SYNTHESIS AND BIOPHYSICAL CHARACTERIZATION OF COLLAGEN-LIKE PEPTIDES: GLYCOSYLATION-INDUCED STABILIZATION OF THE TRIPLE-HELIX

By

James Gerald Bann

A DISSERTATION

Presented to the Department of Biochemistry and Molecular Biology
and the Oregon Health Sciences University
School of Medicine
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

January, 2000

School of Medicine
Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. thesis of

James Gerald Bann

has been approved

[redacted]
Hans Peter Bächinger, Ph.D.

[redacted]
Svetlana Lutsenko, Ph.D.

[redacted]
David Kabat, Ph.D.

[redacted]
Jack H. Kaplan, Ph.D.

[redacted]
David H. Peyton, Ph.D.

TABLE OF CONTENTS

Chapter 1. Introduction. 1

A. Collagen: From Vertebrates to Invertebrates 3

1.1. General Features of Vertebrate Collagens 3
1.2. Structure of the Collagen Triple-Helix 4
1.3. Sweet is Stable: The unusual collagen of *Riftia pachyptila* 6
1.4. Comparison of *R. pachyptila* cuticle to *L. terrestris* and *N. virens* Cuticle collagens 10

B. Towards the Mechanism of Stabilization of Collagen 13

1.1. Hydroxylation and Collagen Stability 13
1.2. Plausible mechanisms for the stabilization of collagen by Hydroxyproline 14
1.3. The role of water in stabilizing the collagen structure 17

C. Glycosylation and Stability 21

1.1 Effect of Glycosylation on Peptide Backbone Conformation 21
1.2 Do carbohydrates act sterically to limit conformational space? 24

Chapter 2. Manuscript #1. Sweet is Stable: Glycosylation Stabilizes Collagen 28

Chapter 3. Manuscript #2. Glycosylation/Hydroxylation-induced stabilization of the collagen triple-helix: 4-trans hydroxyproline in the Xaa position can stabilize the triple-helix 44

Chapter 4. Summary and Conclusions. 62

Literature Cited. 65

Acknowledgements

I would first like to thank my advisor, Dr. Hans Peter Bächinger, for his support and encouragement throughout this thesis, both financially and for my work. He provided an excellent environment in which to explore new ideas, and taught me to think not only independently but critically.

I am especially grateful to Dr. David H. Peyton of Portland State University for his support and encouragement, and for the generous use of his 400 and 500 MHz NMR instruments. I also wish to thank his previous graduate student Ayna Alfadhli for her help and encouragement.

I wish to thank the other members of my committee for their support and for agreeing to evaluate this dissertation. They are: Dr. Jack H. Kaplan, Dr. David Kabat, and Dr. Svetlana Lutsenko. I wish to especially thank Dr. Kabat for his encouragement of me.

I would like to thank the members of the Bächinger laboratory, past and present for their encouragement and support. They are: Bruce Boswell, Dr. Konrad Beck, Dr. J. Randy Macdonald, Dr. B. Kerry Maddox, Diane Mechling, Aswari Mokashi, Dr. Janice Vranka, and Dr. Baifei Zheng. I would also like to thank Jeannine Lawrence at Oregon State University for her help with FT-IR, and to Dr. Curtis Johnson for allowing me to use his instrument.

I wish to thank my children, Seamus, Gabriel and Jonah, for providing the right balance of laughter, fun and "Yea dad, the best scientist in the world!!" needed to complete this thesis.

Last but not least, I am indebted to my wife, Jodie, for without her constant love and encouragement I could not have accomplished any of this.

Abstract

Collagen is the most abundant protein in the human body. It is primarily composed of repeating units of -Gly-Xaa-Yaa-, where Gly is glycine, Xaa is typically proline (Pro) and Yaa is typically 4-(*R*)-hydroxy-L-proline. For most collagens, there is a correlation between the melting temperature (T_m) of the triple-helix and the total content of Pro and Hyp. The cuticle collagen from the deep-sea hydrothermal vent worm *Riftia pachyptila* has been shown, however, to not obey this correlation. The T_m of the triple-helix is quite high (37°C) despite a low content of Pro and Hyp. However, the *R. pachyptila* cuticle collagen was found to have a high content of threonine residues in the Yaa position of the -Gly-Xaa-Yaa- triplet repeat that are *O*-glycosylated, mainly with di and tri-saccharides of galactose.

In this work, I addressed the question, "Does galactose need to be present on threonine in order to achieve a stable triple-helix?" I did this through the synthesis of both glycosylated (with the monosaccharide β-D-galactose) and non-glycosylated peptides. These peptides have the general form Ac-(Gly-Pro-Thr)$_{10}$-NH_2, and serve as models of the triple-helix of the worm collagen. I show, using circular dichroism spectroscopy (CD), analytical ultracentrifugation, and 1 and 2-dimensional nuclear magnetic resonance (NMR) spectroscopy that only in a fully *O*-glycosylated peptide of Ac-(Gly-Pro-Thr)$_{10}$-NH_2 is a stable triple-helix able to form. I also show that hydroxyproline in the Xaa position of the peptide Ac-(Gly-Pro-Thr)$_{10}$-NH_2, previously thought to destabilize the triple-helix, has a stabilizing influence and allows this peptide to form a triple-helix independent of glycosylation.

A plausible mechanism of stabilization by *O*-glycosylation is the possibility that glycosylation might cause the formation of a more extended structure in the unfolded (single-chain) state. The mechanism is based upon the fact that the triple-helix is composed of three extended, left-handed poly-L-proline II helices, and so stabilization may be conferred by the ability of the single polypeptide chain to assume an extended poly-L-proline II helix. Another plausible mechanism is that the carbohydrate hydroxyls are participating in hydrogen bonding interactions with either the peptide backbone or with water molecules in the triple-helical state.

In an attempt to address the latter question, we measured the effect of increasing concentrations of D-(+) galactose on the stability of the triple-helix of Ac-(Gly-Hyp-Thr)$_{10}$-NH_2, and show that while galactose increases the stability of this peptide, no apparent change is observed in the shape of the unfolding transition or in the magnitude of the CD at 223 nm. This suggested that the galactose did not change the conformation of the triple-helix (increase the number of stable hydrogen bonds). Also, this data did not support the idea that the carbohydrates are mediating interactions of the peptide backbone with water, since the effect of increasing concentrations of galactose would be to *remove* water molecules from the peptide backbone. One possibility is that when the carbohydrate is covalently attached to threonine, the local environment (polarity, dielectric constant) around the threonine is changed to favor strengthening of the pre-formed hydrogen bonds.

From these experiments we conclude that glycosylation stabilizes a collagen triple-helix, and that the presence of hydroxyproline in the Xaa position can also stabilize the triple-helix. The exact contribution of glycosylation to the free energy of either the

folded or the unfolded state remains to be determined. However, the mechanism seems to be in contrast to that of vertebrates, where evidence suggests that the influence on the *cis-trans* isomer ratio of proline is a major determinant of the thermal stability of the triple-helix.

Chapter 1

Introduction

One of the most intriguing and exciting problems that has yet to be solved in the realm of biochemistry is the protein folding problem. The late Christian B. Anfinsen, who was awarded the 1972 Nobel Prize for Chemistry along with Stanford Moore and William H. Stein, established that the amino acid sequence of a protein, such as ribonuclease, is solely required for it to achieve its structure (Anfinsen,1973). The final folded structure in general represents the most thermodynamically stable structure, and subsequently understanding the factors that are required to achieve a native structure has been the focus of many laboratories (Creighton, 1990). One such protein that is the focus of the Bächinger laboratory is the protein collagen. This protein is the most abundant protein in the vertebrates (and in some invertebrates). Of the collagens the most prevalent are the fibrillar collagens, and these are characterized by a long chain of approximately 300 repeats of the sequence Gly-Xaa-Yaa, where Xaa and Yaa are frequently the imino acids proline (Pro) and 4-(*R*)-hydroxyproline (Hyp), respectively. The collagen molecules are composed of three chains, each chain assuming an extended, left-handed poly-proline II helix that is staggered by one residue, and these then wrap around one another in a right-handed fashion to form the triple-helix (Ramachandran, 1988). The structure of the collagen triple-helix has been known for a long time, however the forces which govern its formation and stabilization are not understood, after nearly 50 years of work on it, despite its simple structure and sequence. A hypothesis that has had much support and that was established early on was a correlation that, as the total amount of Pro and Hyp residues increased, the stability of the protein should also

increase (Burjanadze,1979). The only collagen thus far to not follow this correlation is the cuticle collagen of the deep-sea hydrothermal vent worm *Riftia pachyptila*. This collagen has a relatively high stability against heat denaturation, with a T_m of 37°C, and is comparable to another deep sea worm cuticle collagen from *Alvinella pompejana*, which has a T_m of 40°C (but nearly 1.5 times longer) (Gaill, Wiedemann et al., 1991). The major difference between these two collagens is the content of imino acids, in particular hydroxyproline. While *A. pompejana* has a total content of Pro/Hyp of 203 (178 Hyp) per 1000 amino acids, the *R. pachyptila* has a total content of only 40 (21 Hyp) per 1000 amino acids. However, the content of threonine residues in *R. pachyptila* was 182 per 1000 amino acids, and strongly suggested that threonine must somehow substitute for Hyp in the Yaa position in order to achieve the high observed thermal stability (Gaill, Wiedemann et al., 1991). Biochemical analyses and partial sequencing of the cuticle collagen further revealed that the threonine residues occupied the Yaa position, and these were almost all glycosylated, mainly with di and tri-saccharides of galactose (Mann, Mechling, et al.,1996). From this it was hypothesized that the carbohydrates were necessary for the cuticle collagen to achieve its thermal stability, since work on short peptides having the sequence –Gly-Pro-Thr- revealed an inability to form a triple-helix in aqueous buffer (Mann, 1996). The main focus of my work then has been to understand *if* galactose is necessary for the formation of a triple-helix, and how this post-translational modification may stabilize such a structure. This was accomplished through the synthesis of collagen-like peptides which have the general sequence Gly-Pro-Thr and Gly-Pro-Thr-(β-D-Galactose). We show, using circular dichroism, NMR and analytical ultracentrifugation, that a stable triple-helical structure is formed if galactose is attached

to threonine. We also have assesed the thermodynamic stability of a peptides which have hydroxyproline in the Xaa position of the Gly-Pro-Thr/Thr(Gal) triplet repeat.

A: Collagen: From Vertebrates to Invertebrates

1.1 General features of Vertebrate collagens

The collagens are a diverse group of proteins which share the common feature of having a very distinct, rod-like structure known as the triple-helix. There are currently 19 different types of vertebrate collagen, and together these molecules form the major components of the extracellular matrix (van der Rest and Garrone, 1991). The collagen triple-helix structure is found in a diverse array of proteins, and includes not only the collagens but also proteins such as the macrophage scavenger receptor, C1q of complement, the acetylcholine receptor, and others (Brodsky and Ramshaw, 1997). The most common of the collagens are the fibrillar collagens, and these include types I, II, III, V, and XI. The fibrillar collagens assemble from single, triple-helical molecules into higher-order aggregates called fibrils. The individual triple-helical molecules (~300 nm in length, 1.5 nm in diameter) are aligned parallel to one another and overlap by a distance of ~67 nm, called a *D*-period. By electron microscopy, this overlap produces a distinctive banded pattern (Kadler, Holmes et al., 1996). The individual chains are also flanked by short non-triple-helical segments called telopeptides, that help to coordinate the assembly of the helices into fibrils. The fibrils then assemble into bundles, which further assemble into tissue-specific macro-aggregates (Kadler, Holmes et al., 1996).

The formation of covalent cross-links between fibrils further increases the overall mechanical strength of the fibril.

1.2 Structure of the collagen triple-helix

The triple-helix structure is composed of three polypeptide chains having the sequence -Gly-Xaa-Yaa- repeated, where Xaa and Yaa are frequently the amino acids proline (Pro) and 4-(*R*)-hydroxyproline (Hyp), respectively. Figure 1 shows a recent high-resolution crystal structure of a collagen-like peptide which has a short sequence from type III collagen in the middle (called the T3-785 peptide) (Kramer, Bella et al., 1999). The short sequence of Type III collagen exhibits the chain dimensions that were originally proposed for collagen in 1955 by Ramachandran and Kartha, as well as Rich and Crick, with 3.3 residues per turn (10/3 or 10_7 helical symmetry) (Ramachandran and Kartha, 1955; Rich and Crick, 1955). Each individual chain assumes the secondary structure of a left-handed polyproline-II helix that is wrapped further in a right-handed helical twist, with the N and C-terminal zones having 3.5 residues per turn (7/2 or 7_5 helical symmetry). This latter symmetry has been found in other crystal structures of collagen-like peptides (Pro-Pro-(Gly-Pro-Pro)$_9$-Gly and Pro-Hyp-(Gly-Pro-Hyp)$_9$-Gly) and indicates that a similar symmetry may occur in a region of collagen that would contain an imino acid-rich sequence (Kramer, Bella et al., 1999). Because of its small size, glycine every third amino acid allows the three chains to pack closely together. Also, glycine is at the correct distance and orentation to contribute an amide hydrogen bond to the carbonyl oxygen of the residue in the Xaa position in a neighboring chain, stabilizing the triple-

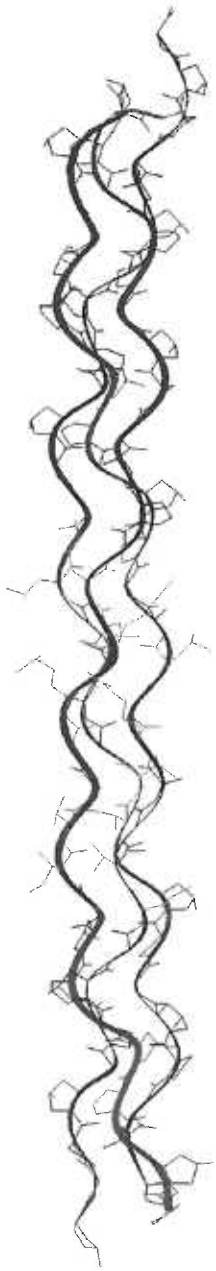


Figure 1. Crystal structure of T3-785. The residues from the Type III sequence are colored magenta. Obtained from the Brookhaven Protein Data Bank (1bkv.ent). The structure was visualized using Insight II (MSI, San Diego, CA.)

helix. Moreover, because of the restriction of the dihedral angle ϕ between the N and Cα bond of the Pro ring (Figure 2), the presence of Pro in the Xaa position and Hyp (or Pro) in the Yaa position further restricts the conformations available to each chain. This in turn enhances the propensity of forming a poly-proline II helix in each chain (Harrington and Rao, 1968). The crystal structure of T3-785 also provides information about the orientation of the various side-chains from the Type III sequence. All of the side chains are pointing out toward the solvent, with very few involved in intra- or inter-molecular contacts other than with solvent. This suggests that within the fibril in which the individual triple-helices are coiled around one another, side-chain side-chain interactions may be important. Indeed, fibrillogenesis, the process of fibril formation, is an entropy-driven process, and a plausible mechanism of fibril formation and stability is the removal of hydrophobic side-chains from the solvent and their subsequent burial within a fibril (Parkinson, Kadler et al., 1995).

