An Investigation on the Application of Mass Spectrometry Protein Footprinting Technique to Study Interactions of Collagen and Collagen Receptors

by

Michele Kirchner

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2013

This manuscript has been read and accepted for the
Graduate Faculty in Biochemistry in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

_____	_____
Date	Chair of Examining Committee
	Dr. Yujia Xu
_____	_____
Date	Executive Officer of Biochemistry
	Dr. Edward Kennelly

Dr. Lynn Francesconi (Hunter College)

Dr. Ranajeet Ghose (City College)

Dr. Peter Lipke (Brooklyn College)

Dr. Shaneen Singh (Brooklyn College)

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

An Investigation on the Application of Mass Spectrometry Protein Footprinting Technique to Study Interactions of Collagen and Collagen Receptors

by

Michele Kirchner

Adviser: Dr. Yujia Xu

Collagen is one of the most abundant proteins in the human body. There are many different types of collagen; they interact with an assortment of other proteins in the extracellular matrix, and with collagen receptors on the surfaces of cells. In this study, oxidative protein footprinting using Fenton chemistry followed by mass spectrometry was explored as a method to investigate collagen interactions. A known collagen binding protein, the A3 domain of the von Willebrand Factor protein (vWF A3), was chosen as a model. This protein association is the first critical step in blood clotting, and has been the target for the development of new drugs to treat von Willebrand disease, a bleeding disorder. Mass spectrometry protein footprinting could potentially aid in these endeavors. In order to accomplish this objective, original studies using Fenton chemistry labeling of triple helical peptides, and mass spectrometry collagen sequencing studies were carried out. The von Willebrand Factor A3 protein was produced from an Escherichia coli expression system.

This work discovered many challenges that have to be overcome for the successful approach of mass spectrometry oxidative protein footprinting by Fenton chemistry to investigate full chain collagen interactions. Additionally, the mix of variable post translational modifications of collagen makes the identification of binding sites by footprinting unreliable. Even sequencing collagen proves problematic because of the many post translational modifications. Despite these setbacks, we did; however, find that the von Willebrand Factor A3 protein binds to a triple helix conformation lacking any of the presumed necessary hydroxyproline sites. We were also able to identify regions in full chain collagen where the oxidation content varied from what was expected, assuming all the prolines in the Y position were hydroxylated.

Acknowledgements

I would like to thank my family for their support. I would like to thank the members in my lab for their assistance. I also would like to thank Dr. Haiteng Deng, Dr. Haiqiang Yu, Dr. Milica Tesic Mark, Dr. Nagarajan Chandramouli, Dr. Henrik Molina, Dr. Chang Geng Ruan, Dr. Jurgen Engle, Dr. Sergey Leikin, and Dr. Elena N. Makareeva for their help. I would like to thank Dr. Cliff Soll for his aid. I would like to thank Dr. Lesley Davenport, Ms. Judy Li, and Dr. Edward Kennelly for their supervision. I would like to thank Dr. Klaus Grohmann, Dr. Thomas Haines, and Dr. Maria Tomasz for their encouragement. I would like to thank Dr. Weigang Qui for his advice. I would like to thank my mentor Dr. Yujia Xu, and my committee members Dr. Lynn Francesconi, Dr. Ranajeet Ghose, Dr. Peter Lipke, and Dr. Shaneen Singh for all of their guidance throughout the years. I am truly grateful to those who have been involved in my education. Thank you.

Table of Contents

List of Tables

List of Figures

Abbreviations

A	alanine
Ala	alanine
Arg	arginine
R	arginine
Asn	asparagine
N	asparagine
Asp	aspartic acid
D	aspartic acid
BK	bradykinin
BSA	bovine serum albumin
CD	circular dichroism
α-CHCA	α-cyano-4-hydroxycinnamic acid
CNBR	cyanogen bromide
C	cysteine
Cys	cysteine
Cyt C	cytochrome C
Da	dalton
DNA	deoxyribonucleic acid
K_d	dissociation constant
ESI	electrospray ionization
E. coli	Escherichia coli
Gln	glutamine
Q	glutamine
Glu	glutamic acid
E	glutamic acid
G	glycine
Gly	glycine
g	gravitational force
HPLC	high performance liquid chromatography
H	histidine
His	histidine
HCl	hydrochloric acid
H_2O_2	hydrogen peroxide
Hyp	hydroxyproline
O	hydroxyproline
I	isoleucine
Ile	isoleucine
kDa	kilodalton
L	leucine
Leu	leucine
LC	liquid chromatography
LB	Luria-Bertani

Lys	lysine
K	lysine
MS	mass spectrometry
m/z	mass-to-charge ratio
MALDI-TOF	matrix-assisted laser desorption ionization-time of flight
Tm	melting temperature
M	methionine
Met	methionine
nm	nanometer
NMR	nuclear magnetic resonance
OD	optical density
F	phenylalanine
Phe	phenylalanine
PAGE	polyacrylamide gel electrophoresis
PTM	post-translational modification
P	proline
Pro	proline
PDB	Protein Data Bank
rpm	revolutions per minute
S	serine
Ser	serine
SDS	sodium dodecyl sulfate
MS/MS	tandem mass spectrometry
T	threonine
Thr	threonine
TOF	time of flight
TRIS	Tris (hydroxymethyl) aminomethane
Trp	tryptophan
W	tryptophan
Tyr	tyrosine
Y	tyrosine
UV	ultraviolet
V	valine
Val	valine
vWF	von Willebrand Factor
vWF A3	von Willenbrand Factor A3

Chapter 1

Introduction

1.1 Collagen

The human body contains a great deal of collagen protein; it is a component of muscles, skin, bones, tendons, ligaments, cartilages, organs, blood vessels, teeth, and hair (Kielty, C., et. al., 2002). There are twenty nine distinct types of collagen, and collagen type I is the most prevalent (Apweiler, R., et. al., 2004). The protein collagen consists of three polypeptide chains, and some collagens are homotrimers; they contain three identical polypeptide chains (Tkocz, C., et. al., 1968). Heterotrimers also occur, they may contain two or three different polypeptide chains. The forty four unique human collagen genes encoding these alpha chains are located on many different chromosomes: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 17, 19, 20, 21 and X (Apweiler, R., et. al., 2004).

The biosynthesis of collagen is complicated, and processing varies by collagen type. Collagen is first synthesized as a larger pre alpha collagen chain, after which its signal peptide is cleaved off in the lumen of the endoplasmic reticulum (Dalbey, R., et. al., 1997). The enzymes prolyl-3-hydroxylase, prolyl-4-hydroxylase and lysyl hydroxylase respectively hydroxylate prolines and lysines (Kivirikko, K., et. al., 1990). The cofactors involved in hydroxylation are ferrous ion, ascorbic acid, molecular oxygen, and α-ketoglutarate (Hutton, J., et. al., 1966 and Miller, R., et. al., 1979). These hydroxylases are only able to hydroxylate prolines and lysines on free alpha chains; so once the three

chains form a triple-helix these amino acids will no longer be modified (Kielty, C., et. al., 2002). Heat shock protein 47 acts as a collagen molecular chaperone involved in folding (Koide, T., et. al., 2006). Peptidyl-prolyl cis-trans-isomerase converts propyl and hydroxypropyl peptide bonds from a cis to a trans conformation, which increases the rate of protein folding; protein disulfide isomerase forms disulfide bonds (Bachinger, H., 1987 and Bachinger, H., et. al., 1980).

Hydroxylysine undergoes further post-translational modification by the enzymes hydroxylysyl galactosyltransferase and galactosylhydroxylysyl-glucosyltransferase where some hydroxylated lysines are O-glycosylated with galactose and glucose-galactose (Spiro, R., et. al., 1971). After three of the helical collagen alpha chains align at their C terminus and coil around each other, the molecule is packaged into secretory vessels and transported out of the cell and into the extracellular matrix (Kivirikko, K., et. al., 1976 and Bachinger, H., et. al., 1980).

Outside the cell, procollagen N-proteinase and procollagen C-proteinase remove any N-propeptides and C-propeptides (Prockop, D., et. al., 1998). The enzyme lysyl oxidase reacts with lysines and hydroxlysines at the ends of the collagen molecule, telopeptidyl lysyl and hydroxylysyl residues, forming aldehydes called allysine and hydroxyallysine which cross-link (Smith-Mungo, L., et. al., 1998). In the extracellular space, collagen can assemble into macromolecular structures such as fibers, and interact with other matrix proteins.

All collagen molecules contain three polypeptide chains with GXY repeat sequences: G is a glycine (Gly), X is often a proline (Pro), and Y is usually a hydroxyproline (Hyp). These three polyproline II-like chains coil around each other to

form a right-handed super helix (Ramachandran, G., 1967). The chains are staggered by one amino acid relative to each other, with their side chains sticking out, and the glycines packing into the center of the triple helix; they form a rod like protein. Because a glycine amino acid residue is so small, it is able to fit into the center of a triple helix; whereas, other amino acids would disrupt the triple helical structure of collagen. Several diseases in humans have been shown to be due to glycine substitutions; osteogenesis imperfecta, a brittle bone disease, can be caused by glycine substitutions in collagen type I, and the severity of the disease can vary with the location of the substitution (Starman, B., et. al., 1989). A model of collagen generated using X-ray crystallography data from a short triple helical peptide containing only Pro-Pro-Gly repeat sequences illustrates the backbone configuration of the three chains; collagen is much longer and contains an assortment of other amino acids (Figure 1). When prolines in the Y position are hydroxylated, they increase the thermal stability of the collagen triple helix (Privalov, P., et. al., 1979 and Persikov, A., et. al., 2004).

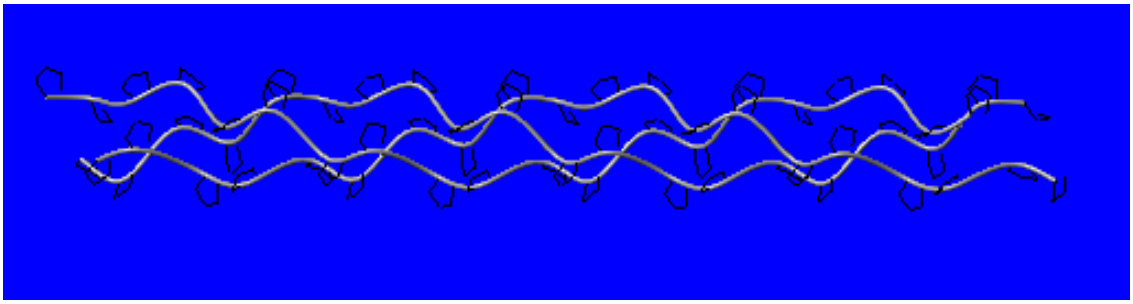


Figure 1: Structure of Collagen Triple Helix. Coordinates for the crystal structure of $[(\text{Pro-Pro-Gly})_{10}]_3$, a thirty amino acid triple helical peptide, were from the Protein Data Bank (PDB) file 1K6F (Berisio, R., et. al., 2002). The proline residues are shown on the surface of the molecule. The model was generated using the molecular graphics program MOLMOL (Koradi, R., et. al., 1996).

The lengths of specific collagen polypeptide alpha chains vary. In humans, collagen polypeptide alpha chains range in size from short having only 421 amino acids (type XXVI alpha 1chain) to as long as 3,152 amino acids in the type VI alpha 3 chain (Apweiler, R., et. al., 2004).

In this study we will focus on the interaction of collagen with collagen receptors. Theses interactions are critical for tissue function and development. Specifically, we will study the interaction of collagen type I and type III with the von Willebrand Factor protein, which is involved in the first step in blood clotting.

1.2 Collagen Type I and Type III

Type I collagen is the most abundant collagen, and it is prevalent in many tissues: bone, cornea, skin, dentine, sclera, vascular tissue, organs, and tendons (Kielty, C., et. al., 2002). It is the major component of bone, where it is found mineralized with calcium hydroxyapatite (Lees, S., 2003). About ten percent of the prolines in collagen type I are hydroxylated (Piez, K., et. al., 1963). Collagen type I is a heterotrimer containing two α1 polypeptide chains and one α2 chain $[(\alpha1(I)]_2\alpha2(I)$. The genes encoding human type I collagen alpha chains are located on different chromosomes. The gene encoding the α1(I) chain is located in humans on chromosome 17 q21.33; the α2(I) gene is located on chromosome 7 q22.1 (Huerre, C., et. al., 1982).

Collagen type I alpha chains are translated as pre procollagens containing a signal peptide and N and C-propeptides which are removed during their biosynthesis. The resulting α1(I) chain contains 1,057 amino acid residues, and its molecular weight is

around 95 kDa; the α2(I) has 1,040 amino acids, and the molecular weight is about 93.5 kDa (Apweiler, R., et. al., 2004). Type I collagen is, therefore, a very large protein with a molecular weight of approximately 282 kDa. Even though the α1(I) and α2(I) chains contain a different number of amino acids, the triple helical domain in both polypeptide chains contain the same number of amino acids (1,014), and it is their N and C-non helical ends called telopeptides which are not the same lengths; the α1(I) N-telopeptide contains seventeen amino acids while α2(I) has only eleven, and the α2(I) C-telopeptide has twenty six amino acids while the α2(I) has just fifteen (Kuivaniemi, H., et. al., 1988 and Tromp, G., et. al., 1988).

Type III collagen is present in the lungs, skin, intestine, liver, lymphoid tisssues, aorta, and blood vessels, and is often in association with type I collagen (Kielty, C., et. al., 2002). Collagen type III is a homotrimer having three identical α1 chains $[\alpha1(III)]_3$. The gene encoding the α1(III) chain is located in humans on chromosome 2q32.2 (Limongi, M., et. al., 1997). Collagen type III was found to contain a high level of 4-hydroxyproline, and a cysteine which is involved in disulfide bond formation (Artimo, P., et. al., 2012).

Just like collagen type I, collagen type III is also translated as a pre procollagen containing a signal peptide and N and C-propeptides which are removed. The final α1(III) chain is longer than either type I alpha chains; it is composed of 1,068 amino acid residues and has a molecular weight around 95 kDa, so the type III triple helical protein has a high molecular weight about 285 kDa (Apweiler, R., et. al., 2004). Collagen type III chain's triple helical domain is fifteen residues larger than type I, and it contains 1,029

amino acid residues; its N and C-telopeptides have fourteen and twenty five amino acids respectively (Ala-Kokko, L., et. al., 1989).

Type I and type III collagen are both fibrillar collagens; in the extracellular matrix these triple helical collagen molecules form fibers (Kielty, C., et. al., 2002). Fiber diffraction and electron microscopy revealed collagen fibrils have a 67 nm repeat pattern, called the D period (Doyle, B., et. al., 1974 and Meek, K., et. al., 1979). A fibril is composed of five 1.4 nm wide triple helical collagen molecules approximately 300 nm in length, 4.4 D periods, that are consecutively staggered 1D period, 234 amino acids, and rotated one-fifth of a circle (Piez, K., et. al., 1974). These fibrils are arranged in a hexagonal fashion in fibers; collagen type I forms fibers which are thicker than type III collagen fibers (Piez, K., et. al., 1974).

Besides providing structural support in many tissues, type I and type III collagen play an important role in the regulation of haemostasis where they are major components in the wall of blood vessels. In the process of blood clotting, they bind to the von Willebrand Factor protein (Fitzsimmons, C., et. al., 1988).

1.3 The von Willebrand Factor

The von Willebrand Factor (vWF) is a large globular glycoprotein found in blood plasma, and the subendothelial matrix; the gene encoding human von Willebrand Factor is located on chromosome 12p13.3 (Ginsburg, D., et. al., 1985). Mutations in the von Willebrand Factor gene cause von Willebrand disease, an inherited bleeding disorder (Ginsburg, D., et. al., 1993).

The biosynthesis of von Willebrand Factor protein is very interesting; it occurs in endothelium and megakaryocytes where it is stored in secretory granules called Weibel-Palade bodies and platelet α-granules respectively (Sadler, J., 1998). The initial translation contains a 22 amino acid signal peptide, a 741 amino acid von Willebrand antigen 2, and the von Willebrand Factor whose protein has 2,050 amino acids; an individual von Willebrand Factor polypeptide has a molecular weight of 225 kDa (Apweiler, R., et. al., 2004). The signal peptide targets the pre propolypeptide to the endoplasmic reticulum where it is removed by a signal peptidase inside the lumen (Dalbey, R., et. al., 1997). The propeptide is post-translationally modified; it undergoes N-linked glycosylation and O-linked glycosylation (Samor, B., et. al., 1986 and Schulte am Esch, J., et. al., 2005).

The von Willebrand Factor propeptide is composed of four types of domains arranged as follows: D1-D2-D'-D3-A1-A2-A3-D4-B1-B2-B3-C1-C2-CK; the first two subunits, D1and D2, form von Willebrand antigen 2 (Sadler, J., 2009). The von Willebrand propeptide dimerizes at its C-terminus, and is transported to the Golgi apparatus; in the trans-Golgi a protease cleaves off the von Willebrand antigen 2, allowing the von Willebrand protein to form multimers whose D3 subunits are linked by disulfide bonds (Vischer, U., et. al., 1994). The multimeric von Willebrand Factor can consist of up to 80 monomers, and is either stored inside secretory granules or released directly into the plasma (Sadler, J., 1998).

The von Willebrand protein is a multifunctional protein consisting of twelve subunits (Figure 2). Its D'/D3 domain binds to the blood clotting protein factor VIII (Takahashi, Y., et. al., 1987). The A1 domain binds to heparin, and to the platelet GP1b

receptor (Fujimura, Y., et. al., 1987 and Miura, S., et. al., 2000). The C1 domain binds to platelet integrin αIIbβ3, (GP IIb-IIIa), at its arginine-glycine-aspartic acid amino acid sequence motif (RGD) (De Marco, L., et. al., 1986 and Ruggeri, Z., 1997). The CK domain dimerizes with another CK domain via disulfide bonds (Katsumi, A., et. al., 2000). After a vascular injury, the von Willebrand Factor is involved in platelet adhesion and aggregation; the A3 domain is considered the collagen binding domain, and binds to collagen during the blood clotting process (Lankhof, H., et. al., 1996).

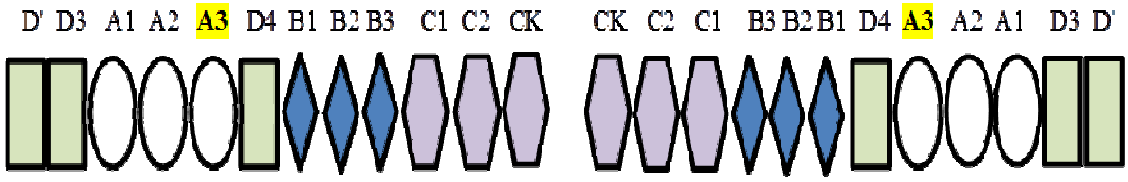


Figure 2: von Willebrand Factor Subunits. The protein has four types of domains: A, B, C and D. The monomer forms a dimer at the C-terminus CK domain and then forms disulfide D3 linked multimers (Sadler, J., 1998). The A3 domain which is involved in collagen binding is highlighted in yellow (Lankhof, H., et. al., 1996).

Human von Willebrand Factor A3 domain has been cloned and can be expressed in Escherichia coli. Our work focuses on the interaction of collagen type I and type III with the human von Willebrand Factor A3 domain, an important association in blood clot formation.

1.4 The von Willebrand Factor A3 Domain

The von Willenbrand Factor A3 (vWF A3) domain binds to exposed collagens type I and type III, after a blood vessel is injured (Fitzsimmons, C., et. al., 1988 and Lankhof,

H., et. al., 1996). This is the initial step in the formation of a platelet plug (Figure 3). When a blood vessel wall is ruptured the von Willebrand Factor protein A3 domain binds to the exposed collagen and conformational changes in the A3 domain allow the von Willebrand Factor A1 domain to then bind to the Gp1b platelet receptor; the platelet is pulled down to plug the injured blood vessel wall, and another platelet receptor protein, α2β1, binds to collagen (Obert, B., et. al., 1999).

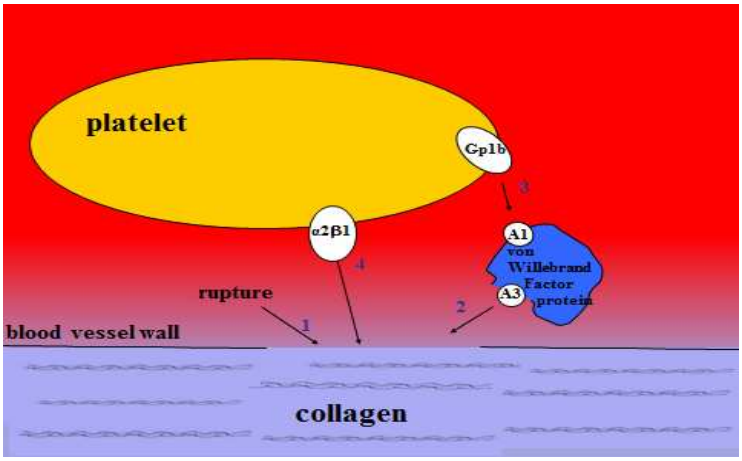


Figure 3: Collagen and von Willebrand Factor Role in Blood Clot Formation. The first four steps in blood clot formation are numbered with arrows. A1 is the von Willebrand Factor A1 domain, A3 is von Willebrand Factor A3 domain, α2β1 and Gp1b are platelet receptor proteins.

The von Willebrand A3 domain is made up of 181 amino acids, and has a molecular weight of 19.3 kDa (Apweiler, R., et. al., 2004). The amino acids in this sequence do not contain any post-translational modifications; of the twelve Asn-linked and ten Thr/Ser-linked oligosaccharide chains on the von Willebrand Factor, none are present in the A3 domain (Titani, K., et. al., 1986).

The structure for the von Willebrand Factor A3 domain has been examined using X-ray crystallography, and coordinates for the crystal structure are available in the Protein Data Bank (Bernstein, F., et. al., 1977). The von Willebrand Factor A3 domain consists of five central beta strands surrounded by six alpha helices (Figure 4). The structure of the von Willebrand Factor A3 domain is similar to integrin I domains, now also called integrin A domains; however, the von Willebrand Factor A3 domain does not bind metal ions and integrin does (Bienkowska, J., et. al., 1997). Integrin proteins are cell receptor proteins, and several of them such as the α1β1 and α2β1 protein bind to many sites on collagen; platelets contain α1β1 and α2β1 (Xu, Y., et. al., 2000).

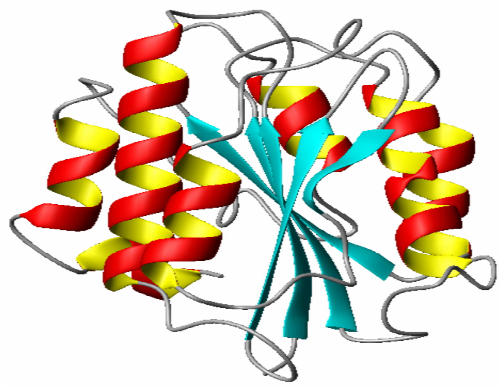


Figure 4: Structure of the von Willebrand Factor A3 Domain. Coordinates for the crystal structure were from the Protein Data Bank (PDB) file 1ATZ (Huizinga, E., et. al., 1997). The model was generated using the molecular graphics program MOLMOL (Koradi, R., et. al., 1996).

How this von Willebrand Factor A3 domain interacts with type I and type III collagen is not fully understood. Many binding sites on collagen type I and collagen type III have been found for the structurally similar integrin proteins, yet, only one von

Willebrand Factor protein A3 binding site on collagen has been located, and that site is on collagen type III (Lisman, T., et. al., 2006). This was done using short synthetic triple helical peptides; studies using full chain collagen have revealed there are multiple binding sites, but the exact location of these binding sites were not identified (Fitzsimmons, C., et. al., 1988). A different approach for finding the position of these binding sites using full chain collagen and the von Willebrand Factor A3 protein, and employing a mass spectrometry oxidative protein footprinting technique, was investigated; in order to oxidize proteins, Fenton chemistry was used to form hydroxyl radicals.

1.5 Fenton Chemistry Protein Footprinting by Mass Spectrometry

Fenton chemistry oxidation followed by mass spectrometry has been used to study globular proteins (Sharp, J., et. al., 2003). Fenton chemistry could be used, as an oxidation method, to map the binding sites of the von Willebrand Factor A3 protein on full chain collagen by oxidative footprinting followed by mass spectrometry. The advantage of using Fenton chemistry to generate hydroxyl radicals is that the technique is inexpensive, and the reagents are easily obtainable. A disadvantage of using Fenton chemistry is potential peptide backbone cleavage (Heyduk, T., et. al., 2001).

Fenton chemistry involves the oxidation of ferrous iron (Fe^{2+}) and the reduction of hydrogen peroxide (H_2O_2). The hydroxyl radicals that are produced can irreversibly modify the side chains of amino acids; a reaction which occurs faster than the cleavage of amide bonds (Heyduk, T., et. al., 2001). The rate at which amino acids in a peptide are

oxidized varies; and the order of reactivity is cysteine, methionine > tryptophan, tyrosine, phenylalanine, cystine > histidine > leucine, isoleucine > arginine, lysine, valine > serine, threonine, proline > glutamine, glutamic acid > aspartic acid, asparagine > alanine > glycine (Takamoto, K., et. al., 2006). Oxidation usually forms an alcohol, aldehyde, or ketone group with a corresponding mass increase of +16 or +14 Da which can then be monitored by mass spectrometry (Figure 5).

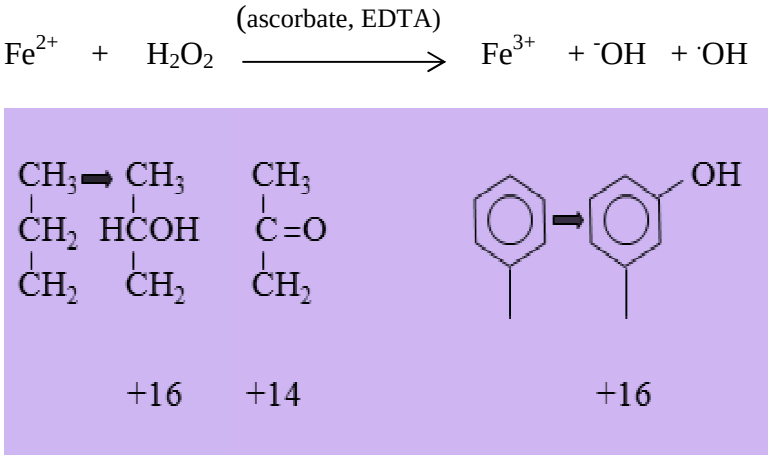


Figure 5: Fenton Chemistry Side Chain Reactions: Oxygen is added to amino acid side chains. Alcohol formation increases the mass by 16 Da. Ketone formation increases the mass by 14 Da.

Even though the oxidation of amino acid side chains occurs faster than the cleavage of the amide bonds, the reaction needs to be controlled (Heyduk, T., et. al., 2001). The efficiency of an aqueous Fenton reaction depends on the Fe^{2+}: H_2O_2 ratio and the pH. (Dercova, K., et. al., 1999). The iron is chelated to prevent nonspecific binding to collagen (Heyduk, E., et. al., 1994). Ascorbate is used to reduce Fe^{3+} back to Fe^{2+} (Elmagirbi, A., et. al., 2012).

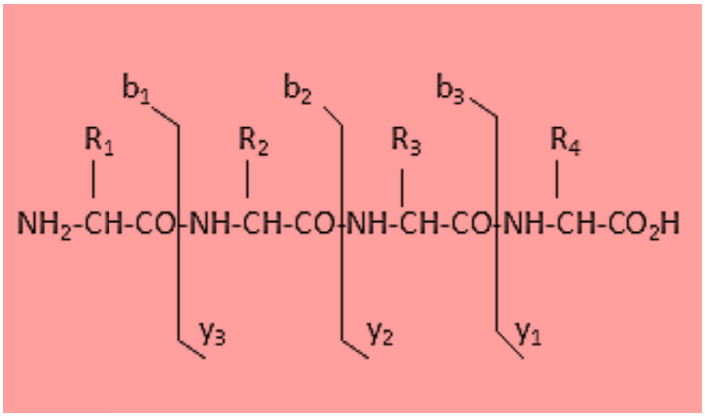
In this study we used Fenton chemistry to generate hydroxyl radicals for protein footprinting. In a typical protein footprinting method, two potential interacting proteins are placed in the same tube and allowed to interact; hydroxyl radicals are then introduced which oxidize any amino acid residues not being protected in the binding complex. In order to determine which residues are involved in the binding interaction and are not oxidized, the complex is first broken down into smaller pieces using an enzyme and these small fragments are sequenced using mass spectrometry.

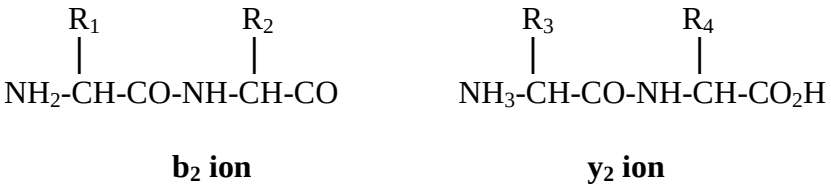
1.6 Mass Spectrometry Oxidative Protein Footprinting

This technique employs mass spectrometers which are instruments that measure the mass to charge ratio of analytes (m/z), and provide structural information. A mass spectrometer consists of an ion source, a mass analyzer, and a detector. The ion source generates charged species called ions. The mass analyzer measures the (m/z) ratio of the ionized analytes, and the detector records the number of ions. Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) and electrospray ionization (ESI) are two techniques used to volatize and ionize proteins and peptides. MALDI-TOF ionizes analytes out of a dry crystalline matrix using laser pulses (Karas, M., et. al., 1988); whereas, ESI ionizes analytes out of a solution (Fenn, J., et. al., 1989).

Collagen is a very large protein, and needs to be broken chemically or enzymatically into peptides for MS/MS analysis. The enzyme trypsin can be used to digest proteins in solution or from a gel (Rosenfeld, J., et. al., 1992). Trypsin cleaves amide bonds after arginines (R) and lysines (K). In MS/MS, peptide bonds are broken with a collision gas

inside a mass spectrometer. The sequence of the peptide can be determined by measuring the fragmented ions m/z ratios. Inside a mass spectrometer, ions are selected for by their m/z, and then in MS/MS, peptide bonds are broken using a neutral collision gas. A tryptic peptide typically breaks apart along the peptide backbone and a series of ions from the N-terminus (b-ions), and a series of ions from the C-terminus (y-ions) are produced (Figure 6). Mass differences between ions in the b series or the y series give the residual mass of an amino acid, and the sequence of the peptide can be determined from this mass.





b_2 ion **y_2 ion**

Figure 6: Peptide Fragmentation Pattern Inside A Mass Spectrometer. Peptide backbone cleavage into b and y ions. The structure of a b_2 ion and a y_2 ion is illustrated. (Roepstreorff, P., 1984).

