Glycobiology and Extracellular Matrices:

Vascular Ehlers-Danlos syndrome mutations in type III collagen differently stall the triple helical folding

GLYCOBIOLOGY AND EXTRACELLULAR MATRICES

PROTEIN STRUCTURE AND FOLDING

Kazunori Mizuno, Sergei Boudko, Juergen Engel and Hans Peter Bachinger

Access the most updated version of this article at doi: 10.1074/jbc.M113.462002

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

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/2013/05/03/M113.462002.DC1.html

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2013/05/03/jbc.M113.462002.full.html#ref-list-1

Vascular Ehlers-Danlos syndrome mutations in type III collagen differently stall the triple helical folding*

Kazunori Mizuno (水野一乗)[1], Sergei Boudko[1,2], Jürgen Engel[3], and Hans Peter Bächinger[1,2]

From [1]Shriners Hospitals for Children Portland Research Center, Portland, OR 97239, USA, [2]Department of Biochemistry and Molecular Biology, Oregon Health and Science University, Portland, OR 97239, USA, and [3]Biozentrum, Universität Basel, CH-4056 Basel, Switzerland

*Running title: *Folding of type III collagen with EDS mutations*

To whom correspondence should be addressed: Hans Peter Bächinger, Shriners Hospitals for Children Portland Research Center, 3101 SW Sam Jackson Park Rd., Portland, OR 97239, USA., Tel.: (503) 221-3433; Fax: (503) 221-3451; Email: hpb@shcc.org

Keywords: collagen; prolyl cis trans isomerization; protein folding; Ehlers-Danlos syndrome

Background: Mutations in collagen may lead to posttranslational overmodifications and a delay in secretion.
Results: Mutations locally delay the propagation of the triple helix.
Conclusion: Position and nature of a mutation are the key factors in the triple helix folding delay.
Significance: Consequences of natural collagen mutations can be now analyzed in terms of rate of folding.

SUMMARY

Vascular Ehlers-Danlos syndrome (EDS) type IV is the most severe form of EDS. In many cases, the disease is caused by a point mutation of Gly in type III collagen. A slower folding of the collagen helix is a potential cause for overmodifications. However, little is known about the rate of folding of type III collagen in patients with EDS. In order to understand the molecular mechanism of the effect of mutations a system was developed for bacterial production of homotrimeric model polypeptides. The C-terminal quarter, 252 residues, of the natural human type III collagen was attached to (GPP)$_7$ with the type XIX collagen trimerization domain (NC2). The natural collagen domain forms a triple helical structure without 4-hydroxylation of proline at low temperature. At 33 °C, the natural collagenous part is denatured, but the C-terminal (GPP)$_7$-NC2 remains intact. Switching to a low temperature triggers the folding of the type III collagen domain in a zipper-like fashion that resembles the natural process. We used this system for the two known EDS mutations (Gly-to-Val) in the middle at G910 and at the C-terminus at G1018. In addition, wild-type and Gly-to-Ala mutants were made. The mutations significantly slow down the overall rate of triple helix formation. The effect of the Gly-to-Val mutation is much more severe compared with Gly-to-Ala. This is the first report on the folding of collagen with EDS mutations, which demonstrates local delays in the triple helix propagation around the mutated residue.

_____

Ehlers-Danlos syndrome (EDS) comprises the wide variety of soft connective tissue disorders (1). Genetic studies have shown that the genes of types I, III, and V collagens (2), lysylhydroxylase 1 (coded by *PLOD1* gene) (3), procollagen-N-proteinase (coded by *ADAMTS2* gene) (4), FKBP22 (coded by *FKBP14* gene) (5), and tenascin X (coded by *TNXB* gene) (6) cause EDS. Lysylhydroxylase 3 (coded by *PLOD3* gene) causes a kind of EDS like phenotype (7). Vascular EDS, type IV, is the most serious form of the disorder due to the lethal rapture of the artery or hollow organs caused by autosomal dominant mutations in the *COL3A1* gene. The skin of the patients is usually thin and translucent; therefore, the veins can be clearly seen. The patients frequently have large eyes, lobeless ears, a small

chin, short stature and thin scalp hair. Extensive bruising is also frequently observed. Type III collagen is ubiquitously distributed as a minor component of collagen fibrils, but relatively abundant in skin, artery, and intestine. Bone (8) and cartilage (9) also contain type III collagen. The type III procollagen molecule secreted from cells is not fully processed at its amino-terminal noncollagenous part. ADAMTS-2 is the processing enzyme of the N-terminal propeptide, which is also linked to dermatosparaxis EDS (type 7C) (10,11). Immunoelectron microscopic analysis with a monoclonal antibody against the triple helical domain suggested that type III collagen can be present on banded collagen fibrils regardless of fibril diameter (12). The type III collagen molecules interact with the cartilage collagen fibril surface (9), and also with type I collagen fibrils in many tissues (13). Type III collagen has a slightly longer major triple helical domain (1,029 residues) compared to type I collagen (1,014 residues). Since the chain stagger of the type I collagen molecule and the molecular packing manner even of pure type I collagen fibril remain unclear (14), the molecular recognition of type III collagen with type I collagen and their role in maintaining the structure and tensile strength of the fibril are largely unexplored. The *col3a1* knock out mouse embryos develop phenotypically normal organs, indicating that type III collagen is not critically important for the early organ morphogenesis stage. At weaning age, the survival rate of *col3a1* null mouse was 5% mainly due to rupture of the aorta (15), indicating that type III collagen is an essential molecule for the maintenance of the mechanical strength of the extracellular matrix of the aorta and to withstand the high blood pressure produced by the heartbeat.

