Investigate the Effects of Osteogenesis Imperfecta Mutations on the Conformation of Collagen Triple Helix

By

KE XU

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

2009

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.

Prof. Yujia Xu

Date

Chair of Examining Committee

Dr. Edward Kennelly

Date

Executive Officer

Dr. Thomas Haines

Dr. Manfred Philipp

Dr. Frida Kleiman

Dr. Weigang Qiu

Supervising Committee

THE CITY UNIVESITY OF NEW YORK

ABSTRACT

Investigate the Effects of Osteogenesis Imperfecta Mutations on the Conformation of Collagen Triple Helix

By

KE XU

Advisor: Professor Yujia Xu

The clinical severity of Osteogenesis Imperfecta (OI), also known as the brittle bone disease, relates to the extent of conformational changes in the collagen triple helix induced by Gly substitution mutations. The lingering question is why Gly substitutions at different locations of collagen cause different disruptions of the triple helix. Here, we describe markedly different conformational changes of the triple helix induced by two Gly substitution mutations placed only 12 residues apart. The effects of the Gly substitutions were characterized using a recombinant collagen fragment modeling the 63-residue segment of the _1 chain of type I collagen containing no Hyp (residues 877 – 939) obtained from Escherichia coli. Two Gly3Ser substitutions at Gly-901 and Gly-913 associated with, respectively, mild and severe OI variants were introduced by site-directed mutagenesis. Biophysical characterization and limited protease digestion experiments revealed that while the substitution at Gly-901 causes relatively minor destabilization of the triple helix, the substitution at Gly-913 induces large scale unfolding of an unstable region C-terminal to the mutation site. This extensive unfolding

is caused by the intrinsic low stability of the C-terminal region of the helix and the mutation induced disruption of a set of salt bridges, which functions to lock this unstable region into the triple helical conformation. The extensive conformational changes associated with the loss of the salt bridges highlight the long range impact of the local interactions of triple helix and suggest a new mechanism by which OI mutations cause severe conformational damages in collagen.

In addition to the biomedical studies, the recombinant collagen fragments have also proven to be a good system to produce nano-templates. The rigidity of the helix backbone, the linear conformation, and the largely exposed side chains of residues at the X and Y positions for chemical modification make the triple helix an ideal template for nanowires.

ACKNOWLEDGEMENTS

First of all, I would like to express my genuine gratitude to my mentor Dr. Yujia Xu who gives me the invaluable support and advice on both academic and personal aspects. This research project would not have been possible without the help of my mentor. It is an honor to have the opportunity to work under her guidance. I could not have asked for more.

I am also thankful to all the committee members of my dissertation defense Dr. Thomas Haines, Dr. Manfred Philipp, Dr. Maria Tomasz, Dr. Frida Kleiman and Dr. Weigang Qiu without their knowledge and assistance this study would not have been successful.

Many thanks go to my wonderful friends and colleagues in the lab, Michele Kirchner, Iwona Nowak and Irina Abramov.

I would like to express my love and gratitude to my beloved parents; for their understanding and endless love, through the duration of my studies.

TABLE OF CONTENTS

List of Tables ……viii

List of Diagrams ……viii

List of Figures ……viii

Introduction ……1

Materials and Methods ……14

Chapter I……22

Chapter II ……71

Chapter III …… 89

Appendixes ……98

Bibliography ……109

LIST OF TABLES

INTRODUCTION

Table 1 Clinical identified Osteogenesis Imperfecta (OI) phenotypes and their characteristics……13

RESULTS

Table 2 The purpose and results of each of the mutations generated in lab……39

LIST OF DIAGRAMS

RESULTS

Diagram 1 The Construction of our recombinant collagen protein……37

Diagram 2 DNA sequence and encoding amino acids sequence of expression range in our reconstructed pET32a (+) plasmid……42

LIST OF FIGURES

INTRODUCTION

Figure 1 The synthesis, processing, and assembly of Type I collagen molecules into fibrils……9

Figure 2 The triple helix structure of type I collagen. ……11

MATERIALS AND METHODS

Figure 3 the mechanism of site-directed nucleotide mutation by PCR……21

CHAPTER I

Figure 4 SDS-PAGE analysis of the expression of F877 fusion protein. ……44

Figure 5 NMR result of F877 sample……46

Figure 6 CD spectra of collagen fragments……48

Figure 7 Temperature melt curves of the collagen fragments……50

Figure 8 The protease digestion of collagen fragments……52

Figure 9 Temperature melt curves of the collagen fragments (2) ……54

Figure 10 Temperature melt curves of the collagen fragments (3) ……56

Figure 11 Temperature melt curves of the collagen fragments (4) ……58

Figure 12 CD spectra of collagen fragments (2) ……60

Figure 13: unfolding steps of recombinant collagen G913S triple helix……70

CHAPTER II

Figure 14: Illustrated structure of the collagen-like triple helix and TEM image of the F877 and G913S triple-helix……82

Figure 15: Construct nanowire using recombinant collagen molecular as template……84

Figure 16: TEM images of the F877 triple-helix peptides coated by Au……86

CHAPTER III

Figure 17: SDS-PAGE of pull-down assay experiment of Hsp47 and F877, G913S……93

Figure 18: Design of long collagen recombinant protein…… 95

Figure 19: SDS-PAGE analysis of the expressed long triple helix recombinant protein and F877 recombinant protein ……97

INTRODUCTION

Overview of type I collagen and its unique triple helix structure:

Collagen is the most abundant protein in human body and one of the major constituents of body's extra-cellular matrix. Its main function is to maintain the structure integrity of tissues and organs of human body (**Kielty, *et al* 1993**). At the same time, collagen also transfers communication signals between the intra- and extra-cellular molecules (**Kadler 1995; Myllyharju and Kivirikko 2001**). More than 25 types of collagens in this family have been explored during the last two decades (**Kielty, *et al.* 1993**). Collagens consist of fibrils with a variety of lengths and diameters. In collagen family, Type I collagen is the classic fibrous collagen, which provides the major mechanical strength of skin, tendon, bone, dentine, cornea and sclera (**Kielty, *et al.* 1993**). Most of the current understanding of collagen's structure, synthesis and assembly is derived from the research on type I collagen**.**

Type I collagen is synthesized in the endoplasmic reticulum (**Kielty, *et al.* 1993**). This biosynthetic process is characterized by several co-translational and post-translational modifications. Intracellular modifications of synthesized collagen chains result in the formation of triple helix procollagen molecules, and extra-cellular processing turns procollagen into collagen and incorporate them into stable, cross-linker fibrils (**Figure 1**) (**Baum and Brodsky 1999**).

The three dimensional structure of type I collagen was first, proposed based on

studies using fiber diffraction and electron microscope in the 1960s (**Rich and Crick 1961; Ramachandran *et al.* 1968**). The first X-ray structure of the triple helix was reported using short synthetic peptides (**Bella *et al.* 1994**). This unique structure of the triple helix is made of three supercoiled polyproline II-like chains (**Figure 2B**). Each amino acid chain requires a Gly at every third residue to form a unique $(Gly\text{-}X\text{-}Y)_n$ repeating sequence pattern, where X and Y can be any amino acids (**Figure 2C**). Gly, the smallest amino acid in every three residues, is essential for the triple helix structure's stability because it is the only amino acid packed into the restricted space in the center of triple helix, while the other residues on X and Y positions are largely exposed to solvent. In type I collagen, the triple helix molecule is a heterotrimer comprising of two identical α1 chains and one α2 chain (**Figure 2A**). The α1 and α2 chains are very similar with over 95% identity on amino acid sequence (**Kielty, e*t al.* 1993**). Each α-chain contains over 1000 amino acids and has a molecular weight of approximately 95,000. These molecules of type I collagen have a length of slightly less than 300nm and diameter of about 1.4nm.

Gly substitution mutations in type I collagen is the major cause of Osteogenesis Imperfecta (OI) disease:

The most common genetic disorder mutation in type I collagen is a single base substitution that leads to the replacement of an obligate Gly by another amino acid (**Byers 1993; Marini *et al.* 2007**). This Gly replacement mutations in type I collagen will destabilize the triple helix structure and cause a series of genetic disorder diseases,

including Osteogenesis Imperfecta (OI) disease. Osteogenesis Imperfecta (OI) disease is a genetic disorder disease characterized by unusually fragile bones that break easily, often under loads that normal bones bear daily. This inherent weakness of the bones is due to a malfunction in the patient's production of the type I collagen. Nearly 200 Gly replacing mutations from both α1 and α2 chains have been linked to OI (**Appendix A**). Recently, the research on synthesized collagen-like peptides indicates that depending on the location and the environment of the Gly substitution, the clinical severity of OI phenotype varies from mild increase of bone fragility (Phenotype I) to the most severe type characterized by death shortly after birth (Phenotype II) (**Table 1**) (**Byers 1993; Marini *et al.* 2007**). However, it remains unclear yet what molecular properties are related to the sequence locations of the Gly substitutions, which in turn affect the properties of bones.

Two established models explaining the intrinsic relationship between OI phenotype and OI mutation site:

While the clear relationship of the mutation to the phenotype remains elusive, recent studies developed several models for relating Gly mutations to OI phenotype, though each has had met with limited success. The renucleation model emphasizes the renucleation action of the mutated chains (**Byer *et al.* 1991**; **Marini *et al.* 2007**). Evidence indicates that the replacement of the Gly by a larger residue delays the folding of the triple helix by interrupting the C-terminus to N-terminus propagation of the helix

conformation (**Liu *et al.* 1998**). The continued folding of the OI collagen requires re-nucleation of the three chains from the mutation site to their N-terminal (**Buevich *et al.* 2004; Hyde *et al.* 2006**). Segments of triple helix consisting of residues with high triple helix propensity and high content of imino acids (Pro and Hyp at the X and the Y positions, respectively) were found to facilitate the renucleation (**Hyde *et al.* 2006**). Thus, it is postulated that the sequence location of the OI mutation relative to an effective "renucleation site" determines the delay time of folding and, in turn, the clinical severity. Due to the propagation of the helix conformation from C-terminus to N-terminus, the disease severity increases with the position of the mutation site along the triple helix from N-terminus to C-terminus.

On the other hand, the domain model emphasizes the stability of the triple helix surrounding the mutation site. Studies using host-guest peptides indicate that residues at the X and Y positions have profound effects on the molecular properties of the triple helix (**Persikov *et al.* 2002; Xu *et al.* 2003; Persikov *et al.* 2005**). Stretches of triple helix consisting of amino acid residues with low triple helix propensity tend to have low stability and flexible conformation and are often referred to as the "micro-unfolding" domains. The presence of multiple flexible, independent micro-unfolding domains along a collagen molecule has been implicated by calorimetric studies (**Privalove, 1979; Bachinger *et al.* 1993**), though their location, size and involvement in the biological functions of collagen are not yet well understood. Mutations disrupting highly stabilizing regions of the helix were predicted to be more severe than those in the micro-unfolding

region (**Bachinger *et al.* 1993**). A recent study using full chain type I collagen has shown that the OI mutation in a region with high triple helix stability, i.e., the N-anchor domain consisting of the first 85 residues of the type I collagen, causes the unfolding of the entire domain (**Makareeva *et al.* 2008**). The effects of similar mutations in a neighboring "flexible" region consisting of residues with low helical propensity are less extensive. The flexible region was also considered the buffering zone to stop the propagation of the effects of the Gly substitution to the next part of the triple helix. However, a correlation of the extent of the effects on the triple helix conformation and clinical severity was not apparent. All the mutations characterized in this study (**Makareeva *et al.* 2008**) are linked to relatively minor OI phenotypes.

Shortages of recent experimental methods to testify these two models:

The difficulties in experimentally examining such correlations reflect, on the one hand, the limitations of studies using full-chain collagen, and on the other hand, the limitation of the current models used to derive local stability of the triple helix. Studies with full-chain collagens are often limited to relying on macroscopic parameters such as the ΔTm, which does not imply any specific changes of the conformation (**Makareeva *et al.* 2008**). Given the long, rope-like structure of the triple helix, a mutation could critically alter the surrounding conformations of the triple helix without causing significant changes in *Tm* of the whole molecule.

Because it is difficult to estimate directly by experiments, the local helix stability of a

segment of the triple helix is usually derived from data obtained from synthetic peptides with limited sizes. Despite extensive peptide studies on the stabilizing/destabilizing effects of isolated X or Y residues, and of the intrachain and interchain molecular interactions involving both X and Y residues in one Gly-X-Y tripeptide or in two adjacent tripeptides (Gly-X-Y-Gly-X'-Y'), how these interactions modulate each other in the context of the long helix of collagen remains largely unknown (**Persikov *et al.* 2002; Persikov *et al.* 2005**) Prediction of local helix stabilities for the peptide data requires assumptions on 1) the range of the stabilizing and/or destabilizing interactions in the context of the long helix of collagen, and 2) the size of a cooperative unit – the size at which a segment of the helix with low stability would undergo cooperative unfolding **(Persikov *et al.* 2005**; **Makareeva *et al.* 2008)**. Both assumptions await experimental validation.

Our novel recombinant collagen protein as a new method to investigate the effect of OI mutation on type I collagen triple helix conformation:

We have developed an *E. coli* expression system for recombinant collagen fragments to overcome the limitations of the studies using full-chain collagens or short synthetic peptides. The recombinant collagen fragments were used to model the 63 residue segment of α1 chain of type I collagen corresponding to residues 877 – 939 (a region containing no hydroxyproline). The optimal size of the fragments allowed us to reveal drastically different effects on the triple helix conformation by two OI mutations placed only 12

residues apart, and to correlate the severe OI to a mutation causing more extensive unfolding of a segment of the triple helix known to be involved in the molecular interactions of type I collagen. Furthermore, we were able to demonstrate that the extensive unfolding is caused by the interruption of a set of interchain salt bridges formed between the K and E residues of a KGE sequence. Such interchain salt bridges (from a KGE or KGD sequence) are known to stabilize the short peptides (**Venugopal *et al*. 1994; Persikov *et al*. 2005**) and provide essential stability in bacteria collagen (**Mohs A *et al*. 2007**). In our model system, the stabilizing interactions of the salt bridges extended through a region containing nearly 20 amino acid residues C-terminus to the KGE sequence. Thus, by revealing the long-range impacts of the salt-bridges, our data highlight the critical roles of the salt bridges in modulating the stabilities of the neighboring regions of the triple helix. A survey of the OI mutation database on Uniprot revealed that 90% of Gly mutations placed within 5 residues of a KGE or KGD sequences in both α1 and α2 chains are linked to lethal cases of OI, suggesting the general nature of the critical effects of the salt bridges in maintaining the conformation and in the functions of collagen. The data presented here supports a novel explanation for the varied effects of mutations at different locations. The conformation of the collagen triple helix is maintained by certain specific stabilizing interactions, such as the interchain salt bridges. Mutations placed in the vicinity of these interactions tend to cause more disruptions of the overall conformation and, possibly, lead to more severe phenotypes of OI.

FIGURE 1: Schematic representation of the intracellular and extracellular steps involved in the synthesis, processing, and assembly of Type I collagen molecules into fibrils (Adapted from **Baum and Brodsky 1999**).

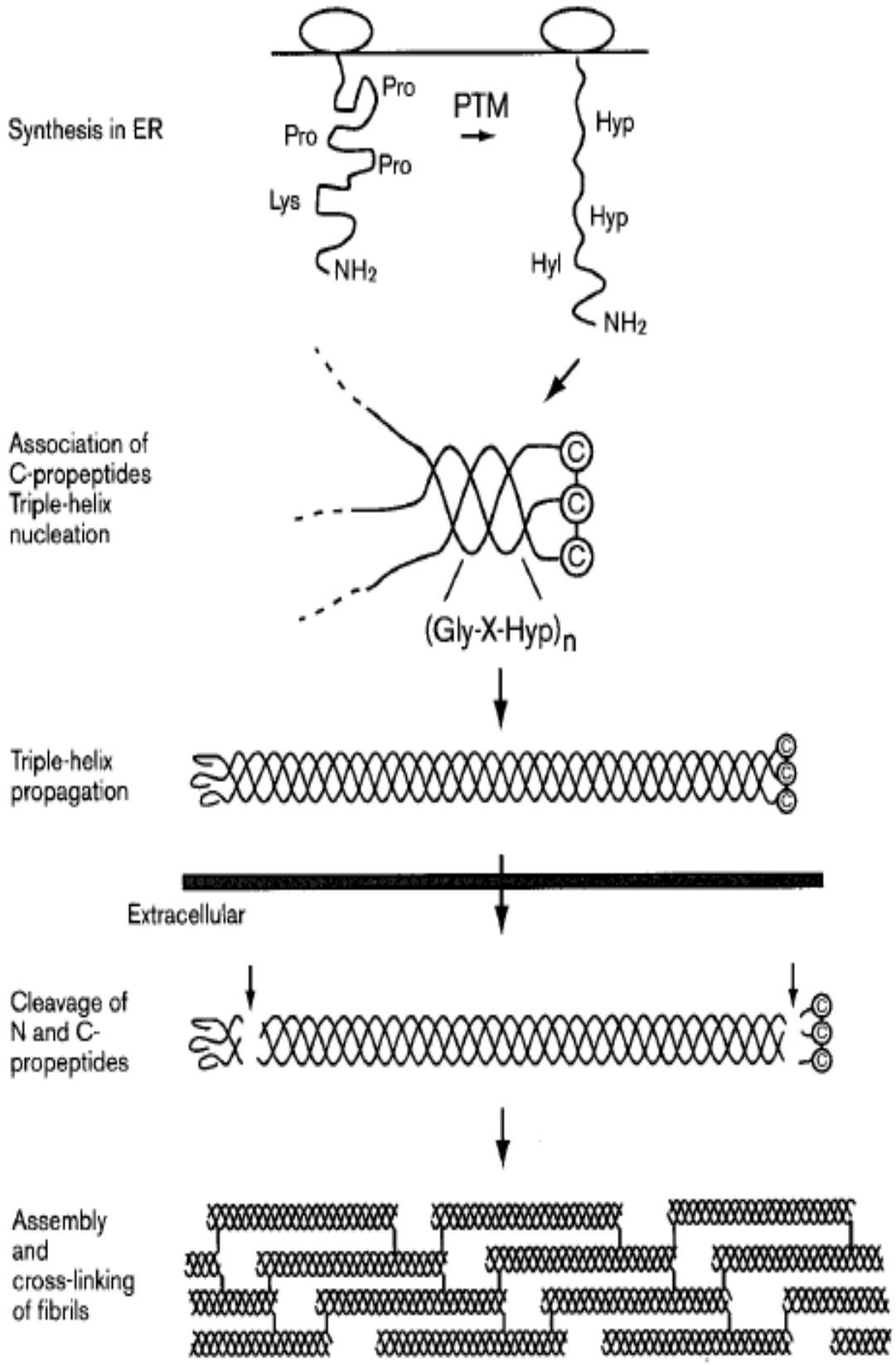
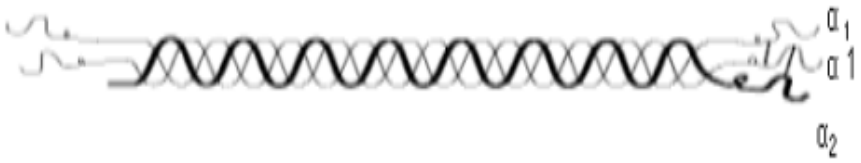
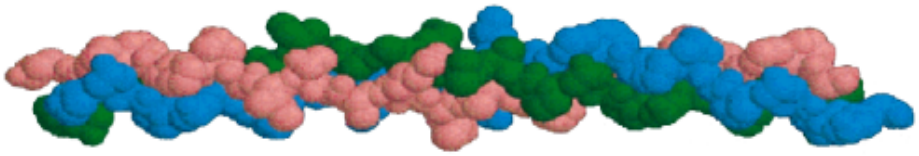

Synthesis in ER
Pro
Pro
Pro
Lys
NH2
PTM
Hyp
Hyp
Hyl
NH2
Association of
C-propeptides
Triple-helix
nucleation
C
C
C
(Gly-X-Hyp)n
Triple-helix
propagation
Extracellular
Cleavage of
N and C-
propeptides
Assembly
and
cross-linking
of fibrils

FIGURE 2: A) Scheme of triple helix structure of type I collagen heterotrimer B) 3D triple helix structure of type I collagen C) Top view scheme of type I collagen triple helix structure

A



B



C

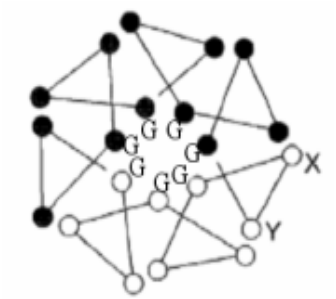


TABLE 1: Clinical identified phenotypes of Osteogenesis Imperfecta (OI) disease and their characteristics (**Sillence *et al.* 1979).**

OI type	Phenotype	Typical Clinical Features
I	Mild	Normal stature, little or no deformity, blue sclerae, hearing loss in 50% of families
II	Lethal	Lethal in the perinatal period; minimal calvarial mineralization, beaded ribs, compressed femurs, marked long bone deformity, platyspondyly
III	Moderately Severe	Progressively deforming bones, usually with moderate deformity at birth. Scleral hue varies, often lightening with age. Dentinogenesis imperfecta common, hearing loss common. Stature very short
IV	Moderately Severe	Mild to moderate bone deformity and variable short stature; dentinogenesis imperfecta is common and hearing loss occurs in some families. White or blue sclerae.