In mass spectrometry oxidative protein footprinting, protein interactions are determined via protein modification, proteolysis, and mass spectrometry analysis. Hydroxyl radicals oxidize exposed amino acid side chains increasing the mass by +16 or +14 Da. The oxidized protein is then digested with a protease into small peptide fragments, separated over a HPLC column, and introduced into a mass spectrometer. Oxidation sites on a peptide are identified by tandem mass spectrometry (MS/MS) analysis. Target amino acid sequences in a protein covered by its binding partner are protected from oxidation and would not be modified. The location of a binding site would be elucidated by comparing the MS/MS data from the binding experiment with controls (Figure 7).

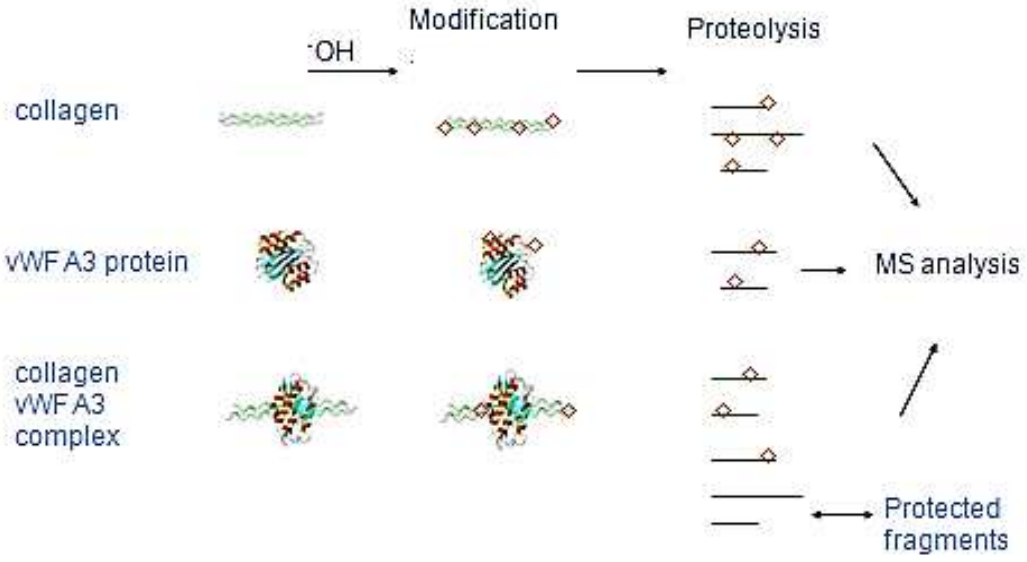


Figure 7: Mass Spectrometry Oxidative Protein Footprinting of Collagen. Collagen and von Willebrand Factor A3 controls are oxidized, digested, and analyzed using mass spectrometry, along with the collagen von Willebrand Factor A3 complex.

1.7 Objectives

The von Willebrand Factor A3 domain is known to bind collagen type I and type III. Studies using short triple helical peptides have identified a single high affinity binding site on collagen type III, but no sites were identified for type I (Lisman, T., et. al., 2006). This approach of using short peptides is limited; binding interactions may require a more extensive structural arrangement, so not all binding sites would be identified. The exact locations of all of the sites where the von Willebrand Factor protein binds to collagen are still unknown. The aim of this research was to investigate mass spectrometry protein footprinting as a way to study full chain collagen interactions with the von Willebrand Factor A3 protein. In order to accomplish this task, we needed to carry out the following original studies:

- Binding of recombinant von Willebrand Factor A3 protein with collagen, and triple helical peptides.
- A procedure for oxidizing a triple helix using Fenton chemistry.
- Mass spectrometry of collagen, which is larger and more modified than most proteins.

Chapter 2

Binding of the von Willebrand Factor A3 Protein to Full Chain Collagen

2.1 Introduction

The binding of the von Willebrand Factor to collagen is the first step in blood clot formation, and there are still many things that are not known about this binding process. Early experiments involving the chemical modification of amino acids showed that arginine residues in collagen and the carboxyl groups in the von Willebrand Factor protein are involved in the binding of collagen to the von Willebrand Factor protein (Fitzsimmons, C., et. al., 1988). Multimeric von Willebrand Factor protein was shown to specifically bind to several sites along the collagen triple helix (Fitzsimmons, C., et. al., 1988). Fitzsimmons demonstrated that this multimeric protein binds to cyanogen bromide fragments of bovine collagen type Iα1 and type IIIα1. Two cyanogen bromide fragments from bovine type III collagen (peptides 4 and 5), and four cyanogen bromide fragments from type Iα1 bovine collagen (peptides 8, 3, 7, and 6b) were identified which bind to the von Willebrand Factor protein, but it was not clear where the von Willebrand Factor protein was binding (Fitzsimmons, C., et. al., 1988). Cyanogen bromide cleaves after methionine (M) residues; the cyanogen bromide cleaved proteins, predicted from their molecular weights, are highlighted in Figure 8 and Figure 9. Bovine and human collagen type I α1 and type III α1 protein sequences are similar; CLUSTAL W alignment of bovine and human collagen type I α1 and bovine and human collagen type III α1 of each chain revealed they are 98.3 % and 90.5 % identical (Thompson, J., et. al., 1994). It

seems likely that the von Willebrand Factor protein would also bind to human collagen at multiple sites.

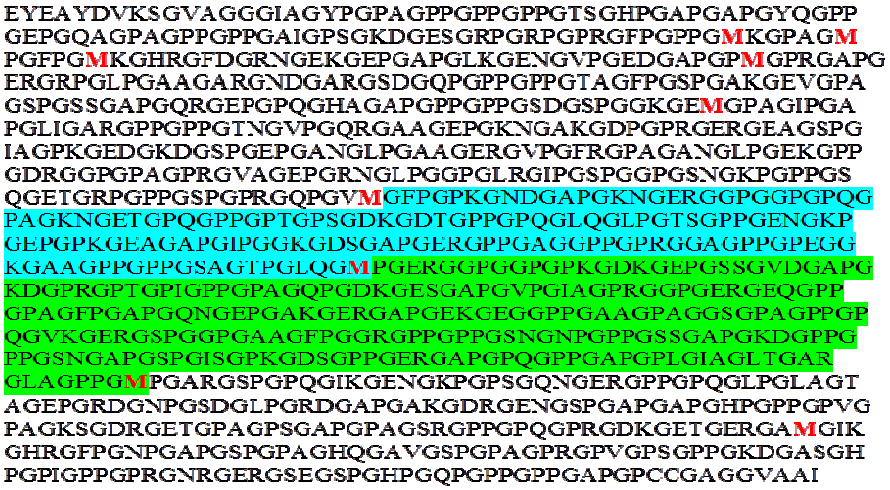


Figure 8: Bovine Collagen III α1 Peptides that Bind to von Willebrand Factor. Bovine collagen type III sequence is from UniProt (Apweiler, R., et. al., 2004). The peptide 4 is highlighted in aqua, and 5 is in green. Methionines (M) are shown highlighted in red.

Figure 9: Bovine Collagen I α1 Peptides that Bind to von Willebrand Factor. Bovine collagen type I α1 sequence is from UniProt (Apweiler, R., et. al., 2004). Peptide 8 is highlighted yellow, 3 is aqua, 7 is green, and 6b is blue. Methionines are highlighted in red.

The von Willebrand Factor domain which binds to type I and to type III collagen was identified as its A3 domain (Lankhof, H., et. al., 1996). Recombinant von Willebrand Factor A3 domain protein was shown to have a lower affinity for collagen than the multimeric von Willebrand Factor protein; the K_d was 1.8 uM instead of being around 2 nM (Van der Plas, R., et. al., 2000 and Cruz, M., et. al., 1995).

X-ray crystallography studies predicted that this von Willebrand Factor A3 domain collagen binding site is small, containing only a few amino acid residues, and a limited sequence pattern (Huizinga, E., et. al., 1997). The predicted number of residues involved in binding is around eight (Romijn, R., et. al., 2003). Since collagen type I and type III contain over 1,000 amino acid residues, this would allow the von Willebrand Factor A3 domain to interact with collagen at several locations along the triple helix as was observed by Fitzsimmons (Figure 8 and 9). A model of the von Willebrand Factor A3 domain binding to a short segment of collagen, approximately one-tenth the size of native collagen type I is depicted in Figure 10.

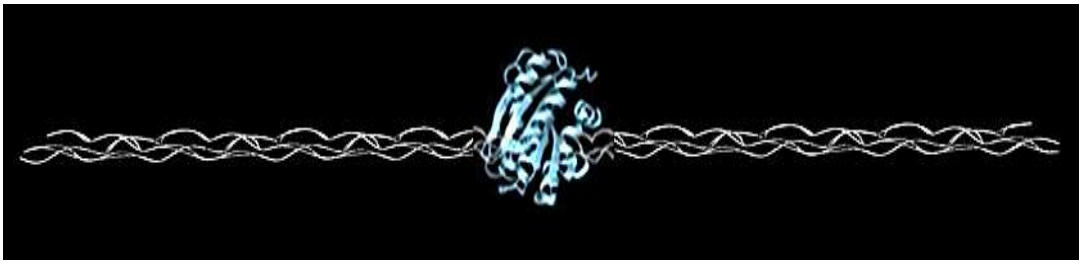


Figure 10: von Willebrand Factor A3 Domain on Collagen. The model was generated using the molecular graphics program MOLMOL (Koradi, R., et. al., 1996) with the coordinates for the crystal structures from PDB files 1ATZ (Huizinga, E., et. al., 1997) and 1K6F (Berisio, R., et. al., 2002).

Recent studies using a cross-saturation technique have revealed that the front face of the von Willebrand Factor A3 domain is where collagen binds (Nishida, N., et. al., 2003). Mutant von Willebrand Factor A3 studies and computer docking work revealed the binding surface displays a shallow groove, negatively charged residues, and a hydrophobic patch; an electrostatic interaction with a positively charged residue on collagen has been predicted, as well as a hydrophobic interaction (Romijn, R., et. al., 2001 and Romijn, and R., et. al., 2003).

Studies using synthetic triple helical peptides have identified only one binding site on collagen III; this site consists of nine amino acids, and contains a hydroxyproline which was considered important (Lisman, T., et. al., 2006). The type III collagen binding site sequence is Arg-Gly-Gln-Hyp-Gly-Val-Meth-Gly-Phe. From these studies, the amino acid residues Arg, Hyp, Val and Phe were considered invariable; whereas, Gln and Met were not (Lisman, T., et. al., 2006). This collagen binding sequence in type III collagen is identical in humans and in cows; the last two amino acid residues Gly-Phe are the first two in bovine type III cyanogen bromide fragment 4 in Figure 8. No sites were identified for type I collagen.

The study by Lisman et. al., utilizes synthetic homotrimer peptides; the peptide length is limited to only twelve amino acid residues that are surrounded by Gly-Pro-Pro sequences needed for triple helical formation. Full chain collagen is long, over 1,000 amino acids, so synthetic peptides may not represent all of the attributes of a collagen molecule, which may be one reason why no other binding sites in human collagen type III have been identified, and none for collagen type I. More than one binding site is expected, since two cyanogen bromide fragments from type III collagen, and four from

type I collagen were implicated in binding (Fitzsimmons, C., et. al., 1988). The von Willebrand Factor A3 domain binds to other sites on collagen type III and type I, and these sites have not been identified. The binding interactions of the von Willebrand Factor A3 domain with full chain collagen are still poorly understood; exactly how many collagen binding sites there are, their affinities, and their locations have not been determined. Understanding the details of this interaction is important for the successful designing of drugs to treat von Willebrand disease, a bleeding disorder. In order to determine the location of any binding sites, the von Willebrand Factor A3 protein was purified for use in binding and oxidative footprinting experiments. The binding of full chain collagen was studied with a recombinant von Willebrand Factor A3 protein. In this chapter, the purification and characterization of the von Willebrand Factor A3 domain protein will be presented first. In the second part, the binding study will be presented.

2.2 Materials and Methods

2.2.1 Materials: Molecular weight markers, Tris-HCl, sodium chloride, sodium hydroxide, sodium phosphate monobasic, sodium phosphate dibasic, acetic acid, urea, lysozyme, bovine serum albumin (BSA), cytochrome C, rat collagen type I, human collagen type III, and human collagen type I were purchased from Sigma. The peptides $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Pro\text{-}Hyp\text{-}Gly)_{10}$ were from Peptides International. Thermo Scientific Precise gels, Talon resin, imidazole, trifluoroacetic acid, and acetonitrile were purchased from Fisher Scientific.

2.2.2 Peptide P23: Peptide P23 was synthesized at The Rockefeller University Proteomics Center by Dr. Nagarajan Chandramouli. Lisman showed this peptide binds to von Willebrand Factor. P23 sequence: Acetyl-Gly-Pro-Hyp-Gly-Pro-Arg-Gly-Gln-Hyp-Gly-Val-Met-Gly-Phe-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-NH_2.

2.2.3 Expression and Protein Purification of von Willebrand Factor A3: The bacteria pET 20b (+) plasmid containing the human von Willebrand Factor A3 insert with a C-terminal His-tag was a kind gift from Dr. Chang Geng Ruan (Suzhou Medical College, Suzhou, China). The plasmid DNA was inserted into DH1 cells, purified with Promega's Wizard Plus SV DNA Mini Prep Purification System, and used to transform Stratagene's E. coli BL21 (DE3) pLysS competent cells for protein expression, according to the manufacturer's protocol.

Three microliters of E. coli BL21 (DE3) pLysS cells containing pET 20b (+) plasmid with the gene for human von Willebrand Factor A3 protein was streaked onto a Luria-Bertani amp^+ plate and grown overnight in an incubator set at 37 °C. For purification, a single bacterial colony was removed, placed in ten milliliters of Luria-Bertani amp^+ media, and grown overnight at 37 °C with shaking. The cells were transferred into one liter of Luria-Bertani amp^+ media; the transformed cells were grown at 37 °C with shaking. At an OD_{600} of 0.5, a sample of the bacteria was removed for gel analysis, and the rest of the one liter transformed cells were induced with 1mM isopropyl-B-D-thiogalactopyranoside overnight. Since the expressed protein is produced and retained inside the bacterial cells, the Escherichia coli cells were harvested by centrifugation for twenty minutes at 10,000 rpm at 4 °C. The cell pellet containing the

von Willebrand Factor A3 protein was resuspended in twenty five milliliters of 50 mM phosphate buffered saline containing 1mM phenylmethylsulfonyl fluoride which is a serine protease and prevents protein degradation. A 100 ul sample of the induced bacteria was removed for gel analysis. The E. coli cells were broken open using one milliliter of lysozyme (20mg/mL), freeze thawed three times, and sonicated four times in order to release the protein. This was done by placing the sample into liquid nitrogen for twenty minutes and then defrosting it in a water bath at 37 °C. The sample was sonicated on ice for thirty seconds; the sonicator machine output was set at three, and the duty cycle was set at fifty percent.

The solution was centrifuged, to clear out any debris and aggregated protein. In order to determine where the von Willebrand Factor A3 protein was and if it had aggregated and formed inclusion bodies which would be insoluble and reside with the cell debris, samples were taken from the supernatant and the pellet. For gel analysis, forty microliters from the supernatant was removed, and a little bit of the pellet was solubilized in 100 ul of sample loading buffer. The von Willebrand Factor A3 protein was found mainly in the cell pellet as an inclusion body so the cell pellet was solubilized overnight in twenty five milliliters of 50 mM phosphate buffered saline containing 8M urea.

The expressed von Willebrand Factor A3 protein contains a C-terminal His-tag. There is a series of histidine amino acids which allow a recombinant protein to be purified with an affinity resin which will selectively bind to histidine. The solubilized von Willebrand Factor A3 protein was affinity purified with Clontec Talon® Metal Affinity resin, which selectively binds to His-tagged proteins. Imidazole was used to

release His-tagged proteins from the talon resin, since it competes with histidine for binding to the resin. The solubilized von Willebrand Factor A3 protein was incubated with 1 ml of talon resin for half an hour, the talon resin was washed three times with ten milliliters of 50 mM phosphate buffer saline, and the recombinant protein was eluted in ten milliliters of 50 mM phosphate buffer saline with 300 mM imidazole following the manufacturer's instructions. Forty microliters were saved for gel analysis. Samples were mixed with loading buffer, the 5X SDS sample loading buffer contained 60 mM Tris-HCl pH 6.8, 25 % glycerol, 2 % sodium dodecyl sulfate, 350 mM dithiothreitol, and 0.1 % bromophenol blue, and were run on a SDS-PAGE gel under reducing conditions, and stained with coomassie blue G-250.

The recovered protein was refolded by dialysis in one liter of 6M urea which was diluted overtime with 50 mM phosphate buffer saline to a final volume of three liters. The sample was further dialyzed three times in one liter of 50 mM phosphate buffer saline. The concentration of the von Willebrand Factor A3 protein was determined to be 1 mg/mL at a UV absorbance of 280 nm, and an extinction coefficient of 0.5 cm^{-1} $(mg/mL)^{-1}$, and the protein was stored at 4 °C.

2.2.4. Expression and Purification of 877 Protein: A pET 32a (+) vector containing an N-terminal His-tag-thioredoxin-collagen I α1 fragment fusion protein (877) was created in the Xu lab by modifying the $(Gly\text{-}Pro\text{-}Pro)_{10}$-foldon insert of Dr. Jurgen Engle's lab (Xu, K., et. al., 2008). Escherichia coli containing the 877 pET 32a (+) plasmid were induced overnight with 1mM isopropyl-B-D-thiogalactopyranoside. The His-tagged fusion protein was purified from the cell extract using Clontec Talon® Metal

Affinity resin, and eluted with imidazole; the 877 fusion protein was cleaved using Thrombin CleanCleave® kit (Sigma) according to the manufacturer. The 877 protein was purified over a C_8 column with a gradient of 0.1 % trifluoroacetic acid and 100 % acetronitrile on a Beckman HPLC, and then lyophilized.

2.2.5 Circular Dichroism: The secondary structure of proteins and peptides were determined by CD measurements using an AVIV 200 CD spectrometer with a 1mm optical path length, a 1nm bandwidth, a Peltier thermal controller, and a wavelength between 190 nm-260 nm. Measurements were performed with 200 ul of 50 mM phosphate buffer saline, and then using the same cuvette, with 200 ul of 0.2 mg/mL von Willebrand Factor A3 protein in 50 mM phosphate buffer saline pH 7.0 at 4 °C. Samples were equilibrated for five minutes at 4 °C.

2.2.6 Pull Down Assay: The His-tagged von Willebrand Factor A3 protein and potential binding proteins were mixed in a one to one mole ratio with the final concentration of each protein being 3 uM, except for cytochrome C which was 12 uM. An excess of cytochrome C was used in a mole ratio of one to four. Twenty six micrograms of the von Willebrand Factor A3 protein (26 kDa) was used in each experiment in a total reaction volume of 300 ul in 50 mM phosphate buffer saline pH 7.0. The eppendorf containing the von Willebrand Factor A3 protein and the potential binding partner was placed on a rotator at 4 °C, for forty five minutes. Each sample was then transferred into a different 0.5 ml eppendorf tubes containing 50 uL of Clontec Talon® Metal Affinity resin, and placed for one hour at 4 °C on a rotator. The supernatant

containing any nonbinding substances was removed, and the talon resin was washed several times with 50 mM phosphate buffer saline. The von Willebrand Factor A3 protein along with any protein interacting with it was eluted from the resin with 50 ul of 300 mM imidazole according to the manufacturer. All pull down assay experiments were repeated three times. Protein samples were analyzed by SDS-PAGE followed by coomassie blue staining. Ten microliters of the collagen elution, and twenty microliters of bovine serum albumin, cytochrome C, and the 877 peptide were loaded onto the gel. Bovine serum albumin, the von Willebrand Factor A3 protein, the 877 protein, and the peptide samples were chromatographed over a C_{18} column on a Beckman HPLC using a gradient of 0.1 % trifluoroacetic acid and 100 % acetonitrile. Fifty microliters of the imidazole elution samples were also chromatographed.

2.3 Results

2.3.1 The von Willebrand Factor A3 Protein is Purified as An Inclusion Body:

The von Willebrand Factor A3 protein was expressed as a recombinant His-tagged protein. It was purified as an inclusion body (Figure 11). pLysS A3 cells do not express the von Willebrand Factor A3 protein, after they are induced overnight with isopropyl-B-D-thiogalactopyranoside, the Escherichia coli cells express the foreign gene, and produce the recombinant protein (Figure 11A). The von Willebrand Factor A3 protein was overexpressed in E. coli and formed inclusion bodies, the protein was found mainly in the pellet and not in the supernatant (Figure 11B). The recombinant His-tagged protein was

purified from the pellet with an affinity resin. On a SDS-PAGE gel the talon resin purified von Willebrand A3 protein runs at a molecular weight of around 29 kDa (Figure 11C).

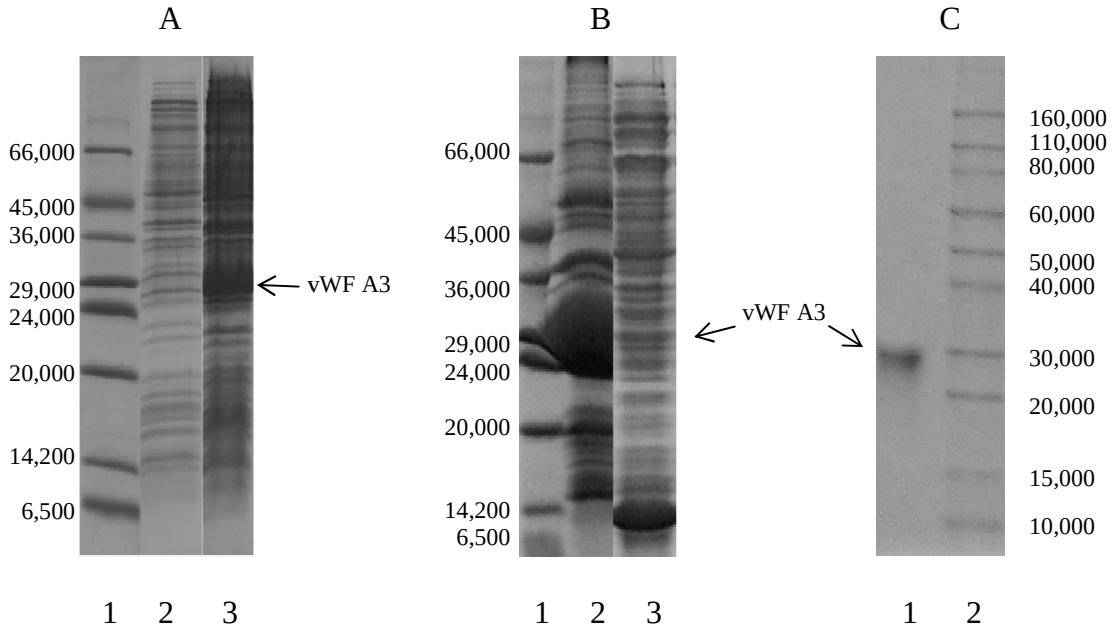


Figure 11: Expression and Purification of von Willebrand Factor A3. A. Lane 1. Sigma marker. Lane 2. A3 pLysS cells before induction. Lane 3. Induced A3 pLysS cells. B. Lane 1. Sigma marker. Lane 2. Lysed induced A3 pLysS cell pellet. Lane 3. Lysed induced A3 pLysS cell supernatant. C. Lane 1. Talon resin purified vWF A3 protein. Lane 2. Marker. All samples were reduced with dithiothreitol, run on a SDS-PAGE gel, and stained with coomassie blue.

In the induced and lysed pLysS A3 cell supernatant there is a protein band running around 29 kDa (Figure 11B). This band is also present in the non induced pLysS A3 cells (Figure 11A). The presence of the 29 kDa band in Figure 11B lane 3 is not indicative of the presence of the von Willebrand Factor A3protein. We therefore, did not purify the induced pLysS A3 cell supernatant.

2.3.2 CD Spectrum of the von Willebrand Factor A3 Protein: The purified recombinant vWF A3 protein folded into the expected conformation. The secondary structure for the von Willebrand A3 protein was determined by circular dichroism measurements using CDNN2.1 spectra deconvolution software. The helical content of the von Willebrand Factor A3 protein was found to be 32.3 % and the beta content was 17.4 %. This is close to the predicted helical content of the von Willebrand Factor A3 domain of 35.6 % and beta content of 20.8 % (Huizinga, E., et. al., 1997). Delta Epsilon, also known as molar circular dichroism, was used to measure circular dichroism, and is defined as $\Delta\varepsilon = \varepsilon L - \varepsilon R$ where εL is the extinction coefficient of left circularly polarized light, and εR is the right (Figure 12).

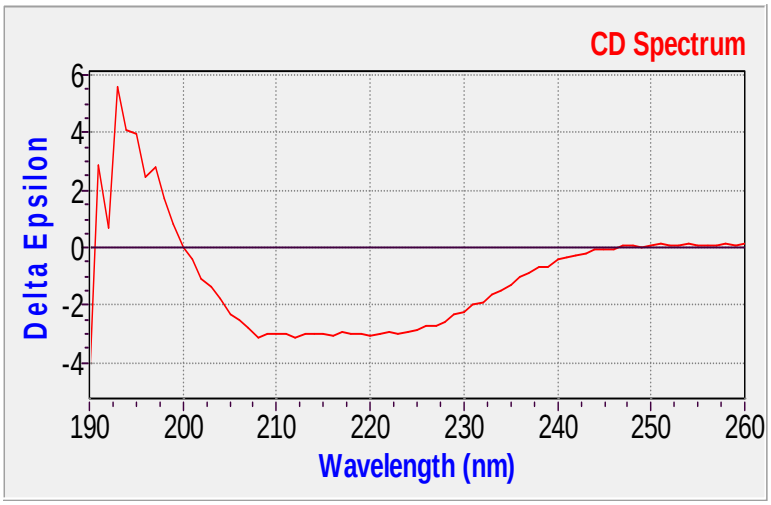


Figure 12: CD Spectrum of von Willebrand Factor A3. Circular dichroism measurements of the recombinant von Willebrand A3 protein were recorded using Delta Epsilon at a wavelength between 190 nm-260 nm for 0.2 mg/ml von Willebrand Factor A3 protein in 50 mM phosphate buffer saline pH 7.0 equilibrated for five minutes at 4 °C.

2.3.3 SDS-PAGE of Pull Down Assay. The binding of type I and type III collagen, and 877 with the von Willebrand Factor A3 protein were analyzed using the pull down assay.

After incubation with the von Willebrand Factor A3 protein and each of its binding partners with the talon affinity resin, the resin was thoroughly washed to remove any non-specific bound protein (Figure 13, Lane 2 of A, B and C) and then eluted with 300 mM imidazole. The fact that the three proteins are all visible in the elution samples indicate they bind specifically to the von Willebrand A3 protein, and thus coeluted with the His-tagged von Willebrand A3 protein (Figure 13). On the contrary, when the same experiment is carried out with the control proteins bovine serum albumin and cytochrome C, the proteins were present only in the non-binding fraction, and are not visible in the imidazole eluent sample; even when the concentration of cytochrome C is increased dramatically, still no cytochrome C could be seen to coelute with the von Willebrand Factor A3 protein (Figure 14).

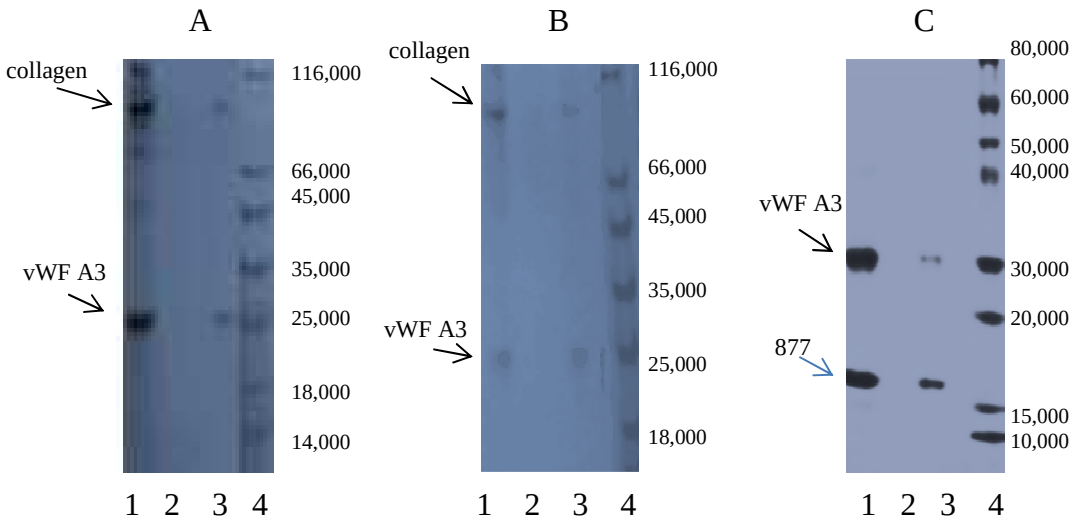


Figure 13: SDS-PAGE of the vWF A3 Protein Pull Down Assay Showing the Interaction of vWF A3 with Collagen Type I (A), Collagen Type III (B), and 877 Protein (C). In all three cases Lane 1. Mixture of vWF A3 protein and binding partner. Lane 2. Wash. Lane 3. Imidazole elution. Lane 4. Marker. All samples were reduced with dithiothreitol, run on a PAGE gel, and stained with coomassie blue.

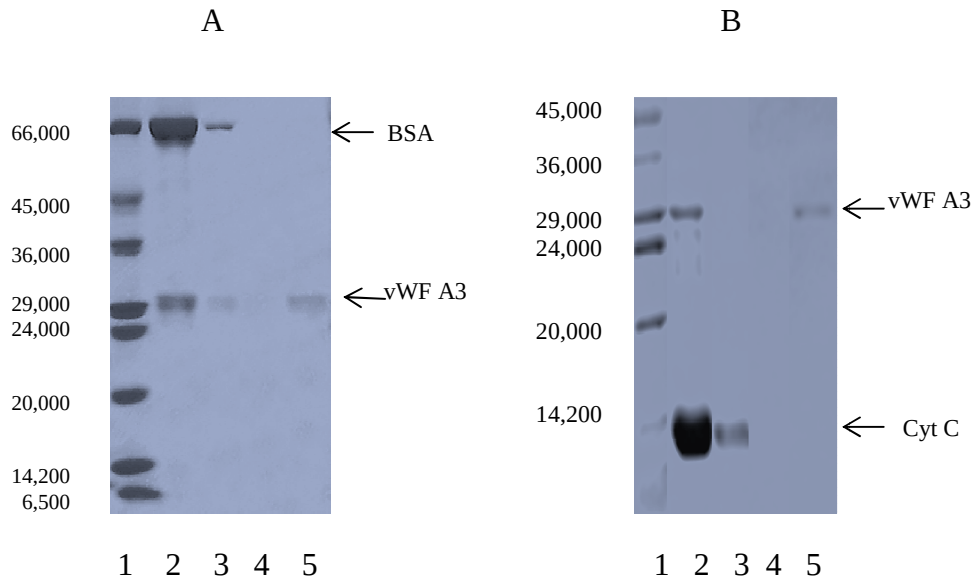


Figure 14: SDS-PAGE of the vWF A3 Protein Pull Down Assay Controls. A: Lane 1. Sigma marker. Lane 2. Bovine serum albumin and vWF A3. Lane 3. Nonbinding. Lane 4. Wash. Lane 5. Imidazole elution. B: Lane 1. Sigma marker. Lane 2. Cytochrome C and vWF A3. Lane 3. Nonbinding. Lane 4. Wash. Lane 5. Imidazole elution. All samples were reduced with dithiothreitol, run on a PAGE gel, and stained with coomassie blue.