As of now, 276 unique DNA mutations in the *COL3A1* gene have been reported in the EDS variant database at the University of Leicester (16,17). About half of them are the point mutations of glycine residues of the G-X-Y tripeptide unit that repeatedly constitutes the triple helical sequence. About a third of the point mutation of repeated Gly residues have been reported to cause the vascular EDS. The phenotypes of patients are somewhat similar. In addition, haploinsufficiency of type III collagen in humans (18) and mice (19,20) also leads to EDS, suggesting that the amount of type III collagen is also important. Therefore, the amount and structural integrity of the type III collagen molecule in the extracellular matrix must be essential for the full function of the aorta and other organs. The mutations of the smallest residue to any other bulkier residue should obviously disturb the molecular structure (21,22). Furthermore, the resulting change of the three-dimensional molecular surface can disturb intermolecular interactions and also change the normal fibril packing and crosslinking patterns. Potentially, the mutation of Gly can also slow down the rate of triple helix folding. Delayed secretion of overglycosylated type I collagen has been previously demonstrated in cultured cells of the osteogenesis imperfecta (OI) patients (23). OI can be caused not only by mutations in the collagen genes but also by the absence or defects in the proteins involved in collagen synthesis and secretion. For example, the lack of prolyl 3-hydroxylase 1 (P3H1) causes the delayed secretion of overglycosylated type I collagen from the cultured skin fibroblasts (24). In one report of vascular EDS, a broadening of the electrophoresis gel band of type III collagen presumably indicated an overmodification as well as an increase of type III collagen inside of fibroblasts from the patient was observed (25).

It is reasonable to speculate that the slower folding of type III collagen molecules leads to a prolonged exposure of single (non-triple helical) procollagen chains to modifying enzymes. However, there are no studies that elucidate the effect of mutations on the kinetics of the triple helix folding.

We have recently found that certain noncollagenous domains of several collagen types, such as XVIII (26), XV (27) and XIX (28), form a stable trimerization domain (29). These trimerization domains are attractive tools for protein engineering (29) and are of special interest for studying collagen structure and folding (30,31). Here we recombinantly attached the C-terminal quarter fragment of human type III collagen to a $(GPP)_7$ stretch stabilized by the type XIX collagen NC2 domain. The natural collagenous domain forms the correctly staggered collagen triple helical structure. Importantly, the $(GPP)_7$-NC2XIX portion of this chimeric molecule is thermally more stable than the collagenous domain. This allows us to thermally unfold the collagenous

2

domain located N-terminally while keeping the C-termini in a correctly pre-staggered trimeric conformation (locked by the (GPP)$_7$-NC2XIX portion). By lowering the temperature, the kinetics of the triple helix propagation can be immediately monitored. This system allows the study of triple helix propagation when it is not hindered by other processes such as chain association and triple helix nucleation. In the type III collagen quarter fragment two EDS mutation sites located either in the middle or close to the C-terminus, G910V[1] (32) or G1018V (33), were chosen to uncover their aberrant role in the collagen folding kinetics. Interestingly, the probability of Gly-to-Ala point mutation is expected to be equal to Gly-to-Val substitution. Nevertheless, the number of reported EDS cases of Gly-to-Ala mutations is much lower than that of Gly-to-Val. Many of the Gly-to-Ala mutations likely do not exhibit EDS symptoms. We made both Val and Ala variants at sites G910 and G1018 sites to uncover the effects of the nature of the mutation and its position on the folding kinetics. For the first time we demonstrate a direct effect of EDS mutations on the rate of triple helical propagation.

EXPERIMENTAL PROCEDURES

Cloning, expression and purification of the C-terminal quarter fragment of human type III collagen with (GPP)$_7$ and a short NC2 domain of type XIX collagen- Human cDNA clone of COL3A1 was kindly gifted by Prof. Takako Sasaki (Oita University, Oita, Japan). The C-terminal quarter fragment, the residue 775-1026 of the major triple helical region was PCR amplified using two oligonucleotides 5'- TGC GGA TCC GGA GCT CCA GGC CCA CTT G -3' (forward primer, the *BamHI* site is underlined) and 5'- AGA GGG CCC AGG GGC ACC AGG AGG TCC A -3' (reverse primer, the *ApaI* site is underlined). The amplified DNA was inserted into the pET23-HisTrx_gpp7-XVIIItR vector (31) with a 6× His tag-thioredoxin-thrombin cleavage site - Gly-Ser- type XVIII collagen trimerization domain (26). The preliminary experiments revealed that an unexpected aggregation problem occurred (data not shown). Thus, we replaced the type XVIII collagen trimerization domain with the NC2 domain of human type XIX collagen. The 5'-*NcoI*-6x His -thioredoxin-thrombin cleavage site-

