Effects of interruptions on collagen triple-helix

Folding Delay and Structural Perturbations Caused by Type IV Collagen Natural Interruptions and Nearby Gly Missense Mutations*

Eileen Hwang¹ and Barbara Brodsky¹²

¹From the Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ 08854

²Department of Biomedical Engineering, Tufts University, Medford, MA 02155

Running title: Effects of interruptions on collagen triple-helix

To whom correspondence should be addressed: Barbara Brodsky, Dept. of Biomedical Engineering, Tufts University, 4 Colby Street, Medford, MA 02155. Tel.: 617-627-0447. Fax: 617-627-3231. Email: barbara.brodsky@tufts.edu.

Keywords: collagen, triple-helix, folding, mutations, interruptions

Background: Non-fibrillar collagens contain natural interruptions in the (Gly-Xaa-Yaa)ₙ sequence pattern.

Results: An interruption within a recombinant triple-helix led to delayed folding, an effect magnified by a nearby Gly missense mutation.

Conclusion: The length and relative positions of interruptions influence triple-helix folding.

Significance: Consequences of natural interruptions can be compared with pathogenic mutations to clarify the basis of collagen disorders.

SUMMARY

The standard collagen triple-helix requires Gly as every third residue in the amino acid sequence, yet all non-fibrillar collagens contain sites where this repeating pattern is interrupted. To explore the effects of such natural interruptions on the triple-helix, a 4- or 15-residue sequence from human basement membrane type IV collagen was introduced between (Gly-Xaa-Yaa)ₙ domains within a recombinant bacterial collagen. The interruptions had little effect on melting temperature, consistent with the high thermal stability reported for non-fibrillar collagens. Although the 4-residue interruption cannot be accommodated within a standard triple helix, trypsin and thermolysin resistance indicated a tightly packed structure. Central residues of the 15-residue interruption were protease susceptible, while residues near the (Gly-Xaa-Yaa)ₙ boundary were resistant, supporting a transition from an alternate conformation to a well-packed triple-helix. Both interruptions led to a delay in triple-helix folding, with the 15-residue interruption causing slower folding than the 4-residue interruption. These results suggest that propagation through interruptions represents a slow folding step. To clarify the relation between natural interruptions and pathological mutations, a Gly to Ser missense mutation was placed three triplets away from the 4-residue interruption. As a result of this mutation, the 4-residue interruption and nearby triple helix became susceptible to protease digestion, and an additional folding delay was observed. Since Gly missense mutations which cause disease are often located near natural interruptions, structural and folding perturbations arising from such proximity could be a factor in collagen genetic diseases.

Collagens are defined as extracellular matrix proteins containing a triple-helix domain. Different subgroups of the human collagen family have specific tissue locations and distinctive higher order structures (1-3). Most abundant are collagens that form fibrils with an axial periodicity of 67 nm in bone, tendon, skin, and many other tissues. The polypeptide chains of these fibrillar collagens have a precise repeating amino acid sequence pattern.
Effects of interruptions on collagen triple-helix

sequence with Gly as every third residue, \((\text{Gly-Xaa-Yaa})_n\). This constraint is imposed by the tight packing of three polyproline II-like helices to form the triple-helix structure (4-6). The replacement of even one Gly by a larger residue leads to pathology. For instance, Gly missense mutations in type I collagen lead to the bone disorder, Osteogenesis Imperfecta (7,8).

The perfect \((\text{Gly-Xaa-Yaa})_n\) sequence pattern seen for fibril forming collagens is not observed in other collagen family members, which are collectively classified as non-fibrillar. Non-fibrillar collagens may form networks, attach to the surface of fibrils or exist as transmembrane proteins. All non-fibrillar collagens contain a few or many sites where the repeating \((\text{Gly-Xaa-Yaa})_n\) sequence pattern is interrupted within the triple-helix (1-3). For example, type VII collagen, a homotrimer in dermal anchoring fibrils, has 19 sites where the repeating tripeptide pattern is disturbed, while each chain of heterotrimeric type IV collagen in basement membrane has more than 20 interruption sites (9). In these collagen types, interruptions are distributed throughout the sequence. The presence of interruptions in the strict repeating tripeptide pattern in non-fibrillar collagen chains indicates that sequence breaks are allowed and may be functional. As intrinsic features of native collagen molecules, these interruption sequences can influence molecular and supramolecular structure and may fill biological roles as sites for degradation, molecular recognition, or self-association (10-13). Rotary shadowing electron microscopy shows molecular flexibility at some interruption sites, and it has been suggested that such sites are important for bends and supercoiling in the higher order structure (10-12,14).