MATERIALS AND METHODS

1. Construct plasmid:

The reconstructed bacterial expression vector pET32a (+) was originally built with the $(GlyProPro)_{10}$-foldon construction (generously provided by Dr. Jürgen Engle at University of Basel, Switzerland). Synthetic genes encoding CoL1A1 877-939 fragment, Cys knot and (Gly-Pro-Pro) repeating sequence at both sides were generated by Gene Oracle, Inc. The synthetic genes were inserted into the BamHI/EcoRI site of reconstructed bacterial expression vector pET32a (+).The inserted gene sequence was verified by DNA sequencing.

2. Introduction Single Nucleotide Mutation by PCR:

Mutations at 901, 913, 920 and 932 amino acid positions were performed by PCR with Quick-change II site-directed mutagenesis kit (Stratagene) (**Figure 3**). The PCR primers were optimized and made by Integrated DNA Technology Company. The PCR was carried out by Eppendorf PCR machine. The repeating progress was followed by site-directed mutagenesis kit manual (**Appendix B**).

G901S mutation primers pair:

5' - CCG GCC GGT CCG GTA AGT CCA GTT GGC GCG CGT GG - 3'

5' - CCA CGC GCG CCA ACT GGA CTT ACC GGA CCG GCC GG - 3'

G913S mutation primers pair:

5' - CCC GCT GGT CCG CAG AGT CCT CGC GGT GAT AAG GGC - 3'

5' - GCC CTT ATC ACC GCG AGG ACT CTG CGG ACC AGC GGG - 3'

E920A mutation primers pair:

5' - CGC GGT GAT AAG GGC GCA ACG GGC GAA CAG GG - 3'

5' - CCC TGT TCG CCC GTT GCG CCC TTA TCA CCG CG - 3'

H932D mutation primers pair:

5' - CGT GGG ATT AAA GGG GAT CGT GGT TTC TCA GG - 3'

5' - CCT GAG AAA CCA CGA TCC CCT TTA ATC CCA CG - 3'

G925S mutation primers pair:

5' – G GGC GAA ACG GGC GAA CAG AGT GAT CGT GGG - 3'

5' – CCC ACG ATC ACT CTG TTC GCC CGT TTC GCC C- 3'

3. Expression and purification of recombinant protein from *E. coli*:

The constructed plasmid was transformed into *E. coli* JM109 (DE3) host strain by heat shock. The competent cell was pre-treated with cold $CaCl_2$-$MgCl_2$ solution and mixed with plasmid at 45 ^{o}C to allow plasmid transfer into *E. coli* cells. After transformation, we used ampicillin resistance to select the successfully transformed cells. After that, the transformed plasmid DNA were extracted and sequenced to confirm the reconstruction and mutations. Then the transformed cells were incubated to optical density (OD) = 0.6 at 600nm and induced by 0.01mM IPTG at 25^{o}C overnight. The following day, we spun down the *E. coli* cells, removed the media, and broke the cells for

the first time by Freeze-Thaw cycles. We put the sample into liquid nitrogen and then 37 ^{o}C water as one cycle for 3 times. After that, we used sonication machine to break the cells 3 times, giving 10 to 20 second for each time with 1 minute intervals in between. We set up sonication machine with output = 3.0 and duty cycle = 50%. Later, we spun down the sample to collect the supernatant of all the proteins, including our recombinant collagen proteins. The 6X His tagged recombinant proteins were first purified by Talon^{TM} affinity chromatography on Co^{2+} - Sepharose resin (BD Biosciences), and the separation of CoL1A1 877-939 fragment from 6X His and thioredoxin tag was performed by thrombin cleavage with Thrombin Cleancleave Kit (Sigma). The further purification to separate CoL1A1 877-939 fragment from 6X His and thioredoxin tag was performed by HPLC (Beckman Coulter) with Phenomenex biosep-sec-s 2000 gel filtration column. The purified recombinant protein was dialyzed to 10mM PBS pH=7.0 buffer and concentrated to proper concentration 1mg/ml for structure characterization. The concentrating was carried on with Ultracel YM-3 centriprep (Millipore). The concentration of protein was measured by OD value at 280nm with DU800 spectrophotometer (Beckman Coulter). The extinction coefficient of our recombinant collagen protein is 8400.units M^{-1} cm^{-1} at 280 nm.

4. Protein characterization by Circular Dichroism (CD):

The CD spectra of purified recombinant proteins were recorded by AVIV model 202-01 CD spectrometer. The far-ultraviolet spectra scanning (190nm-270nm) of protein

samples were measured in 1mm path-length curvet (Hellma 111QS). The mean molar residue ellipticity was calculated based on the concentration, molecular weight and number of residues of protein normalized spectra by the below formula.

$$Mean\quad molar\quad residue\quad ellipticity = \frac{CD\quad Signal \times Molecular\quad Weight}{Concentration \times\quad Number\quad of\quad residues}$$

The thermal transition of recombinant protein was determined by monitoring the changing of the mean molar residue ellipticity at fixed 225nm wavelength as a function of temperature. The CD signal was collected at fixed 225nm wavelength from 5 ^{o}C to 70 ^{o}C. We increased the temperature by 0.3 ^{o}C each step and set up equilibration time of 2 minutes for each step before measurement. The temperature measurement band was from -0.2 ^{o}C to +0.2 ^{o}C. The final heating rate was 9 ^{o}C/hour

The equilibrium melting transition was fit to a two-state monomer to trimer transition. The two-state transition model was applied to the equation below:

$$\text{Fraction of folded} = \frac{\left(\theta_{observed} - \theta_{monomer}\right)}{\left(\theta_{trimer} - \theta_{monomer}\right)}$$

In this equation, θ observed is the observed ellipticity, θ trimer is calculated from the best linear fit equation for the trimer part of the melting curve and θ monomer is calculated from the best linear fit equation for the monomer part of the melting curve.

5. The chymotrypsin and pepsin digestion:

Chymotrypsin digestion experiments were carried out using Chymotrypsin Agarose

resin (Sigma) in the ratio of 6.5 $X10^{10}$ unit/mol proteins. Pepsin digestions were carried out using Immobilized Pepsin (Pierce) resin with activity > 6000 units/ml. The concentration of protein sample was 1 mg/ml. The digestion experiments were carried out following the manufactures' manual with 0.1ml of protein at 1mg/ml for every 100ul agarose-resin. Control experiments of pepsin digestion on F877, G901S, G913S and preheated G913S were carried out at 4 ^{o}C for 1 hour. Control experiments of chymotrypsin digestion on G913S and preheated G913S were carried out at 4 ^{o}C for 15 minutes. All F877, G901S and G913S recombinant proteins were incubated with chymotrypsin at 15 ^{o}C for 15 minutes to check the cleavage activity. The preheated G913S sample was heated to 70 ^{o}C before the control experiments.

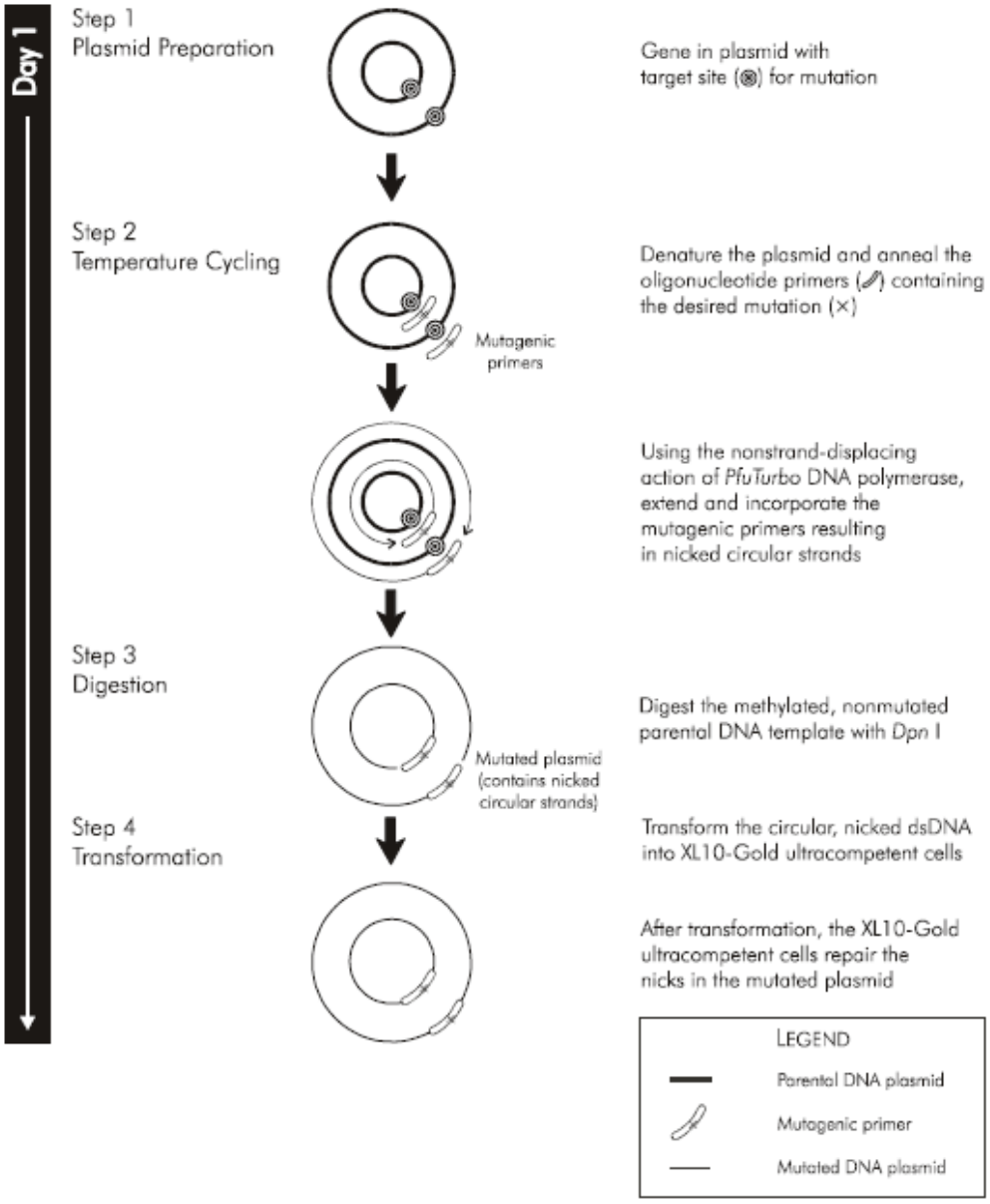
6. Mass spectrometry:

For MALDI-TOF MS analysis (matrix-assisted laser desorption ionization time-of -flight mass spectrometry), α –cyano-4-hydroxycinnamic acid (CHCA) matrix was prepared as a saturated solution in 50% acetonitrile/0.1 % trifluoroacetic acid. Sample and matrix were spotted 1:1 onto a sample plate and allowed to dry. All spectra were acquired using a MALDI time-of-flight mass spectrometer Voyager-DE STR (PE Biosystem, Foster City, CA) in positive, linear, delayed extraction mode with a 337 nm nitrogen laser pulsing at 3Hz. Spectra from 100 individual laser shots were averaged and externally calibrated.

7. Measuring the binding affinity by pull-down assay:

To investigate the binding affinity between collagen and Hsp47, the pull-down assay was carried out using Talon^{TM} affinity Co^{2+} - Sepharose resin. Both F877 and G913S recombinant protein were purified with 6X His-tag attached. Human Hsp47 recombinant protein was ordered from Stressgen. F877 and G913S recombinant proteins were mixed with human Hsp47 recombinant protein separately for 30 minutes at 25^{o}C at same concentration (0.5mg/ml) before binding with resin. The mixed samples were added to resin and combined with resin for another 30 minutes at 25^{o}C. After that, we collected the supernatant, which is called "flowing through" fraction for SDS-PAGE assay. Later we added same volume 1 X wash/extraction buffer (**Appendix B**) twice to wash the resin and collected the sample for SDS-PAGE assay. Finally, we added the same volume 1 X elution buffer (**Appendix B**) three times to elute all the proteins and collected them for SDS-PAGE assay as well.

FIGURE 3: The mechanism and steps of site-directed nucleotide mutation by Polymerase Chain Reaction (PCR) method (Stratagene site-directed mutagenesis kit menu: http://www.stratagene.com/manuals/200523.pdf)


Day 1
Step 1
Plasmid Preparation
Gene in plasmid with
target site (⊗) for mutation
Step 2
Temperature Cycling
Mutagenic
primers
Denature the plasmid and anneal the
oligonucleotide primers (✎) containing
the desired mutation (×)
Using the nonstrand-displacing
action of PfuTurbo DNA polymerase,
extend and incorporate the
mutagenic primers resulting
in nicked circular strands
Step 3
Digestion
Digest the methylated, nonmutated
parental DNA template with Dpn I
Mutated plasmid
(contains nicked
circular strands)
Step 4
Transformation
Transform the circular, nicked dsDNA
into XL10-Gold ultracompetent cells
After transformation, the XL10-Gold
ultracompetent cells repair the
nicks in the mutated plasmid
LEGEND
Parental DNA plasmid
Mutagenic primer
Mutated DNA plasmid

CHAPTER I

Recombinant Collagen Studies Link the Severe Conformational Changes Induced by Osteogenesis Imperfecta Mutations to the Disruption of a Set of Interchain Salt Bridges

(Chapter I was published in *Journal of Biological Chemistry*, Dec, 2008. The full reference is: Xu, K., Nowak, I., Kirchner, M., and Xu, Y. *Recombinant Collagen studies link the severe conformational changes induced by Osteogenesis Imperfecta mutations to the disruption of a set of interchain salt bridges* (2008) *J. Biol. Chem.* 283, 34337-34344.)

INTRODUCTION

Considerable effort has been made to elucidate the mechanisms by which Gly substitution mutations of the collagen triple helix cause Osteogenesis Imperfecta (OI), also known as brittle bone disease. The collagen triple helix consists of three polypeptide chains each in extended polyproline II conformation and with the characteristic $(Gly\text{-}X\text{-}Y)_n$ repeating amino acid sequence (**Rich and Crick 1961; Ramachandran *et al.* 1968; Bella *et al.* 1994**). The Gly at every third position is necessitated by the close packing of the helix; while the X and Y residues (where X and Y can be any amino acids) contribute directly to the stability of the triple helix and confer the sequence dependent properties of collagen **(Persikov *et al.* 2005**). Missense mutations that replace the obligatory Gly by another amino acid residue in type I collagen, the major component of bones, are the most common cause of OI **(Marini *et al.* 2007; Byers 1993**). The triple helix domain of type I collagen is a heterotrimer composed of two α1 chains and one α 2 chain each with more than 1000 amino acids in an uninterrupted (Gly-X-Y)n sequence (**Kielty *et al.* 1993**). Nearly 800 Gly replacing mutations from both α 1 and α 2 chains have been linked to OI, yet, depending on the location and the identity of the Gly substitution, the clinical severity of OI varies from mild increase of bone fragility to the most severe type characterized by death at the prenatal stage (the Type II OI) **(Marini *et al.* 2007)**. It remains unclear what molecular properties are related to the sequence locations of the Gly substitutions, which in turn affect the properties of bones. The

prevailing domain model of OI directly links the OI phenotype to the local helix stability surrounding a mutation. The triple helix is not uniform in structure and stability, but organized into thermally stable and labile domains along the axis of the helix (**Persikov *et al*. 2005; Privalov *et al.* 1979; Privalov 1982; Makareeva *et al.* 2006; Makareeva *et al.* 2008**). Mutations disrupting highly stabilizing regions of the helix are considered to be more damaging to the conformation than those in the more flexible, thermally labile regions and thus, predicted to be more severe (**Bachinger *et al.* 1993**). While the conformational heterogeneity was clearly demonstrated in biophysical characterizations of the full-chain collagens (**Privalov *et al.* 1979; Privalov 1982; Makareeva *et al.* 2008**), the relations between the conformational heterogeneity and the severity of the OI has not yet been established. In fact, a recent study systematically mapping out the effects of 41 OI mutations on the melting temperature ΔTm of type I collagen reported no direct correlation between the *Tm* of the mutants and the severity of OI, and no straight forward relationships between the *Tm* and the local helix stability (**Makareeva *et al.* 2008**). Experimentally examining such correlations is difficult because of the large size, and the rope-like conformation of the collagen. Studies using full-chain collagens are often limited to relying on macroscopic parameters such as the ΔTm, which does not imply any specific changes of the conformation (**Makareeva *et al.* 2008**). Given the long, rope-like structure of the triple helix, a mutation could critically alter the surrounding conformations of the triple helix without causing significant changes in *Tm* of the whole molecule. On the other hand, the local stability of a segment of the triple helix is often

derived from the data obtained from short, synthetic peptides. Despite extensive peptide studies on the stabilizing/destabilizing effects of isolated X or Y residues and of the intra-chain and interchain molecular interactions involving both X and Y residues in one Gly-X-Y tri-peptide or in two adjacent tri-peptides (Gly-X-Y-Gly-X_-Y_), how these interactions modulate each other in the context of the long helix of collagen remain largely unknown (**Persikov *et al.* 2005; Makareeva *et al.* 2008; Persikov *et al.* 2002**). We have developed an *E. coli* expression system for recombinant collagen fragments to overcome the limitations of the studies using full-chain collagens or short synthetic peptides. The recombinant collagen fragments were used to model the 63-residue segment of the 1 chain of type I collagen corresponding to residues 877 – 939 (a region containing no hydroxyproline). The optimal size of the fragments allow us to reveal markedly different effects on the triple helix conformation by two OI mutations placed only 12 residues apart, and to correlate the severe OI to a mutation causing more extensive unfolding of a segment of the triple helix known to be involved in the molecular interactions of type I collagen. Furthermore, we were able to demonstrate that the extensive unfolding is caused by the interruption of a set of interchain salt bridges formed between the Lys and Glu residues of a KGE sequence. Such interchain salt bridges (from a KGE or KGD sequence) are known to stabilize the short peptides (**Persikov *et al.* 2005; Venugopal, *et al.* 1994**) and provide essential stability in bacteria collagen (**Mohs *et al.* 2007**). In our model system, the stabilizing interactions of the salt bridges extended through a region containing more than 20 amino acid residues

C-terminal to the KGE sequence. Thus, by revealing the long range impacts of the salt bridges, our data highlight the critical roles of the salt bridges in modulating the stabilities of the neighboring regions of the triple helix. Based on these findings, a new mechanism by which OI mutations cause severe conformational changes of collagen is proposed.