1.3 Sweet is stable: the unusual collagen from *R. pachyptila*

The collagens found in invertebrates are quite different from the vertebrate collagens (Engel, 1997). Among this diverse group is the smallest of the collagens (14 nm long), found in the nematocysts of hydra, and the longest of collagens (2400 nm) found in the cuticle collagens of annelids (Har-el, 1993). Although the collagens isolated from the interstitial layers of both the vertebrate and invertebrate collagens are similar, the collagens which form the cuticle layer of annelids do not have a similar counterpart in vertebrates (Murray, 1985). One of the most extensively studied and unusual of the cuticle collagens waas isolated from the deep-sea hydrothermal vent worm *Riftia*

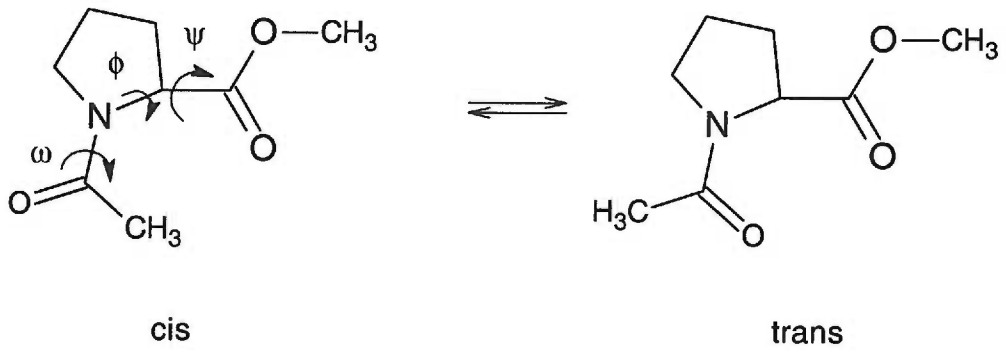


Figure 2: *Cis* and *trans* isomers of *N*-acetyl-proline-*O*-methylester

Pachyptila, illustrated in Figure 3 (Gaill, Wiedemann et al., 1991). This organism lives in hydrothermal vent communities at or below 2600 m beneath the surface of the Pacific, and forms the distinct phylum Vestimentifera (Gaill 1993). Over 90% of the organism is protected by a large tube (0.75 mm to 1.5 m) and is made mostly of chitin. The plume, which functions as a gas-exchange organ, is the only part that is exposed to the surrounding environment, which is hot (~50°C), anoxious, and contains a very high hydrogen sulfide concentration (Gaill, 1993). A thick cuticle protects the plume, and it is from this structure that the cuticle collagen was first isolated (Gaill, Wiedemann et al. 1991). Rather than forming the quarter-staggered fibrils as is found in the interstitial collagens, the cuticle collagen forms a nonstriated plywood-like network of fibrils (Gaill, Herbage et al., 1991). Amino acid analysis of the cuticle collagen revealed that it had a very low content of Pro/Hyp imino acids (~5%), but a very high content of threonine (~18%). Biophysical characterization of the cuticle collagen revealed that the melting temperature (T_m), or the temperature in which half of the molecules are in a triple-helical form, was at 37°C (Gaill, Wiedemann et al., 1991). For most collagens, there is a positive correlation between the melting temperature of the triple-helix and the content of Pro/Hyp residues of a collagen from a particular species (Burjanadze, 1979). However, the *R. pachyptila* collagen, because of the low amount of Pro/Hyp but high T_m, deviated considerably from this trend (Figure 5). Finally, partial sequencing of the cuticle collagen revealed that the threonine residues were glycosylated, mainly with di and tri-saccharides of galactose (Mann, Mechling et al., 1996).

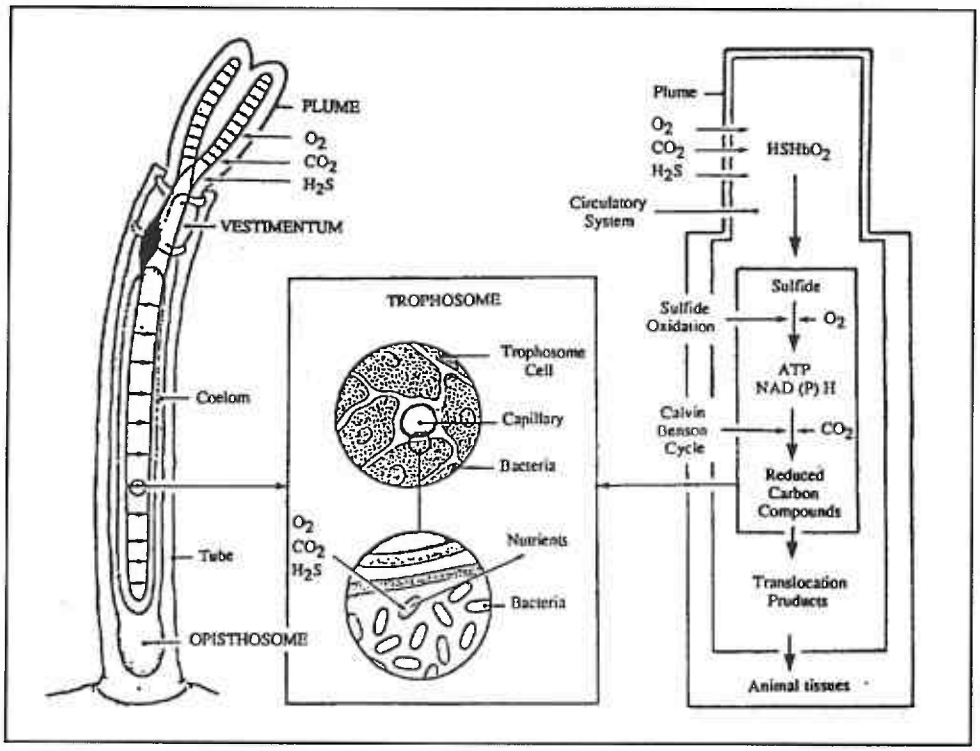


Figure 3: (Top) Schematic of the giant tubeworm *R. pachyptila*. Reprinted without permission from (Gaill, F., 1993). (Bottom) Colony of tubeworms. From http://student.sci.geneseo.edu.

1.4 Similarities between the *R. pachyptila* and annelid cuticle collagens

The observation of di and tri-saccharides of galactose residues in the sequence of *R. pachyptila* was consistent with the carbohydrate pattern observed for the cuticle collagens of the annelids *Lumbricus terrestris* (earthworm) and *Nereis Virens* (clamworm). In the earthworm cuticle collagen, studies by Muir and Lee showed that the carbohydrate content was about 12% D-galactose, and that 70% was present as the di- and tri-saccharide forms, with the remaining 30% being the monosaccharide. These authors investigated the structures of the di- and tri-saccharide forms, and showed that the linkage between the sugars in the tri-saccharide form was *O*-α-D-galactopyranosyl-(1->2)-*O*-α-D-glactopyranosyl-(1-2)-D-galactose (Muir and Lee, 1969). Similar studies by Spiro and Bhoyroo on the carbohydrate composition of both earthworm and clamworm cuticle collagens again showed the presence of di- and tri-saccharides of galactose. However, these authors also found the presence of 6-*O*-α-D-glucuronosyl-D-mannose at about 2% of the total carbohydrate, which was not observed in the cuticle collagen of earthworm (Spiro and Bhoyroo, 1980). Muir and Lee were also the first to show that the carbohydrate linkage to the polypeptide chain of the earthworm cuticle collagen was mediated through either serine of threonine, but the exact residue remained to be determined (Muir and Lee, 1970). Sharma and Tanzer also isolated a short glycopeptide from the cuticle collagen of clamworm after pretreatment of the collagen with collagenase, which has been shown to degrade this collagen only *after* denaturation (Sharman and Tanzer, 1984; Waite, Tanzer et al., 1980). Sequence analysis of the peptide revealed that a monosaccharide of glucuronic acid was attached to threonine (Sharma and Tanzer, 1984).

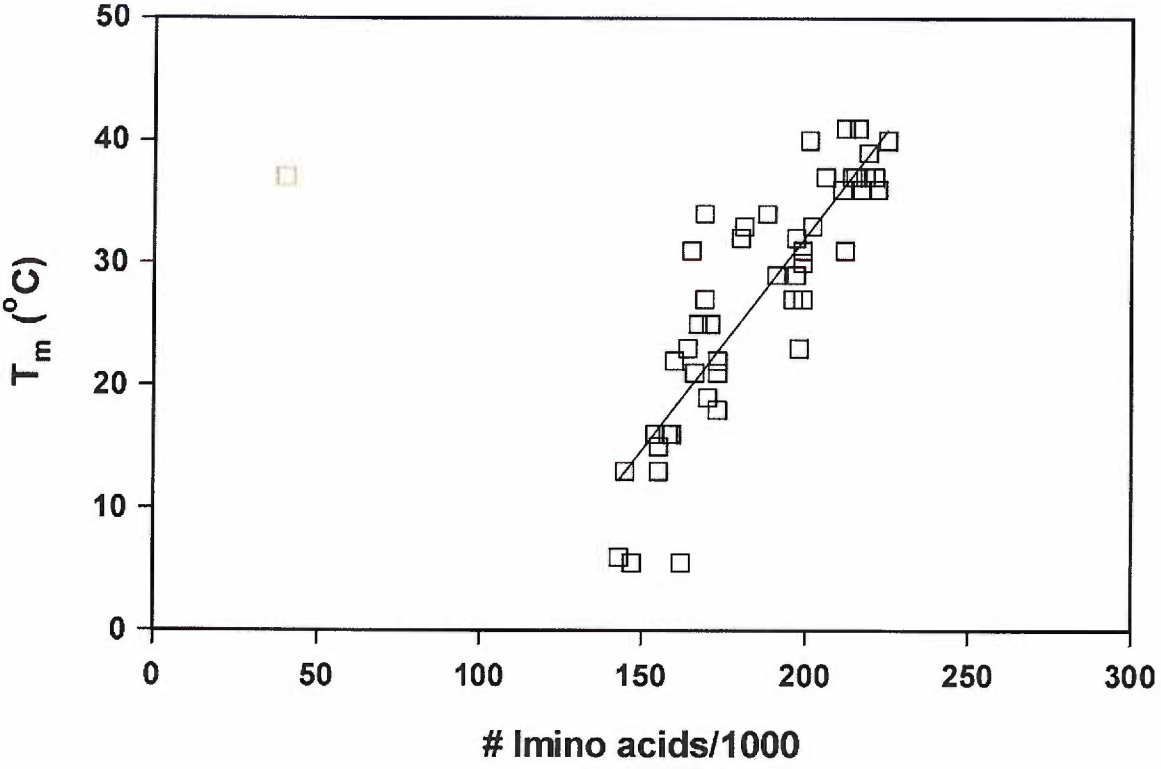


Figure 4. Correlation between the total imino acid content and melting temperature (T_m) of the triple-helix of various collagens (□). Data is from (Burjanadze, 1979). *R. pachyptila* cuticle collagen (□). Data is from (Gaill, et al., 1991).

Only limited sequence information is available for the cuticle collagens of *L. terrestris* and *N. virens*. However, the sequence data that is available for these two collagens show the presence of Hyp in the Xaa position, as is also found for the *R. pachyptila* cuticle collagen. The observation of Hyp in the Xaa position is unique only to the above cuticle collagens, as this is not found in any vertebrate collagen or interstitial collagen from invertebrates. The presence of Hyp in the Xaa position was first demonstrated for the cuticle collagen of *L. terrestris*, with the tripeptide unit Gly-Hyp-Ser representing 4-5% of the total hydroxyproline content (Goldstein and Adams, 1968; Goldstein and Adams, 1970). Hyp in the Xaa position is thought to destabilize the triple-helix, because studies on synthetic peptides of collagen by Inouye and coworkers showed that while the peptides Pro-Pro-(Gly-Pro-Pro)$_9$-Gly and Pro-Hyp-(Gly-Pro-Hyp)$_9$-Gly could form stable triple-helices, the peptide Hyp-Pro-(Gly-Hyp-Pro)$_9$-Gly could not (Inouye, 1982). However, recent studies in our laboratory have shown that while the peptide Ac-(Gly-Pro-Thr)$_{10}$-NH_2 cannot form a triple-helix, the peptide Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 induced the formation of a modestly stable triple-helix, with a T_m of 19°C (Bann and Bächinger, to be submitted). The conclusion from these studies is that while Hyp-Pro is destabilizing, sequences containing Hyp-Thr and possibly Hyp-Yaa (Yaa not including proline) are not. The presence of Hyp in the Xaa position and the presence of glycosylation within the triple-helix suggest that the forces that govern the stabilization of the *L. terrestris* and *N. virens* cuticle collagens may be similar to the forces stabilizing the *R. pac*hyptila cuticle collagen. Although it remains to be determined how these forces are able to achieve a collagen triple-helical structure, perhaps the mechanisms governing

stabilization are similar to those found in vertebrate collagens, where the mechanisms of stabilization are better understood.

B. Towards the mechanism of stabilization of collagen

1.1 Hydroxylation and collagen stability

Gustavson was the first to correlate the stability of a collagen fibril isolated form a particular species to that of the environmental temperature at which the species lived (Gustavson, 1956). Gustavson showed that for species that live at low temperature, for instance cod, the Hyp content and shrinkage temperature (T_s)[1] of the skin collagen is low (53 Hyp residues/per 1000 amino acids, $T_s \sim 15^{\circ}C$). As the temperature of the environment increased, so does the T_s as well as the Hyp content. Since there was no apparent change in the content of Pro residues from one species to another, it was thought that Hyp alone increased the thermal stability of the fibril. Berg and Prockop later showed the influence of hydroxylation on the stability of the individual triple-helix, through a comparison of the melting temperature of procollagen (Types I and III, respectively) purified from cells treated with or without α,α'-dipyridyl (Berg and Prockop, 1973). Procollagen is the precursor form of collagen that has extensions at both the carboxy and amino terminal ends called propeptides; the carboxy-terminal propeptide is essential for the correct association and alignment of the individual chains (Bächinger, 1981). The propeptides are then removed enzymatically by proteases in the extracellular matrix after secretion from the cell. Within a collagen-producing cell, specifically within the endoplasmic reticulum, the hydroxylation of Pro residues occurs post-translationally

[1] The shrinkage temperature was a qualitative measure of the change in the dimensions of the fibril as a function of temperature.

by the enzyme prolyl-4-hydroxylase, and precedes triple-helix formation (Lazarides and Lukens,1971; Lazarides, Lukens et al., 1971). This enzyme is essential, since inhibition of the enzyme by removal of the iron cofactor by the iron chelator α,α-dipyridyl causes the procollagen molecule to accumulate in the rough-endoplasmic reticulum (Jimenez, Dehm et al., 1973; Kivirikko, Myllyla et al., 1989; Beck, Boswell et al., 1996). Procollagen purified from cells treated with α,α-dipyridyl had a T_m that was 15°C lower than the hydroxylated form. Because the molecules were otherwise the same, the only difference being that the Pro residues in one were hydroxylated, it was concluded that hydroxylation stabilizes the individual triple-helix, independent of the fibril. The procollagen molecules extracted from cells treated with α,α-dipyridyl could form triple-helices, but were stable only at lower temperatures (Jimenez, Dehm et al., 1973).

1.2 Plausible mechanisms for the stabilization of collagen by hydroxyproline

Although the aforementioned studies showed that hydroxylation stabilizes the triple-helix, it remained difficult to ascertain the exact mechanism of stabilization, because either hydrogen bonds between the OH of Hyp and water or an unknown intrinsic property of the Hyp residue may affect stability. As mentioned previously, Gustavson first pointed to the correlation between the content of hydroxyproline and thermal stability of various collagens (Gustavson, 1956). However, as the characterization of collagens from different species became available, subsequent analyses revealed a better correlation of T_m to both Pro and Hyp (Harrington and Rao, 1967). Based upon this general correlation between imino acid content and thermal

stability, Harrington and Rao proposed that there should also be a correlation between T_m and the enthalpy and entropy of unfolding through the equation:

$$Tm = \frac{\Delta H^o}{\Delta S^o},$$

assuming that at the T_m the concentration of both folded and unfolded species are equal ($\ln K_{eq} = 0$). These authors found that with an increase in the total imino acid content from the collagens of different species, the enthalpy remained virtually unchanged, however the entropy associated with unfolding *decreased.* This would be expected, since if Pro and Hyp decreased the available number of degrees of freedom of the individual chains, then the greater the number of these imino acids, the more the rigidity would increase. Subsequently, unfolding from a rigid triple-helix to a rigid single chain would not involve a large change in chain dimensions, and thus only a small change in entropy is observed. This analysis assumes that the change in solvent-associated entropy would be same for the folded state as for the unfolded state.