Interruptions can be classified according to the number of residues between \((\text{Gly-Xaa-Yaa})_n\) sequences. The most common types are sites with one residue, usually hydrophobic, between two Gly residues (e.g. \(\text{Gly-Ile-Gly}\)) or four residues between two Gly residues, which often have a small amino acid followed by a hydrophobic one (e.g. \(\text{Gly-Ala-Ala-Val-Met-Gly}\)). Peptide models have been useful in defining the effects of these short natural interruptions on triple-helix structure (15-20). X-ray crystallography of a peptide containing a 1-residue interruption, \(\text{Gly-Pro-Gly}\), showed disruption of the standard hydrogen bonding pattern in the triple-helix and highly localized alterations in dihedral angles at the interruption site (15). NMR studies on a peptide with the 4-residue interruption \(\text{Gly-Ala-Ala-Val-Met-Gly}\) showed the Val residues occupying the central position where Gly would normally be found and forming a small hydrophobic core in the middle of the triple-helix. Interruptions from 1 to 9 residues in length could be incorporated within triple-helical peptides, but a substantial destabilization of the triple-helix was observed (19).

Even though natural interruptions are found within the triple-helices of non-fibrillar collagens, the introduction of a Gly missense mutation often has pathological consequences. Some Gly missense mutations in type VII collagen lead to the dominant form of Dystrophic Epidermolysis Bullosa (21,22), while mutations in the \(\alpha5(IV)\) chain of type IV collagen lead to a kidney disorder, X-linked Alport Syndrome (23,24). The difference between the structural and functional consequences of a natural interruption within the triple-helix and the perturbation caused by a missense mutation is not clear. A better understanding of this difference would help define the effects of pathogenic mutations.

Here, a recombinant bacterial collagen system was used to examine the effect of interruptions in the \((\text{Gly-Xaa-Yaa})_n\) repeating sequence on triple-helix conformation, stability and folding. The recombinant collagen used in this study was based on the collagen-like Sc12 protein from \textit{Streptococcus pyogenes} (25-28), which is easily modified and can be expressed in large amounts. Sc12 has an N-terminal globular trimerization domain, denoted as V, adjacent to a triple-helix domain CL, consisting of \((\text{Gly-Xaa-Yaa})_{80}\). Notably, the thermal stability of the triple-helix is 36 - 37 °C, similar to that of human collagen, even though the protein lacks the stabilizing hydroxyproline (Hyp) post-translational modification (26,27). The CL domain was duplicated to create a recombinant protein, VCLCL, consisting of a longer triple-helix \((\text{Gly-Xaa-Yaa})_{158}\) adjacent to the V domain (28). A single interruption was inserted between the two CL domains, incorporating either a 4- or 15-residue interruption sequence from the \(\alpha5\) chain of type IV collagen. The type IV collagen interruptions were successfully incorporated into
Effects of interruptions on collagen triple-helix

the bacterial triple-helix protein with little impact on overall structure or stability, but folding was significantly delayed. Placement of a Gly missense mutation at a position three triplets away from a 4-residue interruption led to local disruption of the triple-helix and the interruption, together with an additional folding delay.

EXPERIMENTAL PROCEDURES

Recombinant protein construction and expression- pColdIII vectors for expression of the His-tagged recombinant protein VCLCL were generated previously (28,29). pColdIII(VCLCL) contains SmaI and ApaI sites between the CL domains, and using these sites, the amino acid sequence GAAGVM in VCLCL was replaced with sequences GAAVM and GQISEQKRPIDVEFQ to construct the vectors for VCL-X4-CL and VCL-X15-CL (Fig. 1, supplemental Fig. 1). The length of an interruption is defined as the number of residues between Gly-Xaa-Yaa-Gly sequences, so GPPGAAVMGPP is considered to be a 4-residue break. To construct VCL(G-S)-X4-CL, site directed mutagenesis was used to change Gly316 to Ser by mutating G to A using synthesized PCR primers (UMDNJ-RWJMS DNA Core Facility, Piscataway, NJ). All constructs were confirmed by sequencing (Tufts Core Facility, Boston, MA). Amino acid sequences of recombinant proteins are provided in supplemental Fig. 1.

All proteins were expressed in Escherichia coli BL21, purified using a nickel-sepharose affinity column followed by a DEAE-sepharose column (GE Healthcare), and dialyzed in PBS pH7. Protein purity was determined by SDS-PAGE. Concentrations were determined by absorbance at 280 nm using an extinction coefficient of 9970 M⁻¹ cm⁻¹ for all proteins. The purified proteins based on VCLCL contain low amounts of smaller His-tagged triple helical contaminants that may be products of truncated translation, a known phenomenon for bacterial expression of proteins with repetitive sequences (30).