RESULTS

I). Generation of Recombinant Collagen Protein

The complete sequence of the recombinant collagen fragment F877 mimicking the 63 residue region of the type I collagen α1 chain (residue 877- 939) is shown in **Diagram 1**. To best model this region in the fully folded collagen and eliminate the effects of the helix fraying at the ends, repeating sequences of the tri-peptide (Gly-Pro-Pro) with high triple helix stability propensity were added at both the C- and N-terminus. All together, the recombinant triple helix domain consists of 93 amino acid residues. The C-terminal foldon domain, which was taken from bacteria phase T4 fibritin, works as the nucleation domain to initial the folding of the triple helix (**Frank *et al.* 2001**). A short sequence containing two Cys residues (Gly-Pro-Cys-Cys-Gly) was included between the triple helix domain and the foldon domain. When oxidized in folded triple helix conformation, the Cys residues will form a set of interchain disulfide bonds – also known as the Cys knot (**Mechling *et al.* 2000**), which will further increase the stability of the triple helix. Some important positions, such as OI Gly → Ser mutation sites, salt bridges and Pro-free region, were also highlighted in the **Diagram 1**. There are several mutants we introduced into our recombinant protein to investigate the effects of OI mutations on the conformation of triple helix. (**Table 2**)

F877 is called the ‘wild type collagen recombinant protein’ in my experiment, the F877 includes amino acids at 877-939 positions in α1 chain of type I collagen. G901S

and G913S mutants are both OI mutations with single Gly-Ser mutation at 901 and 913 positions (**Diagram 1**). All the mutations introduced are included in **Table 2**.

Diagram 2 showed DNA sequence and encoded amino acid sequence of expression range in our reconstructed pET32a (+) plasmid. T7 RNA polymerase can recognize the specific T7 promoter DNA sequence at C-terminus and initial the transcription. LacI inhibitor, which can block the transcription, will bind the downstream Lac operator DNA sequence. IPTG can remove LacI inhibitor from Lac operator and thus induce the transcription and translation of our recombinant protein. The synthetic gene that encodes the F877 fragment was inserted by Bam HI restriction enzyme.

The recombinant collagen proteins were induced, expressed and purified following the instructions in material and methods (**Figure 5**). The gene product of the expression plasmid is a fusion protein including His-tagged Thioredoxin at the N-terminal end (**Diagram 2**). The exact molecular weight of the recombinant collagen protein was confirmed by mass-spectrometry (**Figure 4**). Comparing the transformed cell sample (Lane 2 and 3) and none-transformed sample (Lane 1) in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) can identify the expressed recombinant protein. After purification, both trimer and monomer recombinant proteins showed up in Lane 5. Thioredoxin and 6 X His-tag were cut away after thrombin cleavage in Lane 6. Lane 7 and Lane 8 are purified recombinant protein samples with and without reducing agent DTT. The different fractions of trimer and monomer protein in Lane 7 and 8

demonstrated that oxidized disulfide bonds were formed by Cys knot at C-terminal of protein.

II). The different effects of OI mutation correlates to their relative location to a set of interchain salt bridges.

Characterize triple helix conformation and stability of recombinant collagen proteins by Circular Dichroism spectroscopy:

The Circular Dichroism (CD) spectra of three fragments, F877, G901S and G913S, at 0.2 mg/ml concentration at 4 ^{o}C are almost identical despite the presence of Gly substitutions (**Figure 6**). The small positive peak at ~ 225 nm and the deep negative peak at ~197 nm are used to identify the triple helix (**Ramachandran *et al.* 1968; Engel *et al.* 2006**). The foldon domain has contributions to the peak at 225 nm (**Frank *et al.* 2001**). Thus, the average molar ellipticity of ~ 4690 deg cm^2/dmol of the three fragments at 225 nm includes ~ 20% contribution from the foldon domain. Small differences in the negative peak at 197 nm of the three fragments were observed. While this deep negative peak is often considered a more sensitive characterization of the triple helix conformation, its accurate estimation is often limited by strong absorption of the peptide backbone at the far UV region, especially for large molecules. The close agreement of the molar ellipticity of all three fragments (within experimental error of ± 2 %) indicates that the inclusion of the (Gly-Pro-Pro)$_6$ and the segment of triple helix N-terminus to the mutation site(s) (residue 877-891) effectively helped the refolding of both G901S and G913S downstream from the mutation sites. The two substitutions appear to cause similarly

limited distortions of the triple helix conformation.

Although they can all form triple helices at low temperature, the thermal unfolding curves of the two mutated fragments are quite different (**Figure 7**). The thermal stability was monitored by the change of the CD signal at 225 nm with the temperature from 4 to 60 ^{o}C. The Gly → Ser substitution at 901 positions decreases the thermal stability of F877 and reduces the T_m by ~ 10 ^{o}C, while the overall unfolding curve remains similar. In contrast, the Gly substitution in G913S affects both the *Tm* and the shape of the unfolding curve. The melting curve reveals two steps in unfolding. The first step occurs at a much lower temperature, between 15 and 25^{o}C, and involves the decrease of about 45% of the ellipticity at 225 nm. The second melting step takes place between 29 and 35^{o}C, causing the complete unfolding of the triple helix domain.

The loss of the signal at 225 nm in **Figure 7** monitors the unfolding of the triple helix domain while the foldon domain remains folded in this temperature range. The *Tm* of foldon at the comparable concentration is ~ 70^{o}C (**Frank *et al.* 2001**). The presence of the Cys knot will further increase its stability. The CD spectrum of G913S taken at 70^{o}C (**Figure 6**), after the triple helix completely unfolds, shows the typical features of a foldon domain.

G901S and G913S recombinant collagen proteins show significantly different resistant to chymotrypsin digestion at 15 ^{o}C.

Proteolytic digestion is often used to assess the conformation of the triple helix

(**Makareeva *et al.* 2006**; **Bruckner *et al.* 1981**). When fully folded, the triple helix is known to be resistant to pepsin and chymotrypsin. The results of the protease digestion experiments are demonstrated in **Figure 8**. The experiments were carried out at both 4^{o}C (**Figure 8A**) and 15^{o}C (**Figure 8B**). Preheated samples were used as controls in the experiments where the fragments of G913S were preheated to 70^{o}C for 2 hours to induce the complete unfolding of the triple helix, followed by quickly cooling to the specified temperature for the proteolytic digestion. Since the refolding of the triple helix is exceedingly slow (**Xu *et al.* 2002; Engel *et al.* 1991; Engel *et al.* 2000**), this pre-heated sample is expected to contain unfolded chains, which are susceptible to the proteases.

All three fragments have the same level of resistance to chymotrypsin at 4^{o}C (**Figure 8A**), consistent with the CD data (**Figure 6**) indicating similar triple helix conformation for all three fragments. Little digestion by chymotrypsin was observed during the 15 minutes of incubation of the three samples at 4^{o}C, while the pre-heated sample was nearly completely digested by the end of the 15 minutes. Similar results were also seen for pepsin digestion experiments at 4 ^{o}C (**Figure 8C**). A small degree of digestion was observed in both G913S and G901S after prolonged incubation in pepsin (60 minutes) reflecting the decreased stability of these two mutated chains.

A significantly higher susceptibility of G913S to chymotrypsin was observed at 15^{o}C (**Figure 8B**) as the G913S starts to get into the first melting step. Clear signs of digestion of G913S were observed after 1 minute incubation with the enzyme, with the digestion nearing completion after 15 minutes of incubation. On the other hand, both F877 and

G901S are resistant to the enzyme after 15 minutes of incubation. The only residue with an aromatic side chain in the triple helix domain (**Diagram 1**), the Phe935, is the expected digestion site of chymotrypsin. The subsequent mass-spec experiments have confirmed the digested samples of G913S contain a 5K species corresponding to the fragment from the C-terminal to the Phe935, and a ~ 7K species corresponding to the fragment N-terminus to the digestion site. Interestingly, after an extended period of incubation (overnight at 15^{o}C) we also observed the presence of fragments that were digested by chymotrypsin at the two Tyr residues inside the foldon domain (**Diagram 1**) in G913S but not in G901S or F877. It appears that the removal or the unfolding of the triple helix domain has also destabilized the foldon domain and promoted the digestion at the Tyr residues. Neither Tyr residues are not assessable in the fully folded foldon conformation (**Diagram 1**).

The stabilization effect of interchain salt bridges KGE (918-920) on Pro free region

The relative local stability of the triple helix is largely contributed by X and Y position amino acids, which are exposed to the solvent. The interchain salt bridges and hydrogen bonds formed by X and Y position amino acid side chains have stabilization effect on the triple helix structure. There is a crucial salt bridge pare KGE at 918- 920 position, which is close to the G913S lethal mutation site in our recombinant protein (**Diagram 1**).

The unfolding of the Pro-free region in G913S must be caused by the interruption of

stabilizing factors based by the Gly913 → Ser substitution. The interchain salt bridges between the K and the E residues at the nearby KGE sequence (Position 918-920, **Diagram 1**) appear to be the most likely candidate. Studies using peptides and molecular modeling revealed that, in the homotrimer environment, the positively charged K residue of a KGE sequence is placed in close contact with the negatively charged E residues of a neighboring staggering chain, and form a set of stabilizing interchain salt bridges.

To test the involvement of the KGE salt bridges in the two-step melting caused by G913S mutation, the temperature melt experiment of the wild type F877 fragment without Gly-substitution mutation was carried out in pH 3 buffer. The carboxyl group of Glu-920 (pK ~ 4.2) is expected to be protonated at pH 3, and the charge-charge interactions involving Glu are expected to be eliminated. Indeed, at pH 3 the F877 fragment follows the same two-step meting profile as that of G913S at pH 7 (**Figure 9**). The titration of the Glu-920 in the F877 functions the same way as the Gly-913 → Ser mutation to cause the unfolding of the Pro-free domain at 18^{o}C.

Such a critical stabilizing effect of the KGE sequence on the Pro-free region is further supported by the study of a new fragment E920A. In E920A, the salt bridges are permanently removed by a Glu920 → Ala substitution. As expected, the E902A fragment forms stable triple helix as that of F877 at low temperature, but follows the two-step temperature melt profile as that G913S (**Figure 9**). Furthermore, E920A shows the same susceptibility to chymotrypsin as G913S does at 15^{o}C (**Figure 9 lower insert**), indicating the same degree of unfolding of the Pro-free region.

From these data we conclude that the Pro-free region in F877 constitutes a "micro-unfolding" region, which has intrinsic low thermal stability and prefers an unfolded conformation. However, this Pro-free region maintains the triple helix conformation under physiological conditions because of the salt bridges of the KGE sequence at its N-terminal end. Without the salt bridges, the 24 residue Pro-free region will unfold as an isolated cooperative unit.

It is also interesting to note the nearly identical temperature melt profiles of F877 at pH 3 with that of G913S and E920A at pH 7 (**Figure 9**). The pH induced conformational changes appear to be dominated by the interactions involving Glu920, despite the presence of several other titratable residues, including the Glu-890 residue in the N-terminal region. Although the Glu-890 is adjacent to another positively charged residue Arg (**Diagram 1**), no stabilizing effects were found in RGE sequence in peptide models (**Persikov *et al.* 2002**), which is consistent with the lack of effects of protonation of Glu-890 seen here.

The effects of a new salt bridge on Pro-free region:

A new salt bridge at 932 position was introduced in the Pro-free region by a His → Asp point mutation (**Diagram 1**). As expected, the opening of Pro-free region caused by G913S mutation can refold again with the new salt bridge (**Figure 10**). The melting curve of double mutation G913S-H932D is highlighted in **Figure 10**. This melting curve is more similar G901S than G913S. This means that the new salt bridge we introduce in the middle of the Pro-free region help to stabilize the C-terminal region opened by G913S

mutation. The new evidence here confirmed that the first melting around 15^{o}C indicated the opening of Pro-free region at C-terminal of collagen. So in one word, introducing an additional salt bridge in the Pro-free region helped to reduce the effects on the helical conformation in this region when the original KGE salt bridge is interrupted. This "rescue" experiment reinforces the idea that the salt bridge acted as "stabilization lock" to sustain the triple helix conformation of the Pro-free region.

Interestingly, when the salt bridge at 932 was introduced in the F877, the fragment H932D (**Table 2**), the collagen protein with this one additional salt bridge did not show significantly higher *Tm* (**Figure 11**). It seems that the salt bridge at 918-920 positions alone provides enough stability to keep the Pro-free domain in folded conformation; introducing the additional salt bridge in the Pro-free domain did not dramatically increase the stability.

Recombinant collagen protein with either E920A or H932D mutations can form the triple helix structure at 4^{o}C. The Circular Dichroism spectra of the three fragments at 4^{o}C are almost identical, with a small positive peak at ~ 225 nm and the deep negative peak at ~197 nm, which are used to identify triple helix (**Figure 12**). The triple helix conformation at low temperature is pre-requirement for further temperature melting data.

DIAGRAM 1: Construction of our recombinant collagen protein modeling 63 amino acids sequence from α 1 type I collagen (residue 877-939)

CONSTRUCTION OF RECOMBINANT COLLAGEN FRAGMENT

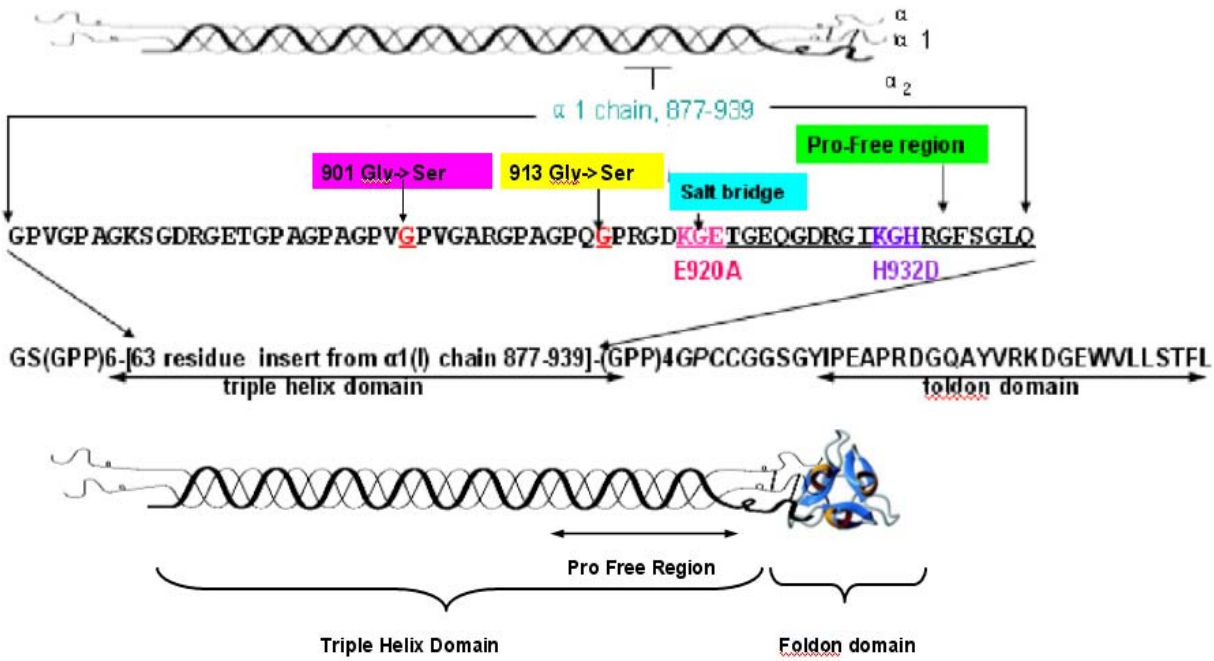


TABLE 2: A summary of mutants generated for this study

Mutation	Purpose	Result
G901S	Investigate the effect of clinically identified Mild OI mutation G901S on the conformation of triple helix	Normal triple helix formation at low temperature with G901S mutation and lower melting temperature than normal one
G913S	Investigate the effect of clinically identified lethal OI mutation G901S on the conformation of triple helix	Normal triple helix formation at low temperature with G913S mutation, two steps melting and lower melting temperature than normal one
E920A	Investigate the stabilizing effect of salt bridge on Pro-free region	Normal triple helix formation at low temperature with E920A mutation, two steps melting and lower melting temperature than normal one
G913S-H932D	Confirm the stabilizing effect of salt bridge on Pro-free region	Normal triple helix formation at low temperature with mutations, one step melting and lower melting temperature than normal one

H932D	Discover the extra stabilizing effect of salt bridge on wild type collagen	Slightly higher melting temperature than normal one
G913S-G925S	Investigate the effect of G925S mutation (within Pro free region) on conformation of triple helix	Characterizing the structure property
G913S-E920A	Disrupt the Pro free region's triple helix conformation by both OI mutation and salt bridge mutation	Didn't get purified protein
G901S-G913S	Investigate the effect of double Gly → Ser mutations on the conformation of triple	Didn't form good triple helix

DIAGRAM 2: DNA sequence and encoding amino acids sequence of expression range in reconstructed pET32a (+) plasmid