Privalov carried out thermodynamic studies using differential scanning calorimetry (DSC) to determine directly the enthalpy of unfolding (Privalov, Tiktopulo et al., 1979), and proposed an alternative hypothesis. He postulated that with an increase in imino acid content, the enthalpy and entropy of the unfolding transition should *decrease* because imino acids cannot form hydrogen bonds. Privalov also performed hydrogen exchange measurements using both tritium exchange (Yee, et al., 1974) and deuterium exchange (following changes in the infrared spectrum in the amide II region) to investigate the correlation of stability to rigidity of the polypeptide backbone as proposed by Harrington (Harrington and Rao, 1967). Previous tritium exchange measurements by

Yee and Englander (Yee, et al., 1974) had shown that there were 1.5 to 1.7 slowly exchangeable hydrogens per tripeptide unit. The hydrogen exchange experiments performed by Privalov also showed that the number of slowly exchangeable hydrogens was approximately 1.7 per tripeptide unit, however this slowly exchanging class could be differentiated into two sub-classes, a *very* slowly exchanging class and a slowly exchanging class. The observation of the slowly exchanging class was dependent on the temperature of the measurement; for those collagens with a low thermal stability, the exchange would occur rapidly as the temperature increased. The very slowly exchanging class, when extrapolated to zero time, revealed approximately 1.0+/- 0.1 hydrogens per tripeptide unit, in close agreement to that of the Rich and Crick model II. The slowly exchanging class of hydrogens (0.7) were thought to arise from water moleculaes that were hydrogen bonded to the triple-helix. Privalov hypothesized that the rigidity from the increased content of imino acids of the folded state decreased the entropy of the solvent through the formation of hydrogen bonds. This would explain the increase in enthalpy that was observed in the folded, triple-helical state (hydrogen bonds to water) and the increase in entropy upon unfolding (release of water). This latter hypothesis substantiated the probable role of Hyp in stabilizing a water structure; several models had been put forward that suggested that the increase in stability due to Hyp was through bridging by water molecules (Ramachandran, Bansal et al., 1973; Traub, 1974).

Based upon the water-binding hypothesis, the mechanism of stabilization by Hyp would be to anchor water molecules to the triple-helix, which would increase the thermal stability by an increase in the number of stabilizing hydrogen bonds. However, solvent entropy would decrease with the ordering of water molecules around the triple-helix,

disfavoring helix formation. Thus, the enthalpic contribution must compensate for the loss in entropy. Several studies have been carried out to attempt to prove or disprove that water does make such hydrogen bonds and thus does contribute to the enthalpic stability of the triple-helix.

1.3 The role of water in stabilizing the collagen structure

Using Fourier-transform infrared spectroscopy (FT-IR), Lazarev and coworkers have examined the effect of hydration on the stability of collagen-like peptides, in particular the peptide Z-(Gly-Pro-Pro)$_8$-OMe (Z = carbobenzoxy) (Lazarev, Grishkovsky et al., 1992). FT-IR was used in this study to monitor the hydration of the backbone, since the stretching frequencies of both the amide NH and carbonyl CO change significantly upon hydration. Lazarev was able to show that upon hydration from a completely dehydrated state, several changes take place in the amide I vibrational region, in which three vibrational modes, at 1640, 1655, and 1693 cm^{-1} are observed, corresponding to the Gly C=O, Pro C=O (Xaa) and Pro C=O (Yaa), respectively. The Xaa Pro C=O is involved in an inter-molecular hydrogen bond, while the other two carbonyls are pointing out into the solvent. During the intitial stages of hydration, the 1655 cm^{-1} component shifts 6 cm^{-1}, and the Gly-NH stretch, observed in the amide A region at 3340-3370 cm^{-1}, changes by 8 cm^{-1}. The change in the Gly-NH stretch is cooperative as the relative humidity increases (from 10-100%), and suggests that upon hydration the hydrogen bonds (the very slowly exchanging class from Privalov) that stabilize the triple-helix form between the three chains. On the other hand, the outermost carbonyls (Gly C=O and Pro (Yaa) C=O) become hydrated quickly (<8% relative

humidity), and a linear change in the amide I stretching frequencies is observed with increasing hydration. The exchange with deuterium oxide for these carbonyls, unlike that of the Pro in the Xaa, is fairly rapid. These measurements suggested that while water molecules do bind to the polypeptide backbone, due to their ease of exchange they could not contribute enthalpically to the stability of the triple-helix (Lazarev, 1992). Engel and coworkers reached a similar conclusion from studies on the synthetic peptides (Pro-Pro-Gly)$_{10}$ and (Pro-Hyp-Gly)$_{10}$. These authors found that the stability of the collagen molecule was not due to a specific binding of water, since in the *anhydrous* solvents 1,2-propanediol or methanol the stability of the triple helix *increases*, rather than decreases (Engel, Chen et al., 1977).

These conclusions, however, were undermined by the recent crystal structure of the peptide "Gly->Ala", which is (Pro-Hyp-Gly)$_{10}$ with a Gly to Ala substitution in the middle (Figure 6). The structure, determined to 1.9 Å, was the first high-resolution crystal structure of a collagen-like peptide (Bella, Eaton et al., 1994; Bella, Brodsky et al., 1995). The authors found that the molecule was coated with water molecules, and that specific contacts could be found between water and Hyp and/or between water and the peptide backbone. This argued strongly that the function of Hyp was to anchor water molecules to the triple-helix. Another crystal structure has been solved recently to 1.9 Å resolution by Nagarajan and coworkers of (Pro-Hyp-Gly)$_{10}$ (Nagarajan, et al., 1999). In contrast to what was observed with the "Gly->Ala" peptide, these authors found a total of only 17 water molecules, versus the 33 in the "Gly->Ala" crystal structure. These authors suggested that the lack of waters in this crystal structure was consistent with the hypothesis of Raines and coworkers. These authors have postulated that the mechanism

of stabilization by Hyp resides rather in the ability of Hyp to stabilize the trans conformation around the prolyl-peptide bond (Holmgren, Taylor et al. 1998; Holmgren, Bretscher et al. 1999). Using the simple molecules Ac-Pro-OMe, Ac-Hyp-OMe, and Ac-4(*R*)-fluoro-Pro-OMe, these authors showed that with increasing electronegativity (from H to OH to F), the amide I band as observed by FT-IR spectroscopy shifted to lower energies (Eberhart, et al., 1996). Subsequently, the *trans* isomer of these compounds increased (F>OH>H). Substitution of Hyp for 4-fluoro-Pro in the peptide (Pro-Hyp-Gly)$_{10}$ resulted in a hyperstable collagen, with a T_m of 91°C. These authors presupposed that since a carbon-fluorine bond cannot form hydrogen bonds, water molecules are not important. However, dipole-dipole interactions between the fluorine and water could not be ruled out, and recent studies on small fluorinated compounds suggest that, while weaker (<2.5 kcal/mol), O-H---F-C hydrogen bonds do exist (Barbarich, et al., 1999; Caminati, et al., 1999; Snyder, et al., 2000).

Although significant progress has been made recently on the role of Hyp in stabilizing the collagen structure, the exact mechanism remains to be determined. However, for glycosylation, understanding the mechanism of stabilizaton may be less complicated. I will review some of the information that is available from researchers who have asked the question, "How does glycosylation affect stability?"

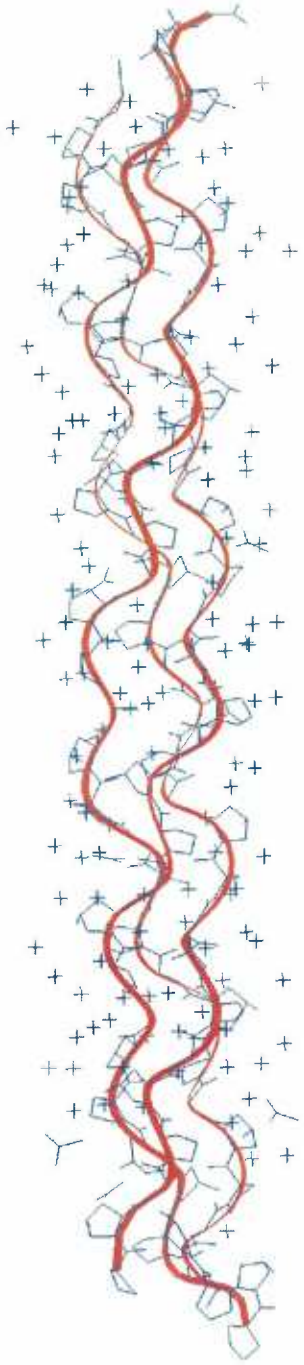


Figure 5: Crystal Structure of the Gly->Ala peptide (Bella, et al., 1995). Obtained from the Brookhaven Protein Data Bank (2cgd.ent).

Glycosylation and stability

1.1 Effect of glycosylation on peptide backbone conformation

The affect of glycosylation on the conformation of the polypeptide backbone has been studied by a number of workers, and in most instances the affect is to increase the rigidity of the polypeptide backbone. How this is achieved is not well understood. The carbohydrate may influence the solvation properties of the backbone, may interact with the polypeptide backbone itself through hydrogen bonds, and, as has been most often discussed, cause the backbone to assume only a limited number of conformations simply for steric reasons. There are two main types of glycosylation, *O*- and *N*-linked, where *O*-linked represents those carbohydrates which can be attached to serine, threonine, and hydroxyproline residues and *N*-linked represents those carbohydrates which can be attached to asparagine residues. Proteins which exhibit *N*-linked glycosylation include many cell-surface receptors, IgG and IgM, extracellular matrix proteins including collagen, and many viral surface glycoproteins (Drickamer and Taylor 1998). *O*-linked proteins also include many cell-surface glycoproteins, including the submaxillary mucous glycoproteins (mucins), the low-density lipoprotein receptor, and the decay accelerating factor (Jentoft 1990). In terms of structure, the most extensively studied glycoproteins in this class are the mucins, in particular the ovine submaxillary mucin (OSM). This large ($>10^6$ Da) protein is heavily glycosylated, almost solely with the disaccharide α-NeuNAc(2-6)α-GalNAc-O-Ser/Thr (Shogren, Gerken et al. 1989). This latter feature makes the OSM tractable in terms of studying the role of carbohydrate on the structure of the protein, as the terminal sialic acid may be removed with neuraminidase or chemically with trifluromethanesulfonic acid without removing the GalNac, generating a partial

asialo form of the protein. Subsequent removal of the GalNac may be achieved with the protein *N*-acetyl-galactosidase or chemically, again without damaging the peptide core.

Early work by Gottschalk demonstrated the effect of carbohydrate on the viscosity of OSM, showing that when the terminal sialic acid is cleaved with the enzyme neuraminidase from *Vibria cholera*, the viscosity was reduced (Gottschalk, 1960). Gottschalk suggested that the negative charge on the carbohydrate may influence the viscoelastic properties, and showed also that the viscosity dropped on going from pH 4.3 to 1.8, and that this was completely reversible. It was concluded that the negative charge on sialic acid must have an important role in maintaining the mucin in a rod-like structure. Interestingly, it was suggested in a communication from Linus Pauling that the carbohydrates may assume a secondary helix around the protein core, which itself assumes an α-helix, "loosely joined by proline residues" (Gottschalk, 1960).

In a striking example of the effect of deglycosylation of the structure of OSM, Rose and colleagues applied electron microscopy and ultracentrifugation to measure the dimensional changes before and after enzymatic removal of the carbohydrates (Rose, Voter et al. 1984). In the native state, highly elongated, filamentous structures could be observed. Conversion of the native glycoprotein to the asialo form with neuraminidase resulted in a less extended structure, and complete removal of carbohydrate with α-*N*-acetyl-D-galactosaminidase yielded a compact structure. Although the observed changes could not exclude the presence of a dimer that could contribute to the filamentous structure, the presence of carbohydrate clearly had a significant impact on the chain dimensions, corroborating the findings of Gottschalk. However, the terminal sialic acid did not seem to change the dimensions significantly compared to the asialo form,

suggesting that the single GalNAc-Thr/Ser, rather than the negative charge on the sialic acid, is the dominant factor influencing the viscoelastic properties of the mucins.

Further evidence that the monomeric chain rather than the dimer could assume the highly extended conformation was from the more recent work of Shogren and coworkers, through the combined use of light-scattering and ^{13}C-NMR measurements (Gerken, Butenhof et al. 1989; Shogren, Gerken et al. 1989). These authors used as a solvent for their studies 5.0 M guanidine hydrochloride (Gdn-HCl), which not only would disrupt the coulombic interactions between chains (preventing dimer formation), but would also disrupt hydrogen-bonding from the carbohydrate to the peptide backbone. These authors found that the glycosylated OSM remained highly extended under these conditions, and that this extended shape persists after removal of the terminal sialic acid, either enzymatically or by chemical treatment. This again suggested that the presence of a negative charge is probably not the main cause of the viscous behavior, and points more toward the *N*-acetyl-galactose as the primary factor influencing the extended shape.

CD analyses of the secondary structure of OSM suggested that the structure was similar to that of the antifreeze glycoprotein from Antarctic cod (Shogren, Gerken et al. 1989). This protein is also highly glycosylated with a primary structure composed of the repeat -Ala-Ala-(Galβ(1-3)GalNAcα1-Thr- (Lin, Duman et al. 1972). The CD spectrum of the antifreeze glycoprotein exhibits a poly-L-proline II like secondary structure similar to that of collagen, with a positive ellipticity at 218 nm $[\theta]_{218}$ = 5000 deg cm^2 dmol^{-1}, and a smaller negative ellipticity at 197 nm $[\theta]_{197}$ = -20000 deg cm^2 dmol^{-1} . Recent evidence from high resolution NMR (Lane, 1998) has confirmed that the conformation of the antifreeze glycoprotein is that of a polyproline II –helix, and not a poly-γ/β-turn repeat

motif, as was also suggested (Drewes and Rowlen 1993). The far-UV CD spectrum of OSM is similar to that of the antifreeze glycoprotein, however the magnitude of the ellipticities at 197 and 218 are smaller; $[\theta]_{197}$ = -13000 deg cm^2 $dmol^{-1}$ and $[\theta]_{218}$ = <100 deg cm^2 $dmol^{-1}$. The magnitudes are also dependent on the state of glycosylation, with the ellipticity at 218 nm decreasing slightly in the asialo form, and in the apo form the CD spectrum more closely resembles that of a random coil. The observed changes in the CD spectrum upon deglycosylation suggest that the native structure is likely a poly-proline-II helix, which undergoes a conformational change to a less extended, random coil-like structure.

1.2 Do carbohydrates act sterically to limit conformational space?

The currently accepted explanation for the ability to achieve the rod-like structure found for the mucins and the antifreeze glycopeptide, as well as for other model peptides, and the explanation given for the changes observed in general upon glycosylation, is that the carbohydrate decreases the number of conformations available to the polypeptide backbone for steric reasons (Shogren, Gerken et al. 1989; O'Conner and Imperiali 1998). This steric effect, although suggested from the studies of OSM, has only recently begun to be understood from a higher resolution perspective. Imperiali and coworkers, through the synthesis of a series of short glycopeptides corresponding to the influenza virus coat protein hemagglutinin, have shown from both fluorescence resonance energy transfer (FRET) studies, as well as 2D-NMR, that glycosylation induces the formation of a beta turn. NMR analyses of nuclear Overhauser effect (NOE) intensities between protons within the peptide provided strong evidence for the presence of a type-I beta turn, and

this correlated well with the structure formed in the native protein. Importantly, only very weak NOEs could be observed between protons of the carbohydrate and the peptide backbone. The presence of an NOE between the N-acetyl group of the carbohydrate and the peptide backbone had been observed previously in a model mucin peptide, suggesting the presence of an intramolecular hydrogen bond that would hold the protons in question at close proximity to each other. However, this NOE was only present in the solvent DMSO, and was absent when measured in aqueous solution (Mimura, Inoue et al. 1989). In order to explain the influence of the carbohydrate, and specifically that of the *N*-acetyl group on the flexibility of the peptide, it was hypothesized that the observed changes result from either steric effects or from an influence by the carbohydrate on the local water structure. In a subsequent study, it was shown that the nature of the carbohydrate itself was critical in determining the dynamics of the peptide; for instance, a peptide containing cellobiose (D-Glc-[β(1-4)]-β-D-Glc-Asn) was less ordered than a peptide containing chitobiose (D-Glc*N*Ac-[β(1-4)]-β-D-Glc*N*Ac-Asn). Although an OH to N-CO-CH3 is a relatively minor change, this result pointed more towards a steric effect of the carbohydrate rather than a change in the solvation properties of the backbone. Indeed, it was shown that with the N-acetyl group present, the dihedral angles of the Asn residue itself and that of the residue prior to the Asn (Pro) were more constrained (O'Conner and Imperiali 1998). However, distinguishing between the two possibilities is difficult, as has been observed with the collagens in relation to the role of Hyp and distinguishing if water or sterics are playing the dominant role.

In a recent NMR study examining the structure of a fragment of the mucin CD43, Danishefsky and coworkers synthesized the peptide Ac-Ser-Thr-Thr-Ala-Val-OH and

compared the effect of the anomeric linkage (α vs. β) of the carbohydrate to the structure of the peptide (Live, Williams et al. 1999). In natural mucins, the core carbohydrate attached to the Ser/Thr residues is α–GalNAc, and this produced a dramatic change in the amide region of the 1-D 1H-NMR spectrum compared to the non-glycopeptide. In particular, the exchange rate of the α-GalNAc acetyl amide hydrogen was significantly slower than that of the β-GalNAc (12 hours versus minutes). The structure of the peptide glycosylated with α–GalNAc was highly elongated and stable, however glycosylation with the β-linked anomer produced almost no changes in the amide region compared to the non-glycosylated peptide. Since previous studies have shown that *O*-glycosylation at a single site induces the formation of a beta turn, the presence of multiple glycosylation sites must force the unkinking of the turn, favoring an extended structure. No reasoning was provided for the observed differences in α versus β linkage .

In terms of the studies on the collagen glycopeptides reported here, the aforementioned studies suggest that the effect of multiple, contiguous residues of Thr(β-Gal) may result in the formation of either an extended structure or the formation a multiple-turn-repeat motif. This would decrease the number of available conformations, and may funnel the polypeptide into a particular fold (conducive to forming a poly-L-proline II helix). This would have not only an influence on the stability of the triple-helix, but also on the rate of triple-helix formation, as the sampling of conformational space would be limited to the already correct fold. Perhaps, however, in the case of the glycopeptides and in the *R. pachyptila* cuticle collagen the OH groups of the galactose are occupying the sites normally occupied by water. In any event, the ability to decipher the role of carbohydrate on the stability of the collagen-like peptide studies here, and an

ultimate understanding of the forces governing the stability of the cuticle collagens of invertegrates, is complicated by the contributions from the single chain conformation and the contribution from the final folded conformation.

Chapter 2

Sweet is Stable: Glycosylation Stabilizes Collagen

James G. Bann[1] , David H. Peyton[3], and Hans Peter Bächinger[1,2]

[1]Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, 3108 SW Sam Jackson Park Road, Portland, Oregon, 97201; [2]Research Department, Shriners Hospital for Children, 3101 SW Sam Jackson Park Road, Portland, Oregon, 97201; [3]Department of Chemistry, Portland State University, Portland, Oregon, 97267

Abstract

For most collagens, the melting temperature (T_m) of the triple-helical structure of collagen correlates with the total content of proline (Pro) and 4-*trans*-hydroxyproline (Hyp) in the Xaa and Yaa positions of the -Gly-Xaa-Yaa- triplet repeat. The cuticle collagen of the deep-sea hydrothermal vent worm *Riftia pachyptila*, despite a very low content of Pro and Hyp, has a relatively high thermal stability. Rather than Hyp occupying the Yaa position, as is normally found in mammalian collagens, this position is occupied by threonine which is *O*-glycosylated. We compare the triple-helix forming propensities in water of two model peptides, Ac-(Gly-Pro-Thr)$_{10}$-NH_2 and Ac-(Gly-Pro-Thr(Galβ))$_{10}$-NH_2, and show that a collagen triple-helix structure is only achieved after glycosylation of threonine. Thus, we show for the first time that glycosylation is required for the formation of a stable tertiary structure and that this modification represents an alternative way of stabilizing the collagen triple-helix that is independent of the presence of Hyp.

Introduction

The collagens of invertebrates have been shown to exhibit unusual and interesting properties [1,2]. An example of this is from the deep-sea hydrothermal vent worm *Riftia pachyptila*. This organism lives under extreme conditions (high pressure, low oxygen, and steep temperature gradients), but is protected from its environment by a thick cuticle [3,4]. The *R. pachyptila* cuticle is mainly composed of a collagen that forms a plywood-like network of fibrils, and exhibits a unique amino acid composition. While having a low content of proline (Pro) and 4-*trans*-hydroxyproline (Hyp) (~5%), this collagen has a very high content of threonine (Thr) residues (>18%) [5]. Biochemical analyses and partial sequencing of the cuticle collagen's triple-helical region revealed that Thr occupied the Yaa position of the -Gly-Xaa-Yaa- triplet repeat and the Thr residues were glycosylated, mainly with di- and tri-saccharides of galactose [6]. In most collagens, there is a positive correlation between the total content of Pro and Hyp imino acids and the melting temperature (T_m) of the triple-helix [7,8]. However, the *R. pachyptila* cuticle collagen, with an observed T_m of 37°C, deviates considerably from this trend (Figure 1).

The finding of a di- and tri- pattern of galactose carbohydrate units is consistent with the pattern found in the cuticle collagens of both *Lumbricus terrestris* (earthworm) and *Nereis virens* (clamworm) [9,10]. Also consistent with the annelid cuticle collagens is the position of Hyp in the -Gly-Xaa-Yaa- triplet repeat, occurring only in the Xaa position [11,12]. In vertebrates, Hyp only occurs in the Yaa position, and Hyp in the Xaa position is thought to either not contribute to or to destabilize the thermal stability of the

triple-helix [11,13]. For example, although the amount of Hyp in the *L. terrestris* collagen is very high (~16-18%), this collagen has a low thermal stability (T_m = 22°C) [14]. Therefore, in *R. pachyptila* glycosylation of Thr must somehow compensate for the lack of Hyp in the Yaa in order to achieve the high-observed thermal stability.

In these studies our main goal was to develop model collagen-like peptides of the *R. pachyptila* cuticle collagen that would allow us to determine the influence of glycosylation on the stability of the triple-helix. In addition, *O*-glycosylation of proteins has been shown to impart an enhanced resistance to enzymatic degradation, and both the *L. terrestris* and *N. virens* cuticle collagens in the native state are resistant to degradation by clostridial collagenase [11,15]. Thus, these collagen-like peptides may also exhibit an enhanced resistance to proteolysis and could lead to the development of novel biomaterials [16].

Materials and Methods

Peptide synthesis. The synthesis of N^{α}-9-fluorenylmethoxycarbonyl (FMOC)-*O*-(2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl)-Thr-OH was done according to [17-19], purified by preparative high-performance liquid chromatography, and characterized by electrospray mass spectrometry, ^{1}H-NMR and ^{13}C-NMR. The glycopeptide and peptide were synthesized on a PAL-PEG-PS resin (Perseptive Biosystems, 0.16 mmol/g) using FMOC dipeptides of Gly-Pro-OH (Novabiochem, 4.0 eq.), FMOC-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-Thr-OH (1.3-2.0 eq.) and FMOC-Thr-OH (4.0 eq.) and HATU (O-(7-azabenzotriazol-1-yl)-1.1.3.3-tetramethyluronium hexafluorophosphate,

Perseptive Biosystems) (4.0 eq.) mediated amino acid couplings on a Milligen 9050 peptide synthesizer. The peptides were cleaved from the resin and purified by semi-preparative high performance liquid chromatography. The glycopeptide was subsequently treated with 6 mM sodium methoxide in methanol overnight to remove the acetyl groups, and again purified by semi-preparative HPLC. The peptides were verified by analytical HPLC, electrospray mass spectrometry, ^{1}H-NMR and amino acid analysis.

Circular dichroism measurements. Circular dichroism spectra were recorded on a Aviv 202 spectropolarimeter using a thermostatted cell of 1 mm pathlength (Hellma). All measurements were performed in water. Concentrations of peptides were determined by amino acid analysis. The thermal transitions were recorded from 5 to 70 °C at 223 nm by raising the temperature at a rate of 12 °C/h.

Analytical ultracentrifugation. Sedimentation equilibrium measurements were carried out on a Beckman Model E analytical ultracentrifuge, using double-sector cells. Peptides were run at two speeds (26,000 and 34,000 rpm), at three concentrations (0.5, 0.35, and 0.2 mg/ml) in water, monitoring at 230 nm at 25.0 °C. Data were fitted using Scientist® (Micromath, Salt Lake City, Utah) with a non-linear least squares algorithm, assuming a partial specific volume of 0.667 cm^3/g based upon the amino acid sequence and the partial specific volume of galactose.

NMR measurements. NMR spectra were recorded on a Bruker AMX-400 spectrometer, operating at 400.14 MHz, using a dedicated 5 mm 1H probe. The 90° pulse width was 8

msec; a low-power 2 s presaturation pulse was applied to suppress the H2O resonance. The spectra were recorded as 16384 points for the 1D spectra, and as 1024 x 512 data point sets for the 2D spectra. The NOESY data were collected with TPPI [20] in the indirect dimension, a mixing time of 150 ms, and a total recording time of about 17 hours. ROESY data [21,22] were collected similarly, but using a 200 ms mixing time. The data were processed with Swan-MR software [23] to 1024 x 1024 real data point sets after application of a 65°-phase-shifted sin2 function and Fourier transformation for the 2D spectra; baselines were straightened with polynomials as needed. Spectra were referenced to 0 ppm via internal 2,2-dimethylsilapentane-5-sulfonate.

Results and Discussion

As an initial step, we synthesized the peptides Ac-(Gly-Pro-Thr)$_{10}$-NH_2, and Ac-(Gly-Pro-Thr(Galβ)$_{10}$-NH_2. Although the peptides (Pro-Pro-Gly)$_{10}$ and (Pro-Hyp-Gly)$_{10}$ have been well characterized as forming stable collagen triple-helices, we chose to use acetyl/amide-terminated peptides since the removal of end charges increases the thermal stability of the triple-helix and also provides a more accurate model of an internal triple-helix motif [24]. The glycopeptide was prepared by first synthesizing *O*-(2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl)-N^{α}-9-fluorenylmethoxycarbonyl (FMOC)-Thr-OH, and then using this as a building block in the solid phase peptide synthesis [17-19]. Monosaccharides of galactose are found in the *R. pachyptila* and annelid collagens, however the anomeric linkage (α vs. β) to the Thr is not known [9,10]. Thus, the facile

synthesis and purification of the β-anomer and the lack of any contaminating α–anomer provided a straightforward starting point for our investigations.

The far-ultraviolet circular dichroism (CD) spectra of collagen molecules typically show a negative ellipticity (θ) around 198-200 nm and a maximum ellipticity near 220-225 nm, indicating the presence of a polyproline II helix [25,26]. In water at 5 °C, the glycopeptide showed these CD spectral features, as is seen in Figure 2(a). In contrast, a weaker negative ellipticity (-2900 deg cm^2 $dmol^{-1}$) around 200 nm and no positive ellipticity in the 220-225 nm range is observed for Ac-(Gly-Pro-Thr)$_{10}$-NH_2, indicating a mostly random-coil conformation. The ellipticity at 223 nm of both peptides was monitored as a function of temperature, and this is shown in Figure 2(b). The melting curve of the glycopeptide in water showed the presence of a cooperative transition, with a T_m of 41 °C. In contrast, no transition was observed for the peptide Ac-(Gly-Pro-Thr)$_{10}$-NH_2, again consistent with the inability of this peptide and a similar peptide (H-(Gly-Pro-Thr)$_{10}$-Gly-Pro-Cys-Cys-OH) to form a triple-helix in aqueous buffer [6]. Sedimentation equilibrium measurements gave a weight-average molecular weight of 12,280 +/- 200 Da, which is close to that predicted for three associated independent test of a triple-helical conformation is the presence of a unique set of inter-chain nuclear Overhauser effects (NOEs) arising from the close packing of the triple-helix [27,28]. NOESY experiments of the glycopeptide revealed a unique set of NOEs (Figure 3a), that disappear upon melting of the peptide as observed in the ROESY spectrum (b).

These results show that glycosylation of threonine is required for the formation of a triple-helix with respect to the non-glycosylated peptide, and thus must be required to

achieve the high thermal stability observed for the *R. pachyptila* cuticle collagen. There are several possible mechanisms, which may account for the stabilization. One possible mechanism is that the addition of galactose may restrict the conformational space available to the polypeptide backbone in a way similar to that of proline, in which the steric restrictions imposed by the pyrrolidine ring stabilizes the conformation of the polyproline II helix [29]. A restriction of conformational space has been observed in other proteins and peptides that exhibit *O*-linked glycosylation [30-34], and in mucins, the high-degree of *O*-glycosylation results in a rigid, expanded structure [35]. Therefore, glycosylation would limit the number of conformations sampled in the unfolded state, and thus *destabilize* this state.

Glycosylation may also influence the stability of the native state, through hydrogen bonds from the sugar to the polypeptide backbone. Crystallographic evidence has suggested that the higher thermal stability of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ vs. $(Pro\text{-}Pro\text{-}Gly)_{10}$ is due to the ability of the γ-OH group of Hyp to bridge water molecules to the amide carbonyls of the peptide backbone [36,37]. Another explanation for how Hyp stabilizes the triple-helix has recently been put forward [38,39]. These authors have substituted Hyp for 4-*trans*-fluoroproline, which does not form hydrogen bonds, and are able to achieve a very stable collagen-like peptide. The conclusion from these experiments is that Hyp, rather than stabilizing hydrogen bonds, stabilizes the *trans* conformation of the imide peptide bond through a stereoelectronic inductive effect; because all peptide bonds in the triple-helix are *trans*, there is a cumulative effect on thermal stability.chains (12,699 Da). Finally, an

In our case, it is also possible that the galactose hydroxyls will mediate interactions with water or that structural waters may be replaced by the sugar hydroxyls. However, recent NH exchange measurements by ^{1}H-NMR indicate that both the Gly NH *and* the Thr NH are well protected from exchange with deuterium in the triple-helix. Since the Thr NH is projected to be solvent exposed [40], this suggests that the sugar prevents access of water to the peptide backbone. Further studies on whether glycosylation influences the *cis-trans* isomer ratio of proline or the conformation of the single chains will help to clarify the mechanism of stabilization by glycosylation.

Acknowledgements

We are grateful to Herbert Waite (Univ. California, Santa Barbara), and Jan Kihlberg (Umea University, Sweden) for their comments on the manuscript. This work was supported by funds from the Shriners Hospitals for Children.

References

[1] Coyne, K.J., Qin, X.X. and Waite, J.H. (1997) Science 277, 1830-1832.

[2] Engel, J. (1997) Science 277, 1785-1786.

[3] Gaill, F., Herbage, D. and Lepescheux, L. (1991) Matrix 11, 197-205.

[4] Gaill, F. (1993) Faseb J 7, 558-565.

[5] Gaill, F., Mann, K., Wiedemann, H., Engel, J. and Timpl, R. (1995) J Mol Biol 246, 284-294.

[6] Mann, K., Mechling, D.E., Bachinger, H.P., Eckerskorn, C., Gaill, F. and Timpl, R. (1996) J Mol Biol 261, 255-266.

[7] Burjanadze, T.V. (1979) Biopolymers 18, 931-938.

[8] Privalov, P.L. (1982) Adv Protein Chem 35, 1-104.

[9] Muir, L. and Lee, Y.C. (1970) J Biol Chem 245, 502-509.

[10] Spiro, R.G. and Bhoyroo, V.D. (1980) J Biol Chem 255, 5347-5354.

[11] Goldstein, A. and Adams, E. (1970) J Biol Chem 245, 5478-5483.

[12] Sharma, Y.D. and Tanzer, M.L. (1984) Anal Biochem 141, 205-212.

[13] Inouye, K., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S. and Prockop, D.J. (1982) Arch Biochem Biophys 219, 198-203.

[14] Josse, J. and Harrington, W.F. (1964) J Mol Biol 9, 269-287.

[15] Waite, J.H., Tanzer, M.L. and Merkel, J.R. (1980) J Biol Chem 255, 3596-3599.

[16] Ramshaw, J.A., Werkmeister, J.A. and Glattauer, V. (1996) Biotechnology and Genetic Engineering Review 13, 335-382.

[17] Kihlberg, J., Elofsson, M. and Salvador, L.A. (1997) Methods Enzymol 289, 221-245.

[18] Salvador, L.A., Elofson, M. and Kihlberg, J. (1995) Tetrahedron 51, 5643-5656.

[19] Elofson, M., Walse, B. and Kihlberg, J. (1991) Tetrahedron Letters 32, 7613-7617.