Circular dichroism spectroscopy- Circular dichroism spectra and thermal denaturation curves were obtained using an Aviv model 62 DS circular dichroism spectrometer as previously described (31). The thermal denaturation curve was obtained by plotting fraction folded (FF) = (CD(T) - CD_{denatured}) / (CD_{native} - CD_{denatured}) vs. temperature. There was no significant temperature dependence of the signal for the low temperature native species or the denatured species, so CD_{native} was taken as the CD signal at the refolding temperature, and CD_{denatured} was taken as the CD signal at 55 °C.

Refolding experiments were performed by denaturing the sample at 55 °C for 20 min, transferring the sample directly to a pre-equilibrated CD machine. Cooling of the sample from 55 °C to the folding temperature showed an exponential decrease in temperature with a t_{1/2} of about 15 sec. The signal at 220 nm was monitored for 16 h (1.5 nm bandwidth; 10 s interval time; 2 s time constant). Refolding rates were found to be independent of the time of denaturation (5 min - 60 min) and the temperature at which the protein was denatured (38 °C - 55 °C). Half time of refolding (t_{1/2}) was calculated as the time at which fraction folded reached 0.5.

Fluorescence- Experiments to monitor the intrinsic fluorescence of the single Trp within the V domain were performed using a 5 mm path-length cuvette in a Hitachi F-4500 fluorescence spectrophotometer equipped with a water-circulator cooled cell jacket. After denaturing the sample at 55 °C for 5 min, it was directly transferred to a pre-equilibrated cuvette in the fluorimeter, and the temperature monitored by a digital temperature probe decreased exponentially to the folding temperature exponentially with a t_{1/2} of about 15 sec. The emission was monitored at 340 nm (5 nm slit width; 0.5 sec response time) with excitation at 295 nm (2.5 nm slit).

Protease digestion, mass spectrometry, and protein sequencing- For digestion of the native proteins with proteases, the recombinant proteins were prepared at an initial concentration of 18 µM in PBS pH 7. 7 µL of protein was combined with 2 µL PBS and 1 µL enzyme for a total reaction volume of 10 µL. TPCK-treated bovine pancreatic trypsins (Sigma) (0.05 mg/ml) was added at a weight ratio of 1:160 and incubated at 25 °C for 5 hours. Thermolysin from Bacillus thermoproteolyticus rokko (Sigma) (0.1 mg/ml) was added at a weight ratio of 1:160 and incubated at 25 °C for 5 hours. Thermolysin from Bacillus thermoproteolyticus rokko (Sigma) (0.1 mg/ml) was added at a weight ratio of 1:160 and incubated at 25 °C for 5 hours. Thermolysin from Bacillus thermoproteolyticus rokko (Sigma) (0.1 mg/ml) was added at a weight ratio of 1:160 and incubated at 25 °C for 5 hours.
Effects of interruptions on collagen triple-helix

sent to the Tufts Core Facility (Boston, MA). For mass spectrometry, trypsin digestion was stopped by adding PMSF to a final concentration of 10 mM, and thermolysin digestion was stopped by adding EDTA to a final concentration of 10 mM. Digested samples were analyzed using a Bruker MicroFlex MALDI-TOF mass spectrometer. Cleavage sites were predicted by ExPASy PeptideCutter www.expasy.org/tools/peptidecutter/ (32).

Refolding experiments using trypsin were performed by denaturing 9 µL aliquots of VCLCL at 55 °C for 5 min and then transferring all aliquots to a 25 °C water bath. At each time point, 1 µL of trypsin (1 mg/mL) was added to an aliquot to reach 10 µL total volume. Digestion was performed at a weight ratio of 1:9 for 5 min at 25 °C, with the reaction stopped by adding 5x SDS sample buffer and boiling for 1 min. All digestion products were run on 12% SDS-PAGE gels and stained with Coomassie blue. Gels were scanned, and band densities were determined using Adobe Photoshop.

Differential scanning calorimetry-Experiments were performed using a Nano DSC II Model 6100 (TA instruments, New Castle, DE). Samples (c=18 µM, overnight dialysis against PBS, pH 7) were loaded in a running instrument and measured using a heating rate of 1 °C/min.

RESULTS

1. Protein Design: Recombinant collagens with interruptions

Recombinant chimeric proteins were created with a defined human type IV collagen interruption sequence inserted between two bacterial collagen triple-helix domains. Recombinant collagens were based on a partial clone of the collagen-like protein Sc12.28 from Streptococcus pyogenes, containing the N-terminal trimerization domain, V, adjacent to a CL (Gly-Xaa-Yaa)80 triple-helix region (28) (Fig. 1, supplemental Fig. 1) (26,27). The construct with two tandem CL domains, VCLCL, was also previously characterized (28). A 4-residue interruption sequence from human α5(IV), residues 390-393, AAVM, was inserted between the two adjacent CL domains to create the protein VCL-X4-CL. This sequence, which fulfills the consensus features of 4-residue interruptions (17), has a known conformation determined from NMR studies on a peptide model (16). Another protein, VCL-X15-CL, was created with a 15-residue interruption, QISEOKRIPDVFEQK, representing the longest interruption found in the human α5(IV) chain at residues 243-257. Finally, a protein was made with a Gly to Ser missense mutation three triplets N-terminal to an AAVM interruption, VCL(G-S)-X4-CL. All recombinant proteins were expressed in a pColdIII vector in Escherichia coli and purified using a nickel column followed by an ion exchange column. The identities of the proteins were confirmed using mass spectrometry, and purity was assessed by SDS-PAGE.

2. Effect of interruptions on structure and stability of recombinant bacterial collagen

To determine the effect of an interruption on triple-helix structure and stability, the purified recombinant proteins were characterized by circular dichroism (CD) spectroscopy and differential scanning calorimetry (DSC) (Fig. 2). The interruption-containing collagens all possessed typical triple-helix spectra, similar to that of the VCLCL control, with a maximum near 220 nm and a minimum near 198 nm (Fig. 2A, Inset). The ratio of the intensity of the positive peak at 220 nm to the negative peak at 198 nm (Rpn), a good measure of triple-helix content (33), was similar for the interruption proteins and the control (Rpn VCLCL = 0.10; Rpn VCL-X4-CL = 0.10; Rpn VCL-X15-CL = 0.11), suggesting that the interruptions did not disrupt a significant portion of the CL triple-helix domains. The introduction of a Gly to Ser mutation near the 4-residue interruption resulted in little change in the Rpn value (Rpn VCL(G-S)-X4-CL = 0.09).

The CD melting curves for the recombinant proteins with interruptions each showed a single sharp thermal transition (Fig. 2A). The thermal stability was decreased by a small, but reproducible amount for the three interruption constructs compared with the control, but these differences fall within experimental error (+/- 0.5°C): T_m = 36.8 °C for VCLCL, compared with 35.1 - 35.8 °C for the proteins with interruptions (Table 1). Thermal transitions seen by DSC showed the same trends (T_m for three interruption constructs 36.9 - 37.3 °C compared with 37.8 °C for control), but the T_m values were slightly higher than the T_m values obtained by CD due to the faster heating rate (Fig. 2B) (31).

Since residues in a standard triple helical conformation are resistant to cleavage by most
Effects of interruptions on collagen triple-helix enzymes, proteins were treated with trypsin and thermolysin to assess the conformation of the interruptions and adjacent triple-helical regions. Trypsin and thermolysin digestion of the control VCLCL resulted predominantly in CLCL bands on SDS-PAGE (Fig. 3). Similar results were found for VCL-X4-CL (Fig. 3). Mass spectrometry and N-terminal sequencing confirmed the V domain was removed by these enzymes, with trypsin cleavage at residue Arg90 and thermolysin cleavage at residue Leu85 (supplemental Fig. 2A, supplemental Fig. 3). The inability of these enzymes to cleave at the interruption site within VCL-X4-CL indicated that the triple-helix conformation within and around the AAVM sequence is tight. Digestion of VCL-X15-CL with trypsin or thermolysin resulted in smaller fragments, the size of a single CL domain, on SDS-PAGE (Fig. 3). The presence of CL length bands indicated enzyme susceptible sites within or near the 15-residue interruption. Mass spectrometry and N-terminal sequencing of the VCL-X15-CL digestion products supported trypsin cleavage at the central Lys residue within the interruption, QISEQK|RPIDVEFKQ, but not at the predicted Lys residue at the C-terminus of the interruption (supplemental Fig. 2C, supplemental Fig. 3). Thermolysin treatment of VCL-X15-CL showed predominantly cleavage at the central Pro-Ile site, with additional cleavage at the Glu-Phe site (supplemental Fig. 2C, supplemental Fig. 3). The predicted thermolysin cleavage site Glu-Ile at the N-terminus of the interruption was not seen. When a Gly to Ser mutation was introduced three triplets N-terminal to the AAVM interruption, trypsin digestion resulted in cleavage at one Lys-Asp site between the mutation and interruption, while thermolysin cleaved at the Ala-Val site within the interruption GKDSK|DGQPGKPGA|VMGPR (supplemental Fig. 2D, supplemental Fig. 3), as indicated by mass spectrometry and N-terminal sequencing of the cleavage products.

3. Refolding of recombinant bacterial collagens

Refolding of recombinant collagens was measured after a temperature jump, using CD to assess triple-helix formation and fluorescence to monitor folding of the V domain via its single Trp residue. Studies were carried out first on the control VCLCL protein. After denaturing the sample at 55 °C for 5 min and cooling to 25 °C, recovery of native Trp fluorescence in VCLCL at 18 µM occurred with a t₁/₂ of 2.1 min (Fig. 4, Table 1), and showed a strong concentration dependence (data not shown), consistent with trimerization of the V domain. Under the same conditions (18 µM, 25 °C), CD refolding studies of VCLCL demonstrated recovery of the triple helical CD signal (maximum at 220 nm) with a t₁/₂ of 47 min, indicating triple-helix formation was much slower than trimerization (Fig. 4, Table 1).

The initial rate of VCLCL triple-helix formation at 25 °C, obtained by linear fit of the refolding curve from 5 to 15 min, ranged from 0.4 - 0.8 x 10⁻³ s⁻¹ over the concentrations studied (1.8 - 17 mM) (supplemental Fig. 4C, supplemental Fig. 5A). The overall refolding curves for VCLCL fit well to a double exponential (supplemental Fig. 5B). At a concentration of 18 µM and a temperature of 25 °C, the fast phase had a rate of 0.5 x 10⁻³ s⁻¹ and amplitude of 33.6 mdeg, and the slow phase had a rate of 0.06 x 10⁻³ s⁻¹ and an amplitude of 8.9 mdeg (supplemental Fig. 5B). The fast phase rate increased with temperature from 0 to 15 °C, and then decreased until 25 °C (supplemental Fig. 5C), suggesting it reflects two steps, one with a positive temperature coefficient (e.g. cis-trans isomerization) and one with a negative temperature coefficient (e.g. nucleation). The amplitude of the slow phase decreased with increasing temperature, similar to the behavior expected for collagen triple-helix misfolding (supplemental Fig. 5D) (34). Consistent with assignment of this slow phase to misfolding, DSC studies on VCLCL samples refolded for 16 h at 0 °C demonstrated a minor peak with lower Tₘ than the native sample along with the native Tₘ transition, whereas a sample refolded at 25 °C only showed a peak at the native Tₘ (supplemental Fig. 6A). After 6 weeks of incubation at 0 °C, VCLCL regained 100% of its native CD signal (supplemental Fig. 6B).

Comparison of VCL with VCLCL showed the V domain (monitored by fluorescence) had the same folding rate, but CD experiments indicated faster folding for the CL triple-helix compared to the longer CLCL triple-helix (Fig. 4). These results suggest that linear propagation through the length of the triple-helix contributes to the observed folding rate. The bacterial construct VCLCL appears to be a reasonable model for collagen folding because its folding is initiated by
a trimerization event at one end of the helix; it exhibits a dependence on the propagation step; and it displays a similar temperature dependence (34).

Refolding of VCLCL was compared with the homologous recombinant proteins containing interruptions. At 25 °C, 18 µM, fluorescence refolding rates were similar for VCLCL (t½ = 2.1 min), VCL-X₄-CL (t½ = 2.0 min), VCL-X₁₅-CL (t½ = 1.9 min), and VCL(G-S)-X₄-CL (t½ = 1.5 min), confirming that trimerization rates are not affected by the presence of interruptions (Fig. 5A, Table 1). In contrast, circular dichroism t½ values at 25 °C indicated that interruptions delayed triple-helix folding, and that the longer 15-residue interruption caused a greater delay than the shorter 4-residue interruption (Fig. 5C, Table 1). When a Gly substitution mutation was added near the 4-residue interruption, it led to an additional folding delay. Proteins with interruptions demonstrated slower initial rates than the control. These refolding curves were also fit to a double exponential, and the primary source of the delay appeared to be a decrease in the amplitude of the fast phase. Folding at 0 °C (18 µM) was faster for all constructs compared with 25 °C (Fig. 5B, Table 1). At 0 °C, the proteins containing interruptions all showed delayed folding compared with VCLCL, but there was little difference between the VCL-X₄-CL, VCL-X₁₅-CL, and VCL(G-S)-X₄-CL proteins (Fig. 5B, Table 1).

CD refolding curves were complemented with protease digestion studies. Trypsin monitoring of VCLCL refolding showed recovery of a large amount of the original full length CLCL triple-helix, with a refolding curve similar to that generated by CD (supplemental Fig. 7). Trypsin could not be applied to the folding of CL-X₁₅-CL and CL(G-S)-X₄-CL because they have trypsin sensitive sites within the interruptions.

DISCUSSION

Because animal non-fibrillar collagens contain multiple interruptions in the (Gly-Xaa-Yaa)ₙ sequences, it has been difficult to use these collagens to dissect out the effects of interruptions on conformation, stability and folding of the triple-helix, and to clarify whether longer interruptions have different consequences from shorter ones. The recombinant model used in this study permit a comparison between a control triple-helical protein and a homologous protein containing a single interruption of defined length. The homotrimeric bacterial collagens studied here contain an interruption in all 3 chains at the same site, making them good models for homotrimeric non-fibrillar collagens such as type VII or type X. Both the 4-residue and the 15-residue interruption were incorporated within the bacterial collagen-like protein without any major disruption of the triple-helix. Atypical sequences are found adjacent to interruptions in type IV collagen, with a higher hydroxyproline content on the N-terminal side and an unusually high content of charged residues on the C-terminal side (20,35), but it appears that such specialized sequences are not needed to accommodate an interruption.

Peptide studies have suggested that the common small 1- and 4-residue interruptions are incorporated into a linear triple-helix with locally perturbed parameters. The resistance of CL-X₄-CL to trypsin and thermolysin indicates that the triple-helix conformation within and around the AAVM interruption is still tightly wound, in agreement with the distorted but closely packed structure seen by NMR (16). This is also consistent with the observations that the triple-helix of type X collagen, which contains only 1 and 4-residue interruptions, is not susceptible to pepsin digestion (36). In contrast to the 4-residue interruption, both thermolysin and trypsin cleave within the 15-residue interruption of the CL-X₁₅-CL protein, indicating that it is not likely to be incorporated within a well-packed triple-helix but rather to be in an alternate or flexible state. A 15-residue peptide containing only this interruption sequence did not show any indication of ordered structure by CD or NMR (Hwang, Thiagarajan, Giddu, Xiao, and Brodsky pers. comm.). It was interesting to find that a predicted trypsin sensitive site at the C-terminus and a predicted thermolysin-sensitive site near the N-terminus of the 15-residue interruption were not cleaved in the recombinant bacterial constructs. This resistance supports a transition from a susceptible state to an enzyme resistant triple-helix state as the boundary between interruption and (Gly-Xaa-Yaa)ₙ sequence is approached. There is no indication of any disruption of triple-helices adjacent to the interruptions, indicating again the highly localized nature of any perturbation. The one very large interruption in homotrimeric type VII collagen (41 residues) is susceptible to both trypsin and pepsin digestion at 0 °C.
Effects of interruptions on collagen triple-helix digestion (37), consistent with the enzyme susceptibility seen here for the longer 15-residue interruption within CL-X_{15}-CL. These studies on homotrimeric provide a basis for thinking about the more complex heterotrimerers, such as type IV collagen, where the majority of interruptions are found at corresponding positions on different chains, and some sites were found to be pepsin sensitive (38-40).

In peptides, the introduction of an interruption within (Gly-Pro-Hyp)_{10} leads to a substantial destabilization of the triple-helix (19), and it is known that the short length of the peptide triple-helix magnifies energetic consequences. The recombinant constructs with a 4- or 15-residue interruption had sharp thermal transitions with T_m values only slightly lower than the control, near 37 °C, indicating little effect on global stability. This is consistent with previous reports that non-fibrillar collagens with many interruptions form triple-helical molecules with a thermal stability of 37 °C or higher (35-37,41). Homotrimeric type VII molecules show a multiphasic thermal transition with peaks near 42 °C and 56 °C (37), while heterotrimeric type IV collagen molecules show a very broad transition centered near 37 °C (41) or multiple transitions between 30 - 44 °C (35). Multiple interruptions within a triple-helix result in more heterogeneous and broader thermal transitions than seen for perfect repeating tripeptide fibrillar collagens, and it was suggested this may reflect the distribution of lengths of the continuous (Gly-Xaa-Yaa)_{n} segments (~5 - 60 tripeptides) between interruptions (41). In contrast, the thermal transitions reported here for recombinant proteins containing one interruption are notable for their sharpness and cooperativity. This may be a consequence of the location of the interruption between two equal CL triple-helix length segments, where the T_m of each CL domain is the same as for the CLCL protein (28), and a sharp thermal transition is expected even if the interruption divides the molecule into two folding units.

In the recombinant bacterial collagen system, the presence of a single interruption decreased the triple-helix folding rate, but did not affect the trimerization step. A break in the (Gly-Xaa-Yaa)_{n} sequence may interrupt the propagation process and could require renucleation. A renucleation step would be expected to be faster than initial nucleation, since the ends of the three chains would be attached to a nearby triple-helix (41). At 25 °C, the longer 15-residue interruption led to much slower folding than the 4-residue interruption. The difference between the 4- and 15 residue interruption folding was much less at 0 °C, possibly because renucleation, like nucleation, is favored at low temperature.

The delay observed in the recombinant bacterial collagen contrasts with studies on isolated type IV collagen showing folding rates similar to those seen for fibrillar collagens (35,41). In addition, the folding rate of type IV collagen could be increased by the addition of prolyl cis-trans isomerase, suggesting cis-trans isomerization was still the limiting step (35). It is plausible that propagation/renucleation at some interruption sites within type IV collagen represent a slow step of the same order of magnitude as cis-trans isomerization, so that both are important factors in folding. It is also possible that the folding delay at interruptions in type IV collagen is moderated by atypical surrounding sequences favorable for renucleation (20,35) or by their heterotrimeric nature. Even though the rate of folding of type IV collagen appeared similar to that of fibrillar collagens, rotary shadowing measurements indicated non-uniform propagation, which is likely to reflect some hindrance due to large interruptions or the variations in (Gly-Xaa-Yaa)_{n} lengths between interruptions (41). Interruptions could play a role in the slow folding of type IV collagen in vivo, as indicated by the high degree of post-translational modification (42) and very slow secretion from cells (43), but it is hard to dissect out factors in the complicated in vivo folding process which requires the molecular chaperone Hsp47 (44,45) as well as post-translational modifying enzymes.

It is interesting to consider why natural interruptions are tolerated in type IV collagen while a new break in the repeating (Gly-Xaa-Yaa)_{n} sequence in the form of a Gly missense mutation leads to pathology. The Gly missense mutations implicated in Alport syndrome lead to an absence of trimer molecules containing mutant chains in the glomerular basement membrane (46,47), suggesting that there could be a folding/secretion problem. In contrast, natural interruptions have persisted through evolution and are compatible with proper assembly of type IV collagen. In the
Effects of interruptions on collagen triple-helix recombinant bacterial system, a Gly to Ser mutation three triplets N-terminal to the 4-residue AAVM interruption resulted in trypsin cleavage at a site within the (Gly-Xaa-Yaa)ₙ sequence, between the mutation and interruption, as well as thermolysin cleavage at a site within the interruption that was previously enzyme resistant in the VCL-X₄-CL protein (Fig. 3, supplemental Fig. 2D, supplemental Fig. 3). Introducing the mutation also caused a further folding delay. The disruptive structural and folding consequences seen in VCL(G-S)-X₄-CL suggest that the distances between natural interruptions, and therefore the lengths of the (Gly-Xaa-Yaa)ₙ segments, influence their evolutionary persistence. Natural interruptions are well spaced throughout the type IV molecule, rarely occurring within 20 residues of each other. In contrast, ~75% of Gly missense mutations in the α5(IV) chain that result in Alport Syndrome are located within 20 residues of a natural interruption. Mutations in close proximity to a natural interruption could amplify deleterious effects on triple-helix conformation and folding.

REFERENCES

Effects of interruptions on collagen triple-helix


Acknowledgements: We are grateful to Ayumi Yoshizumi for construction of the VCL-X4-CL plasmid; to Sumana Giddu for construction of the VCL-X15-CL plasmid; and to Zhuoxin Yu for support in plasmid construction and for helpful discussions. We appreciate the advice of Drs. Smita Patel, Anton Persikov and Collin Stultz regarding the kinetic analysis.

FOOTNOTES

by guest on November 15, 2017 http://www.jbc.org/ Downloaded from
Effects of interruptions on collagen triple-helix

*This work was supported by an NIH F30 MD/PhD fellowship DK089630 (E.S.H.) and NIH grant GM60048 (B.B.).

1Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ 08854
2To whom correspondence should be addressed: Department of Biomedical Engineering, Tufts University, 4 Colby Street, Medford, MA 02155. Tel.: 617-627-0447. Fax: 617-627-3231. Email: barbara.brodsky@tufts.edu.

3The abbreviations used are: Hyp, hydroxyproline; DSC, differential scanning calorimetry; Rpn, ratio of positive to negative peaks; Cp, heat capacity; MRE, mean residue ellipticity; FF, fraction folded; Tm, melting temperature; FL, fluorescence

FIGURE LEGENDS

FIGURE 1. Schematic diagram representing recombinant collagens: VCL, with one copy of the triple-helix domain; VCLCL, with two copies of the triple-helix domain; VCL-X4-CL, with a 4-residue interruption inserted between triple-helix domains; and VCL-X15-CL, with a 15-residue interruption inserted between triple-helix domains. VCL(G-S)-X4-CL is a recombinant collagen with a Gly to Ser mutation three triplets N-terminal to the 4-residue interruption. The N-terminal globular V domain is indicated by the oval, while the black diamond represents the His tag. The black rectangle at the N-terminus of the CL domain represents the junction between V and CL modified to include a protease cleavage site, with the adjacent Gly-Xaa-Yaa triplets altered for convenience of DNA construction.

FIGURE 2. Circular dichroism (CD) and differential scanning calorimetry (DSC) studies of recombinant proteins (18 µM PBS pH7). A. Thermal denaturation curves obtained by monitoring the CD peak at 220 nm with a heating rate of 0.1 °C/min. Inset, CD spectra at 0 °C. Differences between the interruption proteins and the control are reproducible, but fall within experimental error +/- 0.5C. B. DSC curves demonstrating the unfolding peak at a heating rate of 1 °C/min. Red represents VCLCL; orange represents VCL-X4-CL; green represents VCL-X15-CL; and blue represents VCL(G-S)-X4-CL. MRE, mean residue ellipticity. Cp, heat capacity.

FIGURE 3. SDS-PAGE of recombinant collagens after treatment with trypsin or thermolysin. The lanes include digestion of VCLCLN (control in native state); VCLCLD (control in denatured state); VCL-X4-CL; VCL-X15-CL; and VCL(G-S)-X4-CL. VCL is also included for comparison. The VCL contaminant and smaller His-tagged triple helical contaminants are seen in all recombinant collagen samples after purification on two columns. The contaminants are likely due to truncated translation, a known phenomenon for bacterial expression of proteins with repetitive sequences (30). In samples treated with thermolysin, a band representing the protease can be seen above the CL band. Trypsin cannot be seen in any gel lane because of its low concentration.

FIGURE 4. Refolding of the V domain in VCL (solid line) and VCLCL (dotted line) measured by intrinsic tryptophan fluorescence at 340 nm following excitation at 280 nm, and refolding of the triple-helix in VCL (dashed line) and VCLCL (dash dot line) measured by circular dichroism at 220 nm (MRE220nm), after a temperature jump from 55 °C to 25 °C (18µM PBS pH7). The alpha helical V domain has a negative peak at 222 nm, so MRE220nm decreases as it refolds, while the MRE220nm increases with triple-helix folding. For CD refolding of VCL and VCLCL, an early rapid decrease in MRE220nm due to alpha-helical V domain refolding can be observed as an apparent negative fraction folded that is followed by an increase in MRE220nm due to triple-helix refolding.

FIGURE 5. Temperature jump refolding curves for collagens containing interruptions and mutations: VCLCL (red), VCL-X4-CL (orange), VCL-X15-CL (green), and VCL(G-S)-X4-CL (blue). A. Fluorescence (FL) refolding curves obtained by monitoring tryptophan emission at 340 nm (25 °C, 18 µM, PBS, pH7). B. Circular dichroism (CD) refolding curves obtained at 0 °C, monitoring mean residue
ellipticity at 220 nm, MRE$_{220}$ nm (18 µM, PBS, pH7). C. Circular dichroism refolding curves obtained at 25 °C (18 µM, PBS, pH7). For refolding of VCL and VCLCL, an early rapid decrease in MRE$_{220}$nm due to the alpha helical V domain refolding can be observed as an apparent negative fraction folded that is followed by an increase in MRE$_{220}$nm due to triple-helix refolding.

**TABLES**

**TABLE 1.** The thermal stability of the recombinant proteins as determined by circular dichroism (CD) and differential scanning calorimetry (DSC), together with values for half-time of refolding ($t_{1/2}$) monitored by fluorescence (FL) to follow V domain folding and CD to follow triple-helix folding. The fraction folded at the end of the experiment, 16 hours, (FF$_{16h}$) is also given for control collagen and collagens containing interruptions and mutations.

<table>
<thead>
<tr>
<th></th>
<th>T$_{m}$ (°C)</th>
<th>CD(DSC) 25°C (min)</th>
<th>FL $t_{1/2}$ 25°C (min)</th>
<th>CD $t_{1/2}$ 0°C (min)</th>
<th>FF$_{16h}$ 0°C (min)</th>
<th>CD $t_{1/2}$ 25°C (min)</th>
<th>FF$_{16h}$ 25°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCLCL</td>
<td>36.8(37.8)</td>
<td>2.1</td>
<td>17</td>
<td>0.88</td>
<td>135</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>VCL-X$_4$-CL</td>
<td>35.1(37.0)</td>
<td>2.0</td>
<td>37</td>
<td>0.81</td>
<td>447</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>VCL-X$_{15}$-CL</td>
<td>35.8(36.9)</td>
<td>1.9</td>
<td>41</td>
<td>0.81</td>
<td>&gt;1000</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>VCL(G-S)-X$_4$-CL</td>
<td>35.8(37.3)</td>
<td>1.5</td>
<td>41</td>
<td>0.80</td>
<td>&gt;1000</td>
<td>0.45</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1
Figure 2
Figure 3
Figure 4
Folding delay and structural perturbations caused by type IV collagen natural interruptions and nearby Gly missense mutations

Eileen Hwang and Barbara Brodsky

J. Biol. Chem. published online December 16, 2011

Access the most updated version of this article at doi: 10.1074/jbc.M111.269084

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2011/12/16/M111.269084.DC1