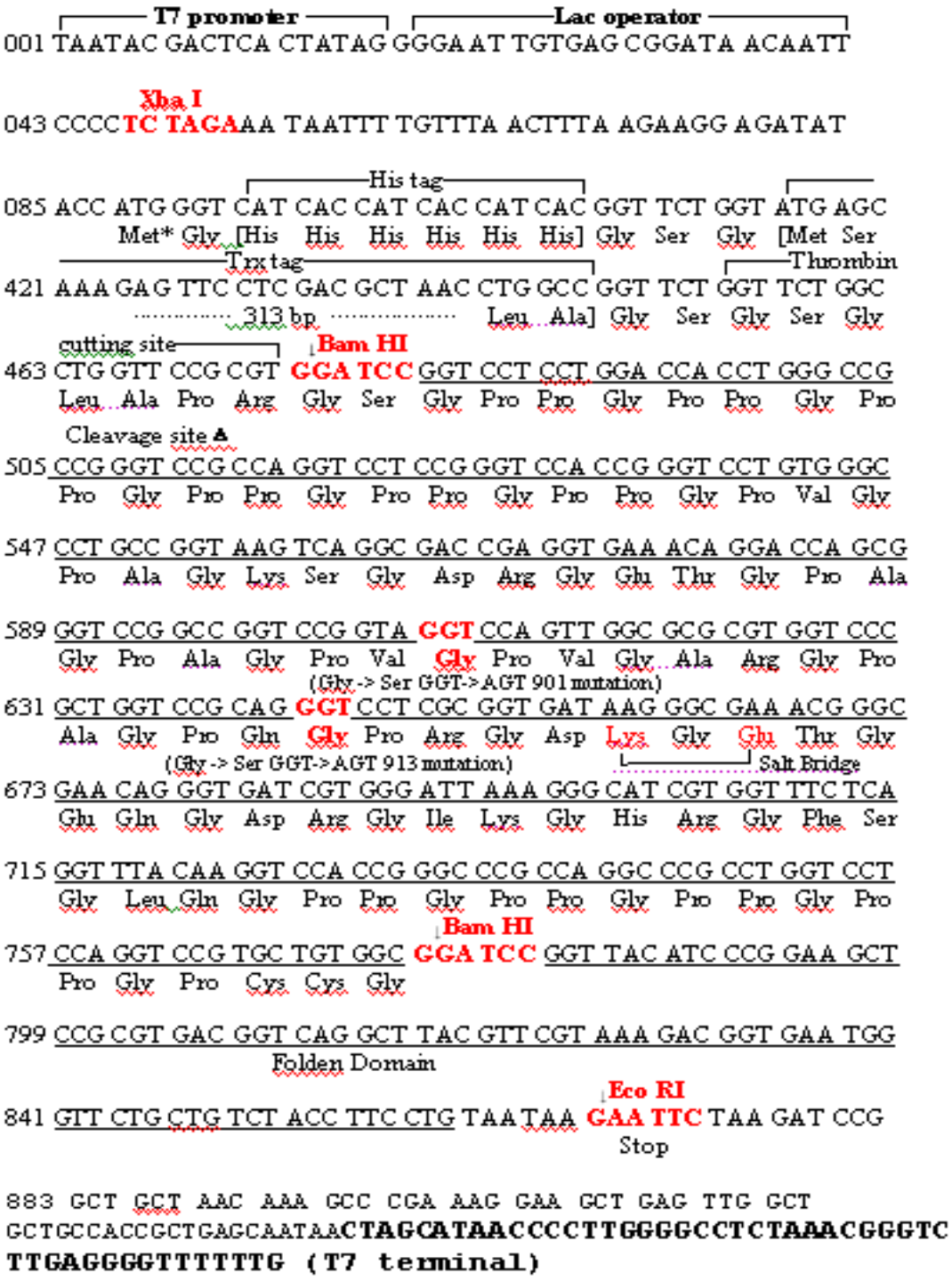

T7 promoter
Lac operator
001 TAATAC GACTCA CTATAG GGGAAT TGTGAG CGGATA ACAATT
Xba I
043 CCCCTC TAGAAA TAATTT TGTTTA ACTTTA AGAAGG AGATAT
His tag
085 ACC ATG GGT CAT CAC CAT CAC CAT CAC GGT TCT GGT ATG AGC
Met* Gly [His His His His His His] Gly Ser Gly [Met Ser
Trx tag
Thrombin
421 AAA GAG TTC CTC GAC GCT AAC CTG GCC GGT TCT GGT TCT GGC
313 bp
Leu Ala] Gly Ser Gly Ser Gly
cutting site
Bam HI
463 CTG GTT CCG CGT GGA TCC GGT CCT CCT GGA CCA CCT GGG CCG
Leu Ala Pro Arg Gly Ser Gly Pro Pro Gly Pro Pro Gly Pro
Cleavage site
505 CCG GGT CCG CCA GGT CCT CCG GGT CCA CCG GGT CCT GTG GGC
Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Val Gly
547 CCT GCC GGT AAG TCA GGC GAC CGA GGT GAA ACA GGA CCA GCG
Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala
589 GGT CCG GCC GGT CCG GTA GGT CCA GTT GGC GCG CGT GGT CCC
Gly Pro Ala Gly Pro Val Gly Pro Val Gly Ala Arg Gly Pro
(Gly -> Ser GGT->AGT 901 mutation)
631 GCT GGT CCG CAG GGT CCT CGC GGT GAT AAG GGC GAA ACG GGC
Ala Gly Pro Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly
(Gly -> Ser GGT->AGT 913 mutation)
Salt Bridge
673 GAA CAG GGT GAT CGT GGG ATT AAA GGG CAT CGT GGT TTC TCA
Glu Gln Gly Asp Arg Gly Ile Lys Gly His Arg Gly Phe Ser
715 GGT TTA CAA GGT CCA CCG GGC CCG CCA GGC CCG CCT GGT CCT
Gly Leu Gln Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro
Bam HI
757 CCA GGT CCG TGC TGT GGC GGA TCC GGT TAC ATC CCG GAA GCT
Pro Gly Pro Cys Cys Gly
799 CCG CGT GAC GGT CAG GCT TAC GTT CGT AAA GAC GGT GAA TGG
Folden Domain
Eco RI
841 GTT CTG CTG TCT ACC TTC CTG TAA TAA GAA TTC TAA GAT CCG
Stop
883 GCT GCT AAC AAA GCC CGA AAG GAA GCT GAG TTG GCT
GCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTC
TTGAGGGGTTTTTTG (T7 terminal)

FIGURE 4: Mass-spectrum result of F877 sample. Two major peaks: left peak with molecular weight (MW) of 12.2K of F877 monomer, and right peak with MW of 36.7K of F877 trimer.

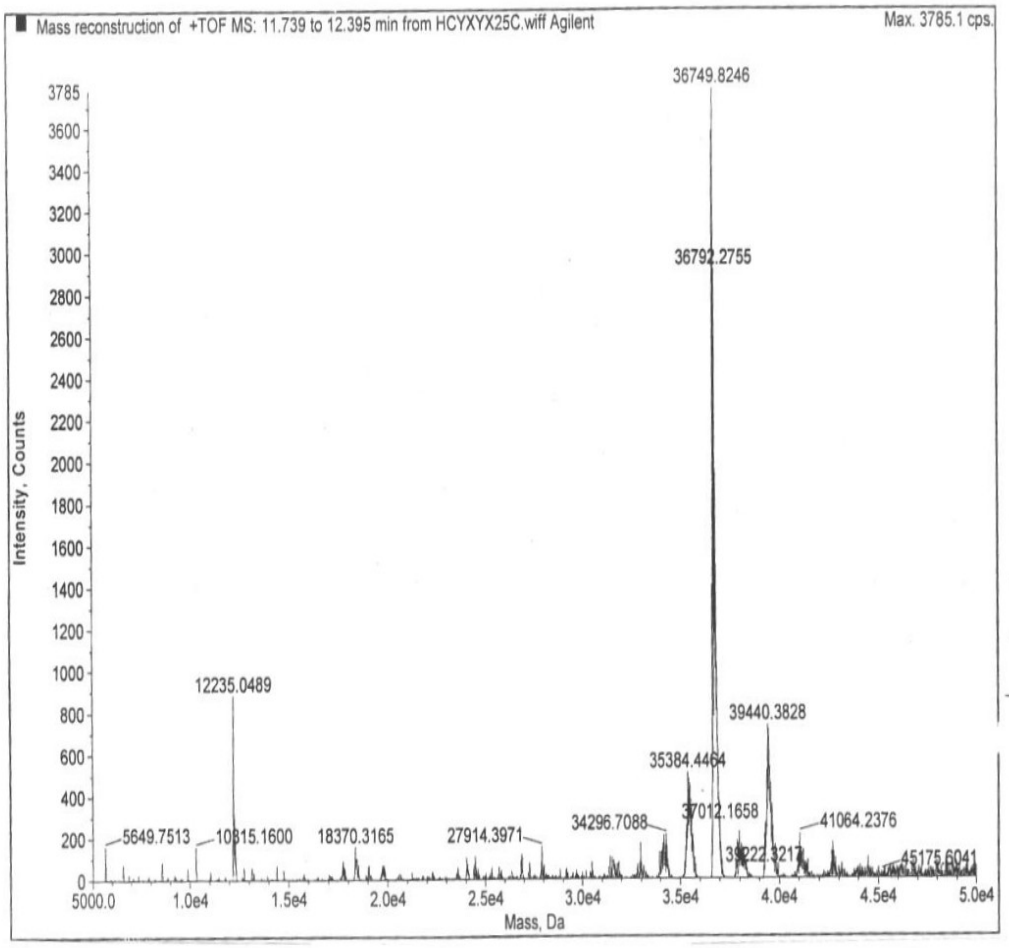

Mass reconstruction of +TOF MS: 11.739 to 12.395 min from HCYXYX25C.wiff Agilent
Max. 3785.1 cps.
Intensity, Counts
Mass, Da
36749.8246
36792.2755
12235.0489
39440.3828
35384.4464
37012.1658
34296.7088
41064.2376
5649.7513
10815.1600
18370.3165
27914.3971
45175.6041

FIGURE 5: SDS-PAGE analysis of the expression of F877 fusion protein

Lane 1: extracted protein sample from non-transformed host cell (JM109) Lane 2: extracted protein sample from transformed and non-induced cell Lane 3: extracted protein sample from transformed and induced cell with 0.1mM IPTG Lane 4 and 9: protein molecular weight marker (66kD, 45kD, 36kD, 29kD, 24kD, 20kD, 14.2kD) Lane 5: purified protein by Co^{++} – His affinity column Lane 6: purified protein from Lane 5 with thrombin cleavage Lane 7: F877 protein with 20 mM DTT Lane 8: F877 protein w/o DTT.

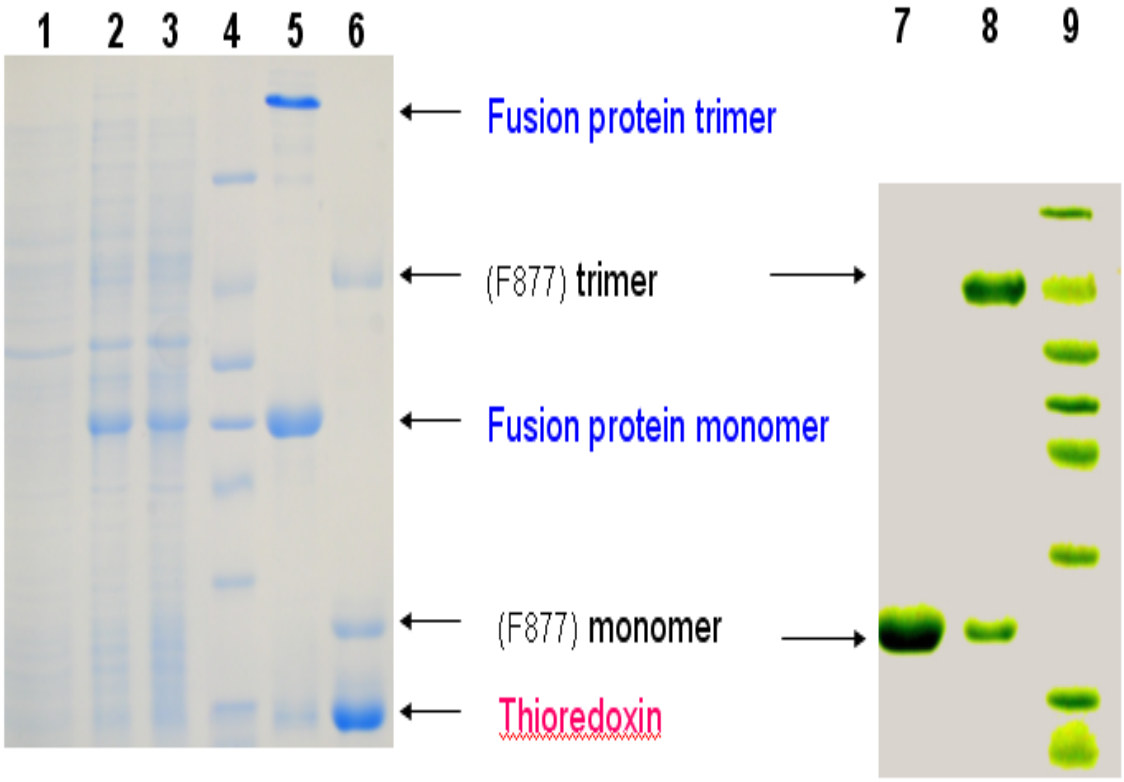

1
2
3
4
5
6
7
8
9
Fusion protein trimer
(F877) trimer
Fusion protein monomer
(F877) monomer
Thioredoxin

FIGURE 6: CD spectra of collagen fragments. The CD spectra of F877 (black square), G901S (red sphere) and G913S (green triangle) were taken at 4 ^{o}C with a concentration of 0.2 mg/ml. The dashed line is the spectrum of G913S (0.2 mg/ml) taken at 70 ^{o}C to show the typical spectrum of foldon domain. All data are normalized to molar ellipticity.

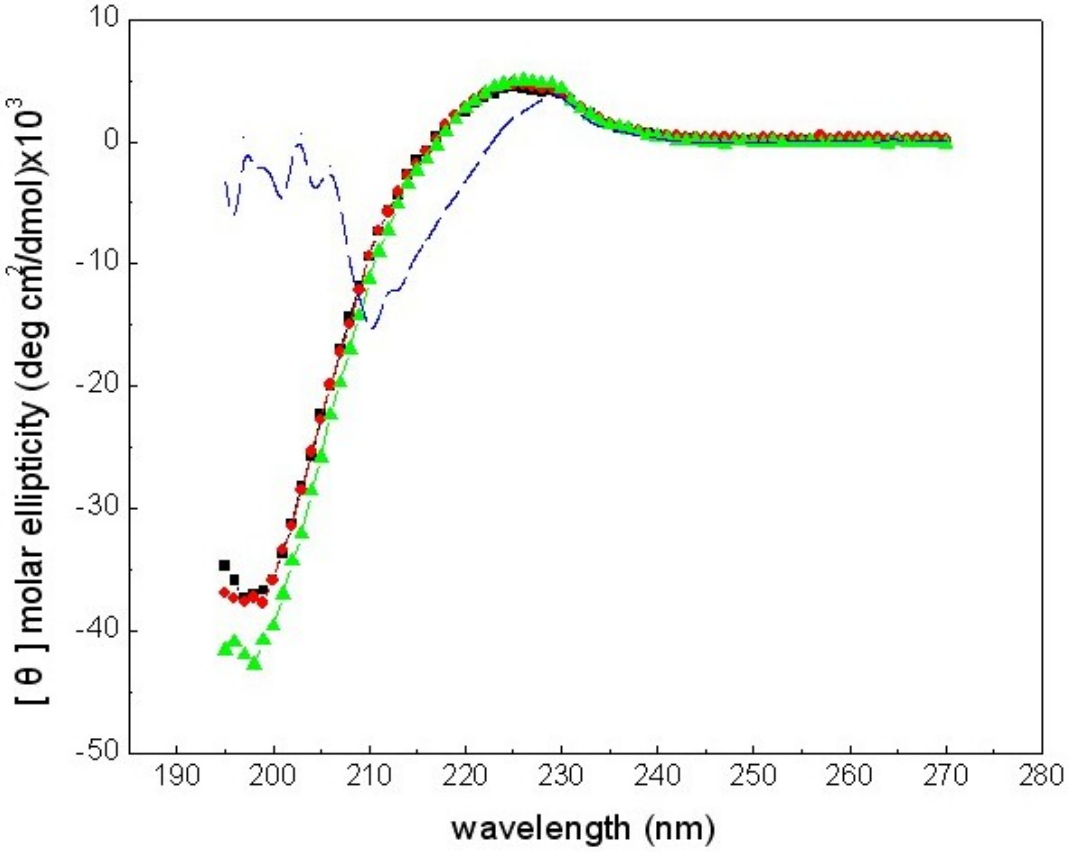

10
0
-10
-20
-30
-40
-50
190
200
210
220
230
240
250
260
270
280
[θ] molar ellipticity (deg cm2/dmol)x10^3
wavelength (nm)

FIGURE 7: Temperature melt curves of the collagen fragments. The thermal unfolding of F877 (red square), G901S (green circle) and G913S (blue triangle) was monitored by the temperature-induced decrease of the CD signal at 225 nm, and normalized to fraction folded. The heating rate is 0.15 ^{o}C/minute and the concentrations of the three fragments are 1 mg/ml.

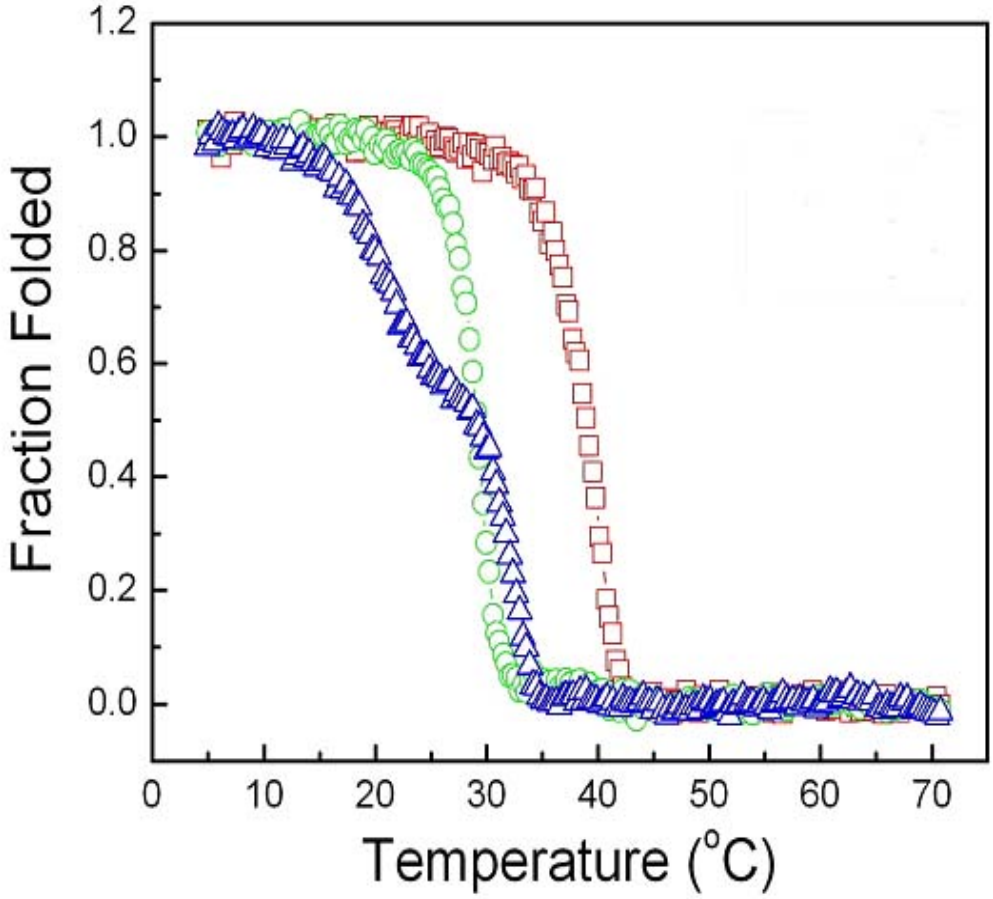

Fraction Folded
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
Temperature (°C)

FIGURE 8: The protease digestion of collagen fragments. A) Chymotrypsin digestion at 4 ^{o}C (similar results of F877 and G901S were not included in the picture due to space limitation). B) Chymotrypsin digestion at 15 ^{o}C. C) Pepsin digestion at 4 ^{o}C. The concentrations of the fragments were 1 mg/ml. The SDS-PAGE was carried out with the presence of 20mM DDT.

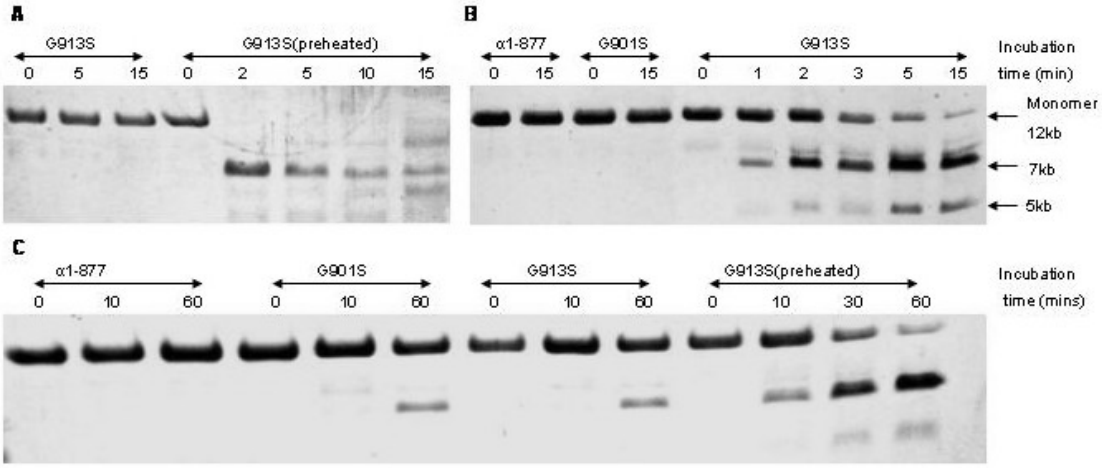

A
G913S
0 5 15
G913S(preheated)
0 2 5 10 15
B
α1-877
0 15
G901S
0 15
G913S
0 1 2 3 5 15
Incubation time (min)
Monomer 12kb
7kb
5kb
C
α1-877
0 10 60
G901S
0 10 60
G913S
0 10 60
G913S(preheated)
0 10 30 60
Incubation time (mins)

FIGURE 9: Temperature melt curves of the collagen fragments. The thermal unfolding of F877 in pH 3 buffer (blue triangle) and of E920A fragment (green triangle) are plotted together with the melting data of F877 (black square) and G913S (red circle) at pH 7. The thermal unfolding curves were monitored by the temperature-induced decrease of the CD signal at 225 nm, and normalized to fraction folded. The heating is 0.15 ^{o}C/minute and the concentrations of the all fragments are 1 mg/ml. The lower insert is the chymotrypsin digestion experiment of E920A under similar conditions as in **Figure 8**.

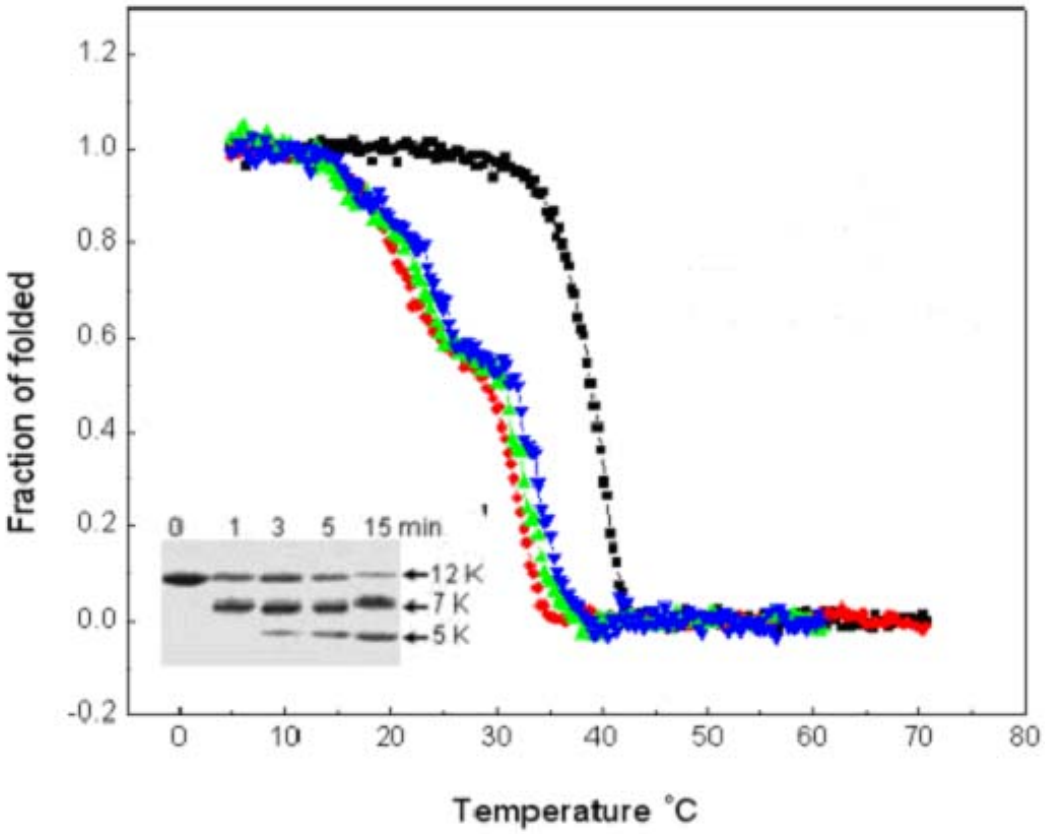

Fraction of folded
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
10
20
30
40
50
60
70
80
Temperature °C
0
1
3
5
15 min
12 K
7 K
5 K

FIGURE 10: Temperature melt curves of the collagen fragments. The thermal unfolding of F877 (black), G901S (red), G913S (green) and G913S-H932D double mutations (blue triangle) were monitored by the temperature-induced decrease of the CD signal at 225 nm, and normalized to fraction folded. The heating rate is 0.15 ^{o}C/minute and the concentrations of the three fragments are 1 mg/ml.

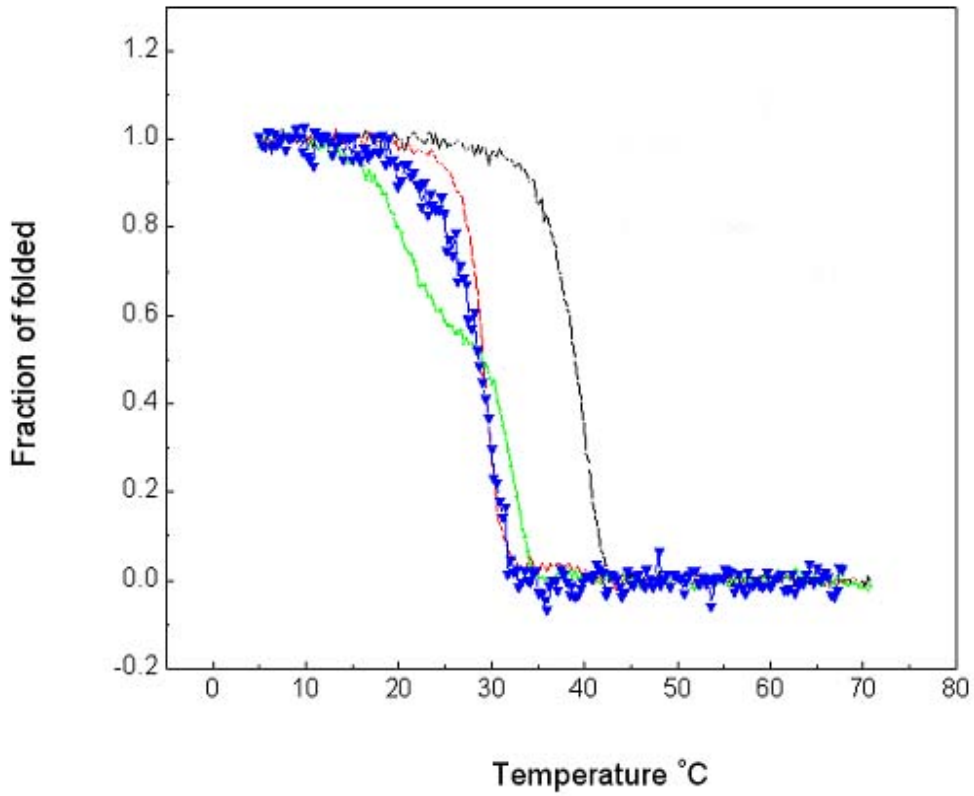

Fraction of folded
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
10
20
30
40
50
60
70
80
Temperature °C

FIGURE 11: Temperature melt curves of the collagen fragments. The thermal unfolding of F877 (black square) and H932D (red circle) were monitored by the temperature-induced decrease of the CD signal at 225 nm, and normalized to fraction folded. The heating rate is 0.15 ^{o}C/minute and the concentrations of the two fragments are 1 mg/ml.

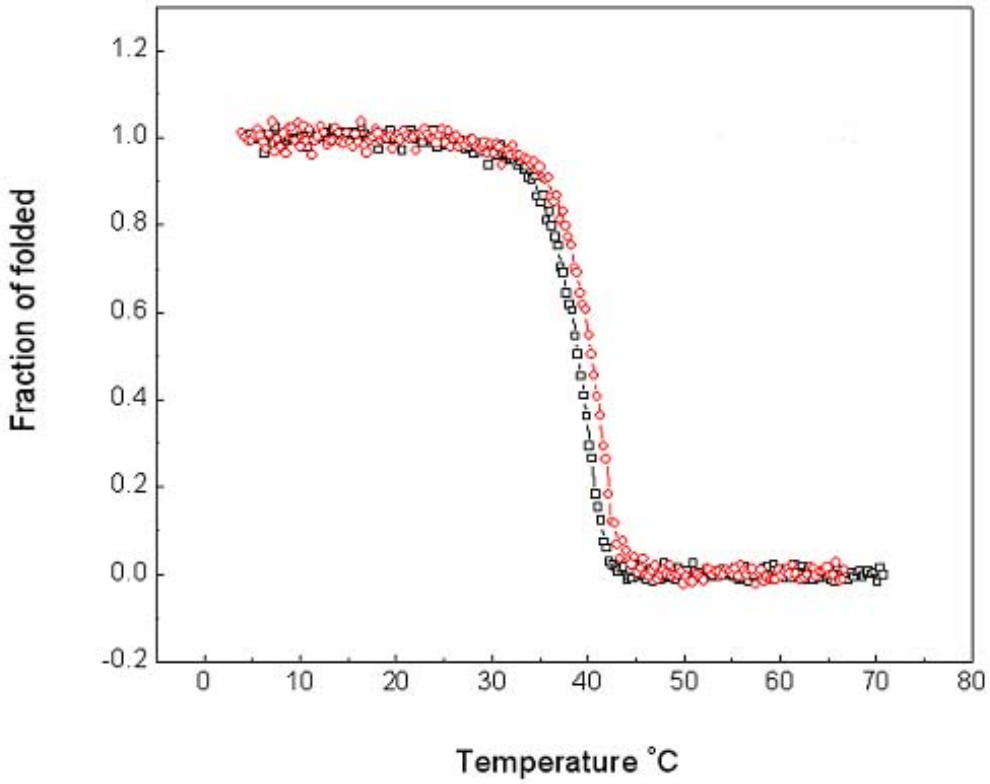
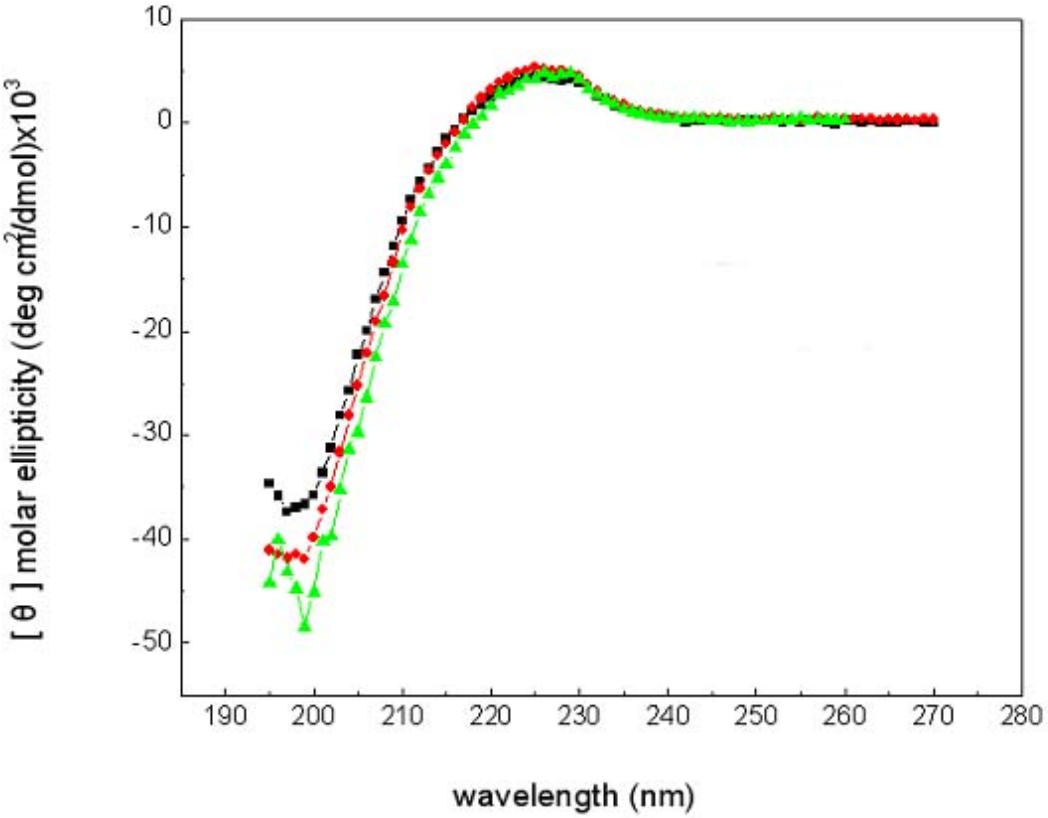

Fraction of folded
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
10
20
30
40
50
60
70
80
Temperature °C

FIGURE 12: CD spectra of collagen fragments. The CD spectra of F877 (square), E920A (sphere) and H932D (triangle) were taken at 4 ^{o}C with a concentration of 0.2 mg/ml.


10
0
-10
-20
-30
-40
-50
190
200
210
220
230
240
250
260
270
280
[θ] molar ellipticity (deg cm²/dmol)x10³
wavelength (nm)

CONCLUSIONS AND DISCUSSION

Using the bacterial expression construction for $(GPP)_{10}$-foldon from Dr. Jürgen Engel's lab, we generated our new recombinant collagen proteins of natural type I collagen alpha one chain fragment with 60 - 200 amino acids. The yield and purity of expressed recombinant collagen protein are good enough for following structure analysis experiments. Based on CD signal, the recombinant protein can form good triple helix structure as natural collagen. This expression system provides us with a perfect platform from which to investigate the relationship between the OI mutations and triple helix structure conformations.

We introduced two adjacently placed Gly → Ser OI mutations (G901S and G913S) into our recombinant collagen protein. However, the effects of these two OI mutations on the triple helix conformation are significantly different. The different effects of the two adjacently placed Gly → Ser mutations on the triple helix conformation highlight the dependence of the regional properties of collagen triple helix on both local amino acid residues and the long-range stabilizing interactions. The different conformational changes are caused by the amino acid sequence from C-terminal region to the substitution sites. Located adjacent to an unstable Pro-free region and to the set of crucial stabilizing salt bridges, the Gly substitution in G913S mutant induces breakage of the salt bridges and an opening of more than 20 amino acids within Pro-free region (**Figure 13**). This extensive opening effect is not seen in G901S because the set of salt bridges is not affected by the

localized conformational changes at G901S, and the salt bridges, in turn, act to limit the spreading of the effects of the substitution at Gly-901 position to reach the Pro-free zone. The effects of the Gly substitutions are thus modulated not only by intrinsic helix propensity of residues in the immediate vicinity of the mutation site but also by the long-range effects of specific stabilizing interactions, such as the interchain salt bridges.

The stabilizing effects of the interchain salt bridges of KGE and KGD sequence were well characterized in synthetic peptides first (**Persikov *et al.* 2005**) and, more recently, in bacteria collagen {**Mohs *et al.* 2007**}. The inclusion of a KGE or KGD sequence increases the melting temperature of triple helix peptides with 24 – 30 amino acid residues by ~ 15^{o}C. The range of this stabilizing effect of the salt bridges in collagen had never been characterized before. Our data has, for the first time, demonstrated experimentally that the stabilizing effects of the salt bridges can extend at least 15 – 20 residues from C-terminal to the KGE sequence. The chymotrypsin digestion site Phe-935 is placed 15 residues from C-terminal to the Glu-920. The effective digestion at Phe-935 site may at least require the unfolding of a few more residues from C-terminus to Phe-935 site. Experiments are underway to investigate if the stabilizing effects of the salt bridges can extend to the similar range N-terminus to the sequence.

Interestingly, the introduction of an additional set of interchain salt bridges in the Pro-free domain helped to maintain the triple helix conformation of this region in G913S but not in F877. It appears that, for a fully folded Pro-free domain, the additional salt bridges are redundant. This finding raises question about the “additive rules” of the triple

helix thermal stability derived from the peptide studies, which predict that the stabilizing effects of the triple helix are "additive" in a linear manner – the *Tm* of a triple helix is the summed effect of all stabilizing interactions of the triple helix. Based on the additive rule, the *Tm* of H932D fragment would be increased by ~ 15^{o}C, which is clearly not the case in our study. More investigation is underway to fully understand the basis for the lack of effects of the additional salt bridges in the triple helix.

Knowledge of the long-range effects of the triple helix's molecular interactions is fundamental to our ability to infer the structures and molecular dynamics of a fragment of the triple helix from its constituent sequences. The Pro-free region has been identified as a "thermal labile" region based on the low helix propensity of the residues in the region (**Bachinger *et al.* 1993; Persikov *et al.* 2005; Makareeva E *et al.* 2006)**, but little is known about its conformation in the context of the long triple helix of type I collagen. Here, we have demonstrated that this Pro-free region *per se* has the potential to undergo cooperative unfolding. In fact, the isolated opening of the 24 residues Pro-free region in E920A mutant (**Table 2**) exemplifies the role of specific micro-unfolding domains of collagen triple helix whose size, location and functions have been implicated in various functions of collagen (**Privalov *et al.* 1979; Makareeva E *et al.* 2006**). However, our data has demonstrated that, in F877, the Pro-free region is kept in triple helix conformation by the nearby salt bridges and unfolds cooperatively only with the rest of the helix.

The nearly identical conformation of F877 and two mutants G913S and G901S at a

low temperature of 4^{o}C highlighted the conformational plasticity of the triple helix (**Bella *at al.* 1994; Hyde TJ *at al.* 2006**). The structural changes of the Gly substitutions appear to be small and localized when embedded in highly stable triple helical peptides (**Bella *et al*. 1994; Hyde TJ *at al*. 2006**). The lowered thermal stability of the mutated chains is also consistent with the structural studies indicating local distortions of the helix and interruption of H-bond networks. Such conformational changes at Gly-913 appear to be enough to affect the orientation and the alignment of the charged residues involved with the salt bridges at the nearby KGE sequence. In contrast, these salt bridges likely remain intact in G901S. The nearly identical backbone dynamics of a Gly located 15 residues C-terminal from Gly-901 in peptides with and without the Gly substitution suggests little conformational changes in this region (**Liu *at al,* 1998**). The KGE sequence located 18 residues away are thus, expected to maintain its normal conformation in G901S mutant and with it, it's stabilizing effects on the Pro-free region. These observations accentuate the pivotal roles of the salt bridges of KGE in maintaining the triple helix conformation and in modulating the effects of the OI mutations. If the opening of the Pro-free region in G913S is the direct consequence of losing the stabilizing salt bridges, the presence of the salt bridges in G901S may act to limit the propagation of the effects of the mutation into the Pro-free region to avoid the large scale unfolding.

The long-range effects of OI mutation were also characterized by a study using full chain type I collagen (**Makareeva E *et al.* 2006**). The Gly substitutions at position 25 and 36 of the α1 chain of type I collagen cause an opening of 70+ residues in the

relatively stable N-anchor domain (the first 85 amino acid residues). It is intriguing how the relatively small changes of a Gly substitution can induce such a large scale unfolding in a region consisting of residues with relatively high helix propensity. The heterotrimeric environment of the type I collagen may be an important factor. The nonequivalent structure may aggravate the conformational effects of the substitution. Other critical stabilizing factors that have not yet to be identified may also play a role. The structural flexibility of a micro-unfolding region, identified by the low helix propensity of constituent residues, next to the N- anchor domain was postulated to mark the boundary of the domain and to help limit the spreading of the mutation-induced conformational effects into the neighboring region (**Makareeva E *et al.* 2006**). Interestingly, there is a KGE sequence at position 108 in both α1 and α2 chains of type I collagen, just 5 residues C-terminus to the postulated micro-unfolding region. It remains interesting to see if the salt bridges of this sequence play any role in limiting the propagation of the conformational effects of the mutations.

The more significant effects of the Gly → Ser substitution on the thermal stability at Gly-913 than at Gly-901 were characterized in synthetic peptides (**Yang *et al.* 1997**). However, since each of the peptides contained very short stretches of the helix, 18 residues surrounding the mutated Gly residue, it could not relate the more severe effects of the Gly-913 substitution to the unfolding of the Pro-free domain and to the disruption of the salt bridges.

The fully folded N-terminal regions of both G901S and G913S mutants appear to

follow the renucleation and bi-directional propagation mechanism of the folding of collagen with Gly mutations (**Hyde TJ *et al.* 2006**). When modeled in short synthetic peptides, the 15 amino acid residues (residues 886-900) immediately N-terminus to G901S failed to refold without the addition of a renucleation domain GPO $(GAO)_3$ (**Hyde TJ *et al.* 2006; Liu *et al.* 1998**). The triple helix formation in the N-terminal region of G901S and G913S can start from the repeating GPP sequence, which shares the same two major features with GPO $(GAO)_3$: a high triple helix propensity and high imino acid content (**Hyde TJ *et al.* 2006**). NMR studies reveal an ordered structure around the Gly-901 substitution site when attached to the renucleation sequence GPO $(GAO)_3$, although with a degree of helix untwisting and alteration of the H-bond network. Similar refolding of the N-terminus of G913 is expected, since the 12 amino acid residues between Gly-901 and Gly-913 are fairly similar to the region N-terminal to Gly-901 in both composition and triple helix propensity. After renucleation, provided the reverse folding (N to C terminus helix propagation) is as effective as the forward propagation (C to N terminus), how well the mutation region resumes the triple helix conformation would be determined by the local helix propensity. The relative position to the renucleation site, however, can affect the rate of the refolding.

The current study is conducted in a homotrimer environment modeling the sequence of the α1 chain of type I collagen only. Significant destabilization of the Pro-free region by the Gly-913 substitution in the hetrotrimeric collagen is expected because of the highly analogous sequences of the α1 and α2 chains of type I collagen in this region.

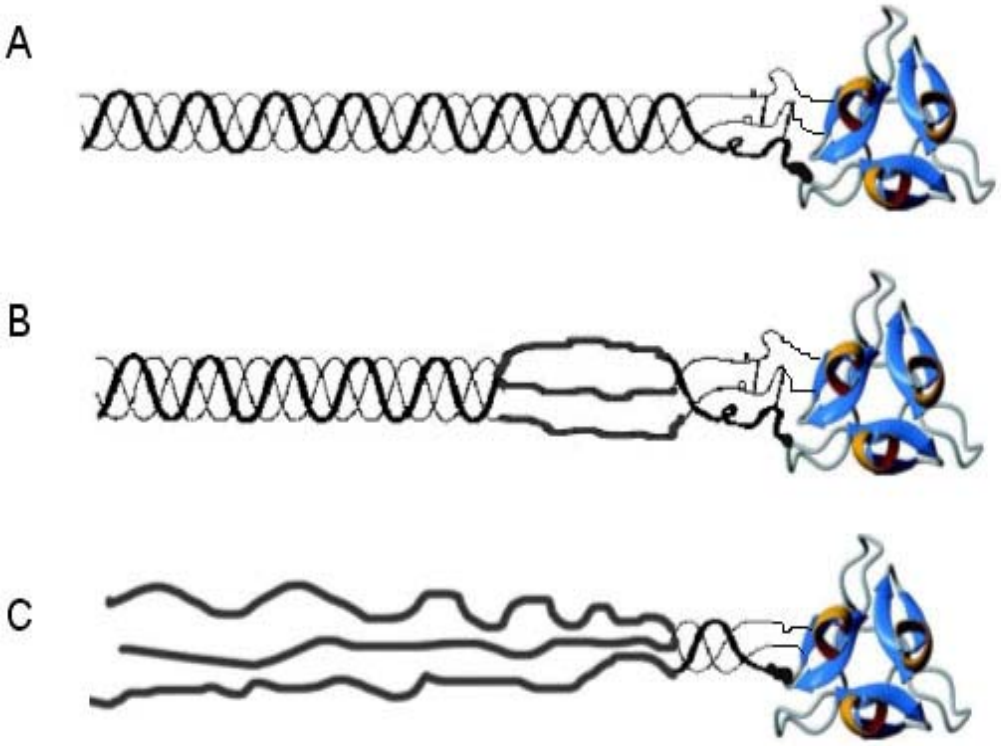
Besides, the KGE sequence for the salt bridges in α1 chain of type I collagen is conserved in α2 chain of type I collagen at the same position. The nonequivalent sequences of heterotrimers are often considered to be more disruptive than homotrimers with the potential to cause the bend or the kink of the triple helix (**Bella *et al.* 1994**). Two altered α1 chains of type I collagen appear to be effective to cause the unfolding of the N-anchor domain even when α2 chain of type I collagen is normal (**Makareeva E *et al.* 2006**). Thus, the salt bridges are expected to play the same critical stabilizing roles in the heterotrimers environment, and their disruption causes similar effects to the structure, as seen in homotrimers.

The crucial position of the Pro-free region suggests that the severe effects of the mutation at Gly-913 correlate with its more pronounced clinical consequences compared to the substitution at Gly-901. The Pro-free region overlaps with the proteoglycan, integrin and heat shock protein binding site (**Di Lullo GA *et al.* 2002**). The region also contains the Lysyl oxidase-binding site Lys-Gly-His-Arg; the Lys-930 is known to be hydroxylated and glycosylated for cross-linking in type I collagen. Conformational changes in this region can potentially affect these interactions and result in detrimental effects on the properties and function of bones. It is therefore no surprise to find that the entire Pro-free region is included in the 'lethal zone' at position 910-964 (**Marini JC *et al.* 2007**).

Our studies have revealed two properties of the triple helix that have not been reported before. First, the stabilizing effects of the interchain salt bridges are modulated

by the location stability. While significant stabilizing effects were observed by introducing the interchain salt bridges to an unstable domain, the effects of adding of such a set of salt bridges to a stable region are rather minor. Second, a segment of triple helix consisting of ~ 20 amino acid residues can unfold as a cooperative unit, independent of the neighboring sections of the helix. Such micro-unfolding domains of collagen have long been postulated to be crucial for molecular recognitions during the functions of collagen, but their independent, cooperative unfolding within the context of a long helix has never been characterized before, and their size has never be specified before.

FIGURE 13: A) Completely folded recombinant collagen G913S triple helix at 4^{o}C. B) Partially unfolded recombinant collagen G913S triple helix around 15 ^{o}C. The unfolded part is Pro-free region. C) Completely unfolded recombinant collagen G913S triple helix after 30 ^{o}C.


A
B
C

Chapter II

Fabrication of Au Nanowires of Uniform Length and Diameter Using a Monodisperse and Rigid Biomolecular Template: Collagen-like Triple Helix

(Chapter II was co-operated paper published on issue of *Angewandte Chem* as presented here. The full reference is Bai, H., Xu, K., Xu, Y.,and Matsui, H. *Fabrication of Au Nanowires of Uniform Length and Diameter Using a Monodisperse and Rigid Biomolecular Template: Collagen-like Triple Helix* (2007) *Angew Chem Int Ed Engl*. **46(18),** 3319-22. I did the recombinant collagen protein synthesis and purification part)

INTRODUCTION

Recently, various metal and semiconductor nanowires have been developed as building blocks for electronics, optics, and sensors. Among these newly developed nanowires, nanowires grown on biomolecular templates such as DNA and peptide assemblies are advantageous since the molecular recognition functions of these biomolecules with specific ligands can be employed to immobilize nanowires onto specific locations to establish desired device geometries (**Gao *et al.* 2005; Whaley *et al*. 2000; Braun *et al.* 1998**). However, most of the biomolecular-nanowire templates made from DNAs or peptides do not possess suitable electric properties for those devices, and therefore there is an extensive effort in the field of bionanotechnology to coat these addressable biomolecular nanowires with metals and semiconductors (**Sharma *et al*. 2006; Deng *et al* 2005; Reches *et al*. 2003; Ryadnov *et al*. 2004; Sugimoto *et al*. 2006; Shenton *et al*. 1999; Tseng *et al*. 2006; Chung *et al*. 2005; Behren *et al*. 2003; Knez *et al*. 2004; Djalaji *et al*. 2003; Banerjee *et al*. 2005; Mao *et al*. 2004**). Recently, the morphology of coating on these peptide – nanotube templates was shown to be controlled by means of changing the peptide sequences and conformations, thus fine-tuning the electronic structures of resulting nanowires for their device applications **(Banerjee *et al*. 2003; Yu *et al* 2003; Yu *et al*. 2004).**

While these biomolecular-nanowire templates appear to be promising building blocks for nanodevices, it is essential to have size monodispersity, strength, and mass

producibility to impact real-world applications. For example, biomolecular templates self-assembled from peptidic monomers tend to yield polydisperse materials with heterogeneous diameters and uncontrolled length through the self-assembly process.

The tobacco mosaic virus (TMV), a rod-shaped biomolecular template, has been applied for various metal coatings, however accurate control of the length with low dispersity is not an easy task. (**Endo *et al.* 2006; Yi *et al*. 2005**) The other type—DNA biomolecular templates—have defined lengths determined by the number of nucleic acids, however they lack conformational rigidity. The tendency of supertwisting of the double-helix DNA structure makes it difficult to obtain rigid and straight nanowires. Their production cost and time may also not be suitable for large-scale syntheses. Herein we report a new application using a collagen-like triple helix as a template nanowire which appears to overcome some of the shortcomings of other biomolecular templates. The collagen-like triple helix is the genetically engineered polypeptide assembly that contains a fragment from the natural collagen sequence. Our study demonstrates that by using the recombinant technology, we can design and amplify a collagen-like triple helix that is monodisperse, easily mineralized with metal ions, and can, thus, be applied as rigid biomolecular templates for metal‐nanowire fabrications. Collagens are the major components of extracellular matrices for bones, cartilages, skins, blood vessels, and corneas, and they are the most abundant proteins in higher organisms with superior mechanical properties. (**Persikov *et al*. 2002; Buehler *et al*.2006; Frank *et al.* 2001**) The collagen-like triple helix is made of three polypeptide chains tightly twisted and bundled

together to form a rigid, rod-shaped molecule that is suitable for applications in building blocks of nanodevices.

MATERIALS AND METHODS

The collagen fragments were cloned into a GPP-foldon vector built on the pET32a (+) plasmid of Novagen (the original GPP-foldon vector was kindly provided by Dr. Jürgen Engel from the University of Basel, Switzerland). The product of this plasmid is a fusion protein with a 6 X His tag and thioredoxin tag as the carrier protein which can be removed by thrombin cleavage to produce the chimaeric protein containing the triple-helix domain and the foldon domain with the Cys knot inserted at the interface of the two domains. The protein was expressed in bacteria *E. coli* JM109 from Promega and purified by His-tag affinity column. After the His-tagged thioredoxin was removed by thrombin digestion and the second round of His-affinity column, the fragments were further purified by gel filtration to isolate the crosslinked triple helix. The final samples are more than 97% pure as estimated by SDS PAGE and gel-filtration experiments.

To grow Au nanocrystals on F807 and G901S triple helices, first we mixed the triple-helix solutions (200 ml, 0.2 mg/ml) with Tris buffer solutions (535 ml, pH 8.6 and 0.01mol/l). The resulting mixture was then vortexed for 10 seconds and left 1 day at 48^{o}C. [$AuPMe_3Cl$] (1.8 mg, Sigma) was incubated for 4 days at 48^{o}C and then the supernatant of the solution containing unattached Au salts was removed by a pipette. Hydrazine hydrate (5 ml, Sigma) was then added to reduce Au salts on the triple helices. To coat the HRE peptides (GenScript Corp., NJ) onto the triple helices, the triple-helix solutions (200 ml) were mixed with the HRE solutions (535 ml, 3.9 X 10^{-4} mg/ml) in the pH 8.6 Tris

buffer solutions for 1 day at 48^{o}C. After immobilization of HRE was confirmed by FTIR, we applied the same Au-growth procedure described above to coat Au on the HRE-immobilized triple helices. After one day of reduction at 48^{o}C, the triple-helix solutions (3 ml) were dried on carbon-coated copper TEM grids as excess solutions were removed by filter papers. These dried samples were then studied by TEM and electron diffraction (JOEL 1200 EX) at an acceleration voltage of 100 kV

RESULTS

To explore the application of a collagen-like triple helix as a nanowire template, we studied the properties of two recombinant triple-helix molecules obtained from an *E. coli* expression system (**Figure 14**). The two recombinant triple helices, F877 and G901S, contain a foldon domain taken from bacterial-phage T4 fibritin, which serves as the nucleation site for the formation of the triple helix. (**Tao *et al.* 1997**) The triple-helix domain of F877 consists of 63 amino acid residues modeling the region starting at position 877 (from the N terminus) of the α1 chain of type I collagen. To increase the thermal stability of the triple helix, repeating $(Gly - Pro - Pro)_n$ sequences were added at the ends of the 63 residues. A pair of Cys residues were inserted at the interface of the foldon and the triple-helix domain to covalently link the three chains of the triple helix through a set of interchain disulfide bonds.(**Frank *et al.* 2001**) Triple helix G901S contains the same sequence as F877 without the Gly → Ser substitution at position G901S (**G in Figure 14**). Replacing the obligatory Gly residue at every third position by any other amino acid residue with bulkier side chains, is known to affect the triple-helix conformation, and such mutations have been implicated in connective tissue diseases. (**Baum *et al.* 1999**) Both recombinant molecules form triple-helix conformations in solution (**Figure 14a**) with denaturation temperatures (*Tm*) of 42 and 30^{o}C, for F877 and G901S, respectively (**Figure 14a**). The triple helix behaves as a trimer with no signs of further association based on the study using analytical ultracentrifugation and gel

filtration.

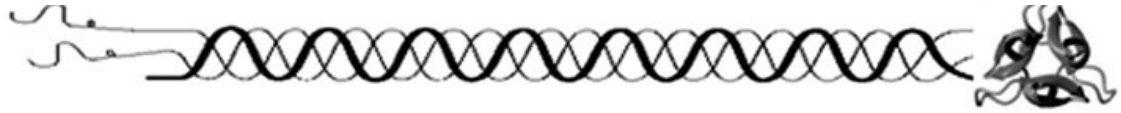
The triple helix appears to be a good template, as it overcomes some of the shortcomings of other bio-molecular templates (**Figure 15**). The structure of the F877 triple helix was imaged by TEM as shown in Figure 1b. Triple-helix F877 formed monodisperse, straight nanowires with an average length of 40 nm and no bending, indicating a rather rigid conformation. The length of the triple helix observed under TEM agrees well with the value of the approximately 35 nm end-to-end distance of a single triple helix consisting of 93 amino acids and a foldon domain estimated from the triple-helix crystal structure. (**Stetefeld *et al*. 2003**) The observed diameter of 4 nm appears to be larger than the 1 – 2 nm predicted from the crystal structure. This slightly larger diameter in the TEM image could be caused by swelling through hydration. (**Kuznetsova *et al*. 1997**) Triple-helix G901S has a similar dimension as the F877 but appears to disperse slightly more as shown in **Figure 14c**. To examine feasibility in their application as building blocks for electronics, these triple helices were coated by Au. When the F877 triple helix (**Figure 16a**) is incubated with trimethylphosphinegold chloride ([$AuPMe_3Cl$]) for four days and then reduced by hydrazine hydrate for one day at 48^{o}C, Au crystals grow on the helix as shown in **Figure 16b**. This TEM image shows that the Au nanocrystalline growth was localized on the helix surface. To grow Au on the triple helix more uniformly, we precoated the triple helix with a Au-mineralizing peptide, Ala-His-His-Ala-His-His-Ala-Ala-Asp (HRE), which has a high affinity for organic Au salts. As we know that the HRE binds to glycine – bolaamphiphile nanotubes through

hydrogen bonding at the amide groups of the nanotube after a simple incubation process. The subsequent Au electroless process on these HRE-bound nanotubes yields a uniform Au nanocrystalline coating. A similarly enhanced, more uniform mineralization is found for the triple helices F877 and G901S, after incubation with HRE. When the F877 triple helix (**Figure 16a**) is incubated in the HRE peptide solution for 24 hours, the triple helix is coated by the HRE peptide, as confirmed by FTIR spectrometry. No significant changes in length, diameter or shape are detected after the triple helix is coated by HRE, as shown in **Figure 16c**.

The HRE peptide coating increases dispersion of the triple helix nanowires, presumably because the coating of the HRE peptide contains clusters of positive charges, which reduce the potential attractive interaction between the triple helices. The reduction of Au on the HRE-coated triple helix produces uniform and highly crystalline Au coating on the surface, as shown in **Figure 16d**. The Au-coated triple helix in the inset of **Figure 16d** appears to be a ricelike shape, which could be due to the inhomogeneous coating of the HRE peptide at the ends. As shown in **Figure 14**, the foldon at the C-terminal end forms a three-stranded b-hairpin propeller, a conformation very different from that of the triple helix, while the helix fray at the N-terminal end of the triple-helix domain has been well documented by structural studies of crystallography and NMR.(**Stetefeld *et al.* 2003**) These conformation differences result in different binding of the HRE peptides and lead to less Au growth on those areas, compared to the middle part of triple helix, consistent with the rice-shape formation. Similarly, Au nanowires with identical features are

obtained when the G901S triple-helix peptides are precoated by HRE.

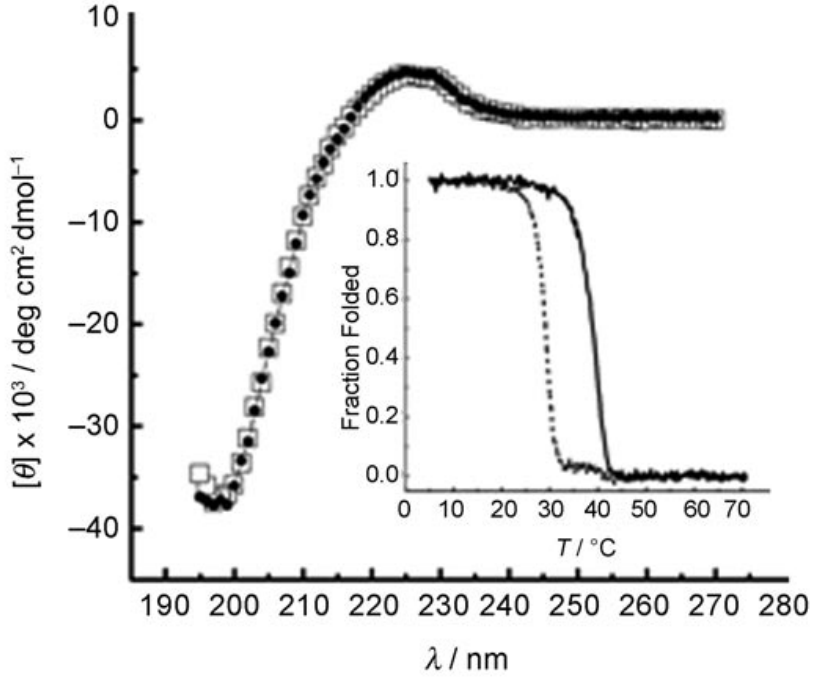
FIGURE 14: Illustrated structure of the collagen-like triple helix (top). An underlined G residue in the wild-type F877 helix was mutated to Ser in G901S. a) Circular Dichroism (CD) spectra of F877 (white rectangle) and G901S (black circle) at 48^{o}C. Inset shows denaturation temperatures of F877 (—) and G901S (·····). b) TEM image of the F877 triple-helix. c) TEM image of the G901S. Scale bars: 40 nm



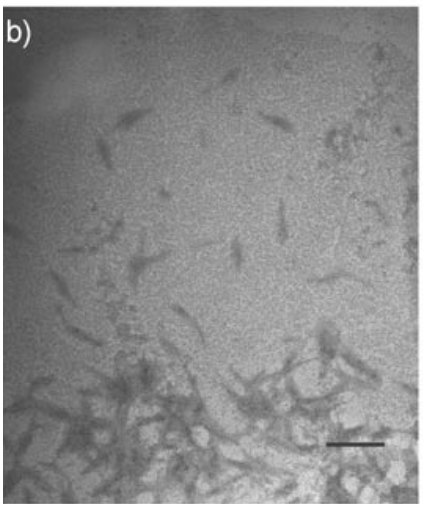
$(GPP)_6$ – [collagen α1 chain, 877-939] – $(GPP)_4$ – (Cys-Cys) – foldon

GPVGPAGKSGDRGETGPAGPA
GPVGPVGARGPAGPQGPRGDK
GETGEQGDRGIKGHRGFSGLQ

a)



b)



c)

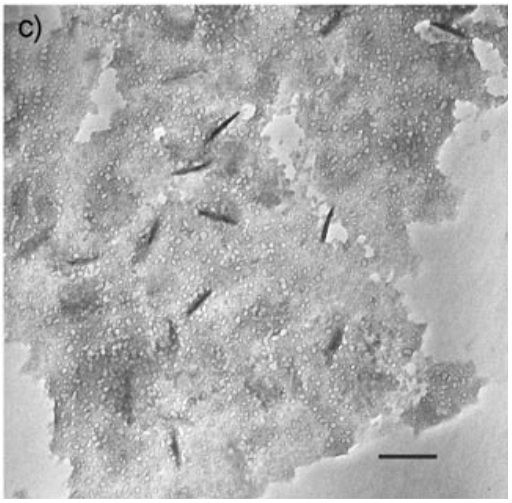


FIGURE 15: Construction of nanowires using recombinant collagen protein as template.

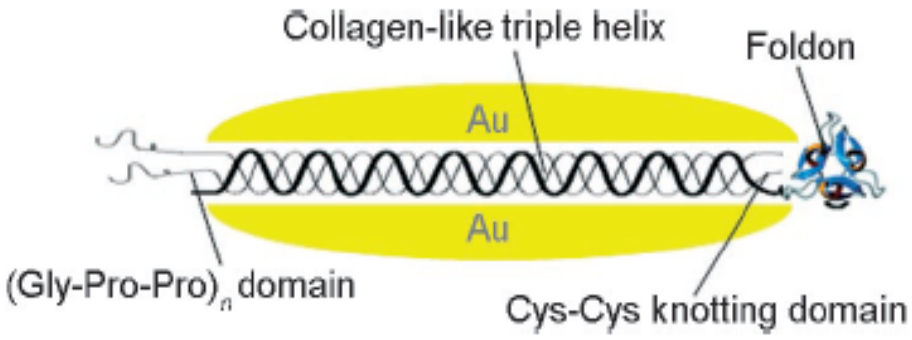
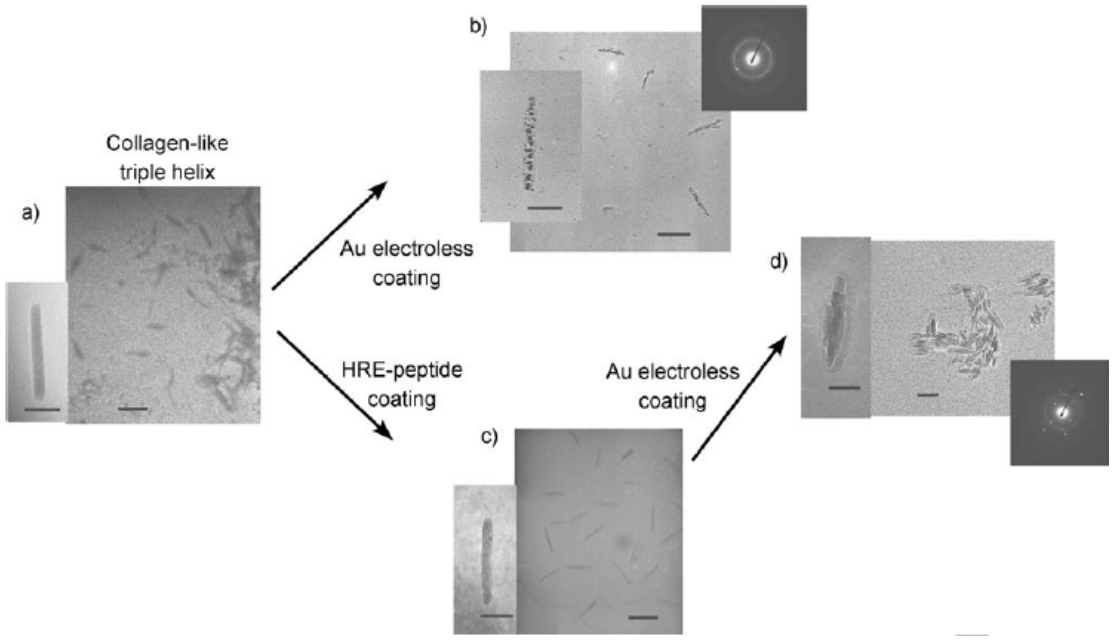

Collagen-like triple helix
Foldon
Au
Au
(Gly-Pro-Pro)n domain
Cys-Cys knotting domain

FIGURE 16: TEM images of the F877 triple-helix peptides: a) neat, b) coated by Au, c) coated by HRE mineralizing peptides. d) TEM image of the HRE-coated triple-helix peptides in (c) coated by Au. Electron diffraction patterns of (b) and (d) are shown next to their TEM images. Scale bars: 40 nm. Insets are their HRTEM images; scale bars: 15 nm.


Collagen-like
triple helix
a)
Au electroless
coating
b)
HRE-peptide
coating
c)
Au electroless
coating
d)

SUMMARY AND CONCLUSIONS

While the Au growth is observed with both the HRE precoated and the neat (i.e. no HRE pretreatment) F877 triple helices, the reduction of Au with the neat G901S helix yields no Au nanowires. The Gly → Ser mutation included in the G901S is known to cause brittle bone disease as a result of the disruption of the triple-helix conformation. (**Baum *et al.* 1999**) While the G901S was shown to adapt to the triple-helix conformation at low temperature, the denaturation temperature (*Tm*) of this triple helix decreases by more than 10^{o}C (**Figure 14a**) compared to F877, indicating the reduced stability of G901S with an altered conformation. It is unclear at this point whether the lack of Au deposit on the G901S is due to the disruption of the charge distribution on the surface of the F877 triple helix or the altering of other structural features of the F877 triple helix caused by the Gly → Ser mutation. Nevertheless, the mutation-caused variations of the Au crystalline morphologies on the triple helix highlight the dependence of the properties of the nanowires on the molecular conformations of the helix template. Such sequence-dependent behavior also offers a practical means to produce nanowires with desired properties by modifying the sequence of the recombinant triple-helix molecules. In summary, monodisperse Au nanowires with defined dimension of 4 nm X 40 nm were obtained by templating recombinant collagen-like triple helices from an *E. coli* expression system. The length of the nanowires can be easily controlled by the number of amino acid residues, and the production of triple helix can be scaled up by means of the

cell multiplication. Thus, the unique molecular properties of collagen-like triple helix combined with the versatility of the recombinant technology offer a promising system to create biomolecular nanowires by design.

Chapter III

Unpublished data and future work

(Chapter III was results and data for future publication. Although works presented here are inconclusive on their own, they lay out the foundation for future studies).

I) Analyzing the binding affinity of Hsp47 with our recombinant collagen protein:

The Pro-free region (916-939) is known to be important for extra-cellular molecular binding **(DiLullo *et al*. 2002)**. The openings of the Pro-free region in the lethal OI mutation G913S may further interrupt the binding between collagen and other signal transudation molecules. Here, we conducted pull-down assay experiments to investigate the effect of G913S mutation on the function of type I collagen. The pull-down assay is used to analyze the binding affinity between collagen binding molecules and collagen triple helix. We picked up Hsp47 as a candidate and applied it on the assay with both wild type F877 and lethal G913S mutant, since the Hsp47 protein is essential for collagen molecular maturation and plays an important role on collagen fiber formation. As the SDS-PAGE showed in **Figure 17**, Hsp47 can bind with F877 and G913S with similar affinity. The concomitant elution of Hsp47 with the recombinant collagen by imidazole was used as an indication of binding. Due to the low sensitivity of such an assay, the results are not conclusive on whether the Gly913 to Ser substitution affects the binding affinity. For the future studies, we will need more negative controls of pull down assay, such as Hsp47 with foldon and Hsp47 with thioredoxin tag to make sure the binding only

happens between Hsp47 and collagen triple helix region. Secondly, we will search the binding map to find other extra-cellular collagen binding molecules to investigate the effects of G913S on the function of the collagen triple helix. There are also uncertainties as to whether the binding of Hsp47 requires a triple helix conformation. Some studies seem to show that Hsp47 binds both the triple helix and the unfolded collagen, which may be the reason that we see the pull down by both the F877 and the G913S with an unfolded binding site (the Pro-free domain). In this case, we need to design more sensitive experiment procedures and search for better extra-cellular candidates to investigate the effect of lethal G913S mutant on the function of collagen triple helix.

II) Design and express new recombinant collagen proteins to generate long nanowires

As for co-operated project using recombinant collagen protein to generate nanowires, we expressed collagen recombinant protein with length of more than 300 amino acids. To make sure the long collagen recombinant protein can form triple helix, we picked up the relative stable regions from α I chain of type I collagen (sequence 118-445) and assembled them into one complete chain. Because of length limitation of DNA synthesis, the DNA fragments encoding the amino acids were split into three fragments and synthesized separately by Gene, Oracle. After synthesis, these three DNA fragments were ligated together and inserted into our pET32a (+) plasmid for expression (**Figure 18**). The expressed long collagen recombinant proteins were identified by SDS-PAGE (**Figure**

19). The long collagen recombinant proteins were identified by transmission electronic microscope (TEM) with length around 120nm. The long collagen recombinant proteins will be coated by Au and used for nano-bridge between two electric poles. The nano-bridge requires at least 100nm on length.

FIGURE17: SDS-PAGE of pull-down assay experiment of Hsp47 and F877, G913S. Lane 1: Mixed protein sample of F877 and Hsp47 before binding with resin Lane 2: Flow through sample of combined F877 and Hsp47 after binding Lane 3 and 4: Collected wash sample of resin with mixed F877 and Hsp47 by 1X wash/extraction buffer; Lane 5, 6 and 7: Collected elution samples of resin with mixed F877 and Hsp47 by 1X elution buffer. Lane 8 Mixed protein sample of G913S and Hsp47 before binding with resin Lane 9: Flow through sample of combined G913S and Hsp47 after binding Lane 10 and 11: Collected wash sample of resin with mixed G913S and Hsp47 by 1X wash/extraction buffer; Lane 12, 13 and 14: Collected elution samples of resin with mixed G913S and Hsp47 by 1X elution buffer Lane 15: Sigma low range marker.

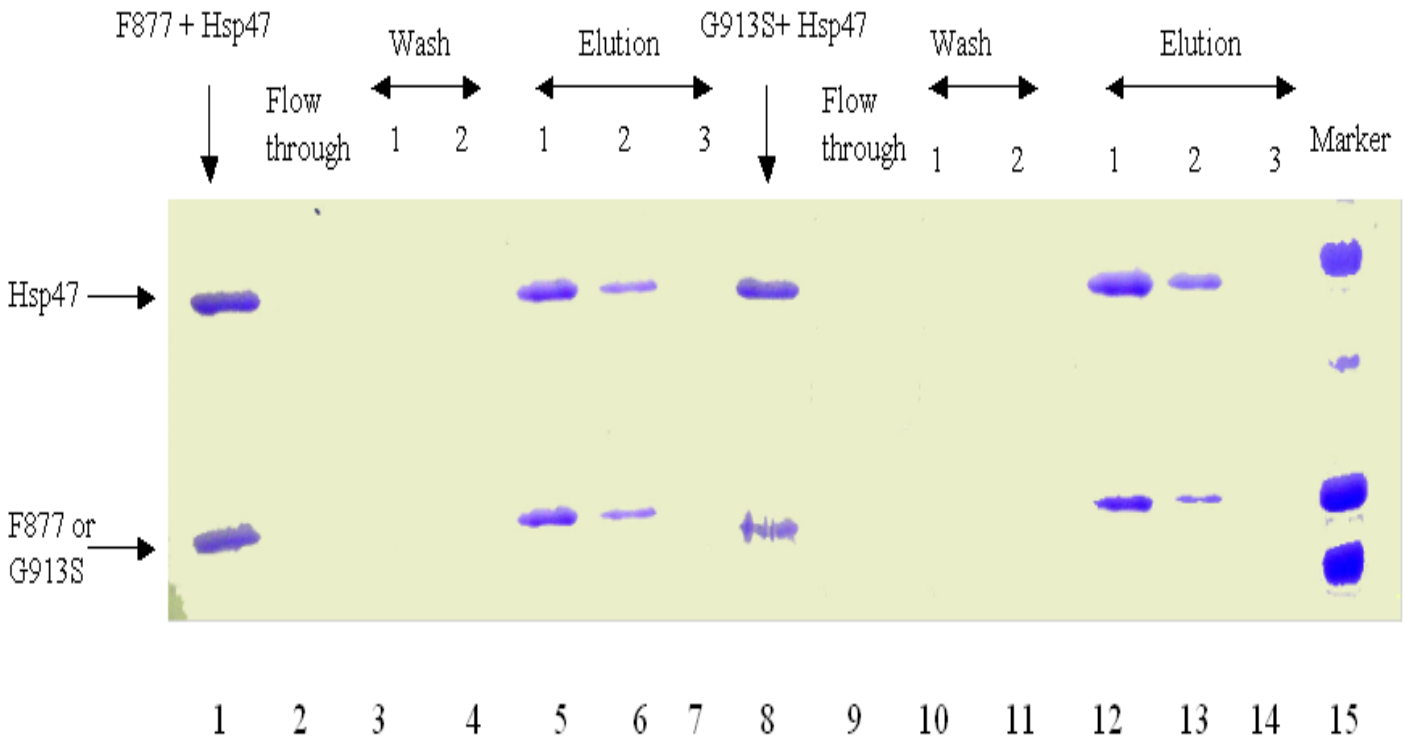

F877 + Hsp47
Wash
Elution
G913S+ Hsp47
Wash
Elution
Flow through
1 2
1 2 3
Flow through
1 2
1 2 3
Marker
Hsp47
F877 or G913S
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

FIGURE 18: Design of DNA fragment of long triple helix recombinant protein. Three fragments with restricted enzymes linker were ligated by T4 ligation enzyme. Extra nucleotides are added beside restricted enzyme cutting site to keep the $(GXY)_n$ pattern.

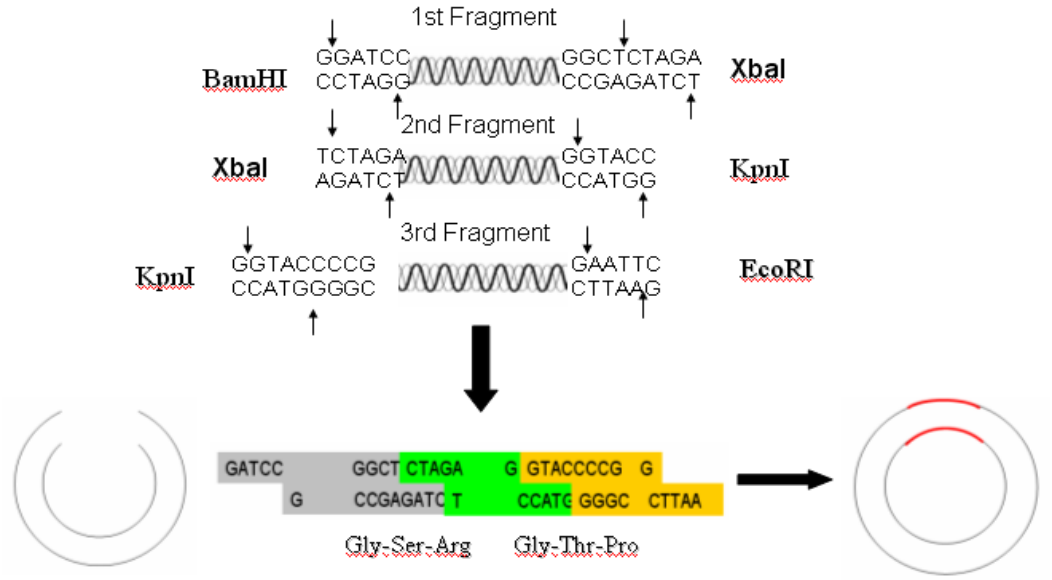
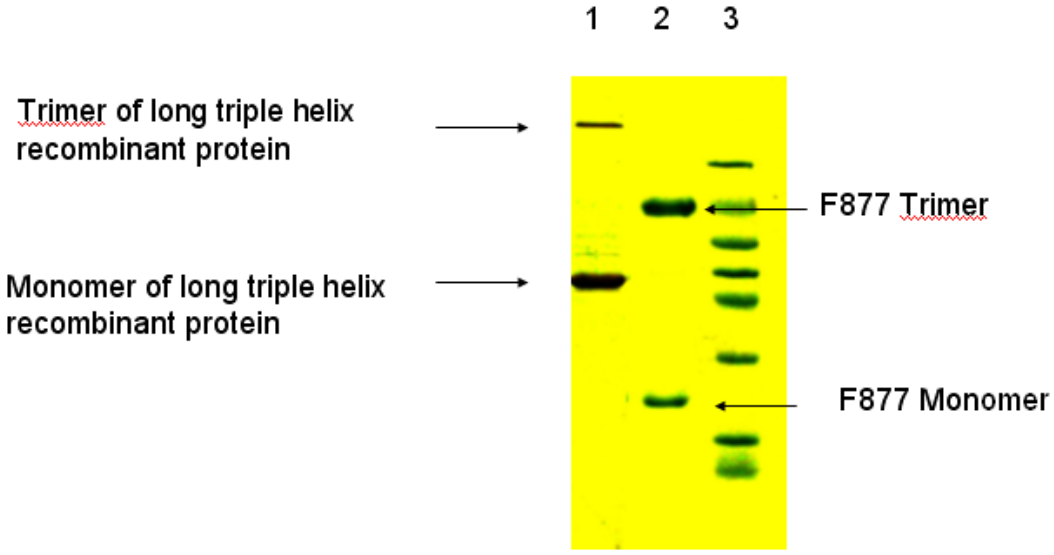

1st Fragment
BamHI
GGATCC
CCTAGG
GGCTCTAGA
CCGAGATCT
XbaI
2nd Fragment
XbaI
TCTAGA
AGATCT
GGTACC
CCATGG
KpnI
3rd Fragment
KpnI
GGTACCCCG
CCATGGGGC
GAATTC
CTTAAG
EcoRI
GATCC
GGCT CTAGA
G GTACCCCG G
G
CCGAGATC T
CCATG GGGC CTTAA
Gly-Ser-Arg
Gly-Thr-Pro

FIGURE 19: SDS-PAGE analysis of the expressed long triple helix recombinant protein and F877 recombinant protein

Lane 1: Purified trimer and monomer of long triple helix recombinant protein expressed from *E. coli* JM109 Lane 2: Purified trimer and monomer of F877 recombinant protein expressed from *E. coli* JM109 Lane 3: protein molecular weight marker from top to bottom (66kD, 45kD, 36kD, 29kD, 24kD, 20kD, 14.2kD)


1
2
3
Trimer of long triple helix recombinant protein
Monomer of long triple helix recombinant protein
F877 Trimer
F877 Monomer

APPENDIX A

Clinically identified Osteogenesis Imperfecta mutation sites and phenotypes (within 877-939 regions) (**http://www.le.ac.uk/genetics/collagen/col1a1_legacy.html**)

Mutation	**Phenotype**	**Reference(s)**
Gly880Ser GGT → AGT	Osteogenesis Imperfecta (OI) IV	Lund *et al.* Hum Mutat 9:378-382 1997 Lee *et al.* Human Mutat Mutation in Brief #894 2006 Online
Gly883Ser GGC → AGC	OI IV	Lightfoot *et al.* J Biol Chem 269:30352-30357 1994
Gly883Asp GGC → GAC	OI II	Cohn *et al.* Am J Hum Genet 46:591-601 1990
Arg888Cys CGT → TGT	OI/EDS	Cabral *et al.* Hum Mutat 28:396-405 2007
Gly898Ser GGT → AGT	OI III	Lund *et al.* Hum Mutat 9:378-382 1997
Gly901Ser **GGC → AGC**	OI I	Mottes *et al.* Hum Genet 89:480-484 1992 Hartikka *et al.* Hum Mutat 24:147-154 2004
Gly904Cys GGC → TGC	OI II	Constantinou *et al.* J Clin Invest 83:574-584 1989
Gly910Ala GGA → GCA	OI II	Valli *et al.* Eur J Biochem 211:415-419 1993
Gly913Ser **GGC → AGC**	OI II	Cohn *et al.* Matrix 10:236 (abst) 1990
Arg915Cys CGT → TGT	OI EDS	Malfait *et al.* Hum Mutat 28:387-395 2007
Gly928Ala GGC → GCC	OI II	Lamande *et al.* J Biol Chem 264:15809-15812 1989

APPENDIX B

Buffers and solutions used in the experiments

Bacterial Media:

1 liter LB media

NaCl	10 g
Yeast Extract	5 g
Tryptone	10 g
ddH2O	1000 ml

1 liter LB media for plate:

NaCl	10 g
Yeast Extract	5 g
Tryptone	10 g
Bacto agar	15 g
ddH2O	1000 ml

Reagents and solutions for plasmid reconstruction, PCR and DNA gel:

Original Plasmid was reconstructed pET32a (+) plasmid from Dr. Jugen Engel's lab.

Inserted expression fragment of recombinant collagen was synthesized by Gene Oracle Company.

T4 ligation enzyme and Bam HI restriction enzyme were bought from sigma.

DNA gel solution:

0.5 X TBE solution

Tris-borate	45 mM
EDTA	1 mM

6 X Gel loading solution

Bromophenol blue	0.25% (Weight/Volume)
Sucrose	40% (Weight/Volume)

Ethidium bromide final concentration = 0.5 ug/ml

PCR kit: Quick-change site-directed mutagenesis kit

Primers were ordered from IDT

PCR program setting:

Segment	Cycles	Temperature	Time
1	1	95^{o}C	2 minutes
2	18	95^{o}C	20 seconds
		60^{o}C	10 seconds
		68^{o}C	6 minutes

3	1	68^{o}C	5 minutes

Reagents and solutions for plasmid transformation and extraction

Transformation solution:

$CaCl_2$	20 mM
$MgCl_2$	80 mM

Plasmid extraction kit: Promega wizard plus miniprep

Cell resuspension solution:

Tris-HCl(pH=7.5)	50 mM
EDTA	10 mM
RNase A	100 ug/ml

Cell lysis solution:

NaOH	0.2 mM
SDS	1% (Weight/Volume)

Neutrilization solution:

1.32 M potassium acetate (pH=4.8).

Column washing solution:

Potassium acetate	80 mM
Tris-Hcl (pH=7.5)	8.3 mM
EDTA	40 uM
Ethanol	55% (Volume/Volume)

TE buffer:

Tris-Hcl(pH=7.5)	10 mM
EDTA	1 mM

Reagents and solutions for SDS-PAGE:

Solution A 100 ml

Acrylamide	29.2g
Bis-acrylamide	0.8g
ddH2O	100 ml

Solution B 100 ml

2M tris-Hcl (pH=8.8)	75 ml
10% SDS	4 ml
ddH2O	21 ml

Solution C 100 ml

1M Tris-HCL (pH=6.8)	50 ml
10% SDS	4 ml
ddH2O	56 ml

Coomassie gel staining solution 1liter

Coomassie blue R-250	1 g
Methanol	450 ml
Glacial acetic acid	100 ml
ddH2O	450 ml

Coomassie gel destaining solution 1liter

Methanol	100 ml
Glacial acetic acid	100 ml
ddH2O	800 ml

Electrophoresis Solution 1L

Tris	3 g
Glycine	14.4 g

SDS	1 g

10% ammonium persulfate (APS) and TEMED solution

15% separation gel 10 ml

Solution A	5 ml
Solution B	2.4 ml
10% SDS	0.1 ml
ddH2O	2.4 ml
APS	50 ul
TEMED	5 ul

Stacking gel 4 ml

Solution A	0.67 ml
Solution C	1 ml
ddH2O	2.3 ml
APS	20 ul
TEMED	4 ul

5 X Sample buffer 10 ml

1 M Tris-Hcl (pH=6.8)	0.6 ml

50% glycerol	5 ml
10% SDS	2 ml
2-mercaptoethanol	0.5 ml
1% bromophenol blue	1 ml
ddH2O	0.9 ml

Loading sample volume and composition:

8 ul protein sample + 2 ul 5 X sample buffer = 10 ul sample

Sigma low range Marker bought from Sigma:

Name of protein marker	Molecular weight
Albumin, bovine serum	66KD
Ovalbumin, chicken egg	45KD
Glyceraldehyde-3-phosphate dehydrogenase, rabbit muscle	36KD
Carbonic anhydrase, bovine erythrocytes	29KD
Trypsinogen, bovine pancreas	24KD
Trypsin inhibitor, soybean	20KD
α-Lactalbumin, bovine milk	14.2KD

Regents and solutions for protein purification:

BD Talon™ metal affinity resin:

1X equilibration/wash buffer (pH=7.0)

Sodium phosphate	50 mM
NaCl	300 mM

1X elution solution

Sodium phosphate	50 mM
NaCl	300 mM
Imidazole	500 mM

MES buffer:

20 mM MES pH=5.0

Storage solution

Ethanol	20% (Volume/Volume)
Azide	1% (Weight/Volume)

Regents and solutions for thrombin cleavage:

Thrombin cleancleave^{TM} kit:

Thrombin-Agarose from bovine plasma (suspension in 50% glycerol 20 mM Tris-HCl pH=8.2)

10 X thrombin cleavage buffer

Tris-HCl (pH=8.0)	500 mM
$CalCl_2$	100 mM

Regents and solutions for HPLC

Column:

Phenomenex biosep-sec-s 2000 gel filtration column

Reverse Phase C8 column from Grace davison discovery science™

Pump solution:

10mM PBS + 6M Urea, Acetonitril + 0.1% TFA (Volume/Volume) and Methanol + 0.1% TFA (Volume/Volume)

Circular Dichroism parameters

For Circular Dichroism wavelength scans:

All samples were scanned at 4°C from 195nm-270nm

For Circular Dichroism Melting experiments:

The fixed 225nm signal was recorded with melting temperature step of 0.15°C/minute

Other machines used in experiment:

Beckman Coulter Avanti J-E centrifuge (high speed centrifuge)

Beckman Coulter micro-centrifuge 18 centrifuge (small centrifuge)

Beckman SC-100 collector

Lyophilize machine from Dr. Francesiconi's lab

BIBLIOGRAPHY

Bachinger, H. P., Morris, N. P., and Davis, J. M. (1993) *American Journal of Medical Genetics* **45**, 152-162

Bai, H., Xu, K., Xu, Y., and Matsui, H. *Fabrication of Au Nanowires of Uniform Length and Diameter Using a Monodisperse and Rigid Biomolecular Template: Collagen-like Triple Helix.* (2007) *Angew Chem Int Ed Engl*. **46(18),** 3319-22

Banerjee, I. A., Yu, L., and Matsui, H. (2003) *Proc. Natl. Acad. Sci. USA.* **100**, 14678

Banerjee, I. A., Yu, L., and Matsui, H. (2005) *J. Am. Chem. Soc.* **127**, 16002

Baum, J., and Brodsky, B. (1999) *Curr. Opin. Struct. Biol.* **9**, 122-128

Behrens, S., Wu, J., Habicht, W., and Unger, E. (2003) *Chem. Mater.* **15**, 3085

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

Braun, E., Eichen, Y., Sivan, U., and Ben-Yoseph, G. (1998) *Nature* **391**, 775

Bruckner, P., and Prockop, D. J. (1981) *Anal Biochem.* **110(2)**, 360-368

Buehler, M. J. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 12285

Buevich, A. V., Silva, T., Brodsky, B., and Baum, J (2004) *J Biol Chem.* **279(45)**, 46890-46895

Byers, P. H. (1993) In *Osteogenesis Imperfecta Connective Tissue and Its Heritable Disorders* (Royce, P. M., and Steinmann, B., eds.), pp. 317-350, Wiley-Liss, New York

Chung, S. W., Ginger, D. S., Morales, M.W., Zhang, Z. F., Chandrasekhar, V., Ratner, M. A., and Mirkin, C. A. (2005) *Small* **1**, 64

Dalgleish, R. (1997) *Nucleic Acids Res.* **25(1)**, 181-187

Deng, Z., Tian, Y., Lee, S. H., Ribbe, A. E., and Mao, C. D. (2005) *Angew. Chem*. **117**, 3648; *Angew. Chem. Int. Ed.* **44**, 3582

DiLullo, G. A., Sweeney, S. M., Körkkö, J., Ala-Kokko, L., and Antonio, J. D. S. (2002) *J. Biol. Chem* **277(6)**, 4223-4231

Djalali, R., Chen, Y.-F., and Matsui, H. (2003) *J. Am. Chem. Soc.* **125**, 5873

Endo, M., Wang, H. X., Fujitsuka, M., and Majima, T. (2006) *Chem. Eur. J.* **12**, 3735

Engel, J., and Prockop, D. J. (1991) *Annu. Rev. Biophys. Biophys. Chem* **20**, 137-152

Engel, J., Chen, H. T., Prockop, D. J., and Klump, H. (1977) *Biopolymers* **16(3)**, 601-622

Engel, J., and Bachinger, H. P. (2000) *Matrix Biology* **19**, 235-244

Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lusting, A., Bachinger, H. P., and Engel, J. (2001) *Journal of Molecular Biology* **308**, 1081-1089

Gao, X., and Matsui, H. (2005) *Adv. Mater* **17**, 2037

Hyde, T. J., Bryan, M. A., Brodsky, B., and Baum, J. (2006) *J. Biol. Chem*. **281(48)**, 36937-36943

Kadler, K. (1995) *Protein Profile*. **2**, 491-619

Kielty, C. M., Hopkinson, I., and Grant, M. E. (1993) in Connective Tissue and Its Heritable Disorders (Royce, P. M., and Steinmann, B., eds), pp. 103 – 147, Wiley-Liss, New York

Knez, M., Sumser, M., Bittner, A. M., Wege, C., Jeske, H., Martin, T. P., and Kern, K. (2004) *Adv. Funct. Mater.* **14**, 116

Kuznetsova, N., Rau, D. C., Parsegian, V. A., and Leikin, S. (1997) *Biophys. J.* **72**, 353.

Liu, X., Kim, S., Dai, Q.-H., Brodsky, B., and Baum, J. (1998) *Biochemistry* **37**, 15528-15533

Makareeva E, Cabral WA, Marini JC, and S., L. (2006) *J Biol Chem.* **281(10)**, 6463-6470

Makareeva, E., Mertz, E. L., Kuznetsova, N. V., Sutter, M. B., DeRidder, A. M., Cabral, W. A., Barnes, A. M., McBride, D. J., Marini, J. C., and Leikin, S. (2008) *J. Biol. Chem*. **283**, 4787 – 4798

Mao, C. B., Solis, D. J., Reiss, B. D., Kottmann, S. T., Sweeney, R. Y., Hayhurst, A., Georgiou, G.., and Iverson, B., and Belcher, A. M. (2004) *Science* **303**, 213

Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., Antonio J. D. S., Milgrom, S., Hyland, J. C., Korkko, J., Prockop, D. J., Paepe A. D., Coucke, P., Symoens, S., Glorieux, F. H., Roughley, P. J., Lund, A. M., Kuurila-Svahn, K., Hartikka, H., Cohn, D. H., Krakow, D., Mottes, M., Schwarze, U., Chen, D., Yang, K., Kuslich, C., Troendle, J., Dalgleish, R., and Byers, P. H. (2007) *Hum Mutat* **28(3)**, 209-221

Mechling, D. E., and Bachinger, H. P. (2000) *J Biol Chem.* **275(19)**, 14532-14536

Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., and Brodsky, B. (2007) *J. Biol. Chem.* **282**, 29757

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

Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A., and Brodsky, B. (2000) *Biochemistry* **39**, 14960-14967

Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A., and Brodsky, B. (2002) *J. Mol. Biol.* **316**, 385-394

Persikov, A. V., and Brodsky, B. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 1101

Persikov, A. V., Ramshaw, J. A. M., and Brodsky, B. (2005) *J. Biol. Chem.* **280**, 19343-19349

Privalov, P. L., Tictopulo, E. I., and Tischenko, V. M. (1979) *J. Mol. Biol.* **127**, 203-216

Privalov, P. L. (1982) *Adv. Prot. Chem.* **35**, 1-104

Ramachandran, G. N., Doyle, B. B., and Boult, E. R. (1968) *Biopolymers.* **6(12)**, 1771-1775

Reches, M., and Gazit, E. (2003) *Science* **300**, 625

Rich, A., and Crick, F. H. C. (1961) *J. Mol. Biol.* **3**, 483-506

Ryadnov, M. G., and Woolfson, D. N. (2004) *J. Am. Chem. Soc.* **126,** 7454

Sharma, J., Chhabra, R., Liu, Y., Ke, Y. G., and Yan, H. (2006) *Angew. Chem.* **118**, 744; *Angew. Chem. Int. Ed.* **45**, 730

Shenton, W., Douglas, T., Young, M., Stubbs, G. and Mann, S. (1999) *Adv. Mater.* **11**, 253

Sillence, D. O., Senn, A., and Danks, D. M. (1979) *J. Med. Genet.* **16**, 101–116

Slocik, J. M., Naik, R. R., Stone, M. O., and Wright, D.W. (2005) *J. Mater. Chem.* **15**, 749

Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S., Landwehr, R., Okuyama, K., and Engel, J. (2003) *Structure* **11**, 339

Sugimoto, K., Kanamaru, S., Iwasaki, K., Arisaka, F., and Yamashita, I. (2006) *Angew. Chem*. **118**, 2791; *Angew. Chem. Int.* Ed. **45**, 2725

Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V., and Rossmann, M. G. (1997) *Structure 5*, 789

Tseng, R. J., Tsai, C., Ma, L., Ouyoung, J., Ozkan, C. S., and Yang, Y. (2006) *Nat. Nanotechnol*. **1**, 72

Venugopal, M. G., Ramshaw, J. A., Braswell, E., Zhu, D., and Brodsky, B. (1994) *Biochemistry* **33(25)**, 7948-7956

Xu, K., Nowak, I., Kirchner, M., and Xu, Y. *Recombinant Collagen studies link the severe conformational changes induced by Osteogenesis imperfecta mutations to the disruption of a set of interchain salt bridges* (2008) *J. Biol. Chem.* **283**, 34337-34344

Xu, Y., Hyde, T., Bhate, M., Lu, X., Brodsky, B., and Baum, J. (2003) *Biochemistry* **42**, 8696-8703

Xu, Y., Bhate, M., and Brodsky, B. (2002) *Biochemistry* **41**, 8143-8151

Yang, W., Battineni, M. L., and Brodsky, B. (1997) *Biochemistry* **36**, 6930-6935

Yi, H. M., Nisar, S., Lee, S. Y., Powers, M. A., Bentley, W. E., Payne, G. F., Rubloff, G. W., Harris, M. T., and Culver, J. N. (2005) *Nano Lett.* **5**, 1931

Yu, L., Banerjee, I. A., Shima, M., Rajan, K., and Matsui, H. (2004) *Adv. Mater.***16**, 709

Yu, L., Banerjee, I. A., and Matsui, H. (2003) *J. Am. Chem. Soc.* **125**, 14837

Whaley, S. R., English, D. S., Hu, E. L., Barbara, P. F., and Belcher, A. M. (2000) *Nature* **405**, 665