BamHI-type III collagen (775-1026) (wild-type, G910A/V, or G1018A/V)-(GPP)$_7$-*XmaI* fragment was recloned into pET23-HisTrx_gpp7-NC2XIXB (31) using the cloning sites *NcoI* and *XmaI*. As a result the sequence encoding the type XVIII trimerization domain was replaced with the human type XIX collagen NC2 domain containing the last COL2 domain tripeptide Gly-Ile-Pro (residues 1007 to 1040 of collagen type XIX human protein, accession number CAI42716 version CAI42716.1 GI:57208771). The exact protein sequence attached after (GPP)$_7$ repeat is GIPADAVSFEEIKKYINQEVLRIFEERMAVFL SQ. The final sequence of the fusion protein is shown in Figure 1. Plasmids encoding point mutations G910A, G910V, G1018A, and G1018V were prepared with the QuikChange® site-directed mutagenesis kit (Agilent). The oligonucleotides were designed at the company's website. For G910A mutation, 5'-CTGGTGCTCCCG<u>C</u>TCCTGCTGGTTC-3' and 5'- GAACCAGCAGGA<u>G</u>CGGGAGCACCAG-3' (mutated nucleotides are underlined) were used as the sense and antisense primers, respectively. For G910V mutation, 5'-GCTGGTGCTCCCG<u>T</u>TCCTGCTGGTTCC-3' and 5'-GGAACCAGCAGGA<u>A</u>CGGGAGCACCAGC -3' were used as the sense and antisense primers. For G1018A, 5'-CAGGGCAACCAGC<u>C</u>CCCTCCTGGACC-3' and 5'- GGTCCAGGAGGG<u>A</u>CTGGTTGCCCT G -3' were used as the sense and antisense primers, respectively. For G1018V, 5'-CAGGGCAACCAGT<u>C</u>CCTCCTGGACC -3' and 5'- GGTCCAGGAGGG<u>A</u>CTGGTTGCCCTG-3' were used as the sense and antisense primers, respectively. All DNA sequences were verified by Sanger dideoxy DNA sequencing.

The recombinant proteins were expressed at 15 °C in the *E.coli* BL21(DE3) host strain (Invitrogen) after isopropyl β-D-1-thiogalactopyranoside induction (final concentration 1 mM) for 3 days in two flasks with 500 ml 2xTY media for each. Cells were chilled and harvested by centrifugation, lysed by ultrasonication, and soluble material was precipitated with saturated solution of ammonium sulfate. Further, the expressed proteins were purified by HiTrap^{TM} chelating HP column (GE Biosciences) with immobilized nickel ions. The

3

His-thioredoxin tag fragment was cleaved by thrombin (BD BaculoGold[TM] BD Biosciences) carried out as described in the manufacturer's instruction. The cleavage was performed with thrombin protease 1 U/mL in 40 mM Tris/HCl buffer, pH 8.2, containing 150 mM NaCl on ice. The resulting fragments had two additional amino acid residues GS before the natural type III collagen sequence. After thrombin cleavage, the protein solutions were treated with 1 mM diisopropyl fluorophosphate for 3 h on ice. Then, the solutions were applied to HisTrap[TM] HP column again to remove the His-thioredoxin tag. The natural type III collagen fragments with the trimerization knot were eluted with approximately 10-20 mM imidazole. After the addition of 0.05% Tween 20, the protein solutions were dialyzed against 20 mM HEPES buffer, pH 7.4, containing 50 mM NaCl, 0.05% Tween 20, and purified with a 5 ml volume of SP Sepharose Fast Flow column (GE Life Sciences). The fractions were collected using a stepwise NaCl gradient. The fractions with the pure proteins (150-200 mM NaCl) were dialyzed and concentrated in 20 mM Tris/HCl buffer, pH 7.0, containing 200 mM NaCl and 0.05% Tween 20 for further analysis. Hereafter, the quarter fragment of human natural type III collagen with $(GPP)_7$ and a short NC2 domain of type XIX collagen protein is indicated as QF.

Rotary Shadowing — The purified proteins (0.1 mg/ml) were dialyzed against 100 mM ammonium bicarbonate at 4 °C, sprayed onto freshly cleaved mica, and shadowed with platinum carbon (34). The images were viewed by transmission electron microscopy. The length of the molecules from the images was analyzed using the software ImageJ (ver. 1.47c; Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2012).

Circular Dichroism (CD) Analysis — CD spectra were recorded on an AVIV model 202 spectrometer (AVIV Biomedical, Inc., Lakewood Township, NJ) with a thermostatted quartz cell of 1 mm path length (Starna cells, Inc. Atascadero, CA). The spectra were normalized for concentration and path length to obtain the mean molar residue ellipticity after subtraction of the buffer contribution. Thermal scanning curves were recorded at 223 nm. Peptide concentrations were determined by amino acid analysis.