This pull down assay demonstrates that type I and type III collagen bind to the von Willebrand Factor A3 protein as expected. Surprisingly, 877 was also found to bind to the von Willebrand A3 protein. The 877 peptide is a triple helical protein containing sixty three amino acid residues, 877-939, from the human collagen type I alpha 1 chain. The 877 protein is obtained from Escherichia coli, and does not contain hydroxyprolines

In Figure 13 A-B, the detection of the von Willebrand Factor A3 protein and collagen is reduced because less imidazole elution sample was loaded onto the gel. To understand if the specific sequence of 877 facilitates binding, or if a general triple helix conformation does, binding experiments were done using the following triple helical peptides: $(Pro\text{-}Pro\text{-}Gly)_{10}$, $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and P23.

2.3.4 High Performance Liquid Chromatography of Pull Down Assay: The von Willebrand Factor A3 binding interactions with triple helical peptides were investigated using the same pull down assay, but the imidazole eluents were analyzed by high performance liquid chromatography, using a C_{18} column, because the peptides do not stain well with commassie blue and are small. For the following experiments, the species in the imidazole eluent were identified based on their retention time on the chromatogram. The retention time of each species was determined using a pure sample on the column and using the same solvent gradient. The results of binding of the von Willebrand Factor A3 protein to bovine serum albumin and 877 are shown in Figure 15. The retention time of pure bovine serum albumin, 877, and the von Willebrand Factor A3 protein were found to be 62.5 minutes, 60 minutes, and 72 minutes respectively (Figure 15 B, E, A and D). In the bovine serum albumin and the von Willebrand Factor A3 binding experiment only the von Willebrand Factor A3 peak was observed in the imidazole eluent (Figure 15C). In the binding with 877, peaks for 877 and the von Willebrand Factor A3 protein were observed in the eluent, indicating 877 is coeluted with the von Willebrand Factor A3 protein. As expected bovine serum albumin did not coelute with von Willebrand Factor A3, and the triple helical protein 877 coeluted with the von Willebrand Factor A3 protein (Figure 15). The imidazole eluent containing the resin purified von Willebrand Factor A3 protein contains only the von Willebrand Factor A3 protein as shown by the labeled sole peak in Figure 15C. Since no bovine serum albumin was found in the von Willebrand Factor A3 protein imidazole eluent sample; it does not bind to the von Willebrand Factor A3 protein (Figure 15C). On the contrary, the

peaks of the von Willebrand Factor A3 protein and the 877 protein are apparent in Figure 15F, and so 877 interacts with the von Willebrand Factor A3 protein.

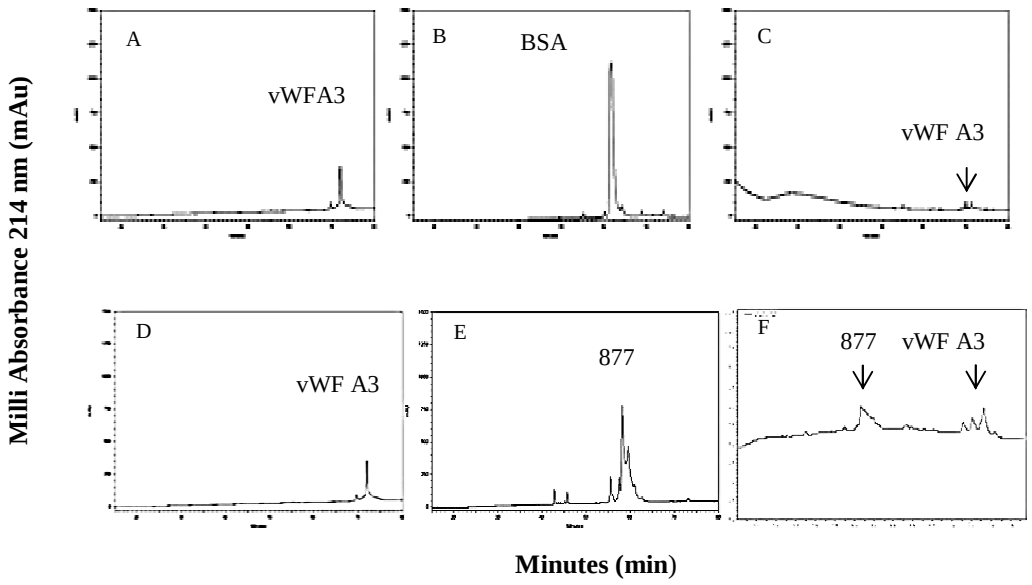


Figure 15: HPLC Pull Down Assay of vWF A3 with BSA (Upper panels) and of vWF A3 with 877 (Lower panels). The upper panels are the chromatograms of pure von Willebrand Factor A3 protein (A), pure bovine serum albumin (B), and the imidazole eluent (C). Lower panels are the chromatograms of pure von Willebrand Factor A3 protein (D), pure 877 protein (E), and imidazole eluent (F). The arrows in C and F mark the retention time of each species. Sample detection was at an absorbance of 214 nm with a gradient of 0.1 % trifluoroacetic acid and 100 % acetonitrile.

When individually chromatographed, the pure peptides had different retention times. P23 had a retention time of 42.5 minutes, $(Pro\text{-}Hyp\text{-}Gly)_{10}$ had a retention time of 36.5 minutes and $(Pro\text{-}Pro\text{-}Gly)_{10}$ had a retention time of 43.0 minutes (Figure 16 A-C). For peptide binding experiments with the von Willebrand Factor A3 protein, the imidazole elutions from each experiment are shown in Figure 16 D, E and F. All three peptides were detected in the imidazole elution indicating all coeluted with the von Willebrand

Factor A3 protein from the affinity resin. The peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ appears to have aggregates as shown in Figure 16B. The additional small peaks in P23 and $(Pro\text{-}Pro\text{-}Gly)_{10}$ could be due to the equilibrium transition between monomer and trimer states (Figure 16 A and C).

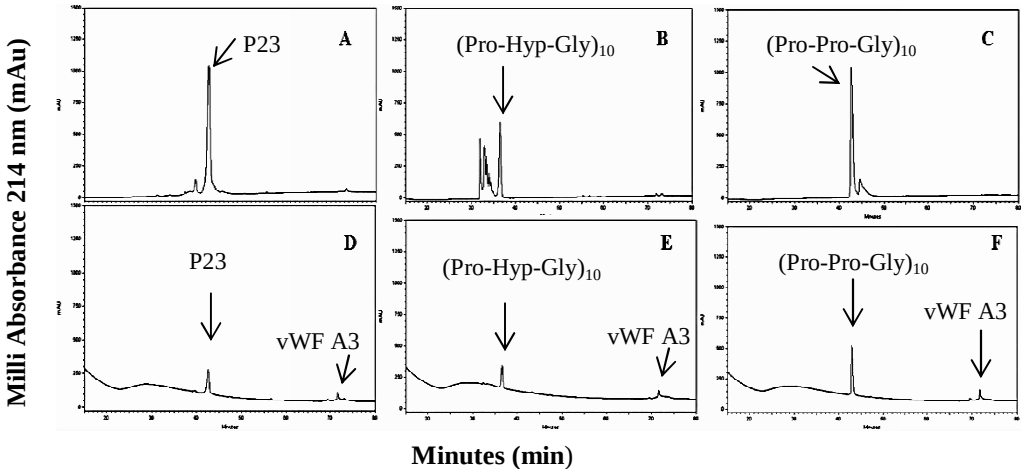


Figure 16: HPLC of von Willebrand Factor A3 Peptide Pull Down Assay: A, B and C are the chromatograms of the pure P23 peptide, pure $(Pro\text{-}Hyp\text{-}Gly)_{10}$ peptide and pure $(Pro\text{-}Pro\text{-}Gly)_{10}$ peptide respectively. D, E and F are the chromatograms of the imidazole elution from each respective experiment. Sample detection was at an absorbance of 214 nm with a gradient of 0.1 % trifluoroacetic acid and 100 % acetonitrile.

P23 peptide, $(Pro\text{-}Pro\text{-}Gly)_{10}$, and $(Pro\text{-}Hyp\text{-}Gly)_{10}$ peptides are also present when the von Willebrand Factor A3 is purified from the talon resin; von Willebrand Factor has a retention time of 72 min, and is eluted from column after seventy two minutes. (Figure 16 D-F). This indicates that they bind to the recombinant von Willebrand Factor A3 protein. A small von Willebrand Factor A3 protein peak was detected in the elutions, and this could be due to a change in concentration or a loss of protein on the affinity resin.

P23 binds to von Willebrand Factor A3 as predicted. The triple helical peptides $(Pro\text{-}Pro\text{-}Gly)_{10}$, and $(Pro\text{-}Hyp\text{-}Gly)_{10}$ also interact with the von Willebrand Factor A3

protein. The peaks of P23, $(Pro\text{-}Pro\text{-}Gly)_{10}$, and $(Pro\text{-}Hyp\text{-}Gly)_{10}$ are clearly visible on the chromatograms of the coeluent in Figure 16 D-F. Even though their binding affinities cannot be determined from these pull down experiments, $(Pro\text{-}Pro\text{-}Gly)_{10}$, and $(Pro\text{-}Hyp\text{-}Gly)_{10}$ seem to interact with von Willebrand Factor A3 at a similar level as the high affinity P23 peptide.

2.4. Discussion and Conclusion

The von Willebrand Factor A3 protein was purified, and circular dichroism confirmed the structure. Binding studies were performed in order to see if the recombinant von Willebrand A3 protein binds to collagen type I and type III, and to the P23 peptide. The purified von Willebrand Factor A3 protein binds to collagen type I and type III, collagen I α1 877 protein, P23 peptide, $(Pro\text{-}Pro\text{-}Gly)_{10}$ and $(Pro\text{-}Hyp\text{-}Gly)_{10}$. As expected, it did not bind to cytochrome C or to bovine serum albumin.

Binding was detected by a pull down assay, so no affinity data was collected. Studies using short peptides have not identified any high affinity binding sites for collagen type I, but our studies show von Willebrand Factor A3 binds to type I collagen in a manner similar to type III. We cannot ascertain the affinity, but recognized there is an association.

In short peptide investigations, the hydroxyproline was considered important for binding, but in our case on two occasions a triple helix without any hydroxyprolines appears to have an affinity for von Willebrand Factor A3 protein. They are the 877 protein type I alpha 1collagen fragment and the $(Pro\text{-}Pro\text{-}Gly)_{10}$ peptide.

The 877 protein contains a segment of human collagen type I alpha 1chain, specifically it contains the sequence of amino acid residues 877-939, and it is expressed in E. coli without hydroxylation. This sequence of amino acids is almost identical to the bovine collagen I α1 sequence. With the exception of not having any hydroxyprolines, there is only one other amino acid difference. At position 900 there is a valine in 877 instead of an isoleucine as in the bovine collagen I α1 chain. This 877 protein is included in the cyanogen bromide fragment 6b shown by Fitzsimmons's study to bind to the von Willebrand Factor protein.

The von Willebrand Factor A3 protein also binds to $(Pro\text{-}Pro\text{-}Gly)_{10.}$ This peptide sequence is not a naturally occurring sequence in collagen. Collagen contains a greater variety of amino acids. This peptide however forms the same backbone configuration as collagen, and is often used as a model for the triple helix (Figure 1). Since the von Willebrand Factor A3 protein binds $(Pro\text{-}Pro\text{-}Gly)_{10,}$ a peptide which doesn't contain hydroxyprolines but has a similar structure as collagen, it would appear that the von Willebrand Factor A3 protein has an affinity to triple helical peptides.

The fact that the von Willebrand Factor A3 protein binds to both 877 and to $(Pro\text{-}Pro\text{-}Gly)_{10}$ peptide indicates that the von Willebrand Factor A3 recognizes the triple helical configurations. This binding may be weak and unspecific. Certain charged residues in a collagen sequence could correlate with a tighter binding.

Research using short peptides have identified only one high affinity binding site on collagen type III; collagen III most likely has other binding sites. It is still not known exactly where the von Willebrand Factor binds to type I collagen, and at how many sites. Clearly, the binding to von Willebrand Factor A3 protein is not only limited to the one

binding site identified by Lisman's peptide study. Additionally the binding may be less specific. Some regions may require specific residues, but the triple helix appears to be a prerequisite. In order to explore the interaction of the von Willebrand Factor A3 domain with full chain collagen by a Fenton footprinting technique, the oxidation of triple helical peptides by Fenton chemistry was then investigated.

Chapter 3

Fenton Chemistry

3.1 Introduction

Fenton chemistry was used many years ago to investigate DNA-protein interactions; this was termed hydroxyl radical footprinting (Tullius, T., et. al., 1986). Hydroxyl radicals generated in the Fenton reaction nonspecifically cleaved the backbone of the DNA in a DNA-protein complex that was not being protected by the binding partner protein, thereby, providing information on the location of the binding interaction (Tullius, T., et. al., 1986). This same approach was also used to study protein-ligand interactions; the peptide backbone of unprotected amino acid sequences was cleaved (Loizos, N., 2004).

Alternatively, hydroxyl radicals can be used to covalently modify the side chains of exposed amino acid residues in a protein (Guan, J., et. al., 2005). Amino acid modification happens before the cleavage of the amide bonds as long as the concentration of hydroxyl radicals is not too high (Heyduk, T., et. al., 2001). The oxidation of amino acid side chains usually forms alcohols, aldehydes, or ketone groups. These modifications can be observed by mass spectrometry because there is a change in the mass of the oxidized amino acid. Alcohol formation increases the mass of the amino acid residue by +16 Da; whereas, with an aldehyde or ketone formation, the increase would be +14 Da. These protein chemical modifications can be used to study protein-protein

interactions (Guan, J., et. al., 2005, and Sharp, J., et. al., 2003). Oxidative Fenton footprinting was explored for use in studying collagen binding reactions.

In the literature, different conditions were reported for oxidizing peptides, and proteins. In certain cases, the peptide backbone was shown to break after several minutes (Heyduk, T., et. al., 2001). Reliable means of monitoring the Fenton reaction were looked at, before investigating its effects on peptides and proteins. Two different methods were used: a methylene blue dye test and a benzoic acid hydroxylation test (Satoh, A., et. al., 2007). In a methylene blue dye test, the formation of hydroxyl radicals is visualized; hydroxyl radicals will cause the blue dye to change its color. The benzoic acid test monitors hydroxyl radical formation indirectly by detecting hydroxylated benzoic acid at its absorbance wavelength of 517 nm (Satoh, A., et al., 2007).

Certain amino acids are known to be more easily oxidized than others; hydroxyl radicals readily oxidize tryptophan, methionine, phenylalanine, tyrosine, cysteine, and cystine (Takamoto, K., et. al., 2006). Collagens usually do not contain tryptophan, cysteine and cystine residues, but it often contains methionine and phenylalanine, and in a few cases tyrosine. The two major amino acid residues in collagen are glycine and proline. Glycines do not oxidize very well, but this is not a concern because in collagen the glycines are located in the center of the triple helix, and are not accessible. Proline residues oxidize well, and in collagen these residues are readily available, and are located on the surface of the molecule. Hydroxyproline residues were not characterized, and how well hydroxyproline residues are oxidized is unclear.

In order to investigate conditions for optimal oxidative footprinting, the peptides bradykinin and $(Pro\text{-}Pro\text{-}Gly)_{10}$ were used. Bradykinin contains two phenylalanines

which are easily oxidized, and three prolines. $(Pro\text{-}Pro\text{-}Gly)_{10}$ is a collagen like peptide which can form a triple helix. Once conditions for oxidizing a triple helical peptide are established, the von Willebrand Factor A3 binding sites on collagen could then be identified by a mass spectrometry oxidative protein footprinting technique.

3.2 Materials and Methods

3.2.1 Materials: Tris-HCl, sodium hydroxide, sodium chloride, sodium phosphate dibasic, sodium phosphate monobasic, acetic acid, ethylenediaminetetraacetic acid, sodium ascorbate, 30 % hydrogen peroxide, methylene blue hydrate, α-cyano-4-hydroxycinnamic acid (α-CHCA), ammonium ferrous sulfate, benzoic acid, the peptide bradykinin, and milli Q water were purchased from Sigma. The peptide $(Gly\text{-}Pro\text{-}Pro)_{10}$ was purchased from Peptides International. Whatman filter paper, hydrochloric acid, methanol, Millipore C_{18} ZipTips, trifluoroacetic acid, acetonitrile, Sequazyme Peptide Mass Standards, and a Maldi plate were purchased from Fisher Scientific. Deionized water (18MΩ) was prepared with a Milli-Q water purification system (Millipore).

3.2.2 Methylene Blue Dye Test: The test was done according to Satoh's procedure (Satoh, A., et. al., 2007). Briefly, whatman filter paper was cut into strips, and a line was drawn with a permanent marker. A paper strip was dipped ten times into a 1 mM methylene blue dye 10 % methanol solution, and allowed to dry overnight in the dark. The Fenton reaction was started by mixing ammonium ferrous sulfate and hydrogen peroxide to the final concentration of, respectively, 0.6 mM and 0.3 %. At 0.5 minutes

and at 2.5 minutes, forty microliters of the reaction sample was directly spotted onto the dyed papers. The bleaching of the paper is an indication that hydroxyl radicals have formed. This test was repeated three times with similar results.

3.2.3 Chemical Oxidation of Benzoic Acid: Benzoic acid was oxidized under different Fenton chemistry conditions, and in each case the solutions were added to a crushed dry benzoic acid powder. The generation of hydroxyl radicals was monitored by the formation of hydroxylated benzoic acid which is detected at 517 nm (Satoh, A., et. al., 2007).

Reactions were done using 9 mM benzoic acid in 0.6 mM ammonium ferrous sulfate, and 3 % hydrogen peroxide at 4 °C. When employing Fenton chemistry as a means to oxidize proteins, it is considered important to add ethylenediaminetetraacetic acid, and sodium ascorbate. Ethylenediaminetetraacetic acid is used to chelate the iron, and the sodium ascorbate recycles the oxidized iron (Heyduk, E., et. al., 1994). For this reason, reactions were done using 9 mM benzoic acid in 0.6 mM ammonium ferrous sulfate, 0.6 mM ethylenediaminetetraacetic acid, and 3 % hydrogen peroxide at 4 °C. Reactions were also done using 9 mM benzoic acid in 0.6 mM ethylenediaminetetraacetic acid, 0.6 mM ammonium ferrous sulfate, 0.6 mM sodium ascorbate, and 3 % hydrogen peroxide at 4 °C. All reactions were quenched with an equal volume of 2M Tris-HCl pH 5.0.

3.2.4 Chemical Oxidation of Bradykinin: Bradykinin was dissolved in water to a concentration of 5 mg/mL. A 1mM stock solution of bradykinin was used for each

oxidation reaction. Reactions were done using 0.25 mM bradykinin in 0.5 mM sodium ascorbate, 0.5 mM ethylenediaminetetraacetic acid, 0.5 mM ammonium ferrous sulfate, 0.5 mM sodium ascorbate, and 0.5 mM hydrogen peroxide at 4 °C. Reactions were quenched with an equal volume of 2M Tris-HCl pH 5.0. The reaction was repeated three times with similar results. Reactions containing 1mM and 10 mM hydrogen peroxide were tested as well as reactions containing all the solutions except bradykinin at those two higher concentrations. The sequence of bradykinin is Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg.

3.2.5 Chemical Oxidation of (Pro-Pro-Gly)$_{10}$: Three stock solutions were prepared. Stock 1 contained 5 mM sodium ascorbate, 50 mM sodium phosphate buffer pH 6.5. Stock 2 contained 5 mM ammonium ferrous sulfate, 5 mM ethylenediaminetetraacetic acid, 50 mM sodium phosphate buffer pH 6.5. Stock 3 contained 88 mM hydrogen peroxide. The lyophilized peptide was dissolved in 50 mM phosphate buffer pH 6.5 to a concentration of 2.25 mg/mL and stored at 4 °C for at least two days. Stock solution 1, 2 and then 3 were added to the (Pro-Pro-Gly)$_{10}$ peptide. The reaction was carried out in a 200 uL solution of 0.2mM (Pro-Pro-Gly)$_{10}$ in 50 mM sodium phosphate buffer pH 6.5, 0.5 mM sodium ascorbate, ammonium ferrous sulfate, ethylenediaminetetraacetic acid, and hydrogen peroxide at 4 °C and at 37 °C. Reactions were also done using 20 mM hydrogen peroxide. An equal volume of 2M Tris-HCl pH 5.0 was used to quench the reaction. The reaction was carried out three times.

3.2.6 MALDI-TOF MS Analysis: The α-cyano-4-hydroxycinnamic acid matrix

(α-CHCA) was used to crystallize samples for MALDI-TOF MS analysis. A saturated solution of α-CHCA matrix was prepared in 50 % acetonitrile, and 0.1 % trifluoroacetic acid. Sequazyme Peptide Mass Standards were prepared according to the manufacturer. Samples were taken from oxidation reactions at several time points before they were quenched, and they were directly spotted onto a MALDI plate and crystallized. Quenched oxidized peptide samples were zip tipped according to the manufacturer, and spotted 1:1 with α-CHCA matrix onto a sample plate, and allowed to dry.

Zip tips are tips that contain a small amount of column material, and are used to clean up samples. The zip tip column is prepared by first wetting the column with 100 % acetonitrile, and then equilibrating the column with 0.1 % trifluoroacetic acid. The sample is bound to the column by pipetting the zip tip up and down in the sample solution. Once the sample is bound to the column , it is cleaned up by washing it many times with 0.1 % trifluoroacetic acid. The peptide sample is then eluted with 0.1 % triflouroacetic acid and 50 % acetonitrile.

A Voyager-DE STR mass spectrometer (PE Biosystems, Foster City, CA) was used to acquire spectra in reflectron, positive ion, delayed extraction mode at The Rockefeller University Proteomic Center. This instrument is equipped with a pulse nitrogen laser (λ =337 nm). The mass spectrometer was calibrated using Sequazyme Peptide Mass Standard STD2, and the calibration file was saved, and used to externally calibrate each sample. The accelerating voltage was set at 20,000 V, the grid voltage at 66 %, the low mass gate at 300 Da, and spectra from 100 individual laser shots were averaged.

3.2.7 ESI MS Analysis: Oxidized peptide samples were zip tipped using C18 zip tips according to the manufacturer's protocol, and submitted to Hunter Mass-spec facility for mass spectrometry analysis. ESI data was used to determine the percentage of oxidized and unoxidized peptide for each Fenton reaction. The percent of unoxidized peptide at 0, 2.5, 5, and 30 minute time points was estimated from its integrated peak area, divided by all the integrated peak areas present. The percentage of oxidized peptide at 0, 2.5, 5, and 30 minute time points was calculated by summing up all of the observed oxidized integrated peak areas, and dividing it by the sum total of all the integrated peak areas.

3.3. Results

3.3.1 Hydroxyl Radical Detection: The bleaching of the methylene blue dye was used as an indicator of hydroxyl radical formation (Satoh, A., et. al., 2007). In this test the Fenton reaction was initiated by adding hydrogen peroxide to iron to form a 0.6 mM ammonium ferrous sulfate and 0.3 % hydrogen peroxide solution, and after different durations of time, samples were spotted onto a strip of dyed paper. For the time points 0.5 minutes and 2.5 minutes, the blue dye was observed to be bleached white, indicating the presence of hydroxyl radicals (Figure 17D-E). This showed us that hydroxyl radicals are forming quickly using millimolar concentrations of ferrous ammonium sulfate. When concentrated Tris-HCl was added to the solution to a final concentration of 1M Tris-HCl the reaction is quenched, and the solution can no longer bleach the dye (Figure 17F). As expected, there was no bleaching observed in the hydrogen peroxide and water controls (Figure 17 A-C). It is apparent from the methylene blue dye test, that hydroxyl

radical formation occurs within thirty seconds, and this reaction can be stopped with a high concentration of Tris-HCl pH 5.0.

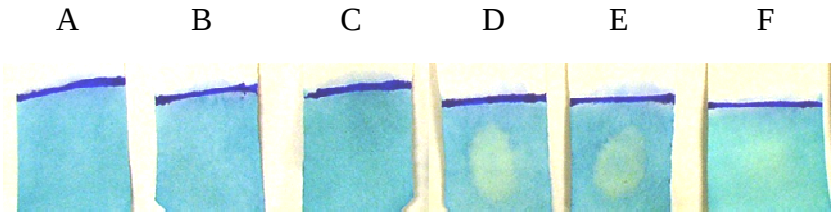


Figure 17: Methylene Blue Dye Test. A. No sample. B. Hydrogen peroxide. C. H_2O. D. 0.5 min. of 0.6 mM ammonium ferrous sulfate, and 0.3 % hydrogen peroxide. E. 2.5 min. of 0.6 mM ammonium ferrous sulfate and 0.3 % hydrogen peroxide. F. 15 min. of ammonium ferrous sulfate, 0.3 % hydrogen peroxide and then quenched with 2M Tris-HCl pH 5.0.

The formation of hydroxyl radicals was also monitored by the hydroxylation of benzoic acid (BA). The hydroxylated benzoic acids have a detected absorbance at 517 nm. Fenton chemistry was carried out under different conditions. Hydroxyl radicals formed very fast using only 0.6 mM ammonium ferrous sulfate, and 0.3 % hydrogen peroxide (Figure 18A). When ethylenediaminetetraacetic acid was added to the standard Fenton reaction it was observed to slow the rate of hydroxyl radical formation, and the addition of sodium ascorbate was shown to further reduce this rate (Figure 18B-C). The use of ethylenediaminetetraacetic acid and sodium ascorbate appears to have lowered the amount of hydroxyl radicals being formed, thereby providing a very mild oxidation environment for peptides. Ascorbate was expected to facilitate the reaction because it recycles the iron.

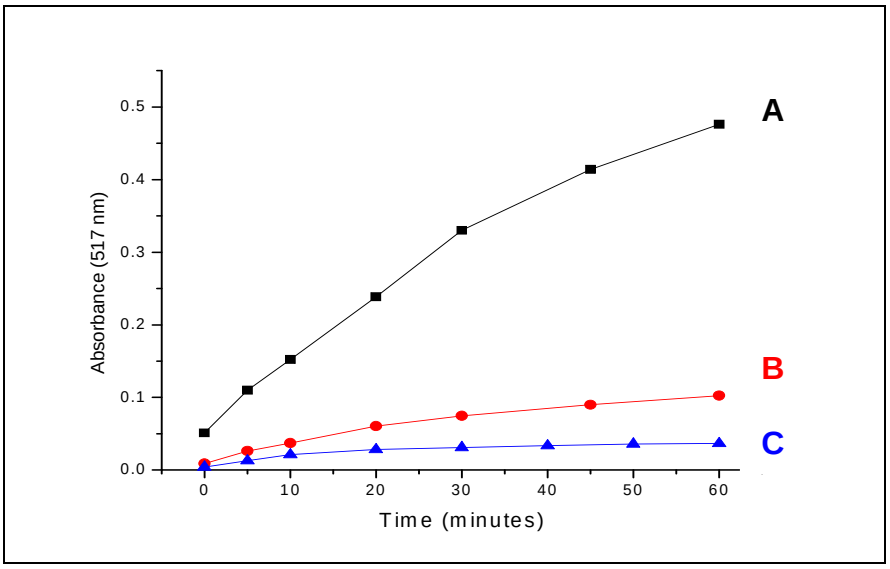


Figure 18: Detection of Hydroxylated Benzoic Acid. A. 0.6 mM ammonium ferrous sulfate and 0.3 % hydrogen peroxide. B. with the addition of 0.6 mM ethylenediaminetetraacetic acid. C. with the addition of 0.6 mM ethylenediaminetetraacetic acid, and 0.6 mM sodium ascorbate. The concentration of benzoic acid was 9 mM. Absorbance was at 517 nm.

3.3.2 Fenton Chemistry Oxidation of Peptides: The oxidation of the peptides by hydroxyl radicals were studied using bradykinin and $(Gly\text{-}Pro\text{-}Pro)_{10}$. The oxidation of bradykinin monitored by MALDI-TOF MS showed that bradykinin is easily oxidized (Figure 19). The best condition for oxidizing bradykinin was found to be 0.25 mM bradykinin in 0.5 mM ethylenediaminetetraacetic acid, 0.5 mM sodium ascorbate, 0.5 mM ammonium ferrous sulfate, and 0.5 mM hydrogen peroxide. Bradykinin was observed at an m/z of 1060^{+1} (Figure 19A-B). After Fenton oxidation, several amino acid residues were oxidized. A significant extent of oxidation of bradykinin was observed, up to six oxygens were added (Figure 19B). Not all of the bradykinin peptide was oxidized, as seen in Figure 19B peak labeled BK.