[20] Marion, D. and Wüthrich, K. (1983) Biochem. Biophys. Res. Commun. 113, 967-974.

[21] Bax, A. and Davis, D.G. (1985) J. Magn. Reson. 65, 355-360.

[22] Bothner-By, A.A., Stephens, R.L., Lee, J., Warren, C.D. and Jeanloz, R.W. (1984) J. Am. Chem. Soc. 106, 811-813.

[23] Balacco, G. (1994) J. Chem. Inf. Comput. Sci. 34, 1235-1241.

[24] Feng, Y., Melacini, G., Taulane, J.P. and Goodman, M. (1996) Biopolymers 39, 859-872.

[25] Engel, J. (1968) Das Leder 19, 21-26.

[26] Brown, F.R.d., Di Corato, A., Lorenzi, G.P. and Blout, E.R. (1972) J Mol Biol 63, 85-99.

[27] Li, M.H., Fan, P., Brodsky, B. and Baum, J. (1993) Biochemistry 32, 7377-7387.

[28] Feng, Y., Melacini, G., Taulane, J.P. and Goodman, M. (1996) J. Am. Chem. Soc. 118, 10351-10358.

[29] Harrington, W.F. and Rao, N.V. (1967) in: Conformation of Biopolymers, pp. 513-531 (Ramachandran, G.N., Ed.) Academic, New York.

[30] O'Conner, S.E. and Imperiali, B. (1998) Chem Biol 5, 427-437.

[31] Live, D.H. et al. (1999) Proc Natl Acad Sci U S A 96, 3489-3493.

[32] Wu, W., Pasternack, L., Huang, D.-H., Koeller, K.M., Lin, C.-C., Oliver Seitz, O. and Wong, C.-H. (1999) J. Am. Chem. Soc. 121, 121-127.

[33] Biondi, L., Filira, F., Gobbo, M., Pavin, E. and Rocchi, R. (1998) J Pept Sci 4, 58-71.

[34] Andreotti, A.H. and Kahne, D. (1993) J. Am. Chem. Soc. 115, 3352-3353.

[35] Jentoft, N. (1990) Trends Biochem Sci 15, 291-294.

[36] Bella, J., Eaton, M., Brodsky, B. and Berman, H.M. (1994) Science 266, 75-81.

[37] Bella, J., Brodsky, B. and Berman, H.M. (1995) Structure 3, 893-906.

[38] Holmgren, S.K., Taylor, K.M., Bretscher, L.E. and Raines, R.T. (1998) Nature 392, 666-667.

[39] Holmgren, S.K., Bretscher, L.Y., Taylor, K.M. and Raines, R.T. (1999) Chem. Biol. 6, 63-70.

[40] Kramer, R.Z., Bella, J., Mayville, P., Brodsky, B. and Berman, H.M. (1999) Nat Struct Biol 6, 454-457.

Figure legends

Figure 1. Correlation between the total imino acid content and melting temperature (T_m) of the triple-helix of various collagens (□). Data taken from ([7,8]). *R. pachyptila* cuticle collagen (□). Data is from ([5]).

Figure 2. (*a*) Circular dichroism spectra of collagen-like peptides. Ac-(Gly-Pro-Thr(β-D-Gal))$_{10}$-NH_2 (blue curve) and Ac-(Gly-Pro-Thr)$_{10}$-NH_2 (black curve) at 5 °C. The red curve represents Ac-(Gly-Pro-Thr(β-D-Gal))$_{10}$-NH_2 at 70 °C. Note the similarity in shape between the glycopeptide at 70°C and the peptide. (*b*) Dependence of ellipticity [θ] at 223 nm as a function of temperature (°C). Ac-(Gly-Pro-Thr(β-D-Gal))$_{10}$-NH_2 (blue curve) and Ac-(Gly-Pro-Thr)$_{10}$-NH_2 (black curve).

Figure 3. (*a*) 1D reference and 2D NOESY ^{1}H NMR spectroscopy (18) of Ac–(Gly–Pro-Thr(β-Gal))$_{10}$–NH_2 (1.8 mM) at 25°C and pH 4.2 (±0.1). NOE cross-peaks from the Thr-γCH3 at 1.47 ppm peak to protons of the Pro aliphatic region, the Gal ring and the amide protons of the Thr and Gly indicate a strong packing of the triple-helix. (*b*) 1D reference and 2D ROESY ^{1}H NMR spectroscopy (18) of Ac–(Gly–Pro-Thr(β-Gal))$_{10}$–NH_2 from (*a*) at 60°C. Note the loss of cross-peaks from the 1.47 ppm peak.

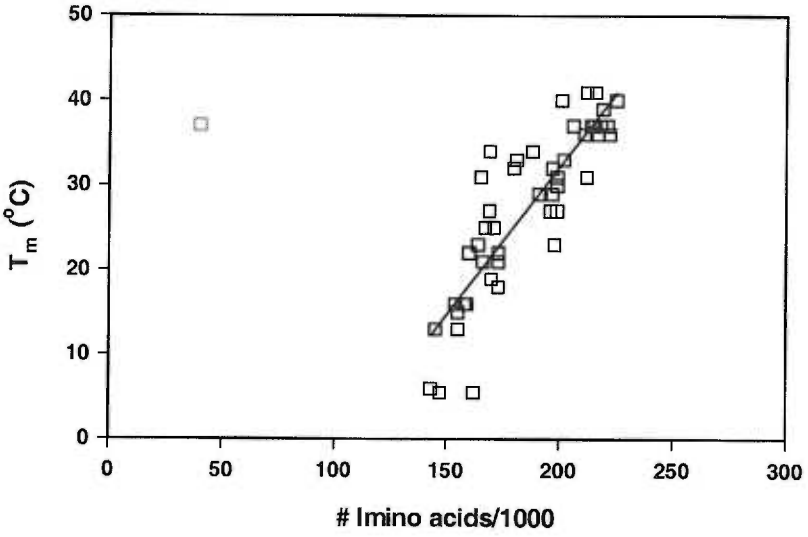


Figure 1

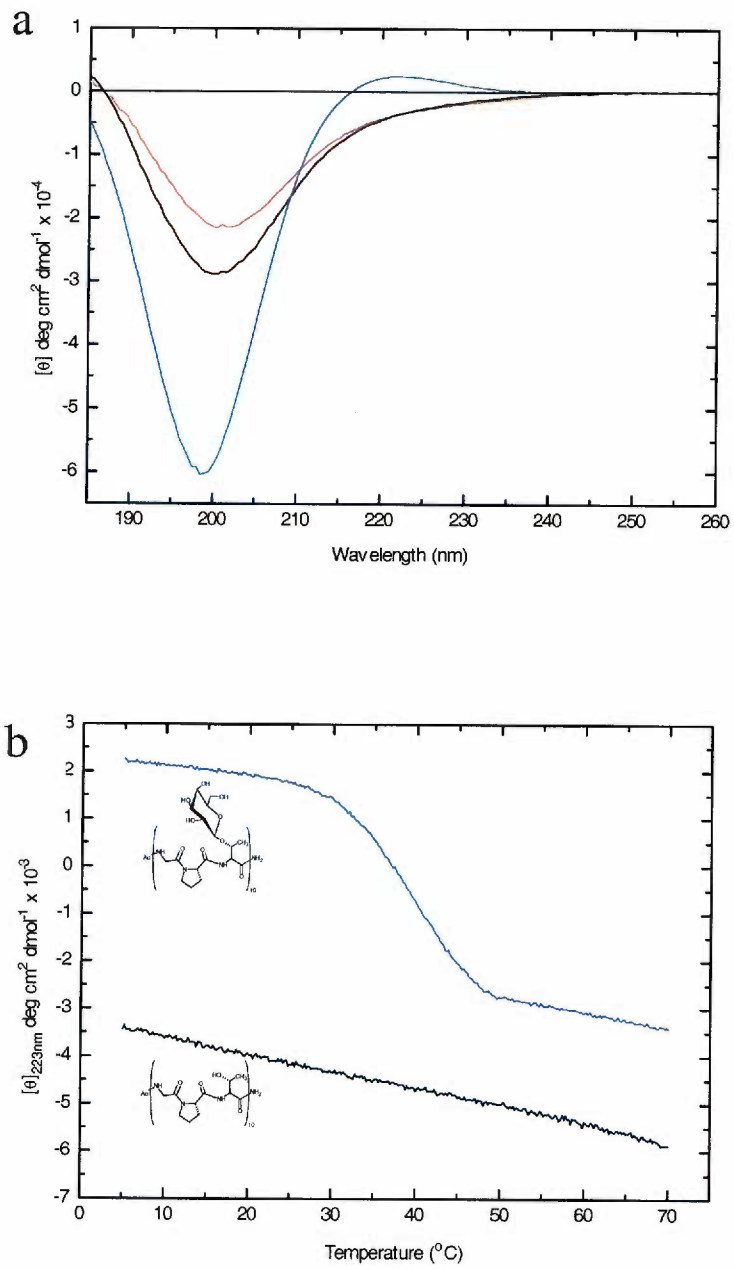

a
[θ] deg cm² dmol⁻¹ x 10⁻⁴
Wavelength (nm)
b
[θ]223nm deg cm² dmol⁻¹ x 10⁻³
Temperature (°C)

Figure 2

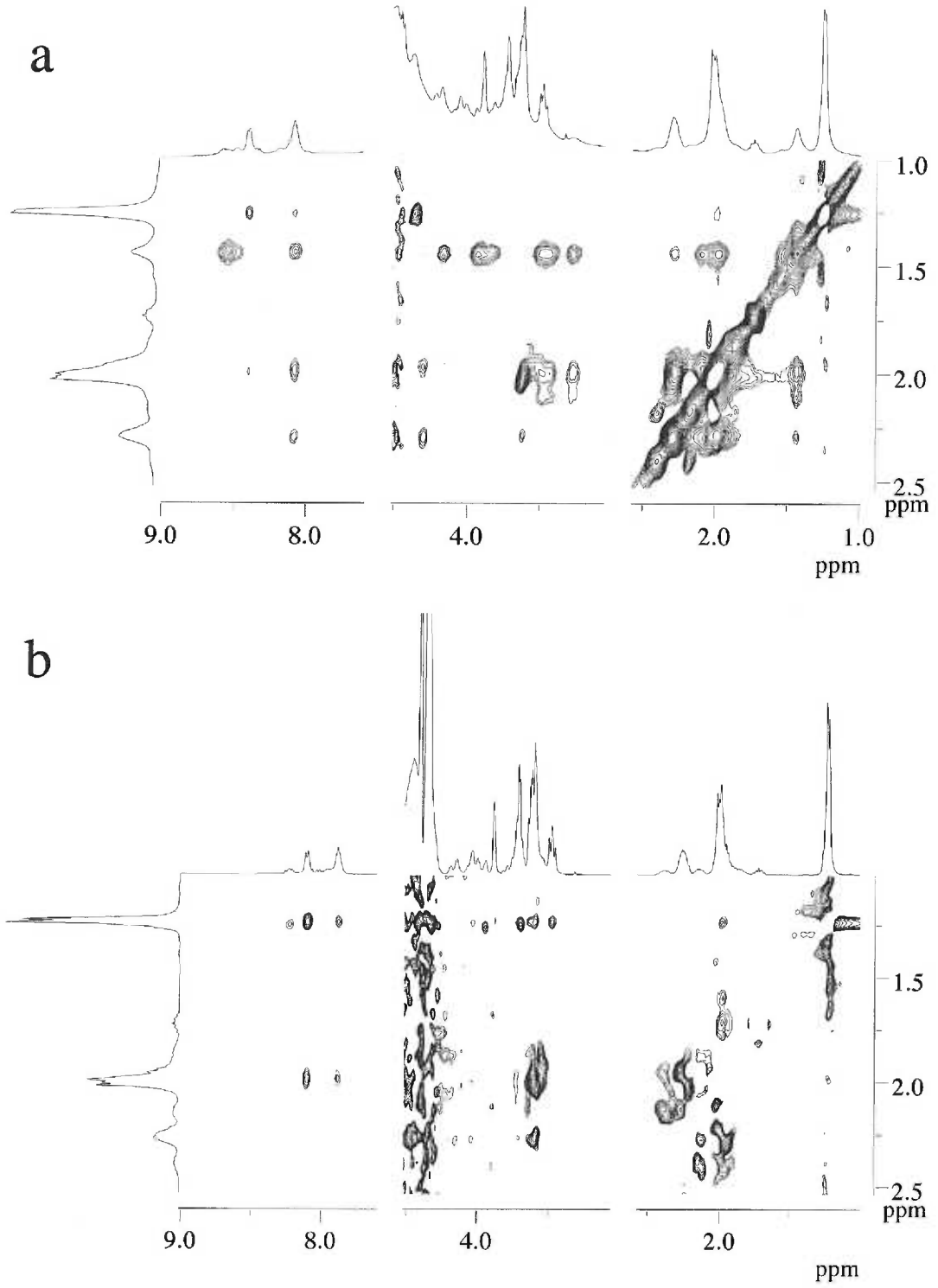


Figure 3

Chapter 3

Glycosylation/Hydroxylation-induced stabilization of the collagen triple helix: 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix

James G. Bann[‡] and Hans Peter Bächinger[‡§¶]

From the [‡]Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, and [§]Shriners Hospital for Children, Portland, Oregon 97201

Running Title: 4-Hyp in the Xaa position can stabilize the collagen triple helix

[¶]To whom correspondence should be addressed:

Shriners Hospital for Children

Research Department

3101 SW Sam Jackson Park Road

Portland, OR 97201

Tel.: 503-221-3433

Fax.: 503-221-3451

E-mail: hpb@shcc.org

Abstract

We have shown recently that glycosylation of threonine in the peptide Ac-(Gly-Pro-Thr)$_{10}$-NH_2 with β-D-galactose induces the formation of a collagen triple helix, while the non-glycosylated peptide does not. In this report, we present evidence that a collagen triple helix can also be formed in the Ac-(Gly-Pro-Thr)$_{10}$-NH_2 peptide, if the proline (Pro) in the Xaa position is replaced with 4-trans-hydroxyproline (Hyp). Furthermore, replacement of Pro with Hyp in the sequence Ac-(Gly-Pro-Thr(β-D-Gal))$_{10}$-NH_2 increases the T_m of the triple helix by 15.7 °C. It is generally believed that Hyp in the Xaa position destabilizes the triple helix, since (Pro-Pro-Gly)$_{10}$ and (Pro-Hyp-Gly)$_{10}$ form stable triple helices but the peptide (Hyp-Pro-Gly)$_{10}$ does not. Our data suggest that the destabilizing effect of Hyp relative to Pro in the Xaa position is only true in the case of (Hyp-Pro-Gly)$_{10}$. Increasing concentrations of galactose in the solvent stabilize the triple helix of Ac-(Gly-Hyp-Thr)$_{10}$-NH_2, but to a much lesser extent than achieved by covalently linked galactose. The data explain some of the forces governing the stability of the annelid/vestimentiferan cuticle collagens.

Introduction

The collagens of both vertebrates and invertebrates share the same characteristic repeating Gly-Xaa-Yaa tripeptide units that are required for the formation of the collagen triple helix. In vertebrates and most invertebrates, Xaa is often proline and Yaa is often 4-trans-hydroxyproline, and the thermal stability of a triple helix from a particular species increases as the total content of these imino acids increases (1). An exception to this correlation is the cuticle collagen of the deep-sea hydrothermal vent worm *Riftia pachyptila*. This collagen has a relatively high thermal stability (T_m = 37 °C), despite a low content of Pro/Hyp residues (~5%) (2). Amino acid analysis and partial sequencing of the cuticle collagen revealed that there is a high content of threonine residues, and that these are glycosylated, mainly with di- and tri-saccharides of galactose (3).

The observation of di- and tri-saccharides of galactose is consistent with the carbohydrate compositions of the cuticle collagens of *Lumbricus terristris* (earthworm) and *Nereis virens* (clamworm). Also consistent is the positional specificity of Hyp, which for these collagens seems to primarily occur in the Xaa position rather than the Yaa position (4, 5). For the *L. terrestris* cuticle collagen, the sequence -Gly-Hyp-Ser- accounts for 4-5% of the total hydroxyproline content (5).

For the vertebrate collagens, 4-Hyp is found exclusively in the Yaa position, and studies of collagen-like peptides show that although the peptides $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Pro\text{-}Hyp\text{-}Gly)_{10}$ formed stable triple-helices, the peptide $(Hyp\text{-}Pro\text{-}Gly)_{10}$ did not (6). From this study it was concluded that Hyp in the Xaa position does not contribute to the stability and is destabilizing to the triple helix. Indeed, even though the Hyp content of

the *L. terrestris* cuticle collagen is high (16%-18%), the thermal stability is low (T_m = 22 °C) (7). Also, for the *R. pachyptila* cuticle collagen, which has a low content of Pro/Hyp residues (~5%), it was suggested that glycosylation of Thr with galactose must compensate for the presence of Hyp in the Xaa position in order to achieve the high observed thermal stability (3).

We have recently shown that substitution of Thr by Thr-*O*-(β-D-galactose) in the peptide Ac-(Gly-Pro-Thr)$_{10}$-NH_2 induces the formation of a collagen triple-helix, with the latter peptide having essentially a random coil structure in aqueous solution (8). In this report we extend these studies to investigate the replacement of Pro by Hyp in the Xaa position of the peptides Ac-(Gly-Pro-Thr)$_{10}$-NH_2 and Ac-(Gly-Pro-Thr(βGal))$_{10}$-NH_2.

Experimental Procedures

Peptide synthesis and purification.

Peptides were synthesized using a Milligen 9050 peptide synthesizer. Couplings were carried out on a PAL-PEG-PS resin (Perseptive Biosystems, 0.16 mmol/g) using *N*α-9-fluorenylmethoxycarbonyl (FMOC) amino acids (Fmoc-Gly-OH, Fmoc-Hyp(tBu)-OH, Perseptive Biosystems, 4.0 eq.) and (O-(7-azabenzotriazol-1-vl)-1.1.3.3.tetramethyluronium hexafluorophosphate (HATU) (Perseptive Biosystems, 4.0 eq.) mediated peptide couplings. The glycopeptide was synthesized using FMOC-*O*-(2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl)-Thr-OH, which was prepared according to Elofson, et al. (9). The peptides were cleaved from the resin and purified by semi-preparative high performance liquid chromatography (HPLC). Treatment of the

glycopeptide with 6 mM sodium methoxide in methanol overnight removed the acetyl protecting groups on the sugar (10). The peptide was again purified by HPLC, and all peptides were characterized by amino acid analysis and MALDI-TOF mass spectrometry.

Circular Dichroism spectroscopy.

Circular dichroism spectra were recorded on an Aviv 202 spectropolarimeter using a Peltier thermostatted cell holder and a 1 mm pathlength rectangular cell. All measurements were performed in water, and peptide concentrations in all cases were 100 μM. Concentrations were determined by amino acid analysis. The wavelength spectra represent an average of 8 scans. Thermal transitions were recorded from 5°C to 70°C at 223 nm, and the temperature was raised at a rate of 12°C/h. The thermal transitions were fit to the all-or-none equation as described (11) in order to calculate the T_m, ΔH^0 (van't Hoff) and ΔS^0. D-Galactose was purchased from Sigma. A stock solution of galactose (0.41 g/ml) was prepared by dissolving galactose in water and measuring the optical rotation in a Perkin Elmer Model 241 MC polarimeter, using a 10 cm cell at room temperature. The concentration was determined based upon an $[\alpha]_D$ value of +80.2 as listed in the Merck Index.

Analytical ultracentrifugation.

Equilibrium sedimentation experiments were carried out on a Beckman Model E analytical ultracentrifuge using double-sector cells. Peptides were dissolved in water to 0.5 and 0.3 mg/ml for the peptides Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 and Ac-Gly-Hyp-

Thr(βGal)$_{10}$-NH$_2$, respectively. The temperature was regulated to 4.3°C, and the wavelength for monitoring was at 230 nm. Weight-average molecular weights were determined by fitting the data using Scientist® (Micromath, Salt Lake City, Utah). Partial specific volumes of 0.71 cm^3/g for the GZT peptide and 0.67 cm^3/g for the GZT(βGal) peptide were calculated from the amino acid composition and galactose content (12). However no value has been published for hydroxyproline, so the value for proline was used.

Differential scanning calorimetry.

The temperature dependence of the partial heat capacity was measured in an MC-2 differential scanning calorimeter (MicroCal Inc., Northampton, MA). The heating rate was 13 °C/h and the data was collected and analyzed using the software provided by the manufacturer.

Results and Discussion

Previous evidence using circular dichroism (CD) spectroscopy showed that the peptide Ac-(Gly-Pro-Thr(β-D-Gal)$_{10}$-NH$_2$ formed a collagen triple helix while the peptide Ac-(Gly-Pro-Thr)$_{10}$-NH$_2$ could not (8). Figure 1A compares the CD spectra of Ac-(Gly-Pro-Thr)$_{10}$-NH$_2$ (8) to that of Ac-(Gly-Hyp-Thr)$_{10}$-NH$_2$ and in water at 5 °C. Only the latter peptide exhibits the CD spectral characteristics of a collagen triple helix, with a positive ellipticity [θ] at 223 nm and a large negative minimum at 198 nm (13). This ellipticity value at 198 nm for Ac-(Gly-Pro-Thr)$_{10}$-NH2 (-30,000 deg cm^2 dmol^{-1}) is

similar to what is observed for the peptide (Pro-Ser-Gly)$_n$, ($[\theta]_{198nm}$ = -27,400 deg cm^2 dmol^{-1}) which, even at fairly high molecular weights (18,000 Da) exhibits a mostly random-coil structure (13). The temperature-dependence of the CD spectrum of Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 is shown in figure 1B, and exhibits a highly cooperative transition with a T_m of 19.2 °C. Equilibrium sedimentation of the peptide in water gave a weight-average molecular weight of 9780 +/- 100 Da, which is slightly higher than the expected mass of the trimer (8315 Da), and is presumably due to an error in the calculated partial specific volume and/or the presence of a small amount of aggregates.

The far-UV CD spectra of Ac-(Gly-Hyp-Thr(βGal))$_{10}$-NH_2 is shown in figure 2A, along with the CD spectrum of Ac-(Gly-Pro-Thr(βGal))$_{10}$-NH_2, which has been reported recently (8). The overall shape of the spectra is the same, with the Hyp peptide having a slightly lower ellipticity at 198 nm and a slightly larger ellipticity at 220 nm. The inset shows the CD spectra at 70°C for both peptides, which is above their respective T_m (see figure 2B). The ellipticity values at 200 nm are similar, suggesting that the differences in the CD at 5 °C are not due to a concentration artifact. Figure 2B shows the dependence of the ellipticity at 223 nm as a function of temperature for both Ac-(Gly-Hyp-Thr(βGal))$_{10}$-NH_2 and Ac-(Gly-Pro-Thr(βGal))$_{10}$-NH_2. The T_m of the Gly-Hyp-Thr(βGal) peptide is 54.8 °C, while the T_m of the Gly-Pro-Thr(βGal) peptide is 39.2 °C. Equilibrium sedimentation of the former peptide gave a weight-average molecular weight of 12,650 +/- 100 Da, slightly less than the calculated mass of the trimer (13,150 Da), probably due to an error in the calculated partial specific volume and/or the presence of a small amount of monomeric chains.

In order to determine if galactose, *by itself*, could have a similar influence as having the galactose covalently linked to threonine, we measured the thermal stability of both Ac-(Gly-Pro-Thr)$_{10}$-NH_2 and Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 in increasing concentrations of galactose in water (Figure 3A and B). Figure 3A shows the influence of galactose at 0, 0.5, 1.0 and 2.0 M on the CD spectra of Ac-(Gly-Pro-Thr)$_{10}$-NH_2 and Ac-(Gly-Hyp-Thr)$_{10}$-NH_2. Little difference in the CD of either peptide is observed at around 220 nm. For the peptide Ac-(Gly-Hyp-Thr)$_{10}$-NH_2, the T_m was measured by again monitoring the change in ellipticity at 223 nm as a function of temperature. A linear increase in T_m is observed as the concentration of galactose increases (Figure 3B). We estimated of the effective concentration of galactose required to match that of having the carbohydrate covalently linked to threonine. Based upon the predicted dimensions from molecular modeling of the peptide Ac-(Gly-Pro-Thr(βGal)$_5$-$NHCH_3$, the effective concentration is about 1.6 M. However, the T_m at 2 M galactose is only about 33 °C, compared to the near 55 °C T_m for Ac-(Gly-Hyp-Thr(βGal))$_{10}$-NH_2.

The above data show the favorable influence of glycosylation on the thermal stability of the triple-helix, as has been shown (8). Previous measurements on the effects of various sugars and polyols on the stability of the triple helix of calf-skin collagen suggested that the observed linear increase in T_m with increasing concentrations of sugars is due to an effect on the structure of water (14). The influence of various solvents on the stability of the peptide (Pro-Ser-Gly)$_n$ was studied (15). Although (Pro-Ser-Gly)$_n$ could not form a triple helix in water, this polypeptide had the *propensity* to form a triple helix, since in the solvent 1,3-propanediol the molecules exhibited the CD spectral features of a

triple helix. A similar observation was made with the peptide H-(Gly-Pro-Thr)$_{10}$-Gly-Pro-Cys-Cys-OH, which in the solvent 1,2-propanediol also forms a triple-helix (3).

The ability of other solvents to stabilize the triple-helix was also investigated by Brown and coworkers, and it was found that both neat trifluoroethanol (TFE) and hexafluoroisopropanol (HFIP) destabilize the triple-helix, whereas the polyhydric alcohols glycerol, ethylene glycol, diethylene glycol, 1,4-butanediol and 1,3-propanediol are stabilizing (15). It was suggested from these studies that the weaker hydrogen bonding solvents provide a more favorable environment than water in which to form the inter-chain hydrogen bonds. Presumably the effect of sugars and polyols would be manifested in a decrease in the activity of water and the ability of water to interact with the polypeptide backbone. Thus, peptide backbone-water hydrogen bonds would be disfavored relative to peptide-peptide hydrogen bonds. A possible mechanism of stabilization by *O*-glycosylation then is to change the local environment of the peptide backbone such that water-backbone interactions are disfavored, and the inter-chain hydrogen bonds become favored.

While the values for ΔH^0 and ΔS^0 for the galactosylated peptides are very similar to the values reported for (Pro-Pro-Gly)$_{10}$ and (Pro-Hyp-Gly)$_{10}$ (see Table 1), the transition curve for (Gly-Hyp-Thr)$_{10}$ is much steeper. This is reflected by a much larger enthalpy change for this peptide. This large change in enthalpy was confirmed by calorimetry. One possible explanation is that when threonine and hydroxyproline are contiguous, there is a strongly enthalpic interaction (probably a hydrogen bond) between the side-chain OH groups within a chain, between chains, or with water molecules. This interaction is compensated by a very large loss in entropy, possibly the side-chain

movement of threonine or in the movement of water molecules (bound to the side-chains).

The stabilizing influence of Hyp in the Xaa position may be explained by either an increase in the number of stable hydrogen bonds with water (16) or through a stereoelectronic inductive effect (17). We favor the latter hypothesis, since based upon the water binding mechanism the peptide Gly-Pro-Thr (and Gly-Pro-Ser) should form a stable triple helix, as was suggested recently (18). The OH group of Hyp is effective at withdrawing electrons away from the imide bond, and results in a change in the equilibrium constant for cis⇆trans isomerization, favoring the trans isomer. Since all peptide bonds in the triple helix are trans, there is a cumulative increase in stability as the number of stable trans isomers increases. Thus, having Hyp in the Xaa position would increase the stability of the triple-helix by also increasing the number of trans isomers.

Why then does $(Hyp\text{-}Pro\text{-}Gly)_{10}$ not form a triple-helix? One possibility is that the stereoconfiguration of the proline ring in the Yaa position is altered if the Pro in the Xaa position is Hyp. The recent crystal structures (19, 20) show that the Pro in the Xaa position is in general puckered down, while the Hyp or Pro in the Yaa position is *always* puckered up. This configuration in puckering between two contiguous prolines or with Pro-Hyp may not be favored if the Hyp is in the Xaa position, and may result in an upward puckering in the Xaa and a downward puckering for Pro in the Yaa. This could then in turn change the Xaa psi angle between the two proline residues, favoring a value other than around 160° as reported for collagen and collagen-like peptides (19, 20).

The presence of Hyp in the Xaa position for the cuticle collagens of *L. terrestris*, *N. virens* and *R. pachyptila* suggest that a novel hydroxylating enzyme is present in these

organisms that is distinct in specificity from the normal vertebrate prolyl-4-hydroxylase (21). Such an enzyme has been described from the subcuticular epithelium of *L. terrestris* (22). From these studies, this enzyme would also be required for the normal stability of the triple helix of these collagens.

References

1. Burjanadze, T.V. (1979) *Biopolymers* **18**, 931-938

2. Gaill, F., Mann, K., Wiedemann, H., Engel, J., and Timpl, R. (1995) *J. Mol. Biol.* **246**, 284-94

3. Mann, K., Mechling, D.E., Bächinger, H.P., Eckerskorn, C., Gaill, F., and Timpl, R. (1996) *J. Mol. Biol.* **261**, 255-266

4. Sharma, Y.D., and Tanzer, M.L. (1984) *Anal. Biochem.***141**, 205-212

5. Goldstein, A., and Adams, E. (1970) *J. Biol. Chem.* **245**, 5478-5483

6. Inouye, K.E., Kobayashi, Y., Kyogoku, Y., Kishida, Y., Sakakibara, S., and Prockop, D.J. (1982) *Arch. Biochem. Biophys.* **219**, 198-203

7. Harrington, W.F., and Rao, N.V. (1967) *Conformation of Biopolymers*, (Ramachandran, G.N., Ed.) Academic, New York, pp. 513-531

8. Bann, J.G., Peyton, D.H., and Bächinger, H.P. (2000) *FEBS Letters*, **473**, 237-240

9. Elofsson, M., Walse, B., and Kihlberg, J. (1991) *Tetrahedron Lett.* **32**, 7613-7616

10. Kihlberg, J., Elofsson, M., Salvador, L.A. (1997) *Methods Enzym.* **289**, 221-245

11. Engel, J., Chen, H-T., Prockop, D.J., and Klump, H. (1977) *Biopolymers.* **16**, 601-622

12. Perkins, S.J. (1986) *Eur. J. Biochem.* **157**, 169-180

13. Brown III, F.R., Hopfinger, A.J., and Blout, E.R. (1972) *J. Mol. Biol.* **63**, 85-99

14. Gekko, K., and Koga, S. (1983) *J. Biochem.* **94**, 199-205

15. Brown III, F.R., Di Corato, A., Lorenzi, G.P., and Blout, E.R.(1972) *J. Mol. Biol.* **63**, 101-115

16. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) *Science* **266**, 75-81

17. Eberhardt, E.S., Panasik, N. Jr., and Raines, R.T. (1996) *J. Am. Chem. Soc.* **118**, 12261-12266

18. Kramer, R.Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H.M. (1999) *Nature Struct. Biol.* **6**, 454-457

19. Nagarajan, V., Kamitori, S., and Okuyama, K. (1999) *J. Biochem.* **125**, 310-318

20. Kramer, R.Z., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A., and Berman, H.M. (1998) *J. Mol. Biol.* **280**, 623-638

21. Kivirikko, K.I., Myllyla, R., and Pihlajaniemi, T. (1989) *FASEB J.* **3**, 1609-1617

22. Adams, E., and Lamon, M. (1977) *J. Biol. Chem.* **252**, 7591-7597

23. Engel, J., and Bächinger, H.P. (2000) *Matrix Biology*, in press

Footnotes

This work was supported by a grant from Shriners Hospital for Children. We thank J. Gambee for help with peptide synthesis and amino acid analysis.

Figure Legends

Figure 1. *A*. The CD spectra of Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 (solid line) and Ac-(Gly-Pro-Thr)$_{10}$-NH_2 (dotted line) in water at 5°C. *B* Melting transition curves for Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 (solid line) and Ac-(Gly-Pro-Thr)$_{10}$-NH_2 (dotted line) recorded at 223 nm. Concentrations in both *A* and *B* were 100 μM.

Figure 2. *A*. The CD spectra of Ac-(Gly-Hyp-Thr-(*O*-β-D-galactose))$_{10}$-NH_2 (solid line) and Ac-(Gly-Pro-Thr-(*O*-β-D-galactose))$_{10}$-NH_2 (dotted line) in water at 5°C and at 70 °C (inset). *B* Melting transition curves for Ac-(Gly-Hyp-Thr-(*O*-β-D-galactose))$_{10}$-NH_2 (solid line) and Ac-(Gly-Pro-Thr-(*O*-β-D-galactose))$_{10}$-NH_2 (dotted line) recorded at 223 nm. Concentrations in both *A* and *B* were 100 μM.

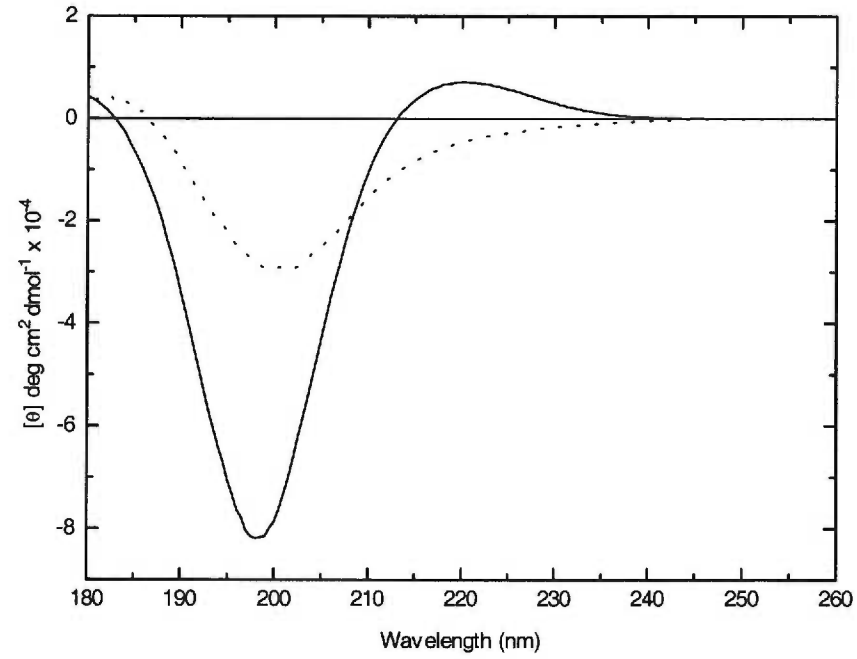
Figure 3. *A*. Dependence of galactose concentration on the far-UV CD spectra of Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 (solid line) and Ac-(Gly-Pro-Thr)$_{10}$-NH_2 (dotted line) at 5°C, 100 μM peptide. The darker lines are the CD traces up to 2M galactose, and no change in the region near 220 nm is observed. *B*. Dependence of the melting curve transition on galactose concentration of Ac-(Gly-Hyp-Thr)$_{10}$-NH_2 (100 μM). A = 0 M, B = 0.5 M, C= 1.0 M, D = 2.0 M galactose.

Table 1

Comparison of thermodynamic parameters for the triple helix ⇆ coil transition of various peptides.

Peptide	T_m (K)[a]	ΔH^0_{VH} (kJ/mol)	ΔS^0 (J/mol/K)	ΔH^0_{cal} (kJ/mol)	c (mM)	Ref
$(Pro\text{-}Pro\text{-}Gly)_{10}$	297.8	-7.91	-22.4	-7.68	2.6	11
$(Pro\text{-}Hyp\text{-}Gly)_{10}$	330.5	-13.4	-36.4	-13.4	2.4	11, 23
$(Gly\text{-}Pro\text{-}Thr(\beta Gal))_{10}$	312.0	-14.0	-39.6		0.1	this study
$(Gly\text{-}Hyp\text{-}Thr(\beta Gal)_{10}$	323.2	-11.1	-29.0		0.1	this study
$(Gly\text{-}Hyp\text{-}Thr)_{10}$	291.2	-27.1	-87.1	-27.5	0.1	this study

[a] Melting temperatures T_m are concentration dependent and are indicated for the concentration c. ΔH^0_{VH} and ΔS^0 values were obtained by fitting the experimental data to equation (2) in reference 11. The thermodynamic parameters are expressed per mol tripeptide units in a triple helix.


2
0
-2
-4
-6
-8
[θ] deg cm² dmol⁻¹ x 10⁻⁴
180
190
200
210
220
230
240
250
260
Wavelength (nm)

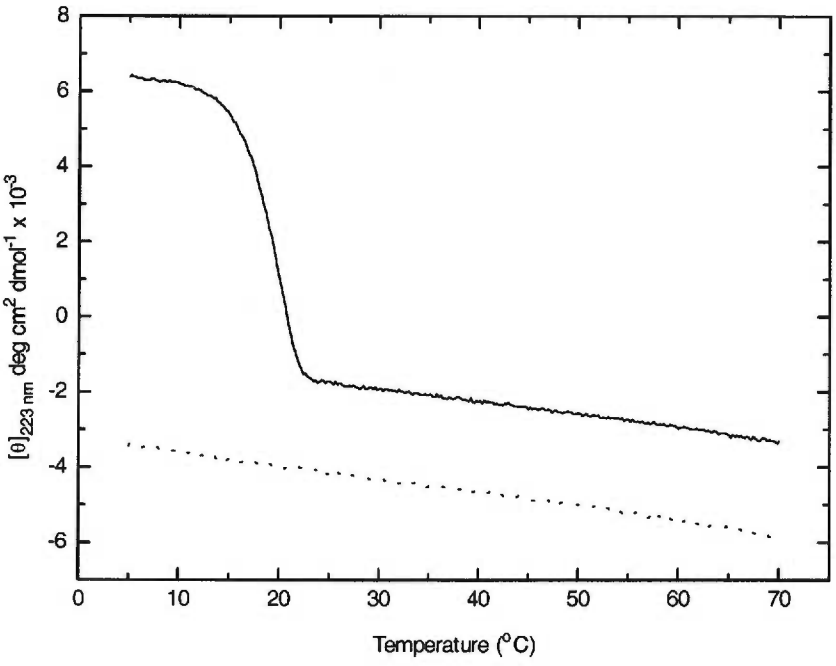
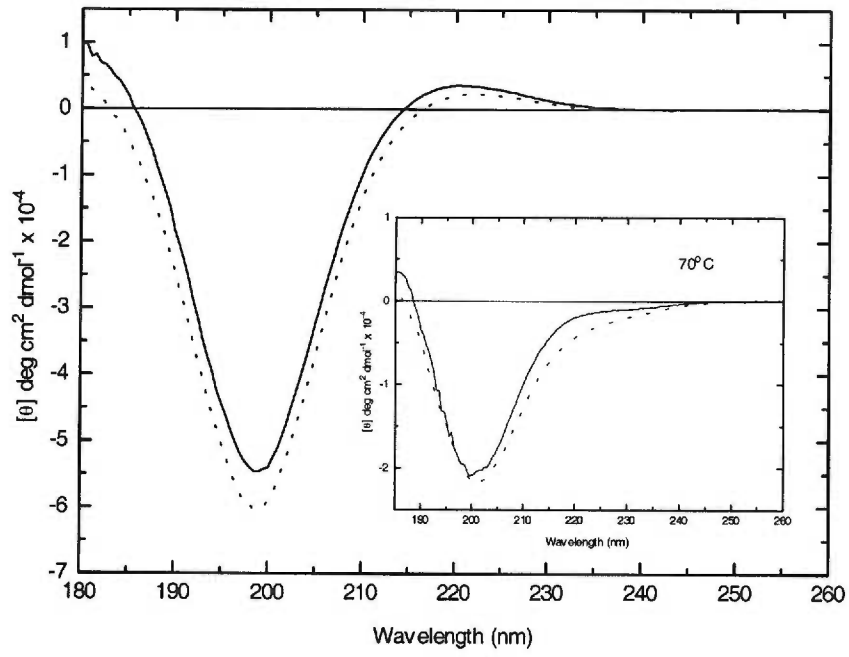

8
6
4
2
0
-2
-4
-6
[θ]223 nm deg cm² dmol⁻¹ x 10⁻³
0
10
20
30
40
50
60
70
Temperature (°C)

Figure 1


[θ] deg cm² dmol⁻¹ x 10⁻⁴
Wavelength (nm)
70°C
[θ] deg cm² dmol⁻¹ x 10⁻⁴
Wavelength (nm)

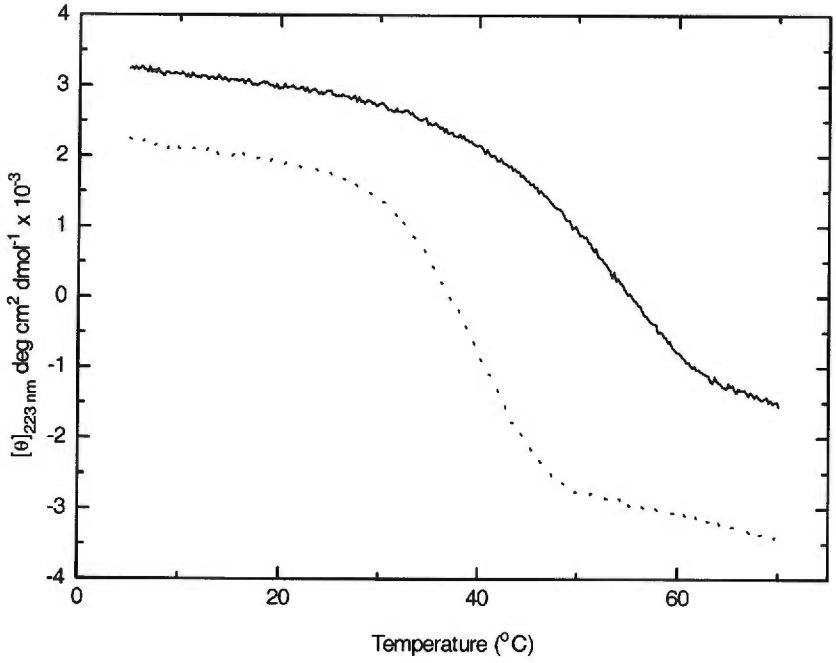
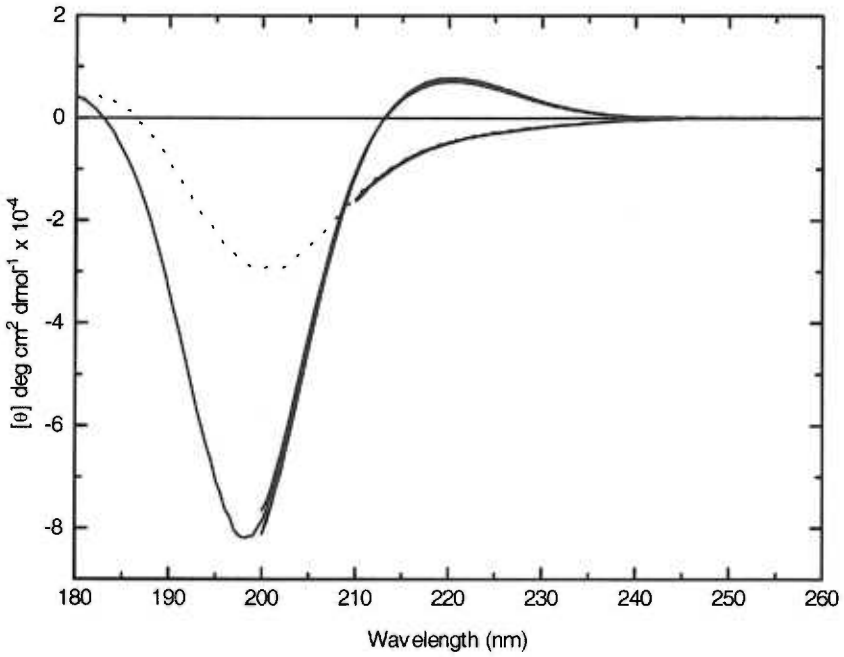

[θ]223 nm deg cm² dmol⁻¹ x 10⁻³
Temperature (°C)

Figure 2


2
0
-2
-4
-6
-8
[θ] deg cm² dmol⁻¹ x 10⁻⁴
180
190
200
210
220
230
240
250
260
Wavelength (nm)

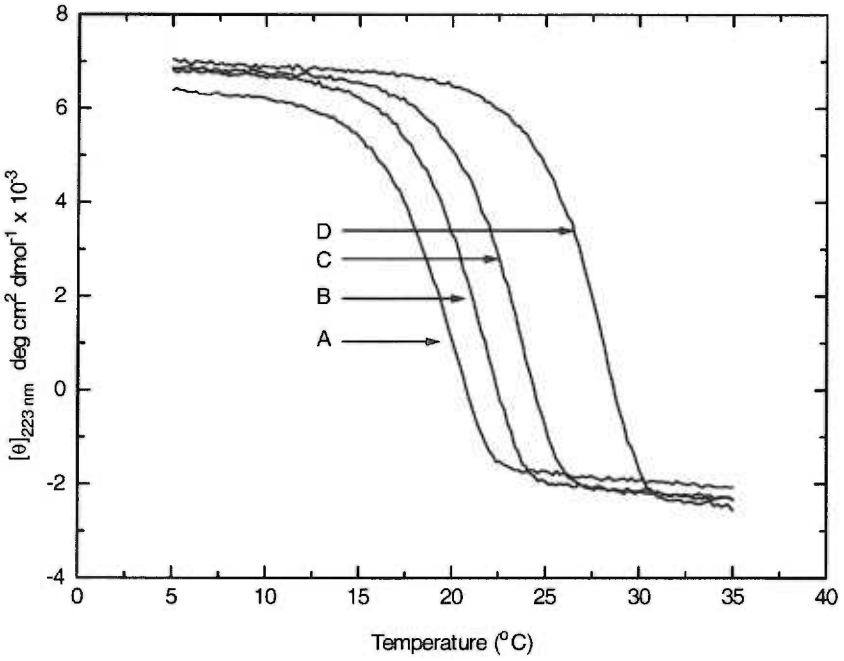

8
6
4
2
0
-2
-4
[θ]223 nm deg cm² dmol⁻¹ x 10⁻³
D
C
B
A
0
5
10
15
20
25
30
35
40
Temperature (°C)

Figure 3

Chapter 4

Summary and Conclusions

The primary aim of these studies was to understand the forces that govern the stabilization of the cuticle collagen of *R. pachyptila.* Through the synthesis of glycosylated, collagen-like peptides and the combined use of CD, analytical ultracentrifugation, and NMR spectroscopy, the results presented strongly suggest that the effect of *O*-galactosylation of threonine in the *R. pachyptila* cuticle collagen is to stabilize the triple-helical structure. A collagen like triple-helical structure was not observed in the peptide Ac-(Gly-Pro-Thr)$_{10}$-NH_2, whereas in the peptide Ac-(Gly-Pro-Thr(Galβ))$_{10}$-NH_2 a triple-helical structure was formed, with a thermal stability of 39°C. Within the native *R. pachyptila* sequence, there has yet to be found a sequence containing a -Gly-Pro-Thr- triplet. However, the sequence -Gly-Hyp-Thr- is observed, and a peptide having this sequence, Ac-(Gly-Hyp-Thr)$_{10}$-NH_2, is able to form a triple-helix independent of having the threonines galactosylated, with a T_m of 19°C. As with the peptide Ac-(Gly-Pro-Thr(Galβ))$_{10}$-NH_2, Ac-(Gly-Hyp-Thr(Galβ))$_{10}$-NH_2 also forms a stable triple-helix, with a T_m of 55°C. Thus, combining both glycosylation in the Yaa position and hydroxylation in the Xaa position greatly enhances the stability of the triple-helix.

Although these studies show that glycosylation stabilizes the triple-helix, we have still an incomplete understanding of why the triple-helix is stable. Perhaps glycosylation restricts the conformational space available to the polypeptide backbone, and decreases the entropy of the unfolded state. Or, perhaps there are hydrogen bonds from the

carbohydrate to the polypeptide backbone, which stabilize the folded state. An important set of experiments that were not shown in these pages are fluorescence energy transfer data indicating that the short peptide DABCYL-(Gly-Pro-Thr)$_5$-EDANS is less extended than the peptide DABCYL(Gly-Pro-Thr(βGal))$_5$-EDANS (Bann, et al, unpublished observations). This would favor the hypothesis that the formation of an extended structure in the unfolded state contributes to stabilizing the collagen triple-helix.

Further studies are required to fully understand the forces governing helix stabilization. Because the triple-helix unfolds very cooperatively as a function of temperature, it is difficult to detect any intermediate structures that may be on the unfolding/folding pathway. The observation of a very sharp transition in the peptide Ac-(Gly-Hyp-Thr)$_{10}$-NH_2, when fit with the all-or-none equation of Engel (Engel, 1977), gave a very high transition enthalpy, despite a lower thermal stability (Bann, JBC, 2000). This transition was also monitored using differential scanning calorimetry, and show the presence of two transitions which, when summed, give the indicated enthalpy of the transition. This is presumably an intermediate in the folding transition, and the equilibrium between the intermediate and the triple-helix is insensitive to increasing concentrations of galactose (no changes in the shape of the transition or in the magnitude of the CD at 223 nm). Perhaps the effect of glycosylation is just a local strengthening of the hydrogen bonds that already exist, rather than shifting the equilibrium towards a more triple-helical structure. Kinetic studies on fully reversible peptides would help to determine if other peptides have intermediate structures, and would hopefully allow a complete determination of the folding pathway. Such studies may be achieved using peptides similar to (Gly-Pro-Thr)$_{10}$-Gly-Pro-Cys-Cys, which has at the carboxyl-terminus

a disulfide knot (Mann, et al., 1996), and where folding/unfolding are not complicated by a nucleation event. Based upon the idea that glycosylation limits conformational space, it is hypothesized that folding of the triple-helix would be faster with glycosylation than without; thus, the peptide $(Gly\text{-}Pro\text{-}Thr)_{10}$-Gly-Pro-Cys-Cys would fold more slowly than $(Gly\text{-}Pro\text{-}Thr(\beta Gal))_{10}$-Gly-Pro-Cys-Cys. An understanding of whether glycosylation also influences the intrinsic properties of threonine (changes in the pKa of the NH and COOH groups), or whether there is an effect on the proline *cis-trans* isomer ratio would also provide clues as to how glycosylation increases triple-helical stability.

Literature Cited

Adams, E., Lamon, M. (1977). Prolyl hydroxylase of earthworms. Substrate specificity of an enzyme from the subcuticular epithelium. J. Biol. Chem. **252**, 7591-7597.

Andreotti, A.H., Kahne, D. (1993). Effects of Glycosylation on Peptide backbone conformation. J. Am. Chem. Soc. **115**, 3352-3353.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. Science **181**, 223-30.

Balacco, G. :(1994). SwaN-MR:A complete and expansible NMR software for the Macintosh. J. Chem. Inf. Comput. Sci. **34**, 1235-1241.

Bann, J.G., Peyton, D.H., Bachinger, H.P. (2000). Sweet is stable: glycosylation stabilizes collagen. Febs Letters, **473**, 237-240.

Bann, J.G., Bachinger, H.P. (2000) Glycosylation/Hydroxylation-induced stabilization of the collagen triple-helix: 4-trans-hydroxyproline in the Xaa position can stabilize the triple-helix. J. Biol. Chem. To be published.

Barbarich, T.J., Rithner, C.D., Miller, S.M., Anderson, O.P., Strauss, S.H. (1999). Significant Inter- and Intramolecular O-H---FC Hydrogen Bonding. J. Am. Chem. Soc. **121**, 4280-4281.

Bax, A., Davis, D.G. (1985). Mlev-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Magn. Reson. **65**, 355-360.

Beck, K., B. A. Boswell, et al. (1996).Triple helix formation of procollagen type I can occur at the rough endoplasmic reticulum membrane. J Biol Chem **271**, 21566-73.

Bella, J., B. Brodsky, et al. (1995). Hydration structure of a collagen peptide. Structure **3**: 893-906.

Bella, J., M. Eaton, et al. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution [see comments]. Science **266**, 75-81.

Berg, R. A. and D. J. Prockop (1973). The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. Biochem Biophys Res Commun **52**, 115-20.

Biondi, L., Filira, F., Gobbo, M., Pavin, E., Rocchi, R. (1998). Conformational investigations on glycosylated threonine-oligopeptides of increasing chain length. J. Peptide Sci. 4, 58-71.

Bothener-By, A.A., Stephens, R.L., Lee, J., Warren, C.D., Jeanloz, R.W. (1984). Strucutre determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. J. Am. Chem. Soc. **106**, 811-813.

Brodsky, B. and J. A. Ramshaw (1997). The collagen triple-helix structure. Matrix Biol **15**, 545-54.

Brown, F.R., Di Corato, A., Lorenzi, G.P., Blout, E.R. (1972). Synthesis and structural studies of two collagen analogues: poly (L-prolyl-L-seryl-glycyl) and poly (L-prolyl-L-alanyl-glycyl). J. Mol. Biol. **63**, 85-99.

Brown, F.R. Hopfinger, A.J., Blout, E.R. (1972). The collagen-like triple-helix to random coil transition: experiment and theory. J. Mol. Biol. **63**, 85-99.

Burjanadze, T. V. (1979). Hydroxyproline content and location in relation to collagen thermal stability. Biopolymers **18**, 931-8.

Caminati, W., Melandri, S., Rossi, I., Favero, P.G. (1999) The C-F---H-O Hydrogen Bond in the Gas Phase. Rotational Spectrum and ab Initio Calculations of Difluoromethane-Water. J. Am. Chem. Soc. **121**, 10098-10101.

Creighton, T. E. (1990). Protein folding. Biochem J **270**, 1-16.

Drewes, J. A. and K. L. Rowlen (1993). Evidence for a gamma-turn motif in antifreeze glycopeptides. Biophys J **65**, 985-91.

Drickamer, K. and M. E. Taylor (1998). Evolving views of protein glycosylation. Trends Biochem Sci **23**, 321-4.

Eberhardt, E.S., Panasik, Jr., N., Raines, R.T. (1996). Inductive Effects on the Energetics of Prolyl Peptide Bodn Isomerization: Implications for Collagen Folding and Stability. J. Am. Chem. Soc. **118**, 12261-12266.

Elofson, M., Walse, B., Kihlber, J. (1991). Building blocks for glycopeptide synthesis: Glycosylation of 3-mercaptopropionic acid and FMOC amino acids with unprotected carboxyl groups. Tetrahedron Letters. **32**, 7613-7617.

Engel, J., Bachinger, H.P. (2000). Cooperative equilibrium transitions coupled with a slow annealing step explain the sharpness and hysterisis of collagen folding. Matrix Biology, to be published.

Engel, J. (1997). Versatile collagens in invertebrates. Science. **277**, 1785-1786.

Engel, J., H. T. Chen, et al. (1977). The triple helix in equilibrium with coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)n and (L-Pro-L-Hyp-Gly)n. Biopolymers **16**, 601-22.

Engel, J. (1968) Das Leder 19, 21-26.

Feng, Y., Melacini, G., Taulane, J.P., Goodman, M. (1996). Acetyl-terminated and template-assembled collagen-based polypeptides composed of Gly-Pro-Hyp sequences. 2. Synthesis and conformational analysis by circular dichroism, ultraviolet absorbance, and optical rotation. J. Am. Chem. Soc. **118**, 10351-10358.

Feng, Y., Melacini, G., Taulane, J.P., Goodman, M. (1996). Collagen-based strucutres containing the peptoid residue N-isobutylglycine (Nleu): synthesis and biophysical studies of Gly-Pro-Nleu sequences by circular dichroism, ultraviolet absorbance, and optical rotation. Biopolymers. **39**, 859-872.

Gaill, F. (1993). Aspects of life development at deep sea hydrothermal vents. Faseb J **7**, 558-65.

Gaill, F., D. Herbage, et al. (1991). A discrete helicoid of collagenous fibrils: the cuticle of deep-sea hydrothermal vent worms (Riftia pachyptila). Matrix **11**, 197-205.

Gaill, F., H. Wiedemann, et al. (1991). Molecular characterization of cuticle and interstitial collagens from worms collected at deep sea hydrothermal vents. J Mol Biol **221**, 209-23.

Gekko, K., Koga, S. (1983) Increased thermal stability of collagen in the presence of sugars and polyols. J. Biochem. **94**, 199-205.

Gerken, T. A., K. J. Butenhof, et al. (1989). Effects of glycosylation on the conformation and dynamics of O-linked glycoproteins: carbon-13 NMR studies of ovine submaxillary mucin. Biochemistry **28**(13): 5536-43.

Goldstein, A. and E. Adams (1968). Occurrence of glycylhydroxypyrolyl sequences in earthworm cuticle collagen. J Biol Chem **243**, 3550-2.

Goldstein, A. and E. Adams (1970). Glycylhydroxyprolyl sequences in earthworm cuticle collagen: glycylhydroxyprolylserine. J Biol Chem **245**, 5478-83.

Gottschalk, A. (1960). Correlation between composition, structure, shape and function of a salivary mucoprotein. Nature. **186**, 949-951.

Har-el, R., Tanzer, M.L. (1993). Extracellular Matrix. 3: Evolution of the extracellualr matric in invertebrates. Faseb J. **7**, 1115-1123.

Harrington, W.F., Rao, N.V. (1967). Pyrrolidine residues and the stability of collagen. in Conformation of Biopolymers (ed. Ramachandran, G.N.) 513-531 (Academic, New York).

Holmgren, S. K., L. E. Bretscher, et al. (1999). A hyperstable collagen mimic. Chem Biol **6**, 63-70.

Holmgren, S. K., K. M. Taylor, et al. (1998). Code for collagen's stability deciphered [letter]. Nature **392**, 666-7.

Inouye, K., et al. (1982) . Synthesis and physical properties of (hydroxyproline-proline-glycine)10: hydroxyproline in the X-position decreases the melting temperature of the collagen triple-helix. Arch. Biochem. Biophys. **219**, 198-203.

Jentoft, N. (1990). Why are proteins O-glycosylated? [see comments]. Trends Biochem Sci **15**, 291-4.

Jimenez, S. A., P. Dehm, et al. (1973). Intracellular collagen and protocollagen from embryonic tendon cells. J Biol Chem **248**, 720-9.

Josse, J., Harrington, W.F. (1964). Role of pyrrolidine residues in the structure and stabilizaton of collagen. J. Mol. Biol. **9**, 269-287.

Kadler, K. E., D. F. Holmes, et al. (1996). Collagen fibril formation. Biochem J **316**(Pt 1), 1-11.

Kihlberg, J., Elofsson, M., Salvador, L.A. (1997). Direct synthesis of glycosylated amino acids form carbohydrate peractetates and Fmoc amino acids: solid-phase synthesis of biomedicinally interesting glycopeptides. Methods in Enzymology. **289**, 221-245.

Kivirikko, K. I., R. Myllyla, et al. (1989). Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. Faseb J **3**, 1609-17.

Kramer, R. Z., J. Bella, et al. (1999). Sequence dependent conformational variations of collagen triple-helical structure. Nat Struct Biol **6**, 454-7.

Lane, A.N., et al. (1998). Conformational and dynamic properties of a 14 residue antifreeze glycopeptide from Antarctic cod. Protein Science. **7**, 1555-1563.

Lazarev, Y. A., B. A. Grishkovsky, et al. (1992). Bound water in the collagen-like triple-helical structure. Biopolymers **32**, 189-95.

Lazarides, E. and L. N. Lukens (1971). Collagen synthesis on polysomes in vivo and in vitro. Nat New Biol **232**, 37-40.

Lazarides, E. L., L. N. Lukens, et al. (1971). Collagen polysomes: site of hydroxylation of proline residues. J Mol Biol **58**, 831-46.

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

Lin, Y., J. G. Duman, et al. (1972). Studies on the structure and activity of low molecular weight glycoproteins from an antarctic fish. Biochem Biophys Res Commun **46**, 87-92.

Live, D. H., L. J. Williams, et al. (1999). Probing cell-surface architecture through synthesis: an NMR-determined structural motif for tumor-associated mucins. Proc Natl Acad Sci U S A **96**, 3489-93.

Mann, K., D. E. Mechling, et al. (1996). Glycosylated threonine but not 4-hydroxyproline dominates the triple helix stabilizing positions in the sequence of a hydrothermal vent worm cuticle collagen. J Mol Biol **261**, 255-66.

Marion, D., Wuthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins. Biochem. Biophys. Res. Commun. **113**, 967-974.

Mimura, Y., Y. Inoue, et al. (1989). N.m.r. study on conformation of Ac-Thr(alpha-GalNAc)-Ala-Ala-OMe as a model for mucin type glycoprotein. Int J Pept Protein Res **34**, 363-8.

Muir, L. and Y. C. Lee (1969). Structures of the D-galactose oligosaccharides from earthworm cuticle collagen. J Biol Chem **244**, 2343-9.

Muir, L. and Y. C. Lee (1970). Glycopeptides from earthworm cuticle collagen. J Biol Chem **245**, 502-9.

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

O'Conner, S. E. and B. Imperiali (1998). A molecular basis for glycosylation-induced conformational switching. Chem Biol **5**, 427-37.

Parkinson, J., K. E. Kadler, et al. (1995). Simple physical model of collagen fibrillogenesis based on diffusion limited aggregation. J Mol Biol **247**, 823-31.

Perkins, S.J. (1986) Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. Eur. J. Biochem. **157**, 169-180.

Privalov, P. L., E. I. Tiktopulo, et al. (1979). Stability and mobility of the collagen structure. J Mol Biol **127**, 203-16.

Ramachandran, G.N. (1988). Stereochemistry of collagen. Int. J. Peptide Protein Res. 31, 1-16.

Ramachandran, G. N., M. Bansal, et al. (1973). A hypothesis on the role of hydroxyproline in stabilizing collagen structure. Biochim Biophys Acta **322**(1): 166-71.

Ramachandran, G.N., Kartha, G. (1955). Structure of collagen. Nature. **176**, 593-595.

Ramshaw, J.A., Werkmeister, J.A., Glattauer, V. (1996). Collagen-based biomaterials. Biotechnology and Genetic Engineering Review. **13**, 335-382.

Rich, A., Crick, F.H.C. (1955). Structure of collagen. Nature. **176**, 780.

Rose, M. C., W. A. Voter, et al. (1984). Effects of deglycosylation on the architecture of ovine submaxillary mucin glycoprotein. J Biol Chem **259**, 3167-72.

Salvador, L.A., Elofson, M., Kihlberg, J. (1995). Tetrahedron. **51**, 5643-5656.

Sharma, Y. D. and M. L. Tanzer (1984). High-performance liquid chromatographic separation of glycopeptides from Nereis cuticle collagen. Anal Biochem **141**, 205-12.

Shogren, R., T. A. Gerken, et al. (1989). Role of glycosylation on the conformation and chain dimensions of O- linked glycoproteins: light-scattering studies of ovine submaxillary mucin. Biochemistry **28**, 5525-36.

Snyder, J.P., et al. (2000). The unexpected Diaxial orientation of cis-3,5-difluoropiperidine in water: a potent CF---NH charge-dipole effect. J.Am. Chem. Soc. Published on the Web.

Spiro, R. G. and V. D. Bhoyroo (1980). Studies on the carbohydrate of collagens. Characterization of a glucuronic acid-mannose disaccharide unit from Nereis cuticle collagen. J Biol Chem **255**, 5347-54.

Traub, W. (1974). Some stereochemical implications of the molecular conformation of collagen. Israel J. of Chem. **12**, 435-439.

van der Rest, M. and R. Garrone (1991). Collagen family of proteins. Faseb J **5**, 2814-23.

Waite, J.H., Tanzer, M.L., Merkel, J.R. (1980). Nereis cuticle collagen. Proteolysis by marine vibrial and clostridial collagenases. J. Biol. Chem. **255**, 3596-3599.

Wu, W., Pasternack, L., Huang, D., Koeller, K.M., Lin, C., Seitz, O., Wong, C. (1999). Structural study on *O*-glycopeptides: Glycosylation-Induced conformational changes of *O*-GlcNAC, O-LacNAc, *O*-Sialyl-LacNAc, and,

and *O*-Sialyl-Lewis-X peptides of the mucin domain of MAdCAM-1. J. Am. Chem. Soc. **121**, 2409-2417.

Yee, R.Y., Englander, S.W., von Hippel, P.H. (1974). Native collagen has a two-bonded structure. J. Mol. Biol. **83**, 1-16.