Detection of Refolding Intermediates by Trypsin Resistance and Polyacrylamide Gel Electrophoresis — Detection of refolding intermediates by analysis of trypsin-resistant fragments on the SDS-PAGE was performed as described (35) with some modification. Twenty μl at 0.2 mg/ml or higher concentration of type III collagen quarter fragment stabilized by the $(GPP)_7$-NC2XIX knot was heat-treated in 0.65 ml thin-wall polypropylene tube at 33 °C for 10 min. The natural type III collagen sequence was unfolded under this condition but the $(GPP)_7$-NC2XIX part remained stable. The sample solutions were then immediately placed on ice. After time periods ranging from 0 to 6 h, 2 μl of 50 mM Tris/HCl buffer, pH 7.4, containing 0.2 M NaCl, 5 mg trypsin/ml was added at 4 °C to the sample. After a 2 min incubation on ice, 2 μl of 50 mM 4-(2-aminoethyl)benzensulfonyl fluoride hydrochloride in cold water was added to the tube. Then, 20% SDS were added and the mixture was brought to a boiling temperature by immersing it into an oil bath heated to 140 °C. The samples boiled for 5 s after the addition of the SDS buffer and were placed into a heat block kept at 96 °C for another 3 min. The samples were analyzed by the Bis-Glycine buffered SDS-PAGE. Photos of the gels taken after the stain with GelCode Blue (Thermo Scientific) were analyzed quantitatively using the software GelQuant.NET provided by biochemlabsolutions.com. The band intensities were integrated and normalized with respect to the undigested sample. The N-terminal sequences of the trypsin-cleaved fragments were analyzed after blotting the SDS-PAGE gel onto a Immobilon[TM] P membrane (Millipore) with a BioRad Mini Trans-Blot Cell® with 40 V for 2.5 h in 50 mM sodium tetraborate. The blot was stained with GelCode Blue. The bands were analyzed on a Procise Protein Sequencer (Applied Biosystems, Inc.).

RESULTS

Rotary shadowing images of the expressed protein — The QF protein (human type III collagen C-terminal quarter fragment with $(GPP)_7$ and the type XIX collagen NC2 domain, Fig. 1) was expressed in BL21(DE3) cells and purified as described in Materials and Methods. Rotary

4

shadowing images showed the rope-like structures with neither significant kinks along the molecules nor globular domains at either end (Fig. 2). From the CD spectrum (28), the secondary structure of the NC2 domain of type XIX collagen is estimated to be mainly α-helical. If all the thirty one residue of the NC2 domain form the α-helical coiled coil, the corresponding length of the domain is estimated to be roughly 5 nm (based on the rise 0.15 nm per residue). The thickness of the trimeric α-helical coiled coil structure is about 2.3 nm as determined by the crystal structure of the trimeric coiled coil leucine zipper GCN4 from yeast (36). On the other hand, the diameter of collagen helix is about 1.4 nm from type III collagen peptide crystal structure (37). The resolution of the rotary shadowing images is about a few nanometers (38). Therefore, the absence of the globular object at the end of the molecules is in accordance with the extended coiled coil structure of the type XIX collagen NC2 domain. The average length of the observed molecules is 73.8 nm (SD 6.9 nm). For the collagen triple helix in hydrated tissues or crystals the residue pitch along the molecular axis is 0.285 nm (39,40). Thus, the expected length of the triple helical domain is about 78 nm including the $(GPP)_7$ segment. The observed average value of the shadowing image is a little smaller than expected (82 nm for the collagen triple helix extended with the coiled-coil). This can possibly be caused by the shrinkage during the sample preparation (dehydration), or a partial unfolding of the N-terminal triple helical domain (breathing). As determined by trypsin digestion (presented below) the amino-terminal part of the collagen sequence is somewhat unstable. As previously reported the unfolded single chain collagenous polypeptide cannot been observed with this technique (41).

Thermal stability of bacterially expressed type III collagen C-terminal quarter fragment with $(GPP)_7$-NC2XIX knot — The QF constructs were purposely designed to keep the C-termini of the collagenous domain of interest in a trimeric pre-staggered conformation even at higher temperature when the collagenous domain is unfolded. This is achieved through the stable $(GPP)_7$-NC2XIX structure at the C-terminus. The melting temperature of the $(GPP)_7$ portion is around 50 °C (31). All temperature experiments were performed below 35 °C to ensure the intact structure of the trimerization and stagger-determining knot.

Figure 3 shows the equilibrium unfolding/refolding experiments with the wild-type (WT) and mutated QF proteins monitored by CD. The heating and cooling rates are 10 K/h. Upon heating the protein samples demonstrate a linear baseline at the lower temperature intervals, which correspond to the triple helical structure. Two transition phases are observed for WT, which indicates the presence of two cooperative domains. The first transition around 20 °C makes up almost 75% of the net signal change and must be assigned to the unfolding of the major N-terminal part of the QF; the second transition around 25 °C points to a more stable C-terminal end of the QF (Fig. 3A). The thermal transitions observed for the mutated QF proteins further supports these assignments. Indeed, the QF proteins with G910A and G910V (mutations in the middle of the triple helix) demonstrate a decrease in the melting temperature only for the first major transition (Fig. 3A). Mutants G1018A and G1018V (located close to the C-terminus) show a destabilization of the second domain, which apparently leads to a shift of the melting temperature to around the position of the first transition (Fig. 3B). The results indicate that the thermal stability of the mutated proteins is globally affected within two separate cooperative domains. Since the proteins were expressed in *E. coli*, all proline residues in the Y position are not posttranslationally modified to 4-Hyp, resulting in a lower thermal stability of the collagenous domain.

Upon cooling the refolding profiles shows a hysteresis for all the QF, a known phenomenon for collagen unfolding-refolding experiments (30). In general, the WT QF has a propensity to refold faster than any mutated QF. Interestingly, the QF proteins with G910A and G910V demonstrate similar refolding profiles in Figure 3A around the second transition (with the higher temperature). It additionally supports the above assignment of the second transition to the C-terminal end, as it should not be affected by the distant mutations. It is also remarkable that the refolding of the mutants with the bulkier Val is significantly diminished after a certain point when compared to the Ala substitution.

The N-terminal sequences of trypsin-resistant

5

fragments of QF — The thermal transition profiles determined by CD showed that the stability of the collagen triple helix of QF is not homogenous. To search for the potential unstable regions we treated the QF with trypsin under mild conditions. Figure 4 shows two examples of blots generated from the transfer of SDS-PAGE gels.. Nine trypsin fragments were observed at different conditions. Even at 4 °C, the amino terminal end of the WT QF was partially cleaved. Apparently, the G775 to R789 stretch does not completely form a trypsin-resistant stable collagen helix, which reflects the observation of shorter fragments in the rotary shadowing experiments.

Refolding kinetics analysis of QF by CD signal — Refolding kinetics of the natural collagenous domain was monitored by the ellipticity change at 223 nm (Fig. 5). As previously reported (35), collagen molecules with a long triple helical domain (above ~250 residues) linked by a trimerization knot at one end demonstrate a zero order refolding (linear gain of refolded structure) at the beginning of refolding. The rate limiting factor in this system is the slow prolyl *cis/trans* isomerization (35,42). The initial apparent first order rates, k_{app}, of WT, G1018A, and G1018V peptides were calculated from the linear fit of the first 500 sec of the refolding data. Table 1 shows the k_{app} of the QF peptides. We compared our bacterial expressed protein data with the tissue-extracted type III collagen quarter fragment. For bovine type III quarter fragment, the k_{app} value was reported to be 0.115 min^{-1} (or 1.9×10^{-3} sec^{-1}) at 25 °C (35). Applying the activation energy, Ea, of *cis/trans* isomerization of 53.5 kJ/mol published for the knotted (Gly-Pro-Pro)$_{10}$ molecules (43), k_{app} at 4 °C can be calculated as 3.7×10^{-4} sec^{-1} if we assume the relative rate constants for the same reaction at two different temperatures are calculated with Arrhenius equation,

$$\ln\left(\frac{k_2}{k_1}\right) = \frac{-Ea}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right)$$

where R is gas constant. Although the natural type III collagen quarter fragment contains the posttranslationally modified 4-hydroxyprolines, its k_{app} value appears basically the same as that for non-hydroxylated molecules (Table 1).

As described in detail by Bächinger et al. (35) the initial apparent rate constant, k_{app}, for the long collagen peptide is related to the number of tripeptide units, n, and the equilibrium constant of [trans]/[cis], K in the following way:

$$\left(\frac{dF}{dt}\right)_{t=0} = k_{app} = k\frac{1+K}{n}$$

where F is the fraction folded, and k is the rate of *cis* to *trans* prolyl-bond isomerization. In 1979, the whole amino acid sequence of bovine type III collagen had not been reported yet. The QF fragment used in this study has the human sequence, which is known. The number of the tripeptide units per chain is 84. Then, the total number of the tripeptide units incorporated into the triple helix is $n = 3 \times 84-2 = 250$. In the QF, the number of proline residues per trimer is $57 \times 3 = 171$. Replacing n with to the number of prolyl bonds seems to be more reasonable, although the distribution of P is not homogeneous.

If we assume that $K = 6.25$ (as has been determined for acetyl-Pro-O-methyl (42)), k for WT is 9.4×10^{-3} sec^{-1}, (calculated using $k_{app} = 4.0 \times 10^{-4}$ sec^{-1} at 4 °C). From the Arrhenius equation, k at 25 °C can be calculated as 4.8×10^{-2} sec^{-1}. For the natural bovine type III collagen quarter fragment used by Bächinger et al. (35), the k value can be calculated as 4.7×10^{-2} sec^{-1} at 25 °C, if we use $n = 153$ (Pro819 was determined as the first N-terminal residue of the bovine type III collagen quarter fragment[2]), 51 proline residues per chain, 153 proline residues per trimer, and K is 6.25. Again, the rates of *cis* to *trans* are essentially similar whether the peptides have 4-Hyp or not. If we used $K = 30$ as (35), the k values would decrease by 23% (1+6.25)/(1+30). These values of k values are still within the range of those previously reported (44,45).

The initial apparent rates of folding of WT, G910A, and G910V are faster than those of the C-terminal mutated peptides, G1018A and G1018V (Fig. 5B). The QF with mutations at G1018 demonstrated slower refolding from the very beginning, whereas the G910 mutants refolded very similar during the initial phase when compared with the WT QF. The delay in refolding is more significant in the G-to-V mutations than in the G-to-A mutations. It is clear that the limiting factor for the apparent rate constants of the peptides G1018A and G1018V is a renucleation of the triple helix behind the mutated G1018. The ratios of k_{app} to that of WT, 0.19 for G1018V, and,

6

0.43 for G1018A (Table 1) demonstrate the link between the nature of mutation and the decrease in the renucleation rate. The Gly-to-Val mutation obviously has a more severe effect and slows down the apparent rate to ~20% of that of WT.

Analysis of Refolding kinetics following trypsin-resistant fragments of QF — The refolding process was monitored by analyzing the appearance/disappearance of trypsin-resistant fragments of varying sizes as previously described (35) (Fig. 6). The type III collagen part was denatured at 33 °C for 10 min. This is possible due to the lack of 4-hydroxyproline, the transition temperature of the natural sequence is around 15 °C (Fig. 3). To minimize the protease treatment time we used trypsin at a high concentration. We found that 0.2 mg/ml trypsin is a suitable concentration to chop off the unfolded chains within 2 min on ice. Due to the high concentration of trypsin and/or the partial instability of the N-terminal triple helix, even without the heat treatment, a part of the natural sequence was cleaved off with trypsin.

At the beginning of the refolding peptide with an apparent molecular mass of 10K was observed for WT on the SDS-PAGE (Fig. 6A). This band corresponds to the cleavage after R1005 as determined by the N-terminal peptide sequencing. This is the most C-terminal potential trypsin cleavage site in the natural collagen part. The calculated mass of this fragment is only 7.6 kDa, but the slower mobility of collagenous peptides is a well-known fact. A band corresponding to a slightly shorter fragment with an apparent molecular mass of 6 kDa was also observed. This small fragment had the N-terminal sequence of the cleaved NC2 domain, which contains four potential sites for trypsin digestion (Fig. 1). The appearance and disappearance of ever larger trypsin resistant fragments in the WT QF is indicative of the zipper-like folding of the triple helix (Fig. 6B). A similar initial (covering only the QF part) pattern was observed for the refolding of the tissue-extracted type III collagen molecule (35).

The overall refolding profiles correspond to the result observed during the CD refolding experiments. The G-to-A mutated peptides refold faster than the G-to-V mutated peptides. Also, the initial refolding rate is faster for the G910 mutated peptides than the G1018 mutated peptides.

There are differences in the band patterns of the mutated proteins reflecting delays at the point of mutation. The G910 mutated peptides have less larger-sized fragments, whereas the G1018 mutated peptides do not show much of the 12 kDa bands. The appearance of band e of all mutants and WT are shown in Figure 6C. It is evident that the nature of the mutation is more important than the location of the mutation for the appearance of fully refolded molecules. These curves differ from the CD measurements (Fig. 5) at earlier folding times which are dominated by partially folded molecules, but converge at later refolding times.

Fitting of the folding kinetics — We attempted to fit the data shown in Figure 5 with the model of linear propagation steps reflecting all *cis*-to-*trans* isomerization events as described (35). However, the calculations required unrealistic computer time when we added a single mutation interruption in the model. Therefore we applied a simpler model. We simplify the refolding kinetics of the QF as schematically drawn in Figure 7. We assume only three different triple helical species, zero, half, and full length. All other intermediate species are ignored to simplify the equations. We assume that all reactions are first order with one direction. The rate for the reaction for propagation of the helix without mutations is defined as k_n. If there is a mutated site between the different lengths of the species, we introduce the rate for that reaction, k_m, which is defined as a renucleation rate without propagation of the triple helix.

For the C-terminal mutated molecular species, we assume four molecular species, *Y*, the quarter fragment is completely unfolded without renucleation at G1018 (*F* (fraction folded)=0), Y1, renucleated at G1018 (*F*=0), *Z1*, a half folded (*F*=0.5), and *W*, fully folded (*F*=1). With the rate constant, *km*, which is assumed to be the rate of the renucleation at the N-terminal side of G1018 to propagate the helix, the reaction can be expressed as follows:

$$[Y] \xrightarrow{k_m} [Y1] \xrightarrow{k_n} [Z1] \xrightarrow{k_n} [W]$$
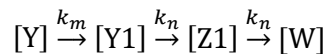
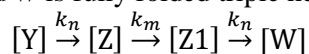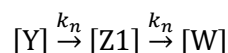The differential equations are shown below:

$$\frac{d[Y]}{dt} = -k_m[Y]$$

$$\frac{d[Y1]}{dt} = k_m[Y] - k_n[Y1]$$

7

$$\frac{d[Z1]}{dt} = k_n[Y1] - k_n[Z1]$$

$$\frac{d[W]}{dt} = k_n[Z1]$$

For the mid-mutated molecular species, G910A and G910V, the equation is changed as bellow. Here, *Y* is the completely unfolded (*F*=0), *Z* and *Z1* are the half-length folded triple helices (*F*=0.5), and *W* is fully folded triple helix (*F*=1).

$$[Y] \xrightarrow{k_n} [Z] \xrightarrow{k_m} [Z1] \xrightarrow{k_n} [W]$$

The differential equations are shown below:

$$\frac{d[Y]}{dt} = -k_n[Y]$$

$$\frac{d[Z]}{dt} = k_n[Y] - k_m[Z]$$

$$\frac{d[Z1]}{dt} = k_m[Z] - k_n[Z1]$$

$$\frac{d[W]}{dt} = k_n[Z1]$$

For WT, the reactions are expressed as simply as below:

$$[Y] \xrightarrow{k_n} [Z1] \xrightarrow{k_n} [W]$$

The differential equations are shown below:

$$\frac{d[Y]}{dt} = -k_n[Y]$$

$$\frac{d[Z1]}{dt} = k_n[Y] - k_n[Z1]$$

$$\frac{d[W]}{dt} = k_n[Z1]$$

With the initial conditions, *Z*(time=0)=0, *Z1*(0)=0, *W*(0)=0, *Y*1(0)=0, *Y*(0)=1, these differential equations can be solved. The integral forms of the solutions are available in the Supplement A. Sets of differential equations for all three models were iteratively solved by the MicroMath Scientist algorithm for Windows (version 2.01; MicroMath, St. Louis, MO). The equations are also available in the Supplement B.

First, we introduced one k_m for the Ala and Val mutated peptide for each as parameters for the fitting (Fig. 7C). Table 2 shows the fitted results using the initial 7,500 sec and also up to 24,000 sec refolding. The k_m for G-to-A mutants is 0.63-0.90 while the k_m for G-to-V is 0.13-0.22. Second, since the renucleation of the mutated sites must depend on the surrounding amino acid sequences, we also introduced a different k_m for the G910 and

G1018 peptides (Fig. 7D). Although the model is highly simplified, the results are reasonable in that the effect of the G-to-V mutation is much more severe than the G-to-A mutation. However, the simplified model ignores the CD contribution of partially formed triple helical intermediates, because only zero, half, or full length triple helices are considered. As a result the model underestimates the rate of folding at the beginning of the refolding kinetics for the G1018 mutants (Fig. 7C and D).

DISCUSSION

We designed a system, which is able to align three polypeptide chains in a specifically staggered manner without the need to nucleate the triple helix. We believe that this is the most advanced system to study the role of mutations in the triple helical folding available to date. An interesting system has been recently developed, which is based on a bacterial collagen (46,47). Although this system provided useful insights into the triple helix folding it has several limitations. First, the triple helical nucleation starts at the N-terminal trimerization domain, which is opposite to most collagen types where the trimerization domain is located C-terminally and the triple helix folding proceeds from the C- to the N-terminus in a zipper-like fashion (35,48,49). Second, the stability of the trimerization domain is insufficient to keep three unfolded chains aligned at the beginning of the experiment, which complicates the refolding mechanism by introducing two additional steps, *i.e.* the trimerization and triple helix nucleation. And third, a collagen site of interest is placed in the alien context of a bacterial sequence, which can significantly disturb any cooperative intra- and inter-chain interactions. Our system is void of these complications. The only drawback of the system reported in this study is the lack of posttranslational modifications, such as hydroxylation of proline and lysine residues as well as glycosylation. Our experimental data indicate that at least the effect of 4-hydroxylation is essentially small on the refolding kinetics. If necessary, this bacterial system could be translated into the appropriate eukaryotic expression system with the desired posttranslational modifications.

We analyzed the triple helix folding of the type III collagen fragment using the model

8

peptides with vascular EDS mutations, G910V (32) and G1018V (33) as well as G910A and G1018A. Our study indicates that the mutations of Gly cause a delay in the triple helical folding. The smaller mutated residue Ala can be easily skipped and the triple helix can be renucleated faster after the mutation site (Table 1). The polypeptides with alanine mutations can thus extend the triple helical region faster than those with a larger valine residue at the mutation sites. Since the delay is more significant in G-to-V mutations than in G-to-A, it can somewhat simulate the autosomal dominant phenotype with only one or two mutated chains in the trimeric molecule. Either one, two, or three chains with a mutation can be associated with their trimerizing C-propeptide in the rough endoplasmic reticulum (ER). Thus the rate of renucleation will vary among these different molecules. Our model describes the worst case scenario with three mutated chains. The substitution of a single nucleotide base for Gly codon may cause nine possible mutations (1/4 for Arg, 1/6 for Ala/Val, 1/12 for Asp/Glu/Ser/Cys, 1/24 for Trp, and 1/24 for Stop codon). If the frequency of the occurrence of mutations is all equal, the ratio of G-to-A mutations cases should be higher than the currently reported in the database (16,17). Presumably some Ala mutations might not be severe enough to cause EDS.

The delayed triple helix formation potentially might cause overglycosylation of hydroxylysine residues as seen in osteogenesis imperfecta (23). The effect of the number of mutated chains as well as staggering positions is interesting factors that should be investigated. Until recently, there were no easy methods to make heterotrimeric collagen model proteins. Now we have found that three chains of the type IX collagen NC2 domain form the $\alpha1(IX)\alpha2(IX)\alpha3(IX)$ heterotrimer (50), which also determines the stagger of the adjacent collagenous sequences (31). This system is suitable for exploring the composition and staggering effects on the structure, stability and folding of various collagen types with any sort of mutations including EDS.

The mechanism of the biosynthesis of collagen molecules is still far from being completely understood. A number of processes, such as transcription, mRNA splicing, translation, transport to ER, trimerization of the propeptides, N-glycosylation of the propeptide, posttranslational modifications of Pro and Lys,, O-glycosylation, ,peptidyl *cis* to *trans* isomerization, triple helix folding, secretion and maturation is orchestrated by a series of enzyme complexes with many chaperone proteins as well as carbohydrate and lipid molecules in the ER lumen and on the ER membrane. The defect of any of these players affects the whole sound of the symphony. Dilation of the rough ER is reported in G910V EDS patient dermal fibroblasts (33). Slower folding rate increases the potential chance of overglycosylation of hydroxylysine residues, and the degradation system might be stimulated. Interestingly, the vascular EDS mutations are found not only C-terminally but also N-terminally of the collagen domain, *i.e.* at Gly13, Gly16, Gly25, Gly34, Gly37, Gly43, Gly49, Gly52, Gly55, Gly58, Gly61, Gly70, Gly73, Gly76, and Gly82 (16,17). Lys84 is the most N-terminal Yaa position Lys in the major Gly-Xaa-Yaa- repeat of human type III collagen. Thus, there are no possibilities for the overmodification of Lys if we assume that the delay of the triple helical formation is only in the N-terminal part starting from the mutation site. We don't have any information about the supramolecular association of type III collagen in tissues. It is possible that the N-terminus of the triple helical chain might have an important role to associate with other collagen molecules or with some chaperones. The occurrence of vascular EDS at the N-terminus of the collagenous domain suggests that the overmodification of the lysine is not the critical factor for the EDS disorder in some cases. Future basic biochemical studies are necessary to understand the molecular mechanism of EDS.

Our study has clearly demonstrated that the triple helix propagation of type III collagen with the EDS mutation is locally delayed around the mutated residue. The use of the thermally stable staggered trimerization domain $(GPP)_7$-NC2XIX has the potential to advance the understanding of molecular mechanisms of the EDS and other collagen related heritable disorders .

9

REFERENCES

1.   De Paepe, A., and Malfait, F. (2012) The Ehlers-Danlos syndrome, a disorder with many faces. *Clin Genet* **82**, 1-11
2.   Malfait, F., and De Paepe, A. (2005) Molecular genetics in classic Ehlers-Danlos syndrome. *Am J Med Genet C Semin Med Genet* **139C**, 17-23
3.   Pinnell, S. R., Krane, S. M., Kenzora, J. E., and Glimcher, M. J. (1972) A heritable disorder of connective tissue. Hydroxylysine-deficient collagen disease. *N Engl J Med* **286**, 1013-1020
4.   Colige, A., Nuytinck, L., Hausser, I., van Essen, A. J., Thiry, M., Herens, C., Ades, L. C., Malfait, F., Paepe, A. D., Franck, P., Wolff, G., Oosterwijk, J. C., Smitt, J. H., Lapiere, C. M., and Nusgens, B. V. (2004) Novel types of mutation responsible for the dermatosparactic type of Ehlers-Danlos syndrome (Type VIIC) and common polymorphisms in the ADAMTS2 gene. *J Invest Dermatol* **123**, 656-663
5.   Baumann, M., Giunta, C., Krabichler, B., Ruschendorf, F., Zoppi, N., Colombi, M., Bittner, R. E., Quijano-Roy, S., Muntoni, F., Cirak, S., Schreiber, G., Zou, Y., Hu, Y., Romero, N. B., Carlier, R. Y., Amberger, A., Deutschmann, A., Straub, V., Rohrbach, M., Steinmann, B., Rostasy, K., Karall, D., Bonnemann, C. G., Zschocke, J., and Fauth, C. (2012) Mutations in FKBP14 cause a variant of Ehlers-Danlos syndrome with progressive kyphoscoliosis, myopathy, and hearing loss. *Am J Hum Genet* **90**, 201-216
6.   Burch, G. H., Gong, Y., Liu, W., Dettman, R. W., Curry, C. J., Smith, L., Miller, W. L., and Bristow, J. (1997) Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nat Genet* **17**, 104-108
7.   Salo, A. M., Cox, H., Farndon, P., Moss, C., Grindulis, H., Risteli, M., Robins, S. P., and Myllyla, R. (2008) A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. *Am J Hum Genet* **83**, 495-503
8.   Keene, D. R., Sakai, L. Y., and Burgeson, R. E. (1991) Human bone contains type III collagen, type VI collagen, and fibrillin: type III collagen is present on specific fibers that may mediate attachment of tendons, ligaments, and periosteum to calcified bone cortex. *J Histochem Cytochem* **39**, 59-69
9.   Wu, J. J., Weis, M. A., Kim, L. S., and Eyre, D. R. (2010) Type III collagen, a fibril network modifier in articular cartilage. *J Biol Chem* **285**, 18537-18544
10.   Colige, A., Ruggiero, F., Vandenberghe, I., Dubail, J., Kesteloot, F., Van Beeumen, J., Beschin, A., Brys, L., Lapiere, C. M., and Nusgens, B. (2005) Domains and maturation processes that regulate the activity of ADAMTS-2, a metalloproteinase cleaving the aminopropeptide of fibrillar procollagens types I-III and V. *J Biol Chem* **280**, 34397-34408
11.   Wang, W. M., Lee, S., Steiglitz, B. M., Scott, I. C., Lebares, C. C., Allen, M. L., Brenner, M. C., Takahara, K., and Greenspan, D. S. (2003) Transforming growth factor-beta induces secretