helix), respectively (adapted from Li et al 1993).
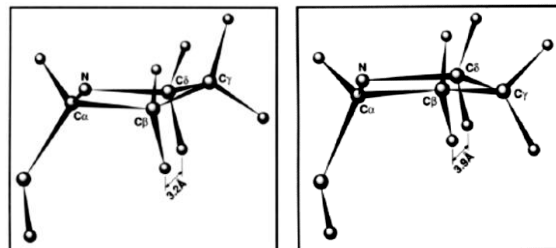

Figure 4. Two orientations of proline depending on the position in the sequence. Depending on placement in the sequence of the triple helix peptides, proline takes on a puckered up or puckered down orientation (bending of the ring), with different spacing between the hydrogens seen in the NOESY spectra (Li *et al* 1993).

The results obtained from homonuclear DQF-COSY and TOCSY were able to distinguish the glycine from proline and hydroxyproline atoms by peaks from the amide hydrogens and two alpha carbon hydrogens, which created different chemical shifting patterns due to downfield de-shielding of nitrogen and hydroxyl group on hydroxyproline (see Figure 3). The through-space NOESY correlation was able to confirm that the peaks seen on TOCSY spectra did in fact originate from the triple helix, as well as the "up" or "down" positions and angles of the proline ring [5]. The ^{1}H-^{15}N heteronuclear spectra of T3-785 peptide showed twelve peaks, nine from the triple helix and three from the monomer peptides, with the nine triple helix peaks disappearing when the temperature is raised above the helix' melting point [5]. The study concluded that NMR can be used to confirm the helical conformation seen in diffraction studies, and the proline ring puckering, bending up or down, is determined by its position in the sequence [Figure 4] [5]. It also established that the T3-785 has a tight packing of the triple helix in the central region, similar to the (POG)$_{10}$ chain, contrary to some beliefs that claimed a non-standard and non-helical flexible conformation. Additionally, NMR was used to confirm that the distances between the three chains are compatible with previous models of collagen helix [5]. ANOESY-documented one amino acid stagger between residues showed that the three-residue repeating sequence is shifted up by one residue in each peptide, and that glycine is found in the center of the helix at each step

(see Figure 5). From this work, it was concluded that collagen binding sites are created mostly by the specificity of side chain residue interactions, without major changes in the backbone of the triple helix. The research conducted by Li et al. developed original ideas and analysis methods that paved the way for collagen studies using nuclear magnetic resonance [5].
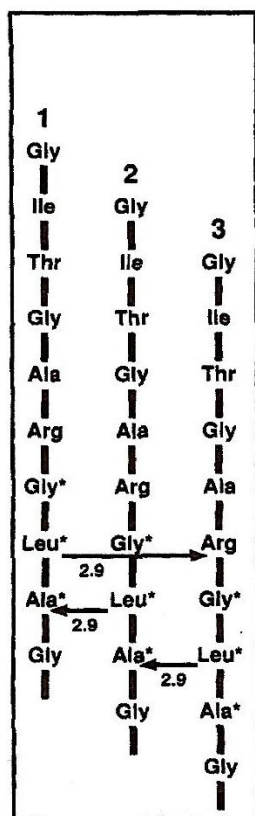

Figure 5. T3-785 staggered conformation. The peptides are arranged in a way that the three residue sequence is offset by one amino acid, and that glycine is found in every cross-section. Residues marked with * are 15N-labeled, and the distances between the residues are determined by NOESY (adapted from Li *et al* 1993).

A subsequent study that further built on the work by Li was performed by Buevich *et al*. [10]. Here, the authors examined the process of folding of a collagen-like triple helix homotrimer, rich in proline, hydroxyproline and glycine residues, in order to better understand the connective tissue diseases that occur due to incorrect folding of the protein. Their research consisted of investigating the effects of *cis* to *trans* conformational changes of proline during folding, as well as the rate of folding of the triple helix. The sequence of the peptides was designed to mimic an eighteen residue sequence of type I collagen, with added four glycine/proline/hydroxyproline sections that

stabilized the helix and acted as nucleation sites - a part of the molecule where the formation of the triple helix is initiated [10]. The results from ^{1}H-^{15}N HSQC data at temperatures above the triple helix melting point revealed the presence of more than one peak for each of the labeled amino acids, indicating the presence of at least two conformational isomers for each residue (see Figure 6) [10]. Also, the volume and intensity of the peaks were not equally distributed between the isomers of the same amino acid residue, signifying that the residue spends more time in one conformation versus the other. In order to confirm that these peaks were in fact originating from the interconverting conformers of the same residue, ^{1}H-^{15}N HSQC-NOESY was performed at the same temperature, which showed that there is a slow exchange from one conformation to the other, in the order of $0.030s^{-1}$ to $0.066s^{-1}$ for the trans to cis conversion and $0.32s^{-1}$ to $0.40s^{-1}$ for the opposite direction conversion [10]. The kinetics of folding was determined to be fast when first initiated, with slowing down after the first data point, to the rate constant of the order of $0.001s^{-1}$. The study was also able to conclude that the folding is initiated at the C-terminal end by examining the rate of disappearance of labeled glycine25 monomer and appearance of labeled glycine25 trimer, which occurred faster than the other labeled amino acids [10]. Another important discovery of this study was the observation of intermediates that formed during folding, with alanine6, for example, having four additional peaks near the trimer peak and one additional peak near the monomer peak, which initially increased and then decreased in size. Similarly, NOESY data showed that only about twenty percent of unfolded monomer residues exist in all trans conformation, which leads to the conclusion that they, therefore, must undergo a cis to trans conformational change before forming the triple helix since residues in all three chains of the triple helix must be in trans form in order for folding to occur (Figure 6). Research performed by Buevich is one of the first studies to confirm the presence of intermediate species during triple helix formation using 2D NMR, as
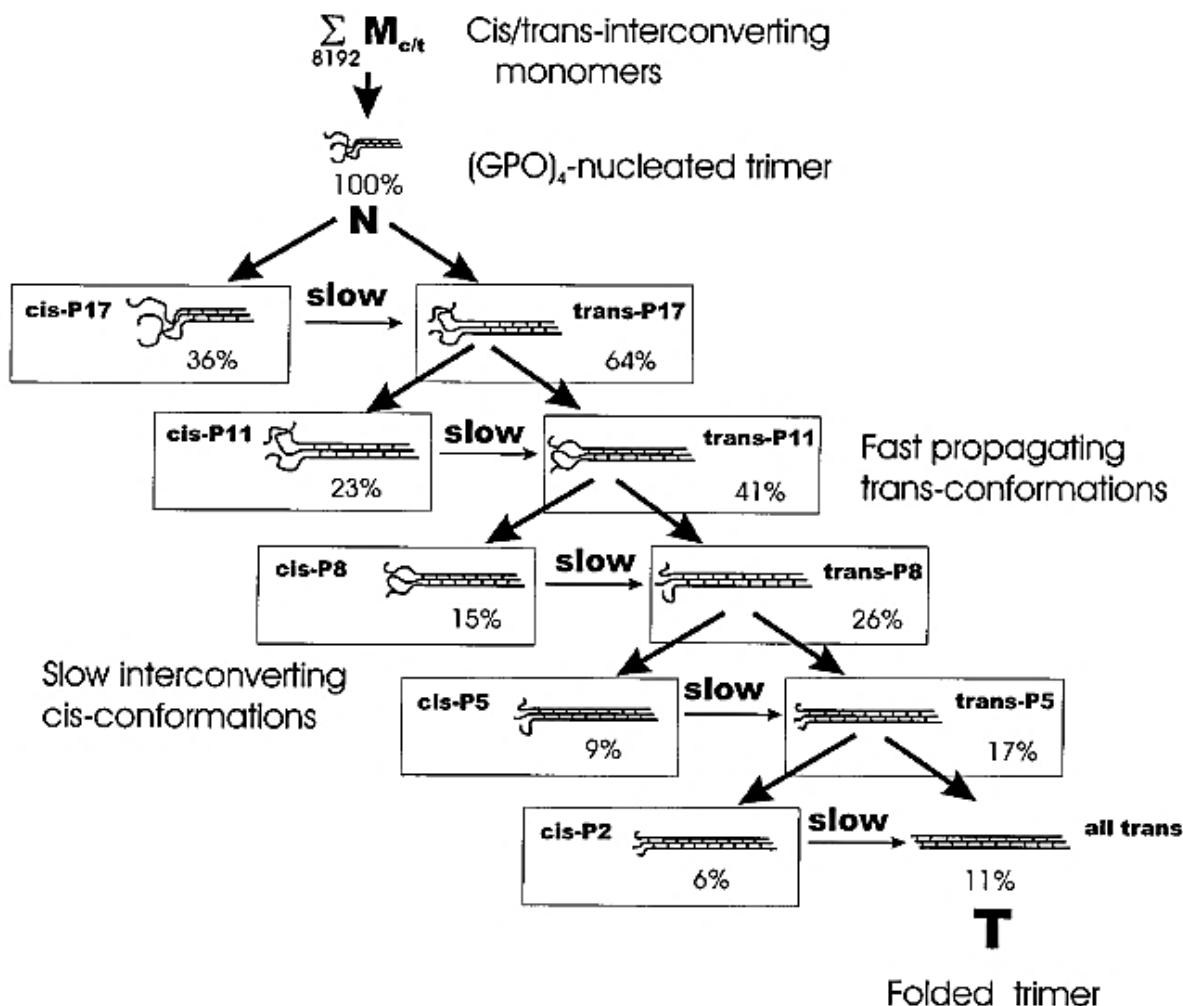
Figure 6. The folding process of the triple helix. The schematic shows the rate limiting step of the *cis* to *trans* conversion, with the fast propagation in a "zipper like" manner occurring at the C terminal when all three residues are in trans conformation (image taken from Buevich *et al* 2000).

well as provide evidence of nucleation being initiated at the C terminal [10]. Also it was shown that the residue conversion from *cis* to *trans* conformation is the rate limiting step for the formation of triple helix [10]. This study is consistent with, and supports the previous findings of Smith's group who noticed that the most severe form of EDS type IV occurs when amino acid substitution in type III collagen sequence arises at the C terminal end [8]. Since the nucleation of triple helix is shown to originate at the C terminal, it is likely that a substitution early on in the folding process would lead to a greater risk of misfolding of the

whole helix and a more severe hindrance of the normal type III collagen function [10].

Research that further investigated the folding of a homotrimer triple helix was performed by Xu *et al*. in 2003 [13]. They looked at changes in the rate of folding of the triple helix when a single proline or hydroxyproline residue at the C terminal is replaced by an alanine residue [10, 13]. The main reason for this research was to investigate the role of proline and hydroxyproline on the nucleation of the helix. The model peptides used for their study were based on those used in Buevich *et al*, an eighteen
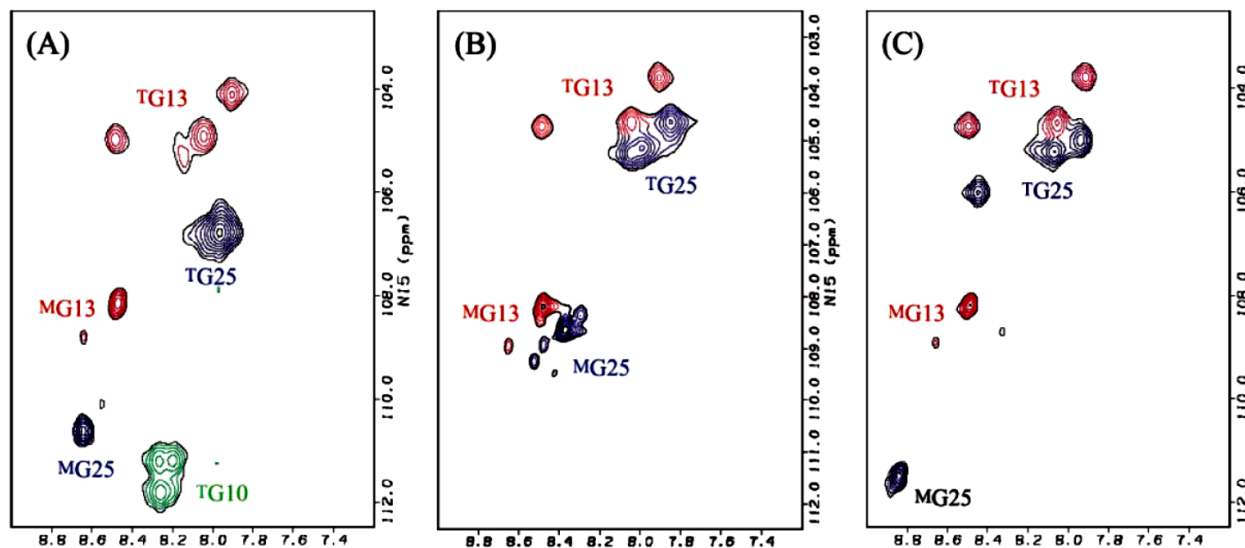
Figure 7. ^{1}H-^{15}N HSQC of the three peptides studied in Xu et al. (2003). The spectra show peptides T1-892 (A), T1-892(O24A) in (B), and T1-892(P26A) (C) at zero degrees Celsius. Peaks are labeled with "T" for trimer and "M" for monomer peak, with the corresponding ^{15}N labeled glycine residues.

amino acid sequence taken from α1 type I collagen chain, with a four time repeating glycine/proline/hydroxyproline section at the C terminal [10, 13]. Additionally, one peptide had a proline replaced by alanine in position 26, and the other had a hydroxyproline replaced by alanine in position 24. The third peptide contained no substitutions but was additionally ^{15}N labeled in positions 6 and 10. Xu *et al* were able to observe differences in the glycine25 peak between the three peptides, which enabled their distinction, and allowed them to monitor the disappearance of monomer peaks in order to study the rates of folding [13]. The ^{1}H-^{15}N NOESY was used to examine the rigidity of monomers and triple helices with and without the substitutions, by looking at distances between labeled residues through space [13] (Figure 7). They were able to conclude that the rate of nucleation at the C terminal is concentration dependent in such a way that at low concentrations (0.14 mM) a replacement of proline and, especially, hydroxyproline by alanine will slow down the nucleation, whereas the rate does not significantly change at higher concentrations [13]. They were also able to conclude that the rigidity of the monomers plays a role in the rate of nucleation, with alanine in

particular making the monomer less rigid. This decreased rigidity slowed the folding rate down, but was equally counteracted by a greater likelihood of containing all *trans* residues which in return increased the rate [13]. Additionally, they observed that further propagation is not affected by any of the substitutions, which does not support previous research mentioned here where residue substitutions early in the sequence had a greater effect on the correct function of the whole molecule [13]. However, their findings are limited to the level of a short model-peptide triple helix.

Researchers have made progress on understanding the folding of collagen. The folding rate of collagen is affected with residue substitutions [13]. The folding of the triple helix is initiated at the C terminal and requires all *trans* residue conformation in order for folding to occur [10]. These results help understand potential causes of diseases such as Ehlers-Danlos syndrome, brittle bone disease, and many others [8].

2D NMR of Collagen-like Heterotrimer Triple Helix

Peptide abbreviations and chemical sequences

Abbreviation	Sequence[a]
K	(PKG)$_{10}$
D	(DOG)$_{10}$
O	(POG)$_{10}$
K*	(PKG)$_4$<u>PKGPKG</u>(PKG)$_4$
D*	(DOG)$_4$DO<u>GD</u>OG(DOG)$_4$
O*	(POG)$_4$PO<u>G</u>POG(POG)$_4$

[a]The underlined amino acids are uniformly ^{15}N, ^{13}C-labeled

Figure 8. The sequence of synthesized peptides and their abbreviations. The peptides K, D and O were synthesized without any ^{15}N and ^{13}C labeled amino acids, whereas peptides K*, D*, and O* have two or three residues labeled in the middle of the sequence (adapted from Fallas *et al* 2009).

Another important investigation in the two dimensional NMR study of collagen-like peptides was performed by Fallas *et al*. where their work was the first that investigated a complete structural determination of a heterotrimer triple helix using two-dimensional nuclear magnetic resonance [3]. Similarly to other NMR studies of collagen-like peptides mentioned earlier, Fallas also used model peptides to explain the structure of natural collagen (Figure 8) [3]. The use of model peptides for this type of research is necessary due to the large size of natural collagen's molecule and the diverse environment it is found in, which contributes to peak overlap and makes it hard to obtain useful NMR data at the atomic level. Researchers were particularly interested in looking at the role of many charged amino acids that are known to exist in the collagen sequence, as well as the role of globular domains like the aforementioned cysteine knot in influencing its formation [3, 10]. However, the effects of globular domains were not analyzed directly, but were implied from other findings about the assembly of a heterotrimer triple helix.

The results obtained from ^{1}H-^{15}N HSQC showed two peaks from the monomer D* and K* chains (labeled M), which were confirmed by TOCSY of individual peptides and remained in solution because of highly charged lysine and aspartate residues [3]. Also, a single peak from the O*/O*/O* homotrimer triple helix was observed in the spectra (labeled GH), which formed because of its high stability [3]. The spectra also contained five additional peaks belonging to the newly formed heterotrimer, whose connectivity's were confirmed by through-bond TOCSY (Figure 9). The presence of only one peak per residue in the newly formed heterotrimer triple helix (5 peaks total) indicates that there is only one preferential arrangement of the triple helix that will form from these three peptides, out of six possible combinations determined by the statistical computer program [3]. The triple helix conformation, with measured distances between hydrogens from neighboring chains, was confirmed by through-space NOESY [Figure 10]. Results concluded that the backbone of the heterotrimer triple helix behaves similarly to a homotrimer backbone, with the preserved

hydrogen bonding that forms between the amide glycine hydrogen of one chain and carbonyl oxygen from X positioned amino acid in the other chain; as well as that the three chains are staggered in a way that glycine is found in every cross-section, consistent with the Li et al. research of homotrimers [3, 5]. This finding supports the idea that heterotrimer model-peptides can be used for the study of charged amino acid residues of type III collagen homotrimer [3]. Another important conclusion came from the discoveries that the charged amino acid residues of lysine and aspartate act as an additional support mechanism to the structure of triple helix. The formation of hydrogen bonding between the oxygen of the carbonyl group of aspartate side chain and the hydrogen from the NH3+ group of lysine side chain acts as an external stabilizing force that wraps around the outside of the helix [3]. Fallas claimed this supplementary hydrogen bonding is the cause for the observed increased thermal stability of this triple helix heterotrimer, and that this discovery is important because collagens have a high proportion of charged residues [3]. For example, a fairly charged NRGERG sequence of type III collagen showed many of intra-chain and inter-chain hydrogen bonds, but it is still unclear whether this bonding is present in solution or is simply a result of crystal packing [1]. Performing the type of NMR research presented in this article on a model peptide with the NRGERG sequence could potentially answer these questions, and provide some explanations to the function of charged amino acids in type III collagen.   Finally, Fallas concluded that the composition of a helix' sequence can be selective in choosing a single arrangement, without having the presence of other globular terminal domains to influence its assembly [3]. This finding is somewhat contradictory to previous studies, which claim that collagen requires small globular proteins, such as small leucine-rich proteoglycans (SLRPs) or a cysteine knot, for proper development of triple helix and fibrils [1, 14].This shows that further research is needed to fully understand the importance of collagen structure in its interactions with other extracellular and cellular proteins, and the roles it plays in the body.
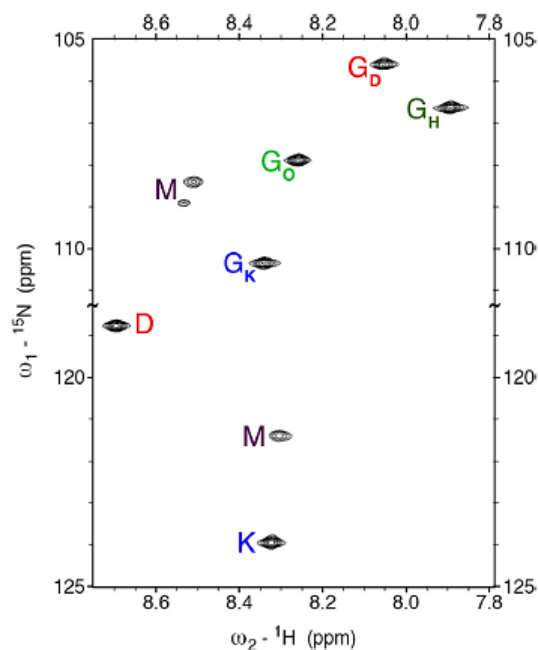

Figure 9. ^{1}H-^{15}N heteronuclear single quantum correlation of K*/D*/O* triple helix. The spectra showed two peaks from the monomer D* and K* chains (labeled "M"), a single peak from the O*/O*/O* triple helix which formed because of its high stability (labeled "G$_H$"), and five peaks from the newly formed heterotrimer (peaks labeled with "D", "K", "G$_D$", "G$_O$", and "G$_K$"). The presence of only one peak per residue in the heterotrimer triple helix indicates that there is only one possible arrangement of the triple helix that will form from these three peptides (spectra from Fallas *et al*, 2009).

Future Direction of Collagen Analysis Using NMR

The methods described in these articles show that NMR can be used to examine the structure of a triple helix trimer at the atomic level, as well as to investigate the roles that different residues play in helix stability and functions of active sites. The difficulties is using nuclear magnetic resonance for structural analysis of collagens lie in the fact that their size presents a challenge in obtaining high resolution data, due to multiple overlapping peaks on NMR spectra. Furthermore, the long acquisition times and multiple scans needed in order to gather two dimensional spectra make the process somewhat inefficient.   Therefore the most
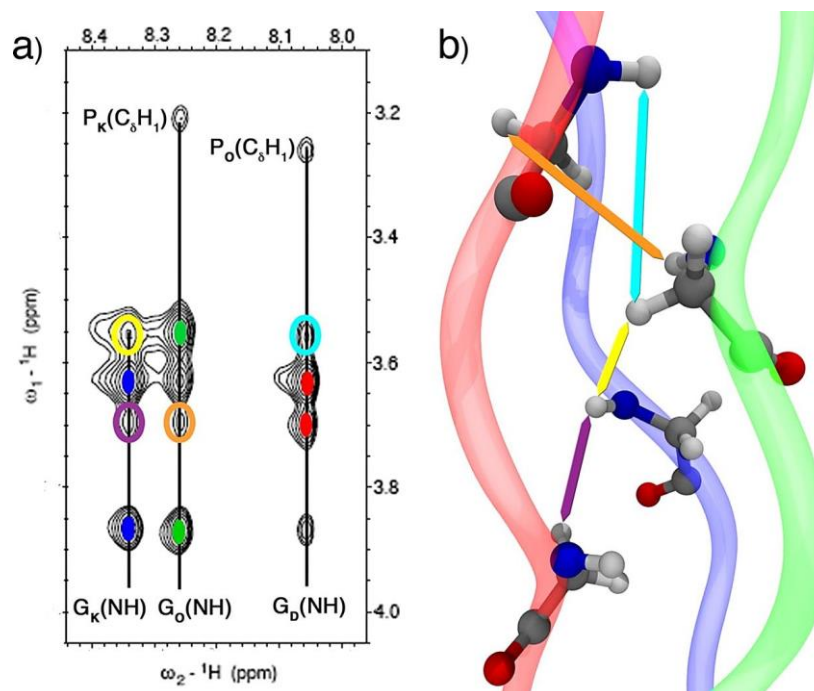
Figure 10. NOESY spectra and model of the heterotrimer triple helix. The differently colored peaks show the through space correlation between hydrogen atoms of neighboring peptides. The volumes of the peaks correspond to their intensities and through-space distance from each other (adapted from Fallas *et al*, 2009).

reliable method for NMR analysis of collagens has been short peptide synthesis, with specifically labeled amino-acid sequences that mimic segments of interest in the natural collagen triple helix. In recent years, subsequent research on model peptides has developed new methods that improved NMR resolution and reduced peak overlap. In addition, these new studies have provided further insight into the folding mechanisms of the triple helix, the role of charged amino acids in structure stability, and the effects of mutations in different locations of collagen sequence. The development of these methods allows for larger peptides with a longer amino acid sequences to be successfully analyzed at the atomic level. Therefore, with a more comprehensive knowledge of the structure and individual amino-acids' roles in collagen, our ability to understand and treat diseases related to type III collagen malfunctions will also be greater.

References

1. Boudko SP, Engel J, Okuyama K, Mizuno K, Bachinger HP, Schumacher MA. 2008. Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. Journal of Biological Chemistry, 283(39):32580-32589.
2. Shoulders MD, Raines RT. 2009. Collagen structure and stability. Annual Review of Biochemistry, 78:929-958.
3. Fallas JA, Gauba V, Hartgerink JD. 2009. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. The Journal of Biological Chemistry, 284(39):26851-26859.
4. Garrett RH, Grisham CM. 2012. Biochemistry (5thed.). Belmont, CA: Cengage Learning. ISBN-13: 978-1-133-10629-6
5. Li MH, Fan P, Brosky B, Baum J. 1993. Two-dimensional NMR assignments and conformation of (Pro-Hyp-Gly)10 and a designed collagen triple-helical peptide. Biochemistry, 32(29):7377-7387.
6. Volk SW, Wang Y, Mauldin EA, Liechty KW, Adams SL. 2011. Diminished type III collagen promotes myofibroblast

differentiation and increases scar deposition in cutaneous wound healing. Cells Tissues Organs, 194(1):25-37.

7. Ala-Kokko L, Kontusaari S, Baldwin CT, Kuivaniemi H, Prockop DJ. 1989. Structure of cDNA clones coding for the entire preproα1(III) chain of human type III procollagen: Differences in protein structure from type I procollagen and conservation of codon preferences.Biochemistry Journal, 260(2):509-516. PMCID: PMC1138697

8. Smith LT, Schwarze U, Goldstein J, Byers. 1997. 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. Journal of Investigative Dermatology, 108(3):241–247.

9. Brodsky B, Persikov AV. 2005. Molecular structure of the collagen triple helix. Advances in Protein Chemistry, 70:301-339.

10. Buevich AV, Dai Q, Liu X, Brodsky B, Baum J. 2000. Site-specific NMR monitoring of cis-trans isomerisation in the folding of the proline-rich collagen triple helix. Biochemistry, 39(15):4299-4308.

11. Nobel Media AB (2014, September 19). The Nobel Prize in Chemistry 1991: Richard R. Ernst – facts. Nobelprize.org.Retrieved from (http://www.nobelprize.org/ nobel_prizes/chemistry/laureates/1991/ernst-facts.html)

12. Simpson JH. 2012. Organic structure determination using 2-D NMR spectroscopy: A problem based approach (2nd ed.). San Diego, CA: Elsevier Science. ISBN 978-0-12-384970-0

13. Xu Y, Hyde T, Wang X, Bhate M, Brodsky B, Baum J. 2003. NMR and CD spectroscopy show that amino acid restriction of the unfolded state leads to efficient folding. Biochemistry, 42(29):8696-8703.

14. Ameye L, Young MF. 2002. Mice deficient in small leucine-rich proteoglycans: Novel in vivo models for osteoporosis, osteoarthritis, Ehlers-Danlos syndrome, muscular dystrophy, and corneal diseases. Glycobiology, 12(9):107R-116R.