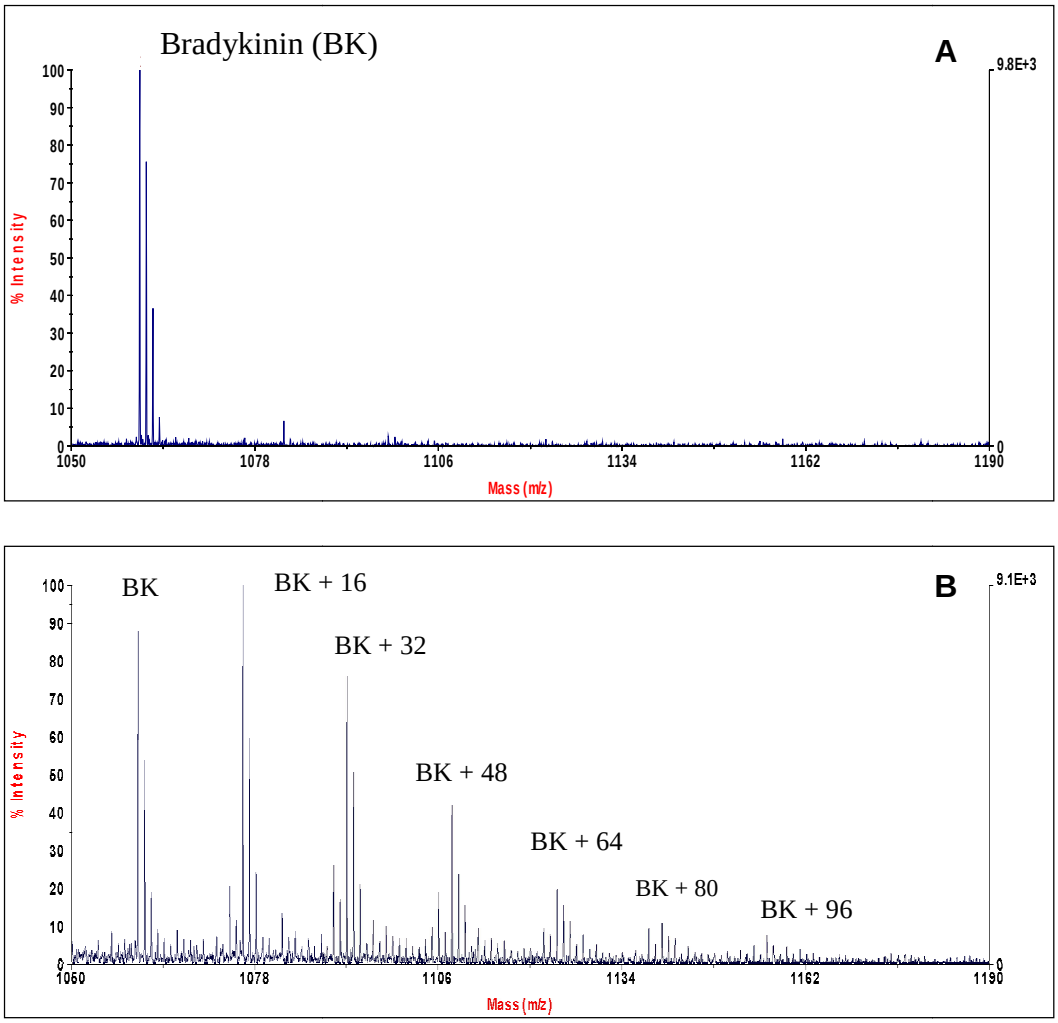


Figure 19: MALDI-TOF MS Spectra of the Oxidation of Bradykinin. A. Bradykinin B. Bradykinin oxidized in 0.5 mM ammonium ferrous sulfate, sodium ascorbate, ethylenediaminetetraacetic, and hydrogen peroxide. The samples were analyzed on a Voyager-DE STR mass spectrometer using α-CHCA matrix.

Similar oxidation conditions were used to oxidize the triple helical peptide (Gly-Pro-Pro)$_{10}$, but only limited oxidation was found (Figure 20B). In the MALDI spectra, (Gly-Pro-Pro)$_{10}$ was seen at an m/z of 2530^{+1} (Figure 19A-B). Sodium and potassium adducts were found associated with the (Gly-Pro-Pro)$_{10}$ peptide as seen as extra peaks with

corresponding increases in mass of twenty three and thirty nine daltons respectively (Figure 19A-B).

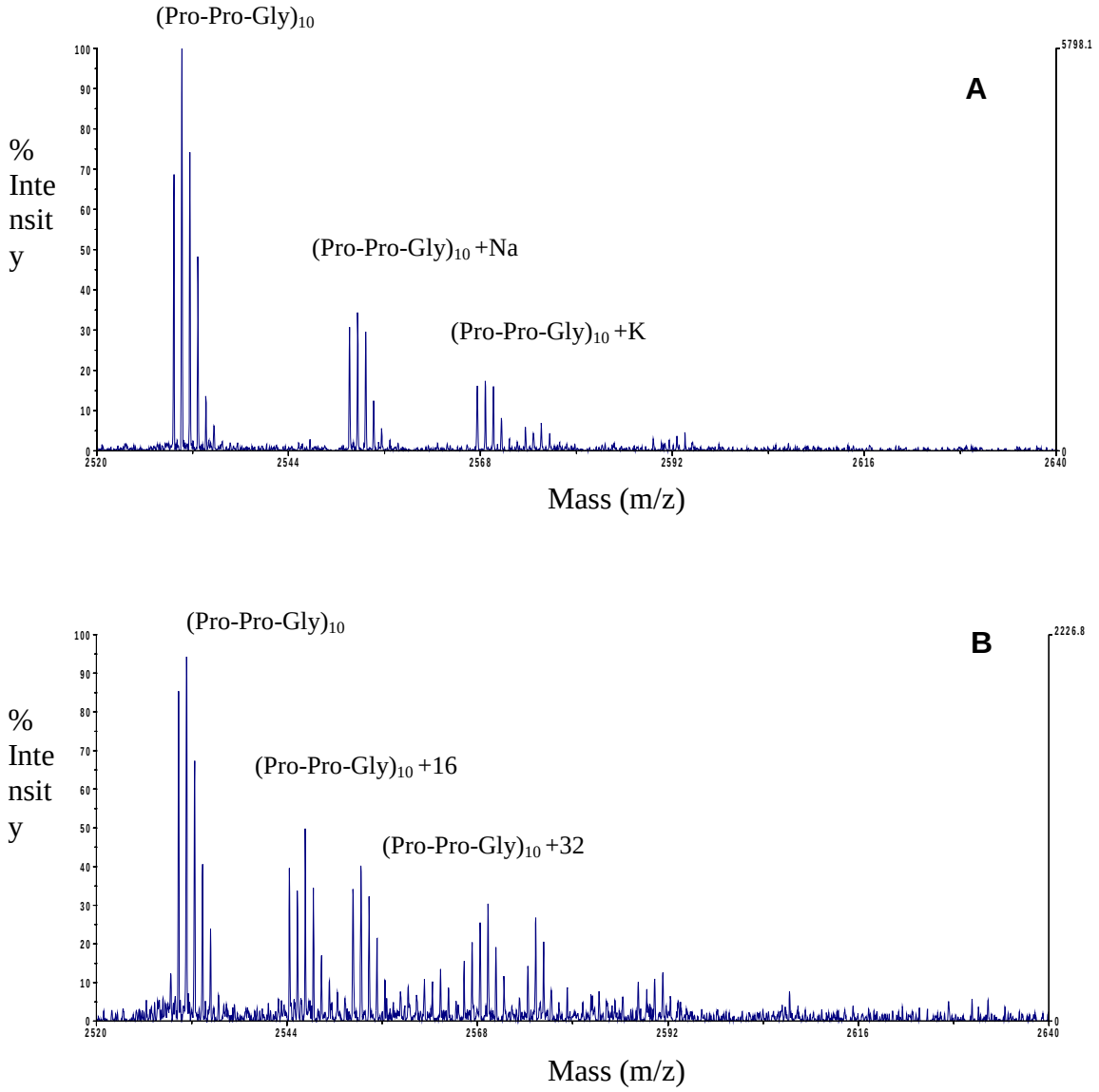


Figure 20: MALDI-TOF Spectra of the Oxidation of $(Pro\text{-}Pro\text{-}Gly)_{10}$. A. $(Pro\text{-}Pro\text{-}Gly)_{10}$. B. $(Pro\text{-}Pro\text{-}Gly)_{10}$ oxidized in 50 mM phosphate buffer pH 6.5, 0.5 mM ammonium ferrous sulfate, ethylenediaminetetraacetic, sodium ascorbate, and hydrogen peroxide. The samples were analyzed on a Voyager-DE STR mass spectrometer using α-CHCA matrix.

Even with extensive cleaning, these sodium (Na+) and potassium (K+) adducts could not be eliminated. It may be due to the conditions of peptide synthesis or $(Pro\text{-}Pro\text{-}Gly)_{10}$ may have a higher affinity to the ions.

Because only limited oxidation of $(Pro\text{-}Pro\text{-}Gly)_{10}$ was observed by MALDI-TOF MS analysis, ESI MS was done just in case some other oxidation m/z peaks were being suppressed by the major peaks in the MALDI-TOF MS. ESI MS of $(Pro\text{-}Pro\text{-}Gly)_{10}$ at 4 °C, revealed that after five minutes twenty two percent of the trimer was oxidized; the peptide continued to be oxidized, and after thirty minutes twenty eight percent of the trimer was oxidized (Figure 21A). The percentage was estimated from the sum of all of the observed oxidized integrated peak areas, divided by the sum total of all the integrated peak areas. The amount of unoxidized peptide present also decreased from ninety seven to forty one percent (Figure 21A). The percentage of unoxidized peptide was estimated from its integrated peak area divided by the sum of all the integrated peak areas. Of the peaks present after thirty minutes, approximately thirty percent of the peaks were unidentified. They are likely cleaved peptides with unusual modifications. Only those modifications with multiples of +16 and +14 increases in mass were considered. After thirty minutes, approximately one percent of the peptide was observed to be unmodified cleavage products. The percentage of cleaved peptide was estimated from its integrated peak area divided by the sum of all the integrated peak areas

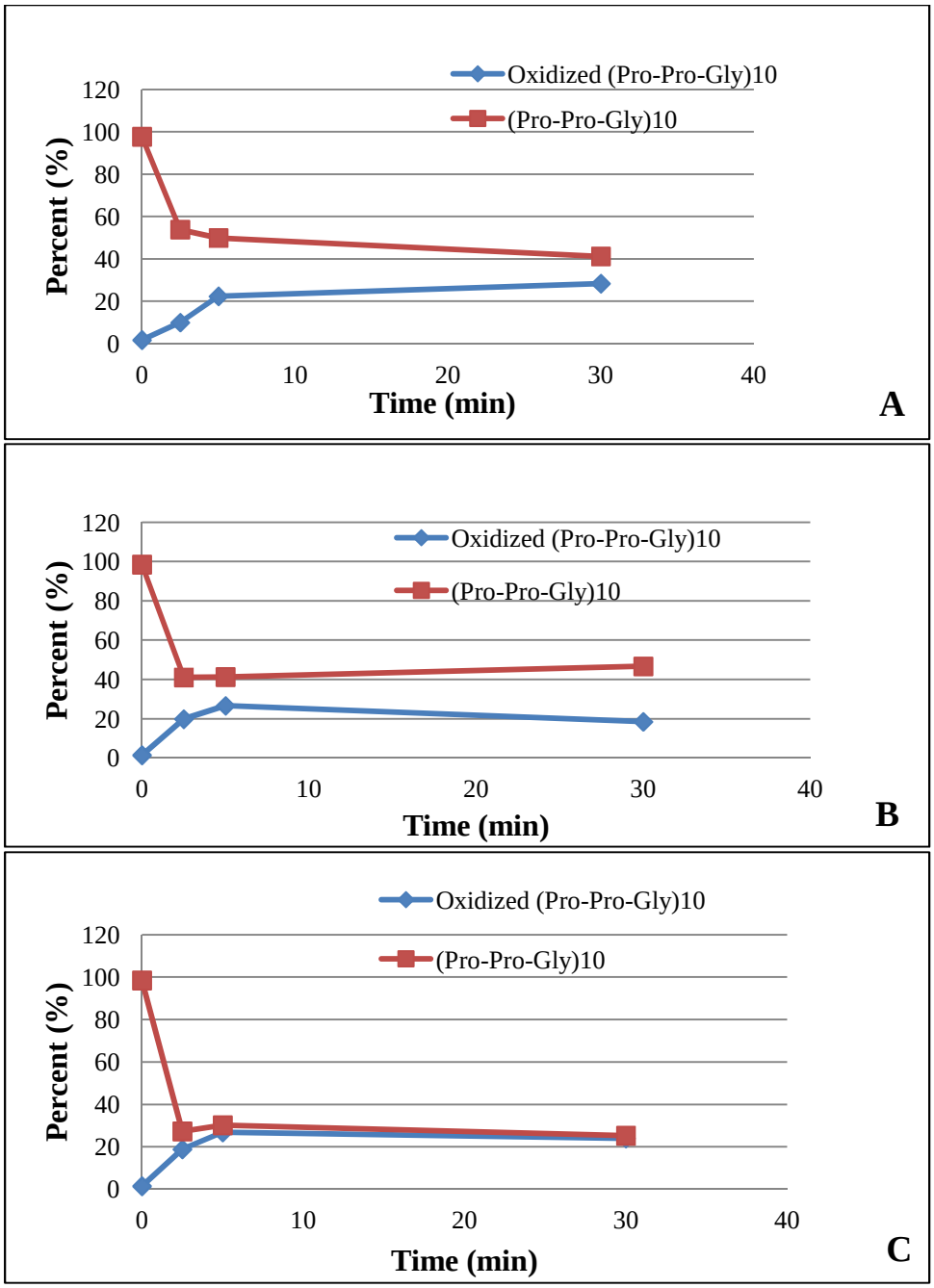


Figure 21. Hydrogen Peroxide Concentration Effect on $(Pro\text{-}Pro\text{-}Gly)_{10}$ Oxidation. A. 4 °C and 0.5 mM hydrogen peroxide. B. 4 °C and 20 mM hydrogen peroxide. C. 37 °C and 20 mM hydrogen peroxide. ESI MS/MS data of 0.2 mM $(Pro\text{-}Pro\text{-}Gly)_{10}$ oxidized in 50 mM phosphate buffer pH 6.5, 0.5 mM ammonium ferrous sulfate, 0.5 mM ethylenediaminetetraacetic, and 0.5 mM sodium ascorbate.

In an attempt to achieve a higher level of oxidation, the hydrogen peroxide concentration was increased from 0.5 mM to 20 mM. The maximum amount of oxidation after five minutes increased slightly to around twenty six percent; however, after thirty minutes only eighteen percent oxidation was detected indicating the peptide is being cleaved at this higher hydrogen peroxide level (Figure 21B).

In order to see if the structure of the triple helix was hindering oxidation, the temperature for the reaction was changed to 37 °C. The $(Pro\text{-}Pro\text{-}Gly)_{10}$ is unfolded at 37 °C, and ESI-MS analysis showed that after 5 minutes, the monomer is only slightly more easily oxidized (Figure 21C).

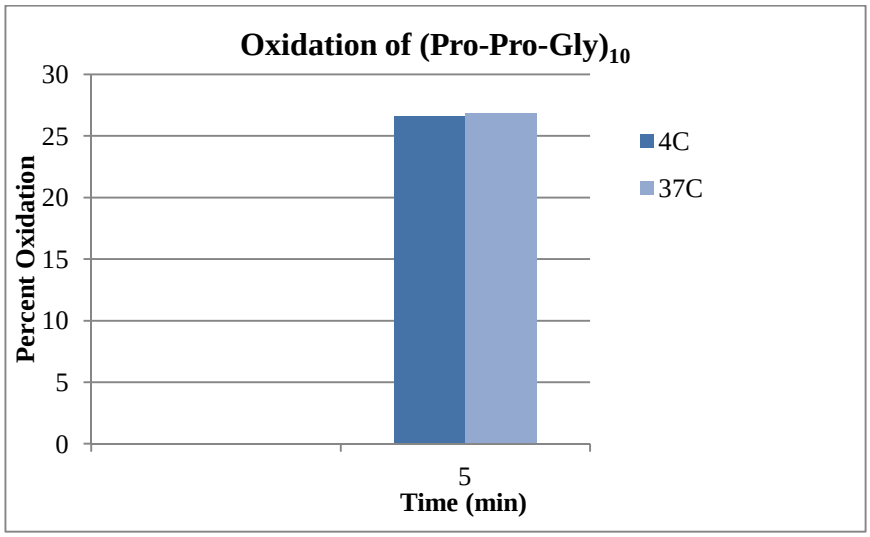


Figure 22: Oxidation of $(Pro\text{-}Pro\text{-}Gly)_{10}$ Monomer and Triple Helical Trimer. $(Pro\text{-}Pro\text{-}Gly)_{10}$ at 4 °C is in dark blue, and $(Pro\text{-}Pro\text{-}Gly)_{10}$ at 37 °C is in light blue. 0.2 mM $(Pro\text{-}Pro\text{-}Gly)_{10}$ oxidized in 50 mM phosphate buffer pH 6.5, 0.5 ammonium ferrous sulfate, 0.5 mM ethylenediaminetetraacetic, 0.5 mM sodium ascorbate, and 20 mM hydrogen peroxide for five minutes and then analyzed by ESI MS.

ESI MS of $(Pro\text{-}Pro\text{-}Gly)_{10}$ at 4 °C, when the peptide is in triple-helix conformation, revealed that after 5 minutes of Fenton chemistry oxidation, twenty six percent of the trimer was oxidized (Figure 22). There wasn't much difference between the folded and unfolded states, and this could be because glycine residues are not easily oxidized.

3.4 Discussion and Conclusion

In this study the benzoic acid experiments showed that both sodium ascorbate and ethylenediaminetetraacetic decrease the initial rate of a Fenton reaction, providing a gentler environment for which to oxidize peptides and proteins. Ethylenediaminetetraacetic acid is added to prevent the binding of excess iron on the protein which may affect both the conformation, and the interaction of the protein, and it prevents the preferential oxidation of any residues (Heyduk, E., et. al., 1994). By removing the iron, ethylenediaminetetraacetic acid also reduces the rate of hydroxyl radical formation and the amount of the hydroxyl radical. Ascorbate was used to recycle the iron, but instead it appears to further suppress the reaction leaving some questions about the effects and role of ascorbate.

Using mild oxidation conditions, six sites on bradykinin were oxidized in one minute. The peptide sequence of bradykinin contains two highly reactive phenylalanine residues which were probably oxidized, as well as three prolines (Takamoto, K., et. al., 2006). Since we did not sequence the oxidized peptide, we do not know the identity of the oxidized amino acids.

The triple helical peptide $(Pro\text{-}Pro\text{-}Gly)_{10}$ Tm is 26.8 °C (Persikov, A., et al., 2004). At 4 °C the $(Pro\text{-}Pro\text{-}Gly)_{10}$ peptide is a triple helix, and at 37 °C it is a monomer. $(Pro\text{-}Pro\text{-}Gly)_{10}$ was oxidized using Fenton chemistry, and approximately twenty six percent of the $(Pro\text{-}Pro\text{-}Gly)_{10}$ triple helix was labeled after five minutes. The $(Pro\text{-}Pro\text{-}Gly)_{10}$ monomer labeled slightly more than the triple helix. Using 0.5 mM hydrogen peroxide at 4 °C about one percent of the peptide was observed to be cleaved after thirty minutes. For the Fenton reaction using 20 mM hydrogen peroxide, at 4 °C approximately one percent of the peptide was detected to have been cleaved. MALDI-TOF MS analysis of $(Pro\text{-}Pro\text{-}Gly)_{10}$ oxidation revealed at most three additional oxygens; whereas, ESI-MS showed a mix of oxidation states with up to ten additional oxygens, but the amount of oxidized species remained low and was difficult to detect in a MALDI spectrum. MALDI-TOF MS analysis may not be the best method for investigating Fenton oxidation reactions because of signal suppression and spot variability in the crystallization process. It appears the best way to study Fenton chemistry reactions is via ESI MS and not MALDI-TOF MS. However, we did not proceed with any protein footprinting experiments because of the concern that the oxidation level and coverage was too low for footprinting purposes.

Chapter 4

Mass Spectrometry of Collagen

4.1 Introduction

The mass spectrometry work was initially planned in order to establish controls for oxidative footprinting experiments, a technique which modifies unprotected amino acids in a complex in order to identify binding sites, and these modifications would need to be identified by mass spectrometry sequencing. The applicability of the mass spectrometry protein footprinting method largely relies on the accuracy of the MS/MS sequencing results. In our study, rat type I collagen and human type I and type III collagen were sequenced using mass spectrometry. Our data revealed many unexpected heterogeneous post-translational modification patterns. How to interpret the data and how to understand the nature of post-translational modifications of collagen became the focus of this work.

Collagen is a ubiquitous protein in humans, and has been studied for many years. It is composed primarily of glycine, and proline amino acid residues. Prolines in collagen are post-translationally modified by two enzymes: prolyl 4-hydroxylase and prolyl 3-hydroxylase (Kivirikko, K., et. al., 1990). The majority of post-translationally modified prolines in collagen are 4-hydroxyproline; there are a few 3-hydroxyprolines (Piez, K., et. al., 1963). Several lysine residues in collagen are also hydroxylated by lysyl hydroxylase, and some of these hydroxylysines are glycosylated by hydroxylysyl

galactosyltransferase and galactosylhydroxylysyl-glucosyltransferase (Kielty, C., et. al., 1993 and Spiro, R., et. al., 1971).

Edman sequence analysis of collagen fragments revealed that hydroxyprolines are found primarily in the Y position of Gly-X-Y sequences where X is any amino acid (Schuppan, D., et. al., 1984). Studies using synthetic peptides determined that the prolyl 4-hydroxylase enzyme hydroxylates prolines in the Y position, except when the X is a glycine (Myllyhararju, J., 2003). A 3-hydroxyproline was found in the X-position in a Gly-X-Hyp collagen sequence, leading to the belief that this was the prolyl 3-hydroxylase motif (Eyre, D., et. al., 2011 and Ogle, J., et. al., 1962). Hydroxylysines were located in X-Lys-Gly collagen sequences suggesting this was a binding motif (Kielty, C., et. al., 2002). Even though the function and recognition motifs of these enzymes are still not fully understood, the assumption that these enzymes are highly selective is widely held.

In collagen there are numerous Gly-Pro-Hyp sites, yet only a few 3-hydroxyprolines have been identified. Collagen purified from bovine tendon was found to contain 0.26 percent 3-hydroxyprolines (Ogle, J., et. al., 1962). The location of a couple of 3-hydroxyprolines in collagen have been identified: α1(I)Pro986, α1(II)Pro986, and α1(V)Pro986, α1(II)Pro944, α2(V)Pro944, α2(I)Pro707, α2(V)Pro707, and α2(V)Pro470; the spacing between some of the sites was observed to be about 234 amino acids, a D period in collagen (Weis, M., et. al., 2010 and Marini, J., et. al., 2007).

Multiple hydroxyprolines in the X-position have been reported in the C-terminal region of collagen where there is a series of Gly-Pro-Pro sequences (Weis, et. al., 2010). In a recent mass spectrometry study of collagen type V, prolines were reported to be hydroxylated in the X position in unexpected Gly-Hyp-Val and Gly-Hyp-Ala sequences

(Yang, C., et. al., 2012). The locations of hydroxylysines have also been investigated. Hydroxylysines have been identified in the N-terminal regions, the N-telopeptide of α1 and α2 chains of various collagens (Barnes, M., et. al., 1971 and Barnes, M., et. al., 1971).

The importance of these hydroxyprolines and hydroxylysines in collagen has been extensively researched. The thermal stability of a collagen triple helix is enhanced when 4-hydroxyprolines are in the Y position (Privalov, P., et. al., 1979). The 3-hydroxyprolines in the X-position may also be important, since a recent study showed osteogenesis imperfecta can be due to mutations in the prolyl 3-hydroxylase enzyme (Cabral, W., et. al., 2007). Hydroxylysines involved in cross-linking increase the stability of collagen fibrils (Kang, A., et. al., 1970).

In this chapter we present our work on the sequencing of collagen type I and type III using mass spectrometry.

4.2 Materials and Methods

4.2.1 Materials: Molecular weight markers, acetic acid, ammonium bicarbonate, acetonitrile, α-cyano-4-hydroxycinnamic acid (α-CHCA), human collagen type III, human collagen type I, and rat collagen type I were purchased from Sigma. Thermo Scientific Precise gels, an Applied Biosystems Sequazyme Peptide Mass Standard Kit, and a Maldi plate were purchased from Fisher Scientific.

4.2.2 Collagen Preparation: Sigma human collagen type I and type III were from placenta, and rat collagen type I was from rat tail. Rat collagen type I was also freshly prepared from rat tail tendon, under the supervision of Dr. Sergey Leikin. All collagens were solubilized in 20 mM acetic acid, pH 3 at 4 °C. The final concentration for the purchased collagens was 2.4 mg/mL; whereas, the purified rat tendon collagen type I was 3.0 mg/mL. Twenty four micrograms of collagen was mixed with 5X SDS sample loading buffer containing 60 mM Tris-HCl pH 6.8, 25 % glycerol, 2 % sodium dodecyl sulfate, 350 mM dithiothreitol, and 0.1 % bromophenol blue. The samples were run on either a 7.5 % SDS-PAGE gel, or a 4-20 % Precise gel from Thermo Scientific. The gels were stained for ten minutes with coomassie blue G-250, destained three times, and rinsed with deionized water. Bands of alpha chains were excised from the gels for in-gel digestion and mass spectrometry analysis.

4.2.3 Rat Type I Collagen Purification: Rat type I collagen was purified according to a procedure provided by Dr. Sergey Leikin (Makareeva, E., et. al., 2006). Briefly, tendons were excised from rat tails, and washed in 3M sodium chloride with enzyme inhibitors overnight. After several water washes, the tendons were then digested 1:50 weight/volume with a 0.1 mg/mL solution of pepsin in 0.5 M acetic acid overnight at 4 °C. At low temperatures, collagens triple helix is resistant to enzyme degradation, and it is also soluble in acid. To remove the insoluble material, the sample was centrifuged at 4 °C for thirty minutes at 15,000 g. The collagen in the supernatant was precipitated with sodium chloride with a final concentration of 0.7M sodium chloride. The collagen was centrifuged at15,000 g at 4 °C, rinsed with 70 % ethanol, and resuspended in 0.5M acetic

acid. After adjusting the collagen concentration to 0.8 mg/mL, the collagen was mixed 1:1 with 50 mM Tris, 0.26M sodium chloride, pH 7.5 and incubated at 32 °C for two hours until the solution became cloudy. The precipitated collagen was spun down and resuspended in 20 mM acetic acid.

4.2.4 Trypsin Solution Digestion: Solutions containing fifty micrograms of human collagen type I, type III and rat collagen type I in 25 mM ammonium bicarbonate pH 8 were individually heat denatured for forty five minutes at 55 °C in order to separate their three alpha chains. The sample was allowed to cool to 37 °C before trypsin was added so as not to denature the enzyme. The sample was digested with trypsin using 1:20 or 1:50 weight/weight at 37 °C overnight; both worked equally well. The reaction was terminated with a final concentration of 0.1 % trifluoroacetic acid, and then stored at -80 °C; at a low pH trypsin is deactivated.

4.2.5 Trysin In-gel Digestion: The excised collagen gel bands were reduced with 10 mM dithiothreitol in 50 mM ammonium bicarbonate for forty five minutes at 55 °C, and alkylated with 100 mM iodoacetamide in 50 mM ammonium bicarbonate for thirty minutes at room temperature in the dark. This will modify the cysteine residues in type III collagen to carboxyamidomethyl cysteine (Sechi, S., et. al., 1998). The bands were then washed with 50 mM ammonium bicarbonate. Ten microliters of 0.02 ug/uL trypsin in 50 mM ammonium bicarbonate, 0.1 % octyl glucopyranoside, and 5 mM calcium chloride was used to digest each sample overnight at 37 °C in 50 mM ammonium bicarbonate. The digestion was stopped with the addition of 5uL of 10 % acetic acid, and

the peptides were extracted first with 30 % acetonitrile and 5 % trifluoroacetic acid, and then with 50 % acetonitrile and 5 % trifluoroacetic acid. The samples were dried down in a Speed Vac to a few microliters.

4.2.6 High Performance Liquid Chromatography Analysis: Fifty microliters of a 0.1 mg/mL trypsin solution digest was injected onto a Beckman HPLC. The tryptic peptides were chromatographed over a C_{18} column using a gradient of 1.8 % solvent B (acetonitrile/0.08 % trifluoroacetic acid) to 50 % B in fifty minutes with solvent A as 0.1 % trifluoroacetic acid. Hydrophilic and small peptides will come off the C_{18} column first; the larger peptides tend to elute later at higher acetonitrile concentrations.

4.2.7 MALDI-TOF MS Analysis: MALDI-TOF MS is a technique which requires samples to be cocrystallized in special matrices. A laser is used to transfer energy to the matrix, which releases the matrix and cocrystallized substance; the time it takes for a particle to travel down a vacuum tube in an electric field is inversely related to its mass to charge ratio. The matrix α-cyano-4-hydroxycinnamic acid was used to cocrystallize peptides; it was prepared as a saturated solution in 50 % acetonitrile and 0.1 % trifluoroacetic acid. Solution digest samples of collagen were spotted 1:1 with matrix onto a sample plate and allowed to dry. All spectra were acquired using a Voyager-DE STR mass spectrometer (PE Biosystems, Foster City, CA) equipped with a pulse nitrogen laser (λ =337 nm) at The Rockefeller University Proteomics Resource Center.

Sequazyme Peptide Mass Standards were prepared according to the manufacturer. The mass spectrometer was calibrated using Sequazyme Peptide Mass Standard STD2,

and the calibration file was saved. Each spectrum was acquired in the reflectron, delayed extraction, positive ion mode with an accelerating voltage of 20,000 V, a grid voltage of 66 %, and a low mass gate of 300 Da. Spectra from 100 individual laser shots were averaged, and externally calibrated.

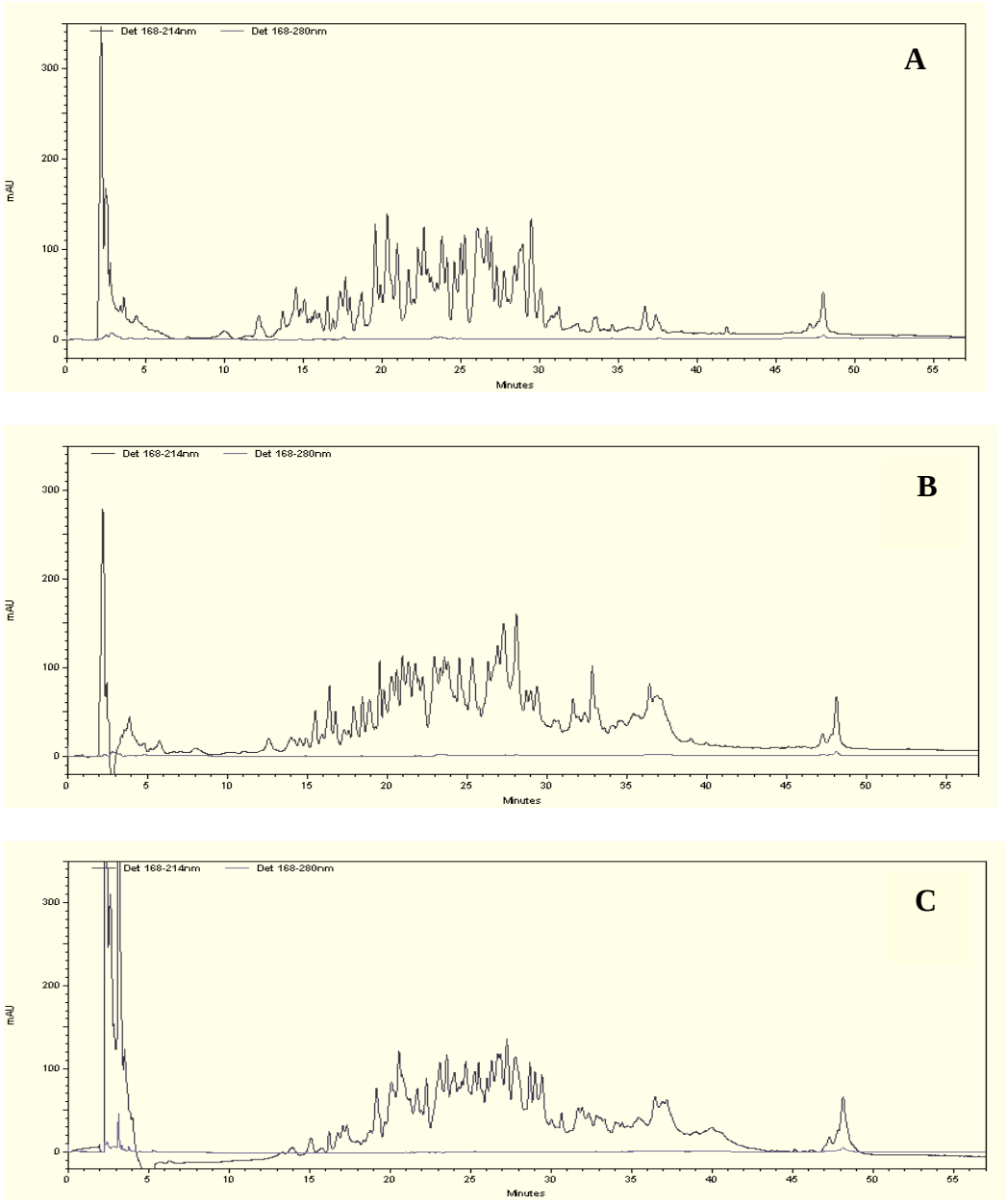
4.2.8 LC ESI-MS/MS Analysis: The in-gel digests and solution digest samples were chromatographed using a C_{18} column on a Dionex HPLC, and eluted with a gradient of 0.1 % formic acid and 100 % acetonitrile. The samples were introduced into the mass spectrometer with the collision energy set at 35, and the capillary temperature set at 275 °C. Tandem mass spectrometry was done at The Rockefeller University Proteomics Resource Center. In-gel digests of commercial human collagen type I, and rat collagen type I alpha chains were sequenced two times using mass spectrometry. In-gel digests of freshly prepared rat tendon collagen type I alpha chains were sequenced once. Two human collagen type III in gel digests, and two solution digests were mass spectrometry sequenced.

4.2.9 Data Base Search and Analysis: For peptide identification a Mascot (Matrix Science) search was performed (Perkins, D., et. al., 1999). Mass spectrometry results in .raw data files were converted to .dta files or .mgf files, and these files were then used to perform a search. The database User 0710 (in-house database at The Rockefeller Proteomics Resource Center) containing human collagen type I and type III, and rat collagen type I, was used with oxidized methionine, proline, and lysine as variable modifications. The Swiss Prot database was used for protein identification confirmation.

The GPM Cyclone (Global Proteome Machine Organization) searches were also performed for peptide identification, and the databases used were Uniprot, Swiss Prot and Ref (Craig, R., et. al., 2004). Peaks6 searches were also done using the following collagen sequences: human type I and type III, and rat type I (Ma, B., et. al., 2003). The selected identified peptides had a Mascot score of 40 or higher, which translates to a probability of false identification of $<10^{-5}$. Peaks6 software was used to generate spectra. The tandem mass spectrometry results for each peptide were examined, and the identified ions were manually checked.

4.3 Results

4.3.1 Collagen Tryptic Solution Digests and SDS PAGE Analysis: The molecular weight of a collagen type I or type III alpha chain is around 95 kDa, and it is too large to sequence using mass spectrometry instruments; these instruments can capture and fragment charged peptides up to 4,000 kDa. It is therefore necessary to digest collagen into smaller peptides within this mass range. Trypsin has become the enzyme of choice for accomplishing this task; it specifically cleaves the amide bond C-terminal to arginines and lysines. The efficiency of trypsin solution digestions of collagen type I and type III were evaluated using high performance liquid chromatography (Figure 23). The observed 214 nm peaks in the HPLC chromatogram are peptides generated from this digestion; their peptide bonds absorb at 214 nm. Peaks observed at 280 nm are tryptic peptides containing tyrosine which absorbs at 280 nm; only a few of these peptides were expected.



Minutes (min)

Figure 23: HPLC of Collagen Trypsin Solution Digests. A. Rat type I collagen containing α1 and α2 chains. B. Human collagen type I containing α1 and α2 chains. C. Human collagen type III. Peptide absorbance monitored at 214 nm (dark grey line on top) and at 280 nm (light grey line on bottom).

A theoretical trypsin digest of a collagen α chain would generate approximately eighty peptides. For human collagen III, a homotrimer, about fifty peaks are discernible

at 214 nm. Type I collagen digestions contains a mix of α1 and α2 chains, and about fifty peaks are visible. For each collagen digest, at fifty percent acetonitrile, all of the tryptic peptides appear to have been eluted. Most peaks in a protein tryptic digest, separated over a C_{18} column by high performance liquid chromatography, generally contain a mixture of peptides, so the actual number of peptides in the digest would be higher than observed in an HPLC chromatogram.

Since type I collagen contains two different alpha chains, the α chains of collagen type I were separated on an SDS-PAGE gel as shown in Figure 24. The chains are well separated, and were individually excised from the gel for in-gel trypsin digestion and mass spectrometry sequencing. Collagen appears as a monomer above 116 kDa, and as a dimer above 200 kDa (Figure 24).

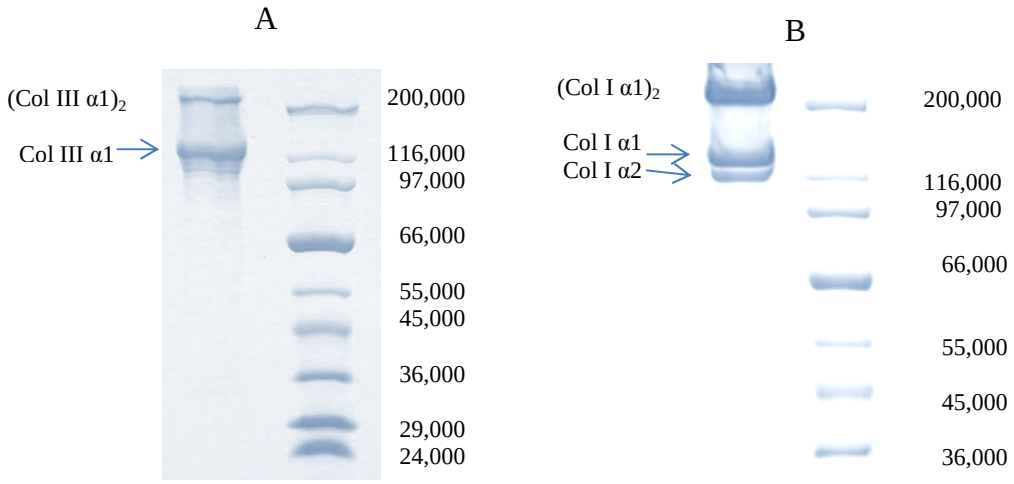


Figure 24: SDS-PAGE of Collagen. A: Lane 1. Human collagen type III. Lane 2. Sigma marker. B: Lane 1. Rat collagen type I. Lane 2. Sigma marker. Gels were stained with coomassie blue.

4.3.2 MALDI-TOF Analysis: Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) spectra was used to further assess the efficiency of the tryptic digests. The peaks observed in the MALDI-TOF MS spectra, are tryptic peptides with a given m/z ratio (Figure 25). In MALDI-TOF MS, the charge on the tryptic peptides is usually +1 so the peptide mass is the m/z ratio. In all the spectra over seventy individual peptides are seen, much more than could be visualized via high performance liquid chromatography; this is further confirmation the digestion was sufficient.

MALDI-TOF MS data is often used for peptide mass spectrometry fingerprinting, where an unknown protein is identified from matching the observed peptide masses to the theoretical masses of a digestion. In these collagen digestions, several tryptic peptides can be identified using the molecular weight of the +1 ions in the spectrum, and comparing it to the generated peptide masses from a theoretical tryptic digest. These theoretical peptides were labeled in Figure 25 as A, B, C, D, E, and F. One feature in the spectra is the multiple number of peaks with a mass variant of 16 Da with regard to the theoretical mass of the labeled peptide assuming all prolines in the Y position are hydroxylated. These mass variants indicate under and over hydroxylation, and were later identified as missed hydroxylation in the Y position, and the unexpected hydroxylation of proline in the X position, hydroxylated lysine, or oxidized methionine.

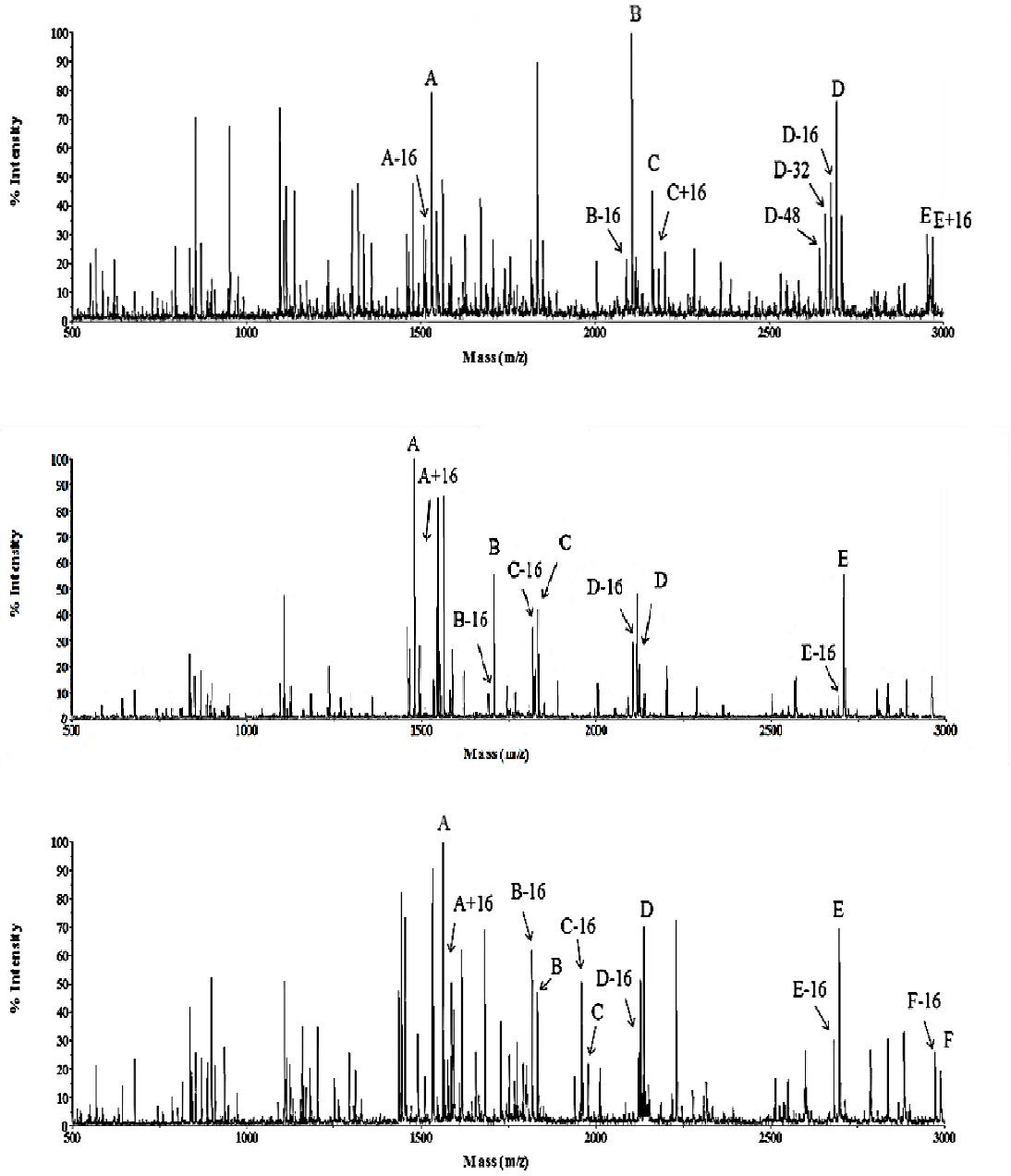


Figure 25: MALDI-TOF MS Spectra of Collagen Trypsin Solution Digests. Top panel. Human collagen type III. Middle panel. Human collagen type I containing α1 and α2 chains. Bottom panel. Rat collagen type I containing α1 and α2 chains. In each panel tryptic peptides of the corresponding collagens identified based on the agreement of their molecular weight (+1 ion) with that of the 'theoretical value' (assuming all Y-Pro as Hyp) are labeled together with their +16 and – 16 variants (see text).

4.3.3 Unusual Hydroxylation in Collagen: Prolines and lysines are the two main amino acids which are modified in collagen. Post-translational modifications of collagen include hydroxylation of proline in the Y-position (Hyp), hydroxylation of proline in the X-position (O_x,) and the hydroxylation of lysine (Hyl). Proteomic software uses both the intact mass of the tryptic peptide, and the ions generated from fragmentation to identify peptides in a protein. The fragmentation of peptides generates a series of b and y ions, and the masses of these ions are used to identify amino acids, and reveal the sites of unexpected hydroxylations. Mass spectrometry sequencing was based on a Mascot database search using hydroxyproline, hydroxylysine and oxidized methionine (M_{ox}) as variable modifications. These modifications increase the mass of the amino acid by +16 Da.

There is some variation between the repeated sequencing results which reflects the level of reproducibility of this technique. Hydroxylated prolines in the X-position (O_x) and unhydroxylated prolines in the Y position (P_y) observed multiple times are shown in bold in Tables 1- 4. The C-terminal lysine residues in tryptic peptides may also be hydroxylated. However, the probability that a tryptic peptide ends in a hydroxylysine is low because trypsin's ability to cleave hydroxylysine is reduced seven fold as compared to lysine (Molony, M., et. al., 1998). Also, glycosylated or cross-linked hydroxylysines would not be observed. We usually do not see fragmented ions to support C-terminal hydroxylysines. We did detect a few hydroxylysines, and in the $\alpha 2$ chain of rat type I collagen we observed Hyl^{87}; this same lysine in human collagen type I is also hydroxylated (Henkel, W., et. al., 2007).

MS/MS spectra of collagen tryptic peptides revealed unusual variations in proline hydroxylation. Unhydroxylated prolines in the Y-position (P_y) were observed as well as hydroxylated prolines in the X-position (O_x). Human collagen type III tryptic peptide 766 contains five prolines, three of them are in the Y-position. An unhydroxylated proline in the Y-position was observed in this peptide as well as the hydroxylated (Figure 26). In Figure 26A the y_{13} ion with an m/z of 1179.6^{+1} indicates the proline in position 777 is not hydroxylated. The expected peptide was also observed as seen in Figure 26B.

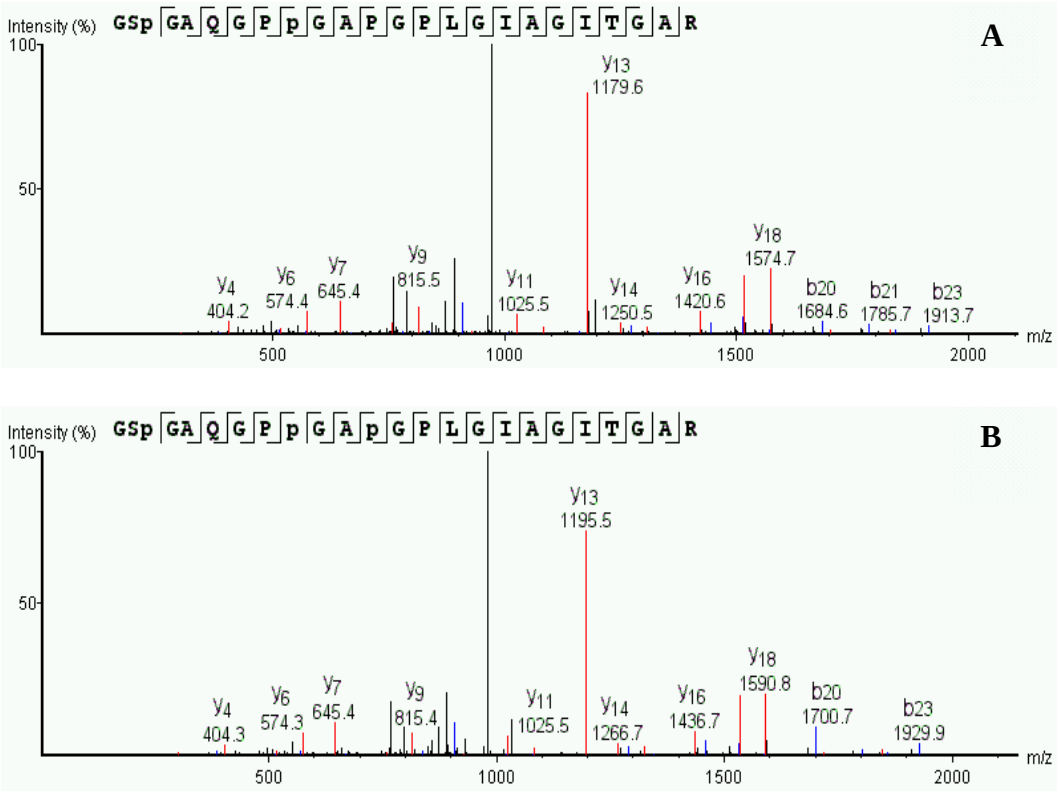


Figure 26: MS/MS Spectra of Human Collagen α1(III) Tryptic Pepide. Sequencing outcomes from ions: A. MS/MS of 1044.54673^{2+} ion (2088.0832^{1+}). B. MS/MS of 1052.546173^{2+} ion (2104.0782^{1+}). The identified peptide is ^{766}GSPGAQGPPGAPGPLGIAGITGAR. The lower case p is hydroxyproline.

The proline in position 777 is hydroxylated as seen from its y_{13} ion whose mass is 16 Da greater and its m/z is 1195.5^{+1}. This peptide was also observed in the MALDI- TOF MS spectrum in Figure 25 top panel labeled B-16 and B.

Another example is shown in Figure 27 where under and over hydroxylation was seen in the same tryptic peptide. Rat tail tendon α_1 chain tryptic peptide 435 contains three proline residues. Two are in the X-position at 446 and 449, and one is in the Y-position 444. Figure 27A is the expected peptide; its proline in position 444 (Hyp^{444}) is hydroxylated, and the peptide has a predicted theoretical m/z of 1680.00^{+1}. This result is supported by y and b-ions, with a strong y_{10} ion. Over hydroxylation was observed in this same peptide; its mass was seen to be 16 Da higher, 1696.75^{+1} (Figure 27B). The extra hydroxylation was located on the proline in the 449 X-position (O_x^{449}). This was supported by the y_5 and y_6 ions and a strong y_{10} in Figure 27B. This peptide was also observed to be under hydroxylated (Figure 27C). The m/z of the -16 species is 1664.76^{+1}, and a strong y_{10} ion, as well as the y_{12} and y_{13} ions indicates the proline in Y-position 444 (P_y^{444}) position was not hydroxylated. Interestingly enough, this tryptic peptide was purified from a single rat tail tendon, and so the heterogeneous hydroxylation observed here is either a species specific variation in its propyl 4-hydroxylase or a variation it accumulated over time.

Methionine residues can also be oxidized, but this is not a normal post- translational modification (Houde, D., et. al., 2006). The same peptide in human collagen type I and rat collagen type I were observed to have their methionines oxidized, but this could just be a coincidence (Figure 28A-B).

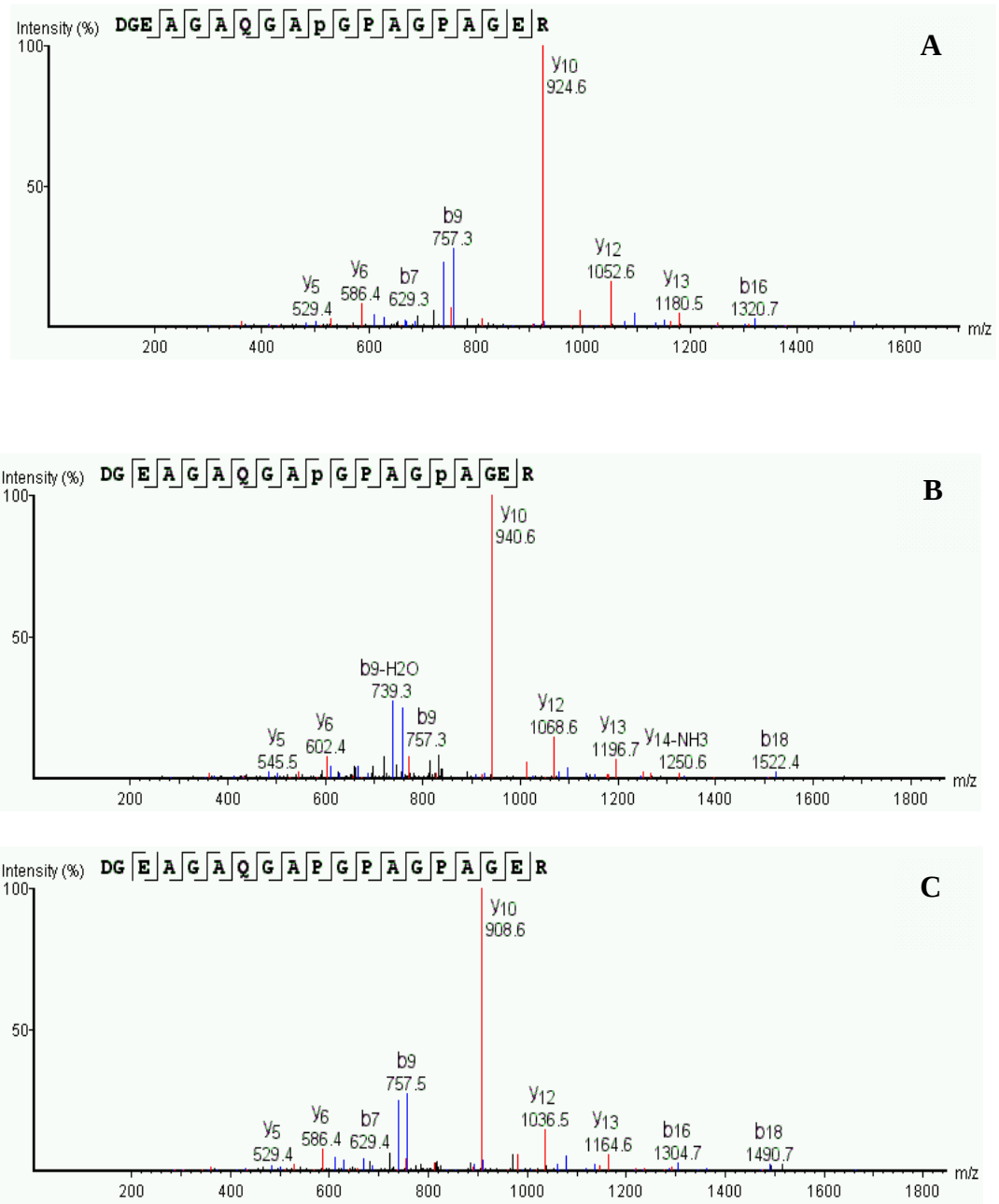


Figure 27: MS/MS Spectra of Rat Collagen α1(I) Tryptic Peptide. Sequencing outcomes from ions: A. MS/MS of 840.8832^{2+} ion (1680.76^{+1}). B. MS/MS of 848.8801^{2+} ion (1696.75^{+1}). C. MS/MS of 832.8864^{2+} ion (1664.76^{+1}). The identified peptide is ^{435}DGEAGAQGAPGPAGPAGER. The hydroxylation sites are shown in lower case.

In rat tail tendon type I alpha 2 chain tryptic peptide 324 a methione is oxidized, this result is supported by the y ions y6, y7, and the b ion series b3, b4 and b5 (Figure 28A).

The human collagen type I alpha 2 chain tryptic peptide 324 also has an oxidized methionine (Figure 28B). This result is supported by the y ion series y6, y7, and y8 and by the b5 ion.

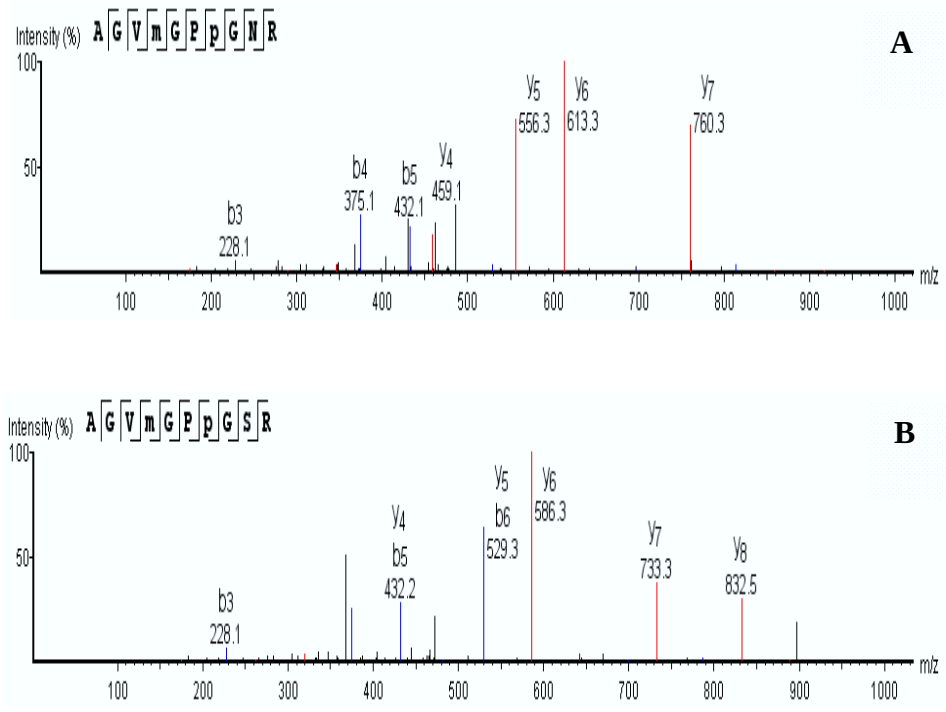


Figure 28: MS/MS Spectra of Rat and Human Collagen Tryptic Peptide. Sequencing outcomes from ions: A. MS/MS of 494.2371^{+2} ion (987.4676^{+1}). B. MS/MS of 480.7322^{+2} ion (960.4567^{+1}). The rat tail tendon α2(I) identified peptide is ^{324}AGVMGPPGNR. The human α2(I) chain identified peptide is ^{324}AGVMGPPGSR. The oxidation and hydroxylation sites are shown in lower case.

Mascot results for collagen, using only peptides whose ion score was 40 or above, shows a sequence coverage of about 30 % for human type III collagen, 45 % for human type I α_1 chain ,40 % for human type I α_2 chain, and 55 % for rat type I collagen. Why

certain regions sequenced better than others is unclear. The peptide may have fragmented poorly, or it may have been further post-translationally modified such as being cross-linked to another α chain.

4.3.4 Collagen Peptide Identification: The complete sequencing results for the five collagen alpha chains have been compiled into Table S1 in the Appendix because of the files large size. The identified unusual hydroxylation sites for the sequencing results have been compiled in Table 1- 4 at end of Chapter 4. A map of these sites was constructed (Figure 29). Figure 29 is at the very end of this Chapter 4. Unusual hydroxylated peptides were identified from their sequencing results and from variations in their mass; collagen was sequenced multiple times, and the combined results are shown in Tables 1- 4.

Table 1 shows the mass spectrometry results from the sequencing of both the α1 and the α2 chains of rat collagen type I that were purchased from Sigma (commercial sample), and revealed a range of variations in peptide hydroxylation. Mass changes of +16 and +32 confirmed the unexpected hydroxylations in six segments of the α1 chain at residues 238-252, 705-725, 375-396, 658-684, 193-219 and 145-174. The α2 chains contained two over-hydroxylated segments at residues 705-725 and 145-174. Most of the hydroxylation sites were in the X-position. The α2(I) chain peptide 145 hydroxylation site could not be confirmed since no y or b ions were found supporting the proline hydroxylation. Either the proline or the lysine was hydroxylated; since lysines are more likely to be cleaved by trypsin than hydroxylysine, the proline in position 173 was

tentatively assigned in italics as being hydroxylated O_x^{173} (Molony, M., et. al., 1998). Under hydroxylation was found in six regions in the α1 chain: residues 793-806, 295-309, 271-291, 757-780, 859-884 and 934-963. Four incomplete hydroxylated regions were detected in the α2 chains: residues 292-309, 757-789 and 265-290. Twelve P_y residues were observed and six of them were found in the triplet Gly-Glu-Pro_y. More than one hydroxylated species was generally observed in the peptides. For example, in the alpha 1 chain in rat collagen type I the 238 peptide, m/z^{+1} of 1306.6386, was observed to contain one or two extra hydroxylations. This is directly visible from the mass change of +16 and +32 m/z^{+1} of 1322.6335, and 1338.6284 respectively. Sometimes the expected hydroxylated peptide was not observed such as in peptide 143 in the α2 chain m/z^{+1} of 2605.2529, but this does not mean it doesn't exist. A peptide may not be observed due to poor fragmentation, additional post-translational modifications or it was not selected for fragmentation.

The wide range of variation in hydroxylation in the commercial sample was unexpected, and rat type I collagen was purified from a single rat tail tendon, to look at whether this variation was due to the commercial samples purity or purification. The mass spectrometry sequencing result of the fresh rat tail tendon sample, were even more variable (Table 2-3). The variations in the hydroxylation content between the commercial and the purified rat tail tendon were similar, and in the seven regions where additional hydroxyprolines were detected in the α1 chain of the rat tail tendon sample, four of them were seen in the commercial sample, and some contained more hydroxylated residues. In the α1 705 peptide the commercial sample was observed to have one to two additional hydroxylation sites (Table 1); whereas, two and three

additional sites were identified for this peptide in the rat tail tendon sample as seen by a mass increase of +32 and +48. The rat tail tendon sample contains additional hydroxylated proline residues in the X-position: O_x^{707}, O_x^{719} and O_x^{722}. In the α2(I) chain, five out of the six peptides with unusual hydroxylations in the commercial sample (Table 1) are also in the rat tail tendon sample (Table 3). The sequenced peptides in Table 1 appear to be a subset of the peptides identified in rat tail tendon (Table 2-3). Because of this similarity, the commercial samples were considered to be representative of the averaged features of post-translational modifications of collagens from a specific tissue and/or organism.

The more heterogeneous hydroxylation in the rat tail tendon sample may be because the sequence coverage was higher (Table S1). A known 3-Hyp O_x^{986} was identified as being hydroxylated in the rat tail tendon α1(I) chain. A proline in the telopeptide region preceding the triple helical region (the P^*) was hydroxylated; since this proline residue proceeds a glycine residue it was expected to be hydroxylated by prolyl 4-hydroxylase; however, it has not been reported before.

Unusual hydroxylation sites detected from both rat samples have been mapped out onto the sequences of the α-chains arranged in D-periodicity in Figure 29. Three regions appear to be highly variable sites. They are two C-terminal regions, residues 705-725 of the α1(I) and the α2(I) chains, respectively, and one N-terminal region, residues 238-252 of the α1 chain. Several O_x residues were detected in these regions. P_y residues were observed in the following regions: residues 941-950 and 821-840 of the α1(I) chain, and 271-300 and 757-780 of both α1 and the α2 chains.

Unusual hydroxylation sites detected in human collagen type I and type III are tabulated in Table 4. Two regions of the α1 chain of the type III collagen were: residues 406-417 and residues 595-627; peptide 595 contained an internal hydroxylated Lys^{612} and peptide 406 had an oxidized methionine, Met^{411}. The oxidation of Met is usually an indication of cellular stress (Houde, D., et. al., 2006). Under hydroxylation was observed in five α1(III) proline residues (P_y) residues, and position 981 (P_y^{981}) was observed in every sequencing attempt as a strong signal.

In human type I collagen only a hydroxylated proline in the X-position was observed (O_x^{986}), and a few under hydroxylations for both the α1 chain and the α2 chain (Table 4). The sequencing of the human type I collagen sample appeared uncomplicated, with only a few unexpected modifications, and the MS/MS sequencing results were mapped out in Figure 29.

The identified O_x residues in the highly variable region in rat collagen are consistent between samples, but other O_x sites are more sporadic. Most of the over hydroxylations seen in the rat type I collagen were not observed in human type I collagen. The Hyp^{986} in the α1 chain and two P_y residues in Gly-Thr-Pro and Gly-Ala-Pro triplets, respectively, at position 771 and 876 of the α1(I) chain were observed in the two type I collagens. In this study we located the sites and the sequence motif surrounding missed hydroxylations which have not been reported before. Unhydroxylated prolines in Y-position were observed in thirty locations, and about 1/6 of these were found in a Gly-Glu-Pro sequence motif, the rest were generally found in Gly-Ala-Pro, Gly-Pro-Pro and Gly-Ser-Pro sequence motifs.

4.4 Discussion and Conclusion

Collagen is composed of GXY repeat units, where Y is usually a hydroxyproline. In the rough endoplasmic reticulum, prolines in an unfolded collagen chain are hydroxylated by collagen prolyl 4-hydroxylase or prolyl 3-hydroxylase, and lysyl hydroxylase hydroxylates lysines (Kivirikko, K., et. al., 1990). These lysines can then be O-glycosylated with Glu-Gal. In general, collagen post-translational modifications vary, as we and others have recently observed (Eyre, D., et. al., 2011, Weis, M., et. al., 2010, and Yang, C., et. al., 2012).

Using mass spectrometry, we detected unusual levels of hydroxylation in rat type I collagen, human type I collagen and human type III collagen, and some regions in collagen appear to be overly hydroxylated. Human collagen type I and type III is less heterogeneic than rat type I collagen, possibly indicating variations in the hydroxylase enzyme in different species and/or tissues.

MALDI-TOF MS analysis of enzymatically digested collagen indicated there is a variation in the extent of proline hydroxylation. Multiple peptides differing by a mass of 16 Da are present in the mass spectra of collagen trypsin digests. To identify these peptides and their post-translational modifications, MS/MS sequencing is needed. MS/MS sequencing has confirmed that there is incomplete hydroxylation of prolines in the Y position (Figure 26). This same peptide was also observed in the MALDI-TOF MS spectrum of human collagen type III trypsin solution digestion as peak B (m/z) of 2104.04 $^{1+}$ and peak B-16 (m/z) 2088.04 $^{1+}$ (Figure 25).

An examination of the tandem mass spectrometry results of collagen reveals collagen has unexpected hydroxylated Pro in the X-positions, and unhydroxylated Pro in the Y-positions. The extent of hydroxylation in the same peptide can differ from the expected level (Tables 1- 4). These discrepancies can occur when there is incomplete hydroxylation of prolines in the Y-position, over hydroxylation of prolines in the X-position, hydroxylation of lysines, or oxidation of methionines. Over all, rat tail type I α chains have more variations in their chains than either human placenta type I or type III. Most of the under hydroxylation and over hydroxylations observed in rat type I collagen were not seen in human type I collagen. In both the rat type I and human type I α1 chain; however, a hydroxylated proline Hyp^{986} was observed, as well as an unhydroxylated proline in the Y-position ($P_{y)}$ at position 771 and 876. The outcomes of the repeated sequencing of the same collagen sample shown in Tables 1- 4 are similar, but the unusual hydroxylations detected vary and only about half of the O_x and/or P_y residues are reproducible. Some O_x or P_y residues, especially those in the highly variable regions of the rat tail tendon type I collagen, appeared to be present even among different samples and may be functionally important.

The over hydroxylation observed in the highly variable regions of rat type I collagen is consistent with the reported three to four 3 Hyp amino acids in their α1(I) and α2(I) chains (Piez, K., et. al., 1963). In rat tail type I collagen, we identified 15 O_x in the α1(I) chain and 11 in the α2(I) chains. These O_x were usually detected as a mixture so some may be present at a very low level. No Edman sequencing was attempted to confirm the identity of the identified O_x residues in this work.

Recent studies have identified a new Gly-Hyp-Ala (O_x) motifs which we have also observed along with a new motif of Gly-Hyp-Ser and one case of Gly-Hyp-Ile, but our observed O_x residues might be from an unexpected hydroxylation by prolyl 4-hydroxylase so it may not be the expected 3Hyp. The amount of missed hydroxylation of proline residues in the Y-positions is also higher in the rat type I collagen than in the human type I collagen. This could be a reflection on the selectivity of the enzyme, and rat prolyl 4-hydroxylase could be less selective that human. Also, sequences with negative charged residues like Gly-Glu-Pro may represent a slippery spot for this rat prolyl 4-hydroxylase because so many Py have been identified with this sequence.

Variation in hydroxylation is difficult to evaluate because very little is known about the selectivity of the hydroxylases and only a few Edman sequencing data have determined the locations of the 4-Hyp and 3Hyp sites. Collagen studies have depended on the amino acid composition of the averaged content of hydroxyproline in an α-chain, and even using MS/MS sequencing, the location of a modification cannot always be resolved. Collagen contains many Gly-Pro-Pro and Pro-Gly-Pro sequences. Since in MS/MS there is a sequence dependent bias to fragment N-terminal to a proline bond; it is generally necessary to assume the hydroxylation of Pro is in the Y-position and not in the X-position in order to resolve mass variations (Henkel, W., et. al., 2007).

An investigation of type I and type III fibrillar collagens revealed a variation in the extent of proline hydroxylation. The stability of collagen is dependent on the level of proline hydroxylation at the Y-position (Privalov, P., et. al., 1979). These differences in the post-translational modification of collagen may be important to consider, when examining the biochemical and biophysical properties of a collagen triple helix. Not only

is hydroxyproline important for stability, it plays a role in immune responses against collagen (Yang, C., et. al., 2012).

4.5 Summary

The purpose of this study was to investigate the application of a mass spectrometry protein footprinting technique to study interactions of collagen and collagen receptors. In this endeavor, we discovered that unhydroxylated triple helical peptides unexpectedly bind to von Willebrand Factor A3 protein. This was a surprise, since it had been assumed that hydroxyproline was an important residue for binding. We found that the oxidation of triple helix experiments using Fenton chemistry were difficult to control, and not sufficient for protein footprinting purposes; for an oxidative footprinting technique to be useful, enough of the molecule needs to be modified to confirm a binding site location.

Our mass spectrometry sequencing results of collagen were very interesting. A few unexpected hydroxylations were detected for human collagen. Proline hydroxylation in rat collagen type I was, however, surprisingly highly variable; we were able to identify two new clusters of hydroxylated prolines in the X-position. Additional studies would be required to determine if these proline hydroxylations impart any function.

The heterogeneity of the post-translational modifications of collagen adds another level of complications for the application of mass spectrometry protein footprinting technique for the study of the binding of collagen and the von Willebrand Factor A3

protein. This and the difficulties in controlling the degree of oxidation by Fenton chemistry are the reasons we did not pursue the binding study as originally proposed.

Table 1. Commercial Rat Type I Collagen Peptides with Mass Variants of 16

	Δm[b]	Score[c]	Unusual Hydroxylation	Sequence of Peptide[d]
The α1(I) chain of rat collagen (commercial sample)				
1306.6386		92		^{238}GPSGPQGPSGAOGPK
1322.6335	+16	65	$\mathbf{O_x^{245}}$	^{238}GPSGPQG$\mathbf{O_x}$SGAOGPK
1338.6284	+32	58	$\mathbf{O_x^{239}}$, $\mathbf{O_x^{245}}$	^{238}G$\mathbf{O_x}$SGPQG$\mathbf{O_x}$SGAOGPK
1328.6481	-16	74	$\mathbf{P_y^{804}}$	^{793}GFOGLOGPSGE$\mathbf{P_y}$GK
1344.6430		41		^{793}GFOGLOGPSGEOGK
1435.6812	-16	43	P_y^{297}	^{295}GEP$_y$GPSGLOGPOGER
1840.9188		50		^{705}VGPOGPSGNAGPOGPOGPVGK
1856.9137	+16	61	$\mathbf{O_x^{707}}$	^{705}VG$\mathbf{O_x}$OGPSGNAGPOGPOGPVGK
1872.9086	+32	49	$\mathbf{O_x^{707}}$, $\mathbf{O_x^{722}}$	^{705}VG$\mathbf{O_x}$OGPSGNAGPOGPOG$\mathbf{O_x}$VGK
1959.9519	-16	87	$\mathbf{P_y^{273}}$	^{271}GE$\mathbf{P_y}$GPAGVQGPOGPAGEEGKR
1975.9468		94		^{271}GEOGPAGVQGPOGPAGEEGKR
2014.9689		41		^{375}TGPOGPAGQDGROGPAGPOGAR
2030.9639	+16	50	O_x^{389}	^{375}TGPOGPAGQDGROGO$_x$AGPOGAR
2120.9955	-16	98	$\mathbf{P_y^{771}}$	^{757}GSOGADGPAGSOGT$\mathbf{P_y}$GPQGIAGQR
2136.9905		84		^{757}GSOGADGPAGSOGTOGPQGIAGQR
2169.0571	-16	57	P_y^{876}	^{859}GETGPAGPOGAOGAOGAP$_y$GPVGPAGK
2185.0520		45		^{859}GETGPAGPOGAOGAOGAOGPVGPAGK
2307.1364		58		^{658}GDAGPOGPAGPAGPOGPIGNVGAOGPK
2323.1313	+16	79	$\boldsymbol{O_x^{683}}$	^{658}GDAGPOGPAGPAGPOGPIGNVGAO$\boldsymbol{GO_xK}$
2316.0487		53		^{193}GEOGPOGPAGAAGPAGNOGADGQOGAK
2332.0436	+16	79	$\mathbf{O_x^{206}}$	^{193}GEOGPOGPAGAAG$\mathbf{O_x}$AGNOGADGQOGAK
2548.2063		101		^{145}GNDGAVGAAGPOGPTGPTGPOGFOGAAGAK
2564.2012	+16	54	$\mathbf{O_x^{155}}$	^{145}GNDGAVGAAG$\mathbf{O_x}$OGPTGPTGPOGFOGAAGAK
2679.2394	-16	42	$\mathbf{P_y^{948}}$	^{934}GFSGLQGPOGSOGS$\mathbf{P_y}$GEQGPSGASGPAGPR
2695.2343		127		^{934}GFSGLQGPOGSOGSOGEQGPSGASGPAGPR
The α2 (I) chain				
1238.6052	+16	58	$\mathbf{Hyl^{87}}$	^{76}GFOGTOGLOGF$\mathbf{K_x}$
1560.8129	-16	102	$\mathbf{P_y^{891}}$	^{889}GE$\mathbf{P_y}$GPAGSVGPVGAVGPR
1576.8078		59		^{889}GEOGPAGSVGPVGAVGPR
1576.7714	-32	79	$\mathbf{P_y^{294}}$, $\mathbf{P_y^{297}}$	^{292}GS$\mathbf{P_y}$GE$\mathbf{P_y}$GSAGPAGPOGLR
1592.7663	-16	58	$\mathbf{P_y^{297}}$	^{292}GSOGE$\mathbf{P_y}$GSAGPAGPOGLR
1608.7612		76		^{292}GSOGEOGSAGPAGPOGLR
1781.7937	-16	69	$\mathbf{P_y^{273}}$	^{271}GE$\mathbf{P_y}$GSAGAQGPOGPSGEEGK
1937.8948	-16	115	$\mathbf{P_y^{273}}$	^{271}GE$\mathbf{P_y}$GSAGAQGPOGPSGEEGKR
2368.0648	-16	41	$\mathbf{P_y^{273}}$	^{265}GETGNKGE$\mathbf{P_y}$GSAGAQGPOGPSGEEGK
2384.0597		60		^{265}GETGNKGEOGSAGAQGPOGPSGEEGK
1833.8905	+16	47	$\mathbf{O_x^{707}}$, $\mathbf{O_x^{719}}$, $\mathbf{P_y^{717}}$	^{705}TG$\mathbf{O_x}$OGPSGITGP$\mathbf{P_y}$G$\mathbf{O_x}$OGAAGK
2605.2529	+16	80	$\boldsymbol{O_x^{173}}$	^{145}GSDGSVGPVGPAGPIGSAGPOGFOGAOG$\boldsymbol{O_xK}$
2971.512	-16	76	P_y^{768}	^{757}GPSGEOGTTGPP$_y$GTAGPQGLLGAOGILGLOGSR
2987.5069		48		^{757}GPSGEOGTTGPOGTAGPQGLLGAOGILGLOGSR

Footnote:

[a]: The mass of the single charged ion used to identify each peptide. The sequencing results of most of the peptides are based on the fragmentation of the double charged species (see below).

[b]: The mass differences compared to the 'theoretical' assuming all Pro in the Y positions are hydroxylated.

[c]: The score of Mascot search engine (see Material and Methods).

[d]: Unexpected hydroxylation of X-Pro and unhydroxylated Y-Pro are shown as O_x and P_y, respectively; those observed in more than one sequencing effort and/or in both the commercial and the rat tendon samples (see text) are in bold; those with uncertainties are shown in italic (see text).

Table 2. Rat Tail Tendon Type I α1 Collagen with Mass Variants of 16

Ion mass[a]	Δ*m*[b]	Score[c]	Unusual	Peptide[d]
α1 chain				
1242.5961	-16	63	P_y^{369}	^{361}GLTGSOGSP$_y$GPDGK
1322.6335	+16	76	**O_x^{239}**	^{238}G**O$_x$**SGPQGPSGAOGPK
1338.6284	+32	76	**O_x^{239},** O_x^{251}	^{238}G**O$_x$**SGPQGPSGAOGO$_x$K
1354.623	+48	49	**O_x^{239}, O_x^{245},** O_x^{251}	^{238}G**O$_x$**SGPQG**O$_x$**SGAOG*O$_x$K*
1452.7264		65		$^{-7}$SAGVSVPGPMGPSGPR
1468.7213	+16	51	M_{ox}^3	$^{-7}$SAGVSVPGPM$_{ox}$GPSGPR
1484.7162	+32	40	P^{-1}*, **M_{ox}^3**	$^{-7}$SAGVSVP*GPM$_{ox}$GPSGPR
1574.8173	-16	86	P_y^{600}	^{586}GLTGPIGPOGPAGAP$_y$GDK
1590. 8122		42		^{586}GLTGPIGPOGPAGAOGDK
1561.7969	+16	40	O_x^{986}	^{975}DGLNGLOGPIGO$_x$OGPR
1872.9086	+32	49	**O_x^{707}**, O_x^{719}	^{705}VG**O$_x$**OGPSGNAGPOGO$_x$OGPVGK
1888.9035	+48	48	**O_x^{707}**, O_x^{719}, **O_x^{722}**	^{705}VG**O$_x$**OGPSGNAGPOGO$_x$OG**O$_x$**VGK
1664.7623	-16	88	P_y^{444}	^{435}DGEAGAQGAP$_y$GPAGPAGER
1680.7551		96		^{435}DGEAGAQGAOGPAGPAGER
1696.7521	+16	92	O_x^{449}	^{435}DGEAGAQGAOGPAGO$_x$AGER
1800.8697	-32	105	P_y^{825}, P_y^{831}	^{817}GPOGPMGPP$_y$GLAGPP$_y$GESGR
1816.8647	-16	125	P_y^{825}, P_y^{831}, M_{ox}^{822}	^{817}GPOGPM$_{ox}$GPP$_y$GLAGPP$_y$GESGR
1803.8508	-16	77	**P_y^{273}**	^{271}GE**P$_y$**GPAGVQGPOGPAGEEGK
1819.8457		82		^{271}GEOGPAGVQGPOGPAGEEGK
2105.0006	-32	78	P_y^{759}, **P_y^{771}**	^{757}GSP$_y$GADGPAGSOGT**P$_y$**GPQGIAGQR
2339.1262	+32	40	O_x^{674}, O_x^{683}	^{658}GDAGPOGPAGPAGPOGO$_x$IGNVGA*O$_x$GO$_x$K*
2548.2063		58		^{145}GNDGAVGAAGPOGPTGPTGPOGFOGAAGAK
2564.2012	+16	96	**O_x^{155}**	^{145}GNDGAVGAAG**O$_x$**OGPTGPTGPOGFOGAAGAK
2663.2444	-32	96	P_y^{945}, **P_y^{948}**	^{934}GFSGLQGPOGSP$_y$GS**P$_y$**GEQGPSGASGPAGPR
2679.2394	-16	105	**P_y^{948}**	^{934}GFSGLQGPOGSOGS**P$_y$**GEQGPSGASGPAGPR
2695.2343		62		^{934}GFSGLQGPOGSOGSOGEQGPSGASGPAGPR

Footnote:

[a]: The mass of the single charged ion used to identify each peptide. The sequencing results of most of the peptides are based on the fragmentation of the double charged species (see below).

[b]: The mass differences compared to the 'theoretical' assuming all Pro in the Y positions are hydroxylated.

[c]: The score of Mascot search engine (see Material and Methods).

[d]: Unexpected hydroxylation of X-Pro and unhydroxylated Y-Pro are shown as O_x and P_y, respectively; those observed in more than one sequencing effort and/or in both the commercial and the rat tendon samples are in bold; those with uncertainties are shown in italic.

Table 3. Rat Tail Tendon Type I α2 Collagen with Mass Variants of 16

Ion mass[a]	Δ*m*[b]	Score[c]	Unusual	Peptide[d]
868.4653		51		^{907}GPSGPQGIR
884.4585	+16	46	O_x^{908}	^{907}GO$_x$SGPQGIR
937.5102		52		^{964}GPAGPSGPIGK
953.5051	+16	51	O_x^{968}	^{964}GPAGO$_x$SGPIGK
1068.5684	-16	71	P_y^{258}	^{253}GLVGEP$_y$GPAGSK
1084.5633		70		^{253}GLVGEOGPAGSK
1187.6280	-16	54	P_y^{981}	^{978}SGHP$_y$GPVGPAGVR
1238.6052	+16	64	**Hyl^{87}**	^{76}GFOGTOGLOGF**K$_{ox}$**
1437.7485	-16	79	P_y^{486}	^{484}GLP$_y$GEFGLOGPAGPR
1453.7434		75		^{484}GLOGEFGLOGPAGPR
1469.7383	+16	110	O_x^{494}	^{484}GLOGEFGLOGO$_x$AGPR
1490.7122		59		^{741}TGEIGASGPOGFAGEK
1506.7071	+16	67	O_x^{749}	^{741}TGEIGASGO$_x$OGFAGEK
1560.8129	-16	102	**P_y^{891}**	^{889}GE**P$_y$**GPAGSVGPVGAVGPR
1576.8078		110		^{889}GEOGPAGSVGPVGAVGPR
2025.0373	-16	82	**P_y^{891}**	^{885}HGNRGE**P$_y$**GPAGSVGPVGAVGPR
2041.0322		76		^{885}HGNRGEOGPAGSVGPVGAVGPR
1576.7714	-32	76	**P_y^{294}, P_y^{297}**	^{292}GS**P$_y$**GE**P$_y$**GSAGPAGPOGLR
1592.7663	-16	85	**P_y^{294}**	^{292}GS**P$_y$**GEOGSAGPAGPOGLR
1608.7612		88		^{292}GSOGEOGSAGPAGPOGLR
1624.7562	+16	94	O_x^{302}	^{292}GSOGEOGSAGO$_x$AGPOGLR
1748.8674	-16	69	**P_y^{297}**	^{291}RGSOGE**P$_y$**GSAGPAGPOGLR
1764.8624		50		^{291}RGSOGEOGSAGPAGPOGLR
1599.8238	-16	87	P_y^{183}	^{175}GELGPVGNP$_y$GPAGPAGPR
1615.8187		105		^{175}GELGPVGNOGPAGPAGPR
1781.7937	-16	85	**P_y^{273}**	^{271}GE**P$_y$**GSAGAQGPOGPSGEEGK
1797.7886		94		^{271}GEOGSAGAQGPOGPSGEEGK
1817.9028		52		^{705}TGPOGPSGITGPOGPOGAAGK
1833.8973	+16	69	**O_x^{707}**, ***O_x^{719}***, ***P_y^{717}***	^{705}TG**O$_x$**OGPSGITGP***P$_y$G**O$_x$***OGAAGK
1849.8926	+32	89	**O_x^{707}, O_x^{719}**	^{705}TG**O$_x$**OGPSGITGPOG**O$_x$**OGAAGK
1865.8876	+48	97	**O_x^{707}**, O_x^{716}, **O_x^{719}**	^{705}TG**O$_x$**OGPSGITGO$_x$OG**O$_x$**OGAAGK
2589.2580		112		^{145}GSDGSVGPVGPAGPIGSAGPOGFOGAOGPK
2605.2529	+16	103	**O_x^{173}**	^{145}GSDGSVGPVGPAGPIGSAGPOGFOGAO***G**O$_x$K***
2621.2478	+32	51	O_x^{164}, **O_x^{173}**	^{145}GSDGSVGPVGPAGPIGSAGO$_x$OGFOGAO***G**O$_x$K***
3045.4257	-16	78	P_y^{552}	^{520}GPSGAOGPDGNKGEAGAVGAOGSAGASGPGGLP$_y$GER
3077.4155	+16	76	O_x^{548}	^{520}GPSGAOGPDGNKGEAGAVGAOGSAGASGO$_x$GGLOGER

Footnote:

[a]: The mass of the single charged ion used to identify each peptide. The sequencing results of most of the peptides are based on the fragmentation of the double charged species (see below).

[b]: The mass differences compared to the 'theoretical' assuming all Pro in the Y positions are hydroxylated.

[c]: The score of Mascot search engine (see Material and Methods).

[d]: Unexpected hydroxylation of X-Pro and unhydroxylated Y-Pro are shown as O_x and P_y, respectively; those observed in more than one sequencing effort and/or in both the commercial and the rat tendon samples are in bold; those with uncertainties are shown in italic.

Table 4. Human Type III and Type I Collagen Peptides with Mass Variants of 16

Ion mass[a] (m/z^{+1})	Δm[b]	Score[c]	Unusual	Peptide[d]
A. Human collagen type III				
949.5102	-16	45	$\mathbf{P_y}^{\mathbf{981}}$	^{973}GPVGPSGP**P$_y$**GK
965.5051		44		^{973}GPVGPSGPOGK
1138.5674		62		^{790}GLAGPOGMOGPR
1154.5623	+16	41	M_{ox}^{797}	^{790}GLAGPOGM$_{ox}$OGPR
1203.5827		57		^{406}GQOGVMGFOGPK
1219.5776	+16	42	O_x^{416}	^{406}GQOGVMGFOGO_xK
1514.7346	-16	50	P_y^{990}	^{984}DGTSGHP$_y$GPIGPOGPR
1530.7295		57		^{984}DGTSGHOGPIGPOGPR
1670.7989		42	P_y^{240}, M_{ox}^{243}	^{229}GEMGPAGIOGAP$_y$GLM$_{ox}$GAR
2088.0832	-16	94	$\mathbf{P_y}^{\mathbf{777}}$	^{766}GSOGAQGPOGA**P$_y$**GPLGIAGITGAR
2104.0782		137		^{766}GSOGAQGPOGAOGPLGIAGITGAR
2283.1000		79	$\mathbf{P_y}^{\mathbf{687}}$	^{667}GEGGPOGVAGPOGGSGPAGP**P$_y$**GPQGVK$_x$
2950.4653		82		^{595}GPTGPIGPOGPAGQOGDKGEGGAOGLOGIAGPR
2966.4603	+16	58	$\boldsymbol{O_x}^{\boldsymbol{605}}$	^{595}GPTGPIGPOG***O$_x$***AGQOGD*K*GEGGAOGLOGIAGPR
B. Human type I collagen				
α1 chain				
1562.7863	+16	65	O_x^{986}	^{975}DGLNGLOGPIGO$_x$OGPR
1832.8596		78	M_{ox}^{822}, P_y^{825}	^{817}GPOGPM$_{ox}$GPP$_y$GLAGPOGESGR
1848.8545	+16	41	M_{ox}^{822}	^{817}GPOGPM$_{ox}$GPOGLAGPOGESGR
2105.0006	-16	66	$\mathbf{P_y}^{\mathbf{771}}$	^{757}GSOGADGPAGAOGTP$_y$GPQGIAGQR
2121.0373		58		^{757}GSOGADGPAGAOGTOGPQGIAGQR
2497.2066	-16	81	$\mathbf{P_y}^{\mathbf{876}}$	^{856}GDRGETGPAGPOGAOGAOGAP$_y$GPVGPAGK
2513.2015		109		^{856}GDRGETGPAGPOGAOGAOGAOGPVGPAGK
2703.2394	-16	63	$\mathbf{P_y}^{\mathbf{639}}$	^{619}GAOGDRGEOGPOGPAGFAGPP$_y$GADGQOG
		73		^{625}GEOGPOGPAGFAGPOGADGQOGAK
α2 chain				
1168.4978	-16	61	P_y^{840}	^{837}DGNP$_y$GNDGPOGR

Footnote:

[a]: The mass of the single charged ion used to identify each peptide. The sequencing results of most of the peptides are based on the fragmentation of the double charged species (see below).

[b]: The mass differences compared to the 'theoretical' assuming all Pro in the Y positions are hydroxylated.

[c]: The score of Mascot search engine (see Material and Methods).

[d]: Unexpected hydroxylation of X-Pro and unhydroxylated Y-Pro are shown as O_x and P_y; those observed in more than one sequencing effort are in bold; those with uncertainties are shown in italic.

The sequence of collagen α-chains in D-periodicity

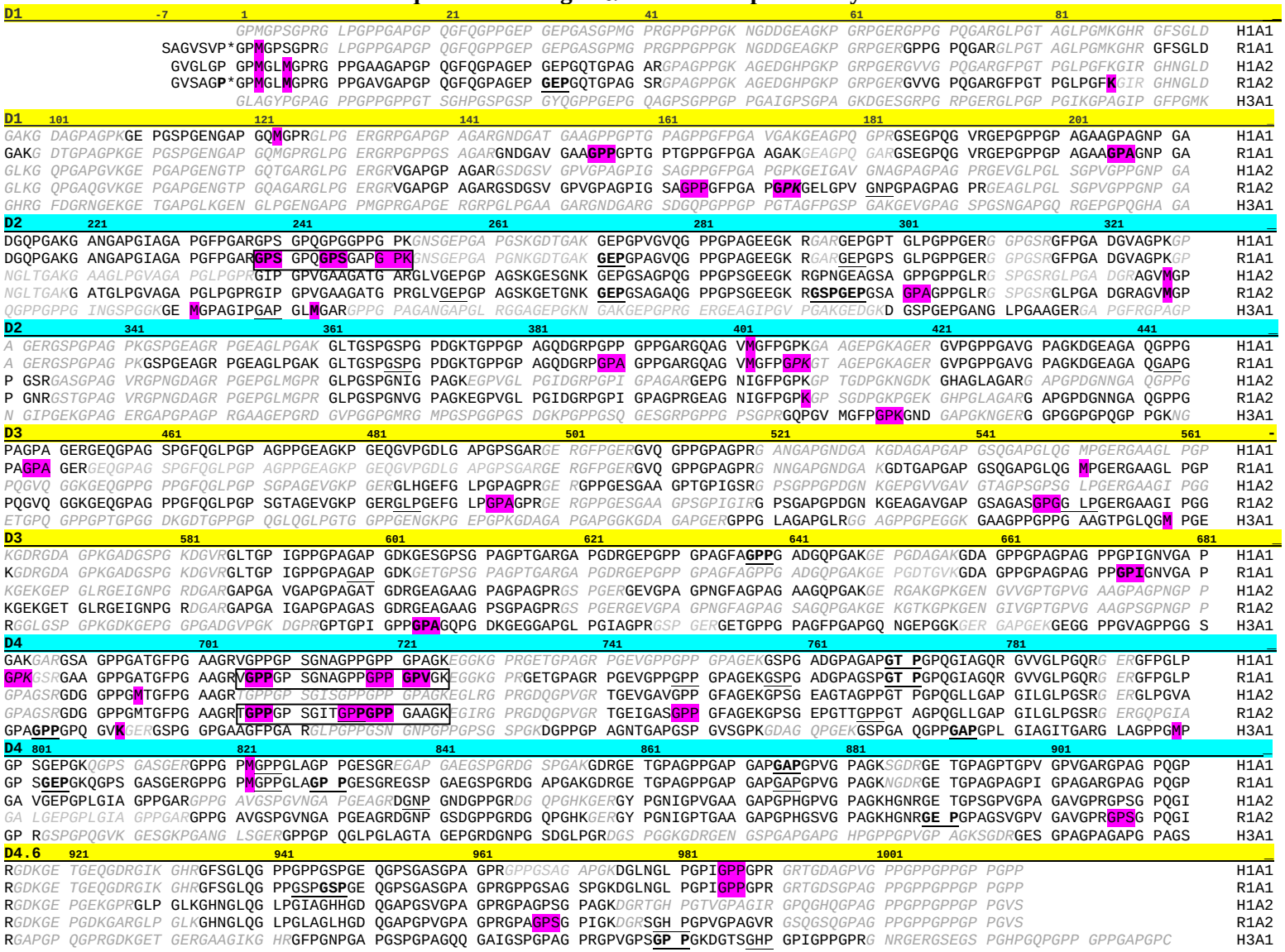


Figure 29. **Unexpected Hydroxylation Sites in Collagen α Chains.** Sequences of the α_1(I) and α_2(I) chains of human, the α_1(I) and the α_2(I) chains of rat and the α1(III) chains of human were arranged by the D-periodicity according to Di Lullo (DiLullo, G., et. al., 2002). The triplets including an O_x are in grey highlight, and in bold for more than one detection. Triplets with P_y are underscored, and in bold for multiple detection. The entire segment of three highly variable regions with multiple O_x residues are boxed. The hydroxylated proline in the telopeptide is P*. Hydroxylysines and oxidized methionines are in magenta highlight, and bold for more than one detection. Not sequenced regions are in grey italic.

Chapter 5

Appendix

Table S1: The complete sequencing results.

RCIα1_a Commercial Rat Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
783.4359	**43**			**561**	^{556}GAAGLOGPK
836.4373	**57**				^{907}GPAGPQGPR
886.4377	**62**				^{184}GSEGPQGVR
1088.5371	**64**			**318**	^{316}GFOGADGVAGPK
1161.5721	**67**			**405**	^{397}GQAGVMGFOGPK
1258.5910	**61**			**366,369**	^{361}GLTGSOGSOGPDGK
1306.6386	**92**			**249**	^{238}GPSGPQGPSGAOGPK
1322.6335	**65**		**245**	**249**	^{238}GPSGPQGOSGAOGPK
1328.6481	**74**	**804**		**795,798**	^{793}GFOGLOGPSGEPGK
1443.6975	**126**			**693,699**	^{688}GAAGPOGATGFOGAAGR
1451.6761	**63**			**297,303,306**	^{295}GEOGPSGLOGPOGER
1532.7816	**92**				^{889}GETGPAGPAGPIGPAGAR
1680.7572	**87**			**444**	^{435}DGEAGAQGAOGPAGPAGER
1832.8596	**51**			**819,825,831**	^{817}GPOGPMGPOGLAGPOGESGR
1840.9188	**50**			**708,717,720**	^{705}VGPOGPSGNAGPOGPOGPVGK
1856.9137	**61**		**707**	**708,717,720**	^{705}VGOOGPSGNAGPOGPOGPVGK
2136.9905	**81**			**759,768,771**	^{757}GSOGADGPAGSOGTOGPQGIAGQR
2185.0520	**74**			**867,870,873,876**	^{859}GETGPAGPOGAOGAOGAOGPVGPAGK
2228.9837	**54**			**537,540,546,552**	^{532}GDTGAOGAOGSQGAOGLQGMOGER
2244.9786	**67**			**537,540,546,552**	^{532}GDTGAOGAOGSQGAOGLQGMOGER
2307.1364	**58**			**663,672,681**	^{658}GDAGPOGPAGPAGPOGPIGNVGAOGPK
2316.0487	**52**			**195, 198,210,216**	^{193}GEOGPOGPAGAAGPAGNOGADGQOGAK

RCIα1_a continued

2323.1313	**79**		***683***	**663,672,681**	[658]GDAGPOGPAGPAGPOGPIGNVGAOG*OK*
2332.0436	**65**		**206**	**195,198,210,216**	[193]GEOGPOGPAGAAGOAGNOGADGQOGAK
2548.2063	**101**			**156,165,168**	[145]GNDGAVGAAGPOGPTGPTGPOGFOGAAGAK
2564.2012	**54**		**155**	**156,165,168**	[145]GNDGAVGAAGOOGPTGPTGPOGFOGAAGAK
2679.2394	**42**	**948**		**942,945**	[934]GFSGLQGPOGSOGSPGEQGPSGASGPAGPR
2695.2343	**132**			**942,945,948**	[934]GFSGLQGPOGSOGSOGEQGPSGASGPAGPR

RCIα1_b Commercial Rat Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
783.4359	**52**			**561**	^{556}GAAGLOGPK
836.4373	**72**				^{907}GPAGPQGPR
851.4258	**46**				^{91}GFSGLDGAK
852.4322	**41**			**69**	^{67}GPOGPQGAR
886.4377	**63**				^{184}GSEGPQGVR
898.5105	**60**			**786**	^{781}GVVGLOGQR
1088.5371	**66**			**318**	^{316}GFOGADGVAGPK
1105.5749	**47**			**513**	^{508}GVQGPOGPAGPR
1161.5721	**67**			**405**	^{397}GQAGVMGFOGPK
1177.5670	**48**			**405**	^{397}GQAGVMGFOGPK
1192.6321	**74**			**423,426**	^{421}GVOGPOGAVGPAGK
1258.5910	**64**			**366,369**	^{361}GLTGSOGSOGPDGK
1306.6386	**82**			**249**	^{238}GPSGPQGPSGAOGPK
1328.6481	**76**	**804**		**795,798**	^{793}GFOGLOGPSGEPGK
1338.6284	**58**		**239,245**	**249**	^{238}GOSGPQGOSGAOGPK
1344.6430	**41**			**795,798,804**	^{793}GFOGLOGPSGEOGK
1435.6812	**43**	**297**		**303,306**	^{295}GEPGPSGLOGPOGER
1443.6975	**127**			**693,699**	^{688}GAAGPOGATGFOGAAGR
1532.7816	**67**				^{889}GETGPAGPAGPIGPAGAR
1680.7572	**86**			**444**	^{435}DGEAGAQGAOGPAGPAGER
1744.7733	**46**			**840,846,852**	^{837}EGSOGAEGSOGRDGAOGAK
1819.8457	**91**			**273,282**	^{271}GEOGPAGVQGPOGPAGEEGK
1856.9137	**44**		**707**	**708,717,720**	^{705}VGOOGPSGNAGPOGPOGPVGK
1872.9086	**49**		**707, 722**	**708,717,720**	^{705}VGOOGPSGNAGPOGPOGOVGK
1959.9519	**87**	**273**		**282**	^{271}GEPGPAGVQGPOGPAGEEGKR
1975.9468	**94**			**273,282**	^{271}GEOGPAGVQGPOGPAGEEGKR

RCIα1_b continued

2014.9689	**41**			**378,387,393**	[375]TGPOGPAGQDGROGPAGPOGAR
2030.9639	**50**		**389**	**378,387,393**	[375]TGPOGPAGQDGROGOAGPOGAR
2120.9955	**98**	**771**		**759,768**	[757]GSOGADGPAGSOGTPGPQGIAGQR
2136.9905	**84**			**759,768,771**	[757]GSOGADGPAGSOGTOGPQGIAGQR
2169.0571	**57**	**876**		**867,870,873**	[859]GETGPAGPOGAOGAOGAPGPVGPAGK
2185.0520	**45**			**867,870,873,876**	[859]GETGPAGPOGAOGAOGAOGPVGPAGK
2216.0578	**57**			**741,747,750**	[733]GETGPAGROGEVGPOGPOGPAGEK
2228.9837	**99**			**537,540,546,552**	[532]GDTGAOGAOGSQGAOGLQGMOGER
2316.0487	**53**			**195,198,210,216**	[193]GEOGPOGPAGAAGPAGNOGADGQOGAK
2332.0436	**79**		**206**	**195,198,210,216**	[193]GEOGPOGPAGAAGOAGNOGADGQOGAK
2513.2015	**142**			**867,870,873,876**	[856]GDRGETGPAGPOGAOGAOGAOGPVGPAGK
2548.2063	**72**			**156,165,168**	[145]GNDGAVGAAGPOGPTGPTGPOGFOGAAGAK
2564.2012	**69**		**155**	**156,165,168**	[145]GNDGAVGAAGOOGPTGPTGPOGFOGAAGAK
2679.2394	**68**	**948**		**942,945**	[934]GFSGLQGPOGSOGSPGEQGPSGASGPAGPR
2695.2343	**127**			**942,945,948**	[934]GFSGLQGPOGSOGSOGEQGPSGASGPAGPR

RCIα1_c Rat Collagen Type I in-gel digestion

m/z^{+1}	**Score**	**Py**	**Ox**	**Oy**	**Peptide**
783.4359	**47**			**561**	^{556}GAAGLOGPK
836.4373	**44**				^{907}GPAGPQGPR
851.4258	**60**				^{91}GFSGLDGAK
852.4322	**43**			**69**	^{67}GPOGPQGAR
886.4377	**62**				^{184}GSEGPQGVR
898.5105	**56**			**786**	^{781}GVVGLOGQR
943.4480	**41**			**966,972**	^{964}GPOGSAGSOGK
945.4384	**56**				^{807}QGPSGASGER
1088.5371	**62**			**318**	^{316}GFOGADGVAGPK
1132.4865	**53**			**840,846**	^{837}EGSOGAEGSOGR
1161.5721	**72**			**405**	^{397}GQAGVMGFOGPK
1177.5670	**44**			**405**	^{397}GQAGV<u>M</u>GFOGPK
1193.5619	**54**			**405**	^{397}GQAGV<u>M</u>GFOG*P*<u>*K*</u>
1242.5961	**63**	**369**		**366**	^{361}GLTGSOGSPGPDGK
1322.6335	**76**		**239**	**249**	^{238}GOSGPQGPSGAOGPK
1338.6284	**76**		**239,251**	**249**	^{238}GOSGPQGPSGAOGOK
1344.6430	**57**			**795,798,804**	^{793}GFOGLOGPSGEOGK
1354.6233	**49**		**239,245,***251*	**249**	^{238}GOSGPQGOSGAOG*OK*
1443.6975	**42**			**693,699**	^{688}GAAGPOGATGFOGAAGR
1451.6761	**46**			**297,303,306**	^{295}GEOGPSGLOGPOGER
1452.7264	**65**	**-1**			$^{-7}$SAGVSVPGPMGPSGPR
1468.7213	**51**	**-1**			$^{-7}$SAGVSVPGP<u>M</u>GPSGPR
1484.7162	**40**			**-1**	$^{-7}$SAGVSVOGP<u>M</u>GPSGPR
1532.7816	**84**				^{889}GETGPAGPAGPIGPAGAR
1561.7969	**40**		**986**	**981,987**	^{975}DGLNGLOGPIGOOGPR
1574.8173	**86**	**600**		**594**	^{586}GLTGPIGPOGPAGAPGDK

RCIα1_c continued

1585.7717	**78**			**225,231,234**	[220]GANGAOGIAGAOGFOGAR
1590.8122	**42**			**594,600**	[586]GLTGPIGPOGPAGAOGDK
1655.7984	**60**			**345,351,357**	[343]GSOGEAGROGEAGLOGAK
1664.7623	**88**	**444**			[435]DGEAGAQGAPGPAGPAGER
1680.7551	**96**			**444**	[435]DGEAGAQGAOGPAGPAGER
1696.7521	**92**		**449**	**444**	[435]DGEAGAQGAOGPAGOAGER
1800.8697	**105**	**825,831**		**819**	[817]GPOGPMGPPGLAGPPGESGR
1803.8508	**77**	**273**		**282**	[271]GEPGPAGVQGPOGPAGEEGK
1816.8647	**125**	**825,831**		**819**	[817]GPOGPMGPPGLAGPPGESGR
1819.8457	**82**			**273,282**	[271]GEOGPAGVQGPOGPAGEEGK
1872.9086	**49**		***707*,719**	**708,720**	[705]VGOOGPSGNAGP*O*GOOG*P*VGK
1888.9035	**48**		**707,719,722**	**708,717,720**	[705]VGOOGPSGNAGPOGOOGOVGK
1975.9468	**69**			**273,282**	[271]GEOGPAGVQGPOGPAGEEGKR
2105.0006	**78**	**759,771**		**768**	[757]GSPGADGPAGSOGTPGPQGIAGQR
2244.9786	**98**			**537,540,546,552**	[532]GDTGAOGAOGSQGAOGLQGMOGER
2339.1262	**40**		**674,*683***	**663,672,*681***	[658]GDAGPOGPAGPAGPOGOIGNVGA*OGOK*
2513.2084	**67**			**867,870,873,876**	[856]GDRGETGPAGPOGAOGAOGAOGPVGPAGK
2548.2063	**58**			**156,165,168**	[145]GNDGAVGAAGPOGPTGPTGPOGFOGAAGAK
2564.2012	**96**		**155**	**156,165,168**	[145]GNDGAVGAAGOOGPTGPTGPOGFOGAAGAK
2663.2444	**96**	**945,948**		**942**	[934]GFSGLQGPOGSPGSPGEQGPSGASGPAGPR
2679.2394	**105**	**948**		**942,945**	[934]GFSGLQGPOGSOGSPGEQGPSGASGPAGPR
2695.2343	**62**			**942,945,948**	[934]GFSGLQGPOGSOGSOGEQGPSGASGPAGPR

RCIα2_a Commercial Rat Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
758.3791	**56**			**318**	^{316}GLOGADGR
868.4635	**50**			**138**	^{135}VGAOGPAGAR
868.4635	**58**				^{907}GPSGPQGIR
937.5102	**41**				^{964}GPAGPSGPIGK
987.4676	**54**			**330**	^{324}AGVMGPOGNR
1084.5633	**71**			**258**	^{253}GLVGEOGPAGSK
1159.5742	**82**			**405**	^{397}GEAGNIGFOGPK
1203.6229	**48**			**981**	^{978}SGHOGPVGPAGVR
1238.6052	**57**			**78,81,84**	^{76}GFOGTOGLOGFK
1239.6328	**80**			**363,366**	^{361}GLOGSOGNVGPAGK
1251.6076	**77**				^{604}GEAGAAGPSGPAGPR
1293.6910	**83**			**240**	^{238}GIOGPVGAAGATGPR
1399.6821	**43**			**-1**	$^{-6}$GVSAGOGPMGLMGPR
1453.7434	**89**			**486,492**	^{484}GLOGEFGLOGPAGPR
1490.7122	**49**			**750**	^{741}TGEIGASGPOGFAGEK
1510.7245	**68**			**588,594**	^{586}GAOGAIGAOGPAGASGDR
1560.8129	**102**	**891**			^{889}GEPGPAGSVGPVGAVGPR
1576.8078	**59**			**891**	^{889}GEOGPAGSVGPVGAVGPR
1592.8391	**83**			**225,231,234**	^{220}GATGLOGVAGAOGLOGPR
1592.7663	**58**	**297**		**294,306**	^{292}GSOGEPGSAGPAGPOGLR
1608.7612	**67**			**294,297,306**	^{292}GSOGEOGSAGPAGPOGLR
1615.8187	**81**			**183**	^{175}GELGPVGNOGPAGPAGPR
1781.7937	**69**	**273**		**282**	^{271}GEPGSAGAQGPOGPSGEEGK
2098.104	**52**			**381,387**	^{375}EGPVGLOGIDGROGPIGPAGPR
2276.1054	**122**			**861,873**	^{859}GYOGNIGPTGAAGAOGPHGSVGPAGK
2384.0597	**51**			**273,282**	^{265}GETGNKGEOGSAGAQGPOGPSGEEGK

RCIα2_a continued

2605.2529	**102**		***173***	**165,168,*171***	[145]GSDGSVGPVGPAGPIGSAGPOGFOGA*OGOK*
2743.3143	**79**			**588,594,**	[586]GAOGAIGAOGPAGASGDRGEAGAAGPSGPAGPR
2836.3496	**73**			**462,468,480**	[454]GEQGPAGPOGFQGLOGPSGTAGEVGKOGER
2971.512	**76**	**768**		**762,780,786**	[757]GPSGEOGTTGPPGTAGPQGLLGAOGILGLOGSR
2987.5069	**48**			**762,768,780,786**	[757]GPSGEOGTTGPOGTAGPQGLLGAOGILGLOGSR
3061.4206	**71**			**525,540,552**	[520]GPSGAOGPDGNKGEAGAVGAOGSAGASGPGGLOGER
4202.0537	**118**		**173**	**165,168,171,183**	[145]GSDGSVGPVGPAGPIGSAGPOGFOGAOGOKGELGPVGNOGPAGPAGPR

RCIα2_b Commercial Rat Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
840.4686	**60**				^{67}GVVGPQGAR
868.4635	**71**			**138**	^{135}VGAOGPAGAR
868.4635	**46**				^{907}GPSGPQGIR
987.4676	**52**			**330**	^{324}AGVMGPOGNR
1057.5636	**42**			**561**	^{556}GAAGIOGGKGEK
1084.5633	**75**			**258**	^{253}GLVGEOGPAGSK
1159.5742	**52**			**405**	^{397}GEAGNIGFOGPK
1238.6052	**58**			**78,81,84**	^{76}GFOGTOGLOGFK
1239.6328	**50**			**363,366**	^{361}GLOGSOGNVGPAGK
1251.6076	**81**				^{604}GEAGAAGPSGPAGPR
1293.6910	**94**			**240**	^{238}GIOGPVGAAGATGPR
1399.6821	**65**	**-1**			$^{-6}$GVSAGPGPMGLMGPR
1453.7434	**87**			**486,492**	^{484}GLOGEFGLOGPAGPR
1510.7245	**81**			**588,594**	^{586}GAOGAIGAOGPAGASGDR
1560.8129	**111**	**891**			^{889}GEPGPAGSVGPVGAVGPR
1576.8078	**91**			**891**	^{889}GEOGPAGSVGPVGAVGPR
1576.7714	**79**	**294,297**		**306**	^{292}GSPGEPGSAGPAGPOGLR
1592.7663	**50**	**297**		**294,306**	^{292}GSOGEPGSAGPAGPOGLR
1592.8391	**88**			**225,231,234**	^{220}GATGLOGVAGAOGLOGPR
1608.7612	**76**			**294,297,306**	^{292}GSOGEOGSAGPAGPOGLR
1615.8187	**93**			**183**	^{175}GELGPVGNOGPAGPAGPR
1781.7937	**69**	**273**		**282**	^{271}GEPGSAGAQGPOGPSGEEGK
1833.8905	**47**	**717**	**707,719**	**708,720**	^{705}TGOOGPSGITGPPGOOGAAGK
1892.8118	**45**			**840,846,852**	^{837}DGNOGSDGPOGRDGQOGHK
1937.8948	**115**	**273**		**282**	^{271}GEPGSAGAQGPOGPSGEEGKR
2041.0322	**97**			**891**	^{885}HGNRGEOGPAGSVGPVGAVGPR

RCIα2_b continued

2098.1040	**56**			**381,387**	[375]EGPVGLOGIDGROGPIGPAGPR
2276.1054	**134**			**861,873**	[859]GYOGNIGPTGAAGAOGPHGSVGPAGK
2368.0648	**41**	**273**		**282**	[265]GETGNKGEPGSAGAQGPOGPSGEEGK
2384.0597	**60**			**273,282**	[265]GETGNKGEOGSAGAQGPOGPSGEEGK
2605.2529	**80**		**173**	**165,168,171**	[145]GSDGSVGPVGPAGPIGSAGPOGFOGAOGOK
2743.3143	**88**			**588,594**	[586]GAOGAIGAOGPAGASGDRGEAGAAGPSGPAGPR
2836.3496	**79**			**462,468,480**	[454]GEQGPAGPOGFQGLOGPSGTAGEVGKOGER
3061.4206	**72**			**525,540,552**	[520]GPSGAOGPDGNKGEAGAVGAOGSAGASGPGGLOGER

RCIα2_c Rat Collagen Type I in-gel digestion

m/z^{+1}	Score	Py	Ox	Oy	Peptide
758.3791	**48**			**318**	^{316}GLOGADGR
840.4686	**54**				^{67}GVVGPQGAR
868.4635	**62**			**138**	^{135}VGAOGPAGAR
868.4653	**51**				^{907}GPSGPQGIR
884.4585	**46**		**908**		^{907}GOSGPQGIR
937.5102	**52**				^{964}GPAGPSGPIGK
953.5051	**51**		**968**		^{964}GPAGOSGPIGK
971.4727	**44**			**330**	^{324}AGVMGPOGNR
987.4676	**54**			**330**	^{324}AGVMGPOGNR
1068.5684	**71**	**258**			^{253}GLVGEPGPAGSK
1084.5633	**70**			**258**	^{253}GLVGEOGPAGSK
1157.4818	**57**			**840,846**	^{837}DGNOGSDGPOGR
1175.5691	**79**			**405**	^{397}GEAGNIGFOGPK
1187.6280	**54**	**981**			^{978}SGHPGPVGPAGVR
1238.6052	**64**			**78,81,84**	^{76}GFOGTOGLOGFK
1239.6328	**85**			**363,636**	^{361}GLOGSOGNVGPAGK
1251.6076	**77**				^{604}GEAGAAGPSGPAGPR
1399.6821	**95**				$^{-6}$GVSAGPGPMGLMGPR
1415.6770	**87**				$^{-6}$GVSAGPGPMGLMGPR
1428.7190	**52**			**579**	^{568}GETGLRGEIGNOGR
1431.6719	**46**			***-1***	$^{-6}$GVSAGOGPMGLMG*P*R
1437.7485	**79**	**486**		**492**	^{484}GLPGEFGLOGPAGPR
1453.7434	**75**			**486,492**	^{484}GLOGEFGLOGPAGPR
1469.7383	**110**		**494**	**486,492**	^{484}GLOGEFGLOGOAGPR
1490.7122	**59**			**750**	^{741}TGEIGASGPOGFAGEK
1506.7071	**67**		**749**	**750**	^{741}TGEIGASGOOGFAGEK

RCIα2_c continued

1510.7245	**86**			**588,594**	[586]GAOGAIGAOGPAGASGDR
1533.6751	**74**			**702,708**	[697]GDGGPOGMTGFOGAAGR
1560.8129	**102**	**891**			[889]GEPGPAGSVGPVGAVGPR
1576.8078	**110**			**891**	[889]GEOGPAGSVGPVGAVGPR
1576.7714	**76**	**294,297**		**306**	[292]GSPGEPGSAGPAGPOGLR
1592.7663	**85**	**294**		**297,306**	[292]GSPGEOGSAGPAGPOGLR
1599.8238	**87**	**183**			[175]GELGPVGNPGPAGPAGPR
1608.7612	**88**			**294,297,306**	[292]GSOGEOGSAGPAGPOGLR
1615.8187	**105**			**183**	[175]GELGPVGNOGPAGPAGPR
1624.7562	**94**		**302**	**294,297,306**	[292]GSOGEOGSAGOAGPOGLR
1748.8674	**69**	**297**		**294,306**	[291]RGSOGEPGSAGPAGPOGLR
1751.8307	**107**			**819,825,831**	[817]GPOGAVGSOGVNGAOGEAGR
1764.8624	**50**			**294,297,306**	[291]RGSOGEOGSAGPAGPOGLR
1781.7937	**85**	**273**		**282**	[271]GEPGSAGAQGPOGPSGEEGK
1797.7886	**94**			**273,282**	[271]GEOGSAGAQGPOGPSGEEGK
1817.9028	**52**			**708,717,720**	[705]TGPOGPSGITGPOGPOGAAGK
1833.8973	**69**	***717***	**707,*719***	**708, 720**	[705]TGOOGPSGITGP*PGO*OGAAGK
1849.8926	**89**		**707,719**	**708,717,720**	[705]TGOOGPSGITGPOGOOGAAGK
1865.8876	**97**		**707,716,719**	**708,717,720**	[705]TGOOGPSGITGOOGOOGAAGK
1953.8897	**52**			**273,282**	[271]GEOGSAGAQGPOGPSGEEGKR
2010.9475	**86**			**540,552**	[532]GEAGAVGAOGSAGASGPGGLOGER
2025.0373	**82**	**891**			[885]HGNRGEPGPAGSVGPVGAVGPR
2041.0322	**76**			**891**	[885]HGNRGEOGPAGSVGPVGAVGPR
2098.1040	**53**			**381,387**	[375]EGPVGLOGIDGROGPIGPAGPR
2145.9908	**76**			**432,444**	[430]GAOGPDGNNGAQGPOGPQGVQGGK
2368.0648	**61**	**273**		**282**	[265]GETGNKGEPGSAGAQGPOGPSGEEGK
2524.1659	**54**	**273**		**282**	[265]GETGNKGEPGSAGAQGPOGPSGEEGKR
2540.1608	**52**			**273,282**	[265]GETGNKGEOGSAGAQGPOGPSGEEGKR

RCIα2_c continued

2589.2580	**112**			**165,168,171**	[145]GSDGSVGPVGPAGPIGSAGPOGFOGAOGPK
2605.2529	**103**		***173***	**165,168,171**	[145]GSDGSVGPVGPAGPIGSAGPOGFOGAO***GOK***
2621.2478	**51**		**164,*173***	**165,168,171**	[145]GSDGSVGPVGPAGPIGSAGOOGFOGAOG***OK***
2786.4081	**77**			**952,964**	[934]GHNGLQGLOGLAGLHGDQGAOGPVGPAGPR
2987.5069	**84**			**762,768,780,786**	[757]GPSGEOGTTGPOGTAGPQGLLGAOGILGLOGSR
3045.4257	**78**	**552**		**525,540**	[520]GPSGAOGPDGNKGEAGAVGAOGSAGASGPGGLPGER
3077.4155	**76**		**548**	**525,540,552**	[520]GPSGAOGPDGNKGEAGAVGAOGSAGASGOGGLOGER
4186.0588	**64**	**183**	**173**	**165,168,171**	[145]GSDGSVGPVGPAGPIGSAGPOGFOGAOGOKGELGPVGNPGPAGPAGPR
4218.0487	**80**		**164,173**	**165,168,171.183**	[145]GSDGSVGPVGPAGPIGSAGOOGFOGAOGOKGELGPVGNOGPAGPAGPR
4459.2012	**41**			**750,762,768,780,786**	[741]TGEIGASGPOGFAGEKGPSGEOGTTGPOGTAGPQGLLGAOGILGLOGSR

HCIα1_a Commercial Human Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
836.4373	**69**				^{907}GPAGPQGPR
886.4377	**50**				^{184}GSEGPQGVR
898.5105	**58**			**786**	^{781}GVVGLOGQR
1088.5371	**48**			**318**	^{316}GFOGADGVAGPK
1105.5749	**40**			**513**	^{508}GVQGPOGPAGPR
1177.5670	**49**			**405**	^{397}GQAGVMGFOGPK
1258.5910	**69**			**366,369**	^{361}GLTGSOGSOGPDGK
1302.6437	**80**			**249**	^{238}GPSGPQGPGGPOGPK
1344.6430	**48**			**795,798,804**	^{793}GFOGLOGPSGEOGK
1459.6924	**135**			**693,699**	^{688}GSAGPOGATGFOGAAGR
1465.6918	**74**			**297,303,306**	^{295}GEOGPTGLOGPOGER
1546.7972	**76**				^{889}GETGPAGPAGPVGPVGAR
1706.7729	**104**			**444**	^{435}DGEAGAQGPOGPAGPAGER
1812.8875	**49**			**708,717,720**	^{705}VGPOGPSGNAGPOGPOGPAGK
1832.8596	**78**	**825**		**819,831**	^{817}GPOGPMGPPGLAGPOGESGR
1848.8545	**41**			**819,825,831**	^{817}GPOGPMGPOGLAGPOGESGR
2003.9781	**100**			**273,282**	^{271}GEOGPVGVQGPOGPAGEEGKR
2056.9795	**67**			**378,387,390,393**	^{375}TGPOGPAGQDGROGPOGPOGAR
2105.0006	**83**	**771**		**759,768**	^{757}GSOGADGPAGAOGTPGPQGIAGQR
2149.9785	**73**			**627, 630,639,645**	^{625}GEOGPOGPAGFAGPOGADGQOGAK
2281.1207	**61**			**663,672,681**	^{658}GDAGPOGPAGPAGPOGPIGNVGAOGAK
2316.0487	**97**			**195, 198,210,216**	^{193}GEOGPOGPAGAAGPAGNOGADGQOGAK
2497.2066	**72**	**876**		**867,870,873**	^{856}GDRGETGPAGPOGAOGAOGAPGPVGPAGK
2703.2394	**40**	**639**		**621,627, 630,645**	^{619}GAOGDRGEOGPOGPAGFAGPPGADGQOGAK
2705.2550	**132**			**942,945,948**	^{934}GFSGLQGPOGPOGSOGEQGPSGASGPAGPR
2869.4075	**73**			**594,600**	^{586}GLTGPIGPOGPAGAOGDKGESGPSGPAGPTGAR
4112.9180	**68**			**462,468,474,480,486,492**	^{454}GEQGPAGSOGFQGLOGPAGPOGEAGKOGEQGVOGDLGAOGPSGAR

HCIα1_b Commercial Human Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
836.4373	**67**				^{907}GPAGPQGPR
886.4377	**58**				^{184}GSEGPQGVR
1192.6320	**71**				^{421}GVOGPOGAVGPAGK
1459.6924	**101**			**693, 699**	^{688}GSAGPOGATGFOGAAGR
1465.6918	**78**			**297,303,306**	^{295}GEOGPTGLOGPOGER
1546.7972	**67**				^{889}GETGPAGPAGPVGPVGAR
1562.7863	**65**		**986**	**981,987**	^{975}DGLNGLOGPI**G**OOGPR
1585.7717	**72**				^{220}GANGAOGIAGAOGFOGAR
1706.7729	**93**			**444**	^{435}DGEAGAQGPOGPAGPAGER
1758.7347	**79**			**111, 114,120**	^{109}GEOGSOGENGAOGQ<u>M</u>GPR
2003.9781	**72**			**273,282**	^{271}GEOGPVGVQGPOGPAGEEGKR
2105.0006	**66**	**771**		**759,768**	^{757}GSOGADGPAGAOGTPGPQGIAGQR
2121.0373	**58**			**759,768,771**	^{757}GSOGADGPAGAOGTOGPQGIAGQR
2149.9785	**57**			**627, 630,639,645**	^{625}GEOGPOGPAGFAGPOGADGQOGAK
2281.1207	**84**			**663,672,681**	^{658}GDAGPOGPAGPAGPOGPIGNVGAOGAK
2497.2066	**81**	**876**		**867,870,873**	^{856}GDRGETGPAGPOGAOGAOGAPGPVGPAGK
2513.2015	**109**			**867,870,873,876**	^{856}GDRGETGPAGPOGAOGAOGAOGPVGPAGK
2703.2394	**63**	**639**		**621,627, 630,645**	^{619}GAOGDRGEOGPOGPAGFAGPPGADGQOGAK
2705.2550	**111**			**942,945,948**	^{934}GFSGLQGPOGPOGSOGEQGPSGASGPAGPR
2869.4075	**97**			**594,600**	^{586}GLTGPIGPOGPAGAOGDKGESGPSGPAGPTGAR

HCIα2_a Commercial Human Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
809.4377	**48**				^{421}GHAGLAGAR
868.4635	**73**			**138**	^{135}VGAOGPAGAR
868.4635	**40**				^{907}GPSGPQGIR
895.4632	**54**				^{964}GPAGPSGPAGK
960.4567	**62**			**330**	^{324}AGVMGPOGSR
1084.5633	**42**			**258**	^{253}GLVGEOGPAGSK
1201.5848	**67**			**399,405**	^{397}GEOGNIGFOGPK
1253.6484	**89**			**363,366**	^{361}GLOGSOGNIGPAGK
1267.6753	**51**			**240**	^{238}GIOGPVGAAGATGAR
1427.7134	**56**				$^{-6}$GVGLGPGPMGLMGPR
1477.7546	**67**			**492**	^{484}GLHGEFGLOGPAGPR
1488.7329	**49**			**750**	^{741}TGEVGAVGPOGFAGEK
1510.7245	**69**			**588,594**	^{586}GAOGAVGAOGPAGATGDR
1549.6700	**80**			**693, 699**	^{688}GDGGPOGMTGFOGAAGR
1562.7921	**85**				^{889}GETGPSGPVGPAGAVGPR
1580.7663	**81**			**504**	^{502}GPOGESGAAGPTGPIGSR
2027.0166	**119**				^{885}HGNRGETGPSGPVGPAGAVGPR
2115.1193	**134**			**795,804,813**	^{793}GLOGVAGAVGEOGPLGIAGPOGAR
2284.1469	**106**			**861,873**	^{859}GYOGNIGPVGAAGAOGPHGPVGPAGK
2727.3193	**88**			**588,594**	^{586}GAOGAVGAOGPAGATGDRGEAGAAGPAGPAGPR
2957.4963	**76**			**768,771,780,786**	^{757}GPSGEAGTAGPOGTOGPQGLLGAOGILGLOGSR
2959.3565	**47**			**12,18,30,33**	^{10}GPOGAAGAOGPQGFQGPAGEOGEOGQTGPAGAR

HCIα2_b Commercial Human Collagen Type I in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
868.4635	**78**			**138**	^{135}VGAOGPAGAR
895.4632	**65**				^{964}GPAGPSGPAGK
1168.4978	**61**	**840**		**846**	^{837}DGNPGNDGPOGR
1201.5848	**59**			**399,405**	^{397}GEOGNIGFOGPK
1267.6753	**96**			**240**	^{238}GIOGPVGAAGATGAR
1477.7546	**78**			**492**	^{484}GLHGEFGLOGPAGPR
1580.7663	**75**			**504**	^{502}GPOGESGAAGPTGPIGSR
1619.7772	**102**			**303,306**	^{292}GPNGEAGSAGPOGPOGLR
1775.8783	**77**			**303,306**	^{291}RGPNGEAGSAGPOGPOGLR
1823.8042	**89**			**273,282**	^{271}GEOGSAGPQGPOGPSGEEGK
2027.0166	**89**				^{885}HGNRGETGPSGPVGPAGAVGPR
2050.9941	**66**			**645**	^{625}GEVGPAGPNGFAGPAGAAGQOGAK
2115.1193	**124**			**795,804,813**	^{793}GLOGVAGAVGEOGPLGIAGPOGAR
2284.1469	**106**			**861,873**	^{859}GYOGNIGPVGAAGAOGPHGPVGPAGK
2396.0597	**93**			**273,282**	^{265}GESGNKGEOGSAGPQGPOGPSGEEGK
3397.7108	**67**			**930,933,942,954**	^{928}GLOGLOGHNGLQGLOGIAGHHGDQGAOGSVGPAGPR

HC3α1_a Commercial Human Collagen Type IIII in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
949.5102	**47**	**981**			^{973}GPVGPSGPPGK
965.5051	**73**			**981**	^{973}GPVGPSGPOGK
1094.5953	**62**			**519,525**	^{517}GPOGLAGAOGLR
1154.5623	**41**			**795,798**	^{790}GLAGPOGMOGPR
1173.5131	**69**			**849,855**	^{846}DGNOGSDGLOGR
1207.5702	**76**			**432,435,441**	^{430}GGOGGOGPQGPOGK
1303.6026	**48**			**699,702,708**	^{697}GSOGGOGAAGFOGAR
1508.7088	**51**			**909**	^{898}GESGPAGPAGAOGPAGSR
1530.7295	**40**			**990,996**	^{984}DGTSGHOGPIGPOGPR
1702.7887	**58**			**237,240**	^{229}GEMGPAGIOGAOGLMGAR
1833.9090	**110**			**828,834,843**	^{826}GPOGPQGLOGLAGTAGEOGR
1921.8886	**92**			**738,747,750**	^{735}DGPOGPAGNTGAOGSOGVSGPK
2104.0782	**103**			**768,774,777**	^{766}GSOGAQGPOGAOGPLGIAGITGAR
2180.0037	**72**			**546,549,555,561**	^{541}GAAGPOGPOGAAGTOGLQGMOGER
2283.1000	**79**	**687**		**672,678**	^{667}GEGGPOGVAGPOGGSGPAGPPGPQGVK
2690.2553	**78**			**945,948,951,954,966**	^{943}GFOGNOGAOGSOGPAGQQGAIGSOGPAGPR
2950.4653	**81**			**603,609,618,621**	^{595}GPTGPIGPOGPAGQOGDKGEGGAOGLOGIAGPR

HC3α1_b Commercial Human Collagen Type IIII in-gel digestion

m/z $^{+1}$	Score	Py	Ox	Oy	Peptide
949.5102	**67**	**981**			^{973}GPVGPSGPPGK
965.5051	**40**			**981**	^{973}GPVGPSGPOGK
1094.5953	**77**			**519,525**	^{517}GPOGLAGAOGLR
1138.5674	**65**			**795,798**	^{790}GLAGPOGMOGPR
1173.5131	**86**			**849,855**	^{846}DGNOGSDGLOGR
1203.5827	**49**			**408,414**	^{406}GQOGVMGFOGPK
1207.5702	**75**			**432,435,441**	^{430}GGOGGOGPQGPOGK
1303.6026	**57**			**699,702,708**	^{697}GSOGGOGAAGFOGAR
1508.7088	**69**			**909**	^{898}GESGPAGPAGAOGPAGSR
1530.7295	**55**			**990,996**	^{984}DGTSGHOGPIGPOGPR
1631.8024	**80**			**603,609**	^{595}GPTGPIGPOGPAGQOGDK
1670.7989	**72**			**237,240**	^{229}GEMGPAGIOGAOGLMGAR
1833.9090	**117**			**828,834,843**	^{826}GPOGPQGLOGLAGTAGEOGR
1921.8886	**90**			**738,747,750**	^{735}DGPOGPAGNTGAOGSOGVSGPK
2104.0782	**125**			**768,774,777**	^{766}GSOGAQGPOGAOGPLGIAGITGAR
2164.0088	**122**			**546,549,555,561**	^{541}GAAGPOGPOGAAGTOGLQGMOGER
2283.1000	**71**	**687**		**672,678**	^{667}GEGGPOGVAGPOGGSGPAGPPGPQGV<u>K</u>
2690.2553	**94**			**945,948,951,954,966**	^{943}GFOGNOGAOGSOGPAGQQGAIGSOGPAGPR
2950.4653	**82**			**603,609,618,621**	^{595}GPTGPIGPOGPAGQOGDKGEGGAOGLOGIAGPR

HC3α1_c Commercial Human Collagen Type IIII solution digestion

m/z $^{+1}$	**Score**	**Py**	**Ox**	**Oy**	**Peptide**
949.5102	**45**	**981**			^{973}GPVGPSGPPGK
965.5051	**44**			**981**	^{973}GPVGPSGPOGK
1203.5827	**57**			**408,414**	^{406}GQOGVMGFOGPK
1219.5776	**42**		***416***	**408,414**	^{406}GQOGVMGFOG*OK*
1514.7346	**50**	**990**		**996**	^{984}DGTSGHPGPIGPOGPR
1530.7295	**57**			**990,996**	^{984}DGTSGHOGPIGPOGPR
1631.8024	**96**			**603,609**	^{595}GPTGPIGPOGPAGQOGDK
1670.7989	**51**			**237,240**	^{229}GEMGPAGIOGAOGLMGAR
1756.7733	**88**			**303,306,312**	^{300}DGSOGEOGANGLOGAAGER
1833.9090	**125**			**828,834,843**	^{826}GPOGPQGLOGLAGTAGEOGR
2088.0832	**94**	**777**		**768,774**	^{766}GSOGAQGPOGAPGPLGIAGITGAR
2104.0782	**137**			**768,774,777**	^{766}GSOGAQGPOGAOGPLGIAGITGAR
2164.0088	**107**			**546,549,555,561**	^{541}GAAGPOGPOGAAGTOGLQGMOGER
2180.0037	**97**			**546,549,555,561**	^{541}GAAGPOGPOGAAGTOGLQG<u>M</u>OGER
2690.2553	**98**			**945,948,951,954,966**	^{943}GFOGNOGAOGSOGPAGQQGAIGSOGPAGPR
2966.4603	**58**		***605***	**603,609,618,621**	^{595}GPTGPIGPOG$\boldsymbol{O_x}$AGQOGD*K*GEGGAOGLOGIAGPR

HC3α1_d Commercial Human Collagen Type IIII solution digestion

m/z $^{+1}$	**Score**	**Py**	**Ox**	**Oy**	**Peptide**
949.5102	**42**	**981**			^{973}GPVGPSGPPGK
965.5051	**41**			**981**	^{973}GPVGPSGPOGK
1094.5953	**52**			**519,525**	^{517}GPOGLAGAOGLR
1138.5674	**44**			**795,798**	^{790}GLAGPOGMOGPR
1173.5131	**44**			**849,855**	^{846}DGNOGSDGLOGR
1203.5827	**52**			**408,414**	^{406}GQOGVMGFOGPK
1303.6026	**68**			**699,702,708**	^{697}GSOGGOGAAGFOGAR
1337.6808	**42**			**618,621**	^{613}GEGGAOGLOGIAGPR
1508.7088	**77**			**909**	^{898}GESGPAGPAGAOGPAGSR
1530.7295	**66**			**990,996**	^{984}DGTSGHOGPIGPOGPR
1631.8024	**90**			**603,609**	^{595}GPTGPIGPOGPAGQOGDK
1670.7989	**42**	**240**		**237**	^{229}GEMGPAGIOGAPGL<u>M</u>GAR
1833.9090	**136**			**828,834,843**	^{826}GPOGPQGLOGLAGTAGEOGR
2088.0832	**117**	**777**		**768,774**	^{766}GSOGAQGPOGAPGPLGIAGITGAR
2104.0782	**126**			**768,774,777**	^{766}GSOGAQGPOGAOGPLGIAGITGAR
2164.0088	**105**			**546,549,555,561**	^{541}GAAGPOGPOGAAGTOGLQGMOGER
2196.9792	**52**			**639,645,648,654**	^{634}GETGPOGPAGFOGAOGQNGEOGGK
2690.2553	**68**			**945,948,951,954,966**	^{943}GFOGNOGAOGSOGPAGQQGAIGSOGPAGPR
2966.4603	**60**		***605***	**603,609,618,621**	^{595}GPTGPIGPOG***O***$_x$AGQOGD*K*GEGGAOGLOGIAGPR

Chapter 6

Bibliography

Ala-Kokko, L., Kontusaari, S., Baldwin, C., Helena Kuivaniemi, H., and Prockop, D. (1989). Structure of cDNA clones coding for the entire prepro α1(III) chain of human type III procollagen. Biochem. J. 260, 509-516.

Apweiler, R., Bairoch, A., Wu, C., Barker, W., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M., Natale, D., O'Donovan, C., Redaschi, N., and Yeh, L. (2004). UniProt: The Universal Protein Knowledgebase. Nucleic Acids Research. 32, 115-119.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40(W1), 597-603.

Bachinger, H. (1987). The Influence of Peptidyl-Prolyl Cis-Trans Isomerase on the in Vitro Folding of Type III Collagen. The Journal of Biological Chemistry. 262(35), 17144-17148.

Bachinger, H., Bruckner, P., Timpl, R., Prockop, D. and Engel, J. (1980). Folding Mechanism of the Triple Helix in Type-III Collagen and Type-III pN-Collagen: Role of Disulfide Bridges and Peptide Bond Isomerization. Eur. J. Biochem. 106, 619-632.

Barnes, M., Constable, B., Morton, L., and Kodicek, E. (1971). Hydroxylysine in the N-terminal regions of α1 and α2-chains of various collagens. Biochem. J. 125(2), 433–437.

Barnes, M., Constable, B., Morton, L., and Kodicek, E. (1971). Hydroxylysine in the N-Terminal Telopeptides of Skin Collagen from Chick Embryo and Newborn Rat. Biochem. J. 125(3), 925-928.

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

Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112, 535-542.

Bienkowska, J., Cruz, M., Atiemo, A., and Handin, R., and Liddington, R. (1997). The von Willebrand Factor A3 Domain Does Not Contain a Metal Ion-dependent Adhesion Site Motif. The Journal of Biological Chemistry. 272(40), 25162–25167.

Cabral, W., Chang, W., Barnes, A., Weis, M., Scott, M., Leikin, S., Makareeva, E., Kuznetsova, N., Rosenbaum, K., Tifft, C., Bulas, D., Kozma, C., Smith, P., Eyre, D., and Marini, J. (2007). Prolyl 3-hydroxylase 1 deficiency causes recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfect. Nature genetics. 39, 359-365.

Craig, R., and Beavis, R. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 20, 1466-1467.

Cruz, M., Yuan, H., Lee, J., Wise, R., and Handin, R. Interaction of the von Willebrand factor (vWF) with collagen: localization of the primary collagen binding site by analysis of recombinant vWF A domain polypeptides. (1995) J. Biol. Chem. 270, 10822–10827.

Dalbey, R., Lively, M., Bron, S., and van Dijl, J. (1997). The chemistry and enzymology of the type I signal peptidases. Protein Sci. 6, 1129–1138.

De Marco, L., Girolami, A., Zimmerman,T., and Ruggeri, Z. (1986). von Willebrand Factor Interaction with the Glycoprotein IIb/IIIa Complex Its Role in Platelet Function as Demonstrated in Patients with Congenital Afibrinogenemia. J. Clin. Invest. 77, 1272-1277.

Dercova, K., Branislav, V., Tandlich, R., and Subova, L. (1999). Fenton's type reaction in chemical pretreatment of PCBs. Chemosphere. 39, 2621-2628.

DiLullo, G., Sweeney, S., Ala-Kokko, J., Antonio, J. (2002). Mapping the Ligand-binding Sites and Disease-associated Mutations on the Most Abundant Protein in the Human, Type I Collagen. J. Biol. Chem. 277(6), 4223-4231.

Doyle, B., Hulmes, D., Miller, A., Parry, D., Piez, K., and Woodhead-Galloway, J. (1974). A D-periodic narrow filament in collagen. Proc. R. Soc. Lond. B. 186, 67-74.

Elmagirbi, A., Sulistyarti, H., and Atikah. (2012). Study of Ascorbic Acid as Iron(III) Reducing Agent for Spectrophotometric Iron Speciation. J. Pure App. Chem. Res. 1(1), 11-17.

Eyre, D., Weis, M., Hudson, D., Wu, J., and Kim, L. (2011). A Novel 3-Hydroxyproline (3Hyp)-rich Motif Marks the Triple-helical C Terminus of Tendon Type I Collagen Comprehensive mass spectrometric mapping of the hydroxylated amino acid residues of the α1(V) collagen chain. The Journal of Biological Chemistry. 286(10), 7732–7736.

Fenn, J., Mann, M., Meng, C., Wong, S., and Whitehouse, C. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science. 246, 64-71.

Fitzsimmons, C., Cockburn, C., Hornsey, V., Prowse, C., and Barnes, M. (1988). The Interaction of von Willebrand Factor (vWf) with Collagen: Investigation of vWf-Binding Sites in the Collagen Molecule. Thrombosis and Haemostasis. 2, 186-192.

Fujimura, Y., Titanif, K., Holland, L., Roberts, J., Kostel, P., Ruggeri, Z., and Zimmerman, T. (1987). A Heparin-binding Domain of Human von Willebrand Factor Characterization and Localization to a Tryptic Frament Extending From Amino Acid Residue Val-449 To Lys-728. The Journal of Biological Chemistry. 262(4), 1734-1739.

Ginsburg, D., Handin, R., Bonthron, D., Donlon, T., Bruns, G., Latt, S., and Orkin, S. (1985). Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) clones and chromosomal localization. Science. 228(4706),1401-6.

.

Ginsburg, D., and Sadler, J. (1993). von Willebrand disease: a database of point mutations, insertions, and deletions. For the Consortium on von Willebrand Factor Mutations and Polymorphisms, and the Subcommittee on von Willebrand Factor of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Thromb. Haemost. 69(2), 177-184.

Guan, J., and Chance, M. (2005). Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. Trends in Biochemical Sciences. 30, 583-592.

Henkel, W., and Dreisewerd, K. (2007). Cyanogen bromide peptides of the fibrillar collagens I, III, and V and their mass spectrometric characterization: detection of linear peptides, peptide glycosylation, and cross-linking peptides involved in formation of homo- and heterotypic fibrils. Journal of Proteome Research. 6, 4269-4289.

Heyduk, T., Baichoo, N., and Heyduk, E. (2001). Hydroxyl radical footprinting of proteins using metal ion complexes. Met. Ions Biol. Syst. 38, 255-87.

Heyduk, E. and Heyduk, T. (1994). Mapping Protein Domains Involved in Macromolecular Interactions: A Novel Protein Footprinting Approach. Biochemistry. 33, 9643-9650.

Houde, D., Kauppinen, P., Mhatre, R., Lyubarskaya, Y. (2006). Determination of protein oxidation by mass spectrometry and method transfer to quality control. Journal of chromatography A. 1123, 189-198.

Huerre, C., Junien, C., Weil, D., Chu, M., Morabito, M., Van Cong, N., Myers, J., Foubert, C., Gross, M., Prockop, D, Boué, A., Kaplan, J., de la Chapelle, A., and Ramirez, F. (1982). Human type I procollagen genes are located on different chromosomes. Proc. Natl. Acad. Sci. 79(21), 6627–6630.

Huizinga, E., Martijn van der Plas, R., Kroon, J., Sixma, J., and Gros, P. (1997). Crystal structure of the A3 domain of human von Willebrand factor: Implications for collagen binding. Structure. 5, 1147-1256.

Hutton, J., Tappel, A., and Udenfriend, S. (1966). Requirements for a-ketoglutarate, ferrous ion and ascorbate by collagen proline hydroxylase. Biochemical and Biophysical Communications. 24(2), 179-184.

Kang, A., and Gross, J. (1970). Relationship between the Intra and Intermolecular Cross-links of Collagen. Proceedings of the National Academy of Sciences. 67(3), 1307-1314.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Anal. Chem. 60, 2299-2301.

Katsumi, A., Tuley, E., Bodo, I., and Sadler, J. (2000). Localization of disulfide bonds in the cystine knot domain of human von Willebrand factor. J. Biol. Chemistry. 275(33), 25585-25594.

Kielty, C., and Grant, M. (2002). Collagen: The Collagen Family: Structure, Assembly, and Organization in the Extracellular Matrix. Connective Tissue and Its Heritable Disorders. Wiley-Liss. New York, 159-221.

Kivirikko, K., Risteli, L. (1976). Biosynthesis of collagen and its alterations in pathological states. Med. Biol. 54, 159-86.

Kivirikko, K., Helaakoski, T., Tasanen, K., Vuori, K., Myllylä, R., Parkkonen, T., and Pihlajaniemi, T. (1990). Molecular biology of prolyl 4-hydroxylase. Ann. N.Y. Acad. Sci. 580, 132–142.

Koide, T., Asada, S., Takahara, Y., Nishikawa, Y., Nagata, K., and Kitagawa, K. (2006). Specific Recognition of the Collagen Triple Helix by Chaperone HSP47. Minimal Structural Requirement and Spatial Molecular Orientation. Journal of Biological Chemistry. 281(6), 3432-3438.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graphics. 14, 51-55.

Kuivaniemi, H., Tromp G., Chu M., Prockop, D. (1988). Structure of a full-length cDNA clone for the prepro a2(I) chain of human type I procollagen. Comparison with the chicken gene confirms unusual patterns of gene conservation. Biochem. J. 252, 633-640.

Lankhof, H., van Hoeij, M., Schiphorst, M., Bracke, M., Wu, Y., Ijsseldijk, M., Vink, T., de Groot, P., and Sixma, J. (1996). A3 domain is essential for interaction of von Willebrand factor with collagen type III. Thromb Haemost. 75(6), 950-8.

Lees, S. (2003). Mineralization of Type I Collagen. Biophysical Journal. 85, 204–207.

Limongi, M., Pelliccia, F., and Rocchi, A. (1997). Assignment of the human nebulin gene (NEB) to chromosome band 2q24.2 and the alpha 1 (III) collagen gene (COL3A1) to chromosome band 2q32.2 by in situ hybridization; the FRA2G common fragile site lies between the two genes in the 2q31 band. Cytogenet. Cell Genet. 77(3-4), 259-60.

Lisman, T., Raynal, N., Groeneveld, D., Maddox, B., Peachey, A., and Huizinga, E. (2006). A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. Blood. 108, 3753-3756.

Loizos, N. (2004). Identifying protein interactions by hydroxyl-radical protein footprinting. Methods Mol. Bio. 261, 199-210.

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectormetry. Rapid Commun. Mass Spectrom. 17(20), 2337-42.

Makareeva, E., Cabral, W., Marini, J., and Leikin, S. (2006). Molecular Mechanism of α1(I)–Osteogenesis Imperfecta/Ehlers-Danlos Syndrome: Unfolding of an N-Anchor Domain at the N-Terminal End of the Type I Collagen Triple Helix. Journal of Biological Chemistry. 281, 6463-6470.

Marini, J., Cabral, W., Barnes, A., and Chang, W. (2007). Components of the collagen prolyl 3-hydroxylation complex are crucial for normal bone development. Cell Cycle. 6(14), 1675-81.

Meek, K., Chapman, J., and Hardcastle, R. (1979). The staining pattern of collagen fibrils. Improved correlation with sequence data. Journal of Biological Chemistry. 254(21), 10710-10714.

Miller, R., Louis, B., Elsas, L., and Priest, R. (1979). Ascorbate Action on Normal and Mutant Human Lysyl Hydroxlases from Cultured Dermal Fibroblasts. The Journal of Investigative Dermatology. 72(5), 241-247.

Miura, S., Li, C., Cao, Z ., Wang, H., Wardell, M., and Sadler, J. (2000). Interaction of von Willebrand Factor Domain A1 with Platelet Glycoprotein Iba-(1–289). Slow Intrinsic Binding Kinetics Mediate Rapid Platelet Adhesion. J. Biol. Chem. 275, 7539-7546.

Molony, M., Quan, C., Mulkerrin, M., and Harris, R. (1998). Hydroxylation of Lys residues reduces their susceptibility to digestion by trypsin and lysyl endopeptidase. Analytical Biochemistry. 258, 136-137.

Myllyharju, J. (2003). Prolyl-4-hydroxylases, the key enzymes of collagen biosynthesis. Matrix Biology. 22, 15-24.

Nishida, N., Sumikawa, H., and Sakakura, M. (2003). Collagen-binding mode of vWf-A3 domain determined by a cross saturation experiment. Nat. Structural Biology. 10, 53-58.

Obert, B., Houllier, A., Meyer, D., Girma, J. (1999). Conformational changes in the A3 domain of von willebrand factor modulate the interaction of the A1 domain with platelet glycoprotein Ib. Blood. 93(6), 1959–1968.

Ogle, J., and Arlinghaus, R. (1962). Collagen 3-Hydroxyproline, a New Amino Acid of Collagen. J. Biol. Chem. 237, 3667-3673.

Perkins, D., Pappin, D., Creasy, D., and Cottrell, J. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 20, 3551-3567.

Persikov, A., Xu, Y., and Brodsky, B. (2004). Equilibrium thermal transitions of collagen model peptides. Protein Science. 13, 893-902.

Piez, K., Eigner, E., and Lewis, M. (1963). The chromatographic separation and amino acid composition of the subunits of several collagens. Biochemistry. 2, 58–66.

Piez, K., and Miller, M. (1974). The Structure of Collagen Fibrils. Journal of Supramolecular Structure. 2, 12 1–137.

Privalov, P., Tiktopulo, E., and Tischenko, V. (1979). Stability and mobility of the collagen structure. Journal of Molecular Biology. 127(2), 203-216.

Prockop, D., Sieron, A., and Shi-Wu, L. (1998). Procollagen N-proteinase and procollagen C-proteinase. Two unusual metalloproteinases that are essential for procollagen processing probably have important roles in development and cell signaling collagen. Matrix Biology. 16(7), 399-408.

Ramachandran, G. (1967). Structure of collagen at the molecular level. Treatise on collagen. Academic Press. New York, 103-183.

Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. 11(11), 601.

Romijn, R., Bouma, B., Wuyster, W., Gros, P., Kroon, J., Sixma, J., and Huizinga, E. (2001). Identification of the collagen-binding site of the von Willebrand factor A3-domain. J. Biol. Chem. 276, 9985-9991.

Romijn, R., Westein, E., Bouma, B., Schiphorst, M., Sixma, J., Lenting, P., and Huizinga, E. (2003). Mapping the Collagen-binding Site in the von Willebrand Factor-A3 Domain. The Journal of Biological Chemistry. 278, 15035-15039.

Rosenfeld, J., Capdevielle, J., Guillemot, J., and Ferrara, P. (1992). In gel digestion of protein for internal sequence analysis after one or two dimensional gel electrophoresis. Anal. Biochem. 203, 173-179.

Ruggeri, Z. (1997). Von Willebrand Factor. J. Clin. Invest. 99(4), 559–564.

Sadler, J. (1998). Biochemistry and genetics of von Willebrand factor. Annu. Rev. Biochem. 67, 395- 424.

Sadler, J. E. (2009). von Willebrand factor assembly and secretion. Journal of Thrombosis and Haemostasis. 7(Suppl. 1), 24–27.

Samor, B., Michalski, J., Debray, H., Mazurier, C., Goudemand, M.,Van Halbeek, H., Vliegenthart, J., and Montreuil, J. (1986). Primary structure of a new tetraantennary glycan of the N-acetyllactosaminic type isolated from human factor VIII/von Willebrand factor. Eur. J. Biochem. 158, 295-298.

Satoh, A., Trosko, J., and Masten, S. (2007). Methylene Blue Dye Test for Rapid Qualitative Detection of Hydroxyl Radicals Formed in a Fenton's Reaction Aqueous Solution. Environ. Sci. Technol. 41, 2881-2887.

Schulte am Esch, J., Robson, S., Knoefel, W., Eisenberger, C., Peiper, M., and Rogiers, X. (2005). Impact of O-linked glycosylation of the VWF-A1-domain flanking regions on platelet nteraction. Br. J. Haematol. 128(1), 82-90.

Schuppan, D., Glanville, R., Timpl, R., Dixit, S., and Kang, A. (1984). Sequence comparison of pepsin-resistant segments of basement-membrane collagen al(IV) chains from bovine lens capsule and mouse tumour. Biochem. J. 220, 227-233.

Sechi, S., and Chait B. (1998). Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. Anal. Chem. 70, 5150–5158.

Sharp, J., Becker, J., and Hettich, R. (2003). Protein surface mapping by chemical oxidation: Structural analysis by mass spectrometry. Analytical Biochemistry. 313, 216-225.

Smith-Mungo, L., and Kagan, H. (1998). Lysyl oxidase: Properties, regulation and multiple functions in biology. Matrix Biology. 16(7), 387-398.

Spiro, R., and Spiro, M. (1971). Studies on the Biosynthesis of the Hydroxylysine-linked Disaccharide Unit of Basement Membranes and Collagens: III. Tissue and Subcellular Distribution of Glycosyltransferases and the Effect of Various Conditions on the Enyzme Levels. J. Biol. Chem. 246, 4919-4925.

Starman, B., Eyre, D., Charbonneau, H., Harrylock, M., Weis, M., Weiss, L., Graham, J., and Byers, P. (1989). Osteogenesis Imperfecta: The Position of Substitution for Glycine by Cysteine in the Triple Helical Domain of the Pro αl(I) Chains of Type I Collagen Determines the Clinical Phenotype. J. Clin. Invest. 84, 1206-1214.

Takahashi, Y., Kalafatis, M., Girma, J., Sewerin, K., Andersson, L., Meyer, D. (1987). Localization of a factor VIII binding domain on a 34 kilodalton fragment of the N-terminal portion of von Willebrand factor. Blood. 70, 1679-1682.

Takamoto, K., and Chance, M. (2006). Radiolytic Protein Footprinting with Mass Spectrometry to Probe the Structure of Macromolecular Complexes. Annu. Rev. Biophys. Biomol. Struct. 35, 251-76.

Thompson, J., Higgins, D., and Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22), 4673-80.

Titani, K., Kumar , S., Takio, K., Ericsson, L., Wade, R., Ashida, K., Walsh, K., Chopek, M., Sadler, J., and Fujikawa, K. (1986). Amino acid sequence of human von Willebrand factor. Biochemistry. 25, 3171–3184.

Tkocz, C., and Kühn, K. (1968). The formation of Triple-Helical Collagen Molecules from α1 or α2 Polypeptide Chains. European J. Biochem. 7, 454-462.

Tromp, G., Kuivaniemi, H., Stacey, A., Shikata, H., Baldwin, C., Jaenisch, R., Prockop, D. (1988). Structure of a full-length cDNA clone for the prepro αl(I) chain of human type I procollagen. Biochem. J. (1988). 253, 919-22.

Tullius, T., and Dombroski, B. (1986). Hydroxyl radical "footprinting": High-resolution information about DNA-protein contacts and application to X repressor and Cro protein iron(II) EDTA/Fenton reaction/DNA-protein interaction. Proc. Natl. Acad. Sci. 83, 5469-5473.

Van der Plas, R., Gomes, L., Marquart, J. A., Vink, T., Meijers, J., de Groot, P., Sixma, J., and Huizinga, E. (2000). Thromb. Haemost. 84, 1005–1011.

Vischer, U., and Wagner, D. (1994). von Willebrand factor proteolytic processing and multimerization precede the formation of Weibel–Palade bodies. Blood. 83, 3536–44.

Weis, M., Hudson, D., Kim, L., Scott, M., Wu, J., and Eyre, D. (2010). Location of 3-hydroxyproline residues in collagen types I, II, III and V/XI implies a role in fibril supramolecular assembly. The Journal of Biological Chemistry. 285(4), 2580–2590.

Xu, K., Nowak, I., Kirchner, M., and Xu, Y. (2008). Recombinant Collagen Studies Link the Severe Conformational Changes Induced by Osteogenesis Imperfecta Mutations to the Disruption of a Set of Interchain Salt Bridges. The Journal of Biological Chemistry. 283(49), 34337–34344.

Xu, Y., Gurusiddappa, S., Rich, R., Owens, R., Keenei, D., Mayne, R., Agneta Höök, A., and Höök, M. (2000). Multiple Binding Sites in Collagen Type I for the Integrins α1b1and α2b1. The Journal of Biological Chemistry. 255(50), 38981–38989.

Yang, C., Park, A., Davis, N., Russell, J., Kim, B., Brand, D., Lawrence, M., Ge, Y., Westphall, M., Coon, J., and Greenspan, D. (2012). Comprehensive mass spectrometric mapping of the hydroxylated amino acid residues of the α1(V) collagen chain. The Journal of Biological Chemistry. 287(48), 40598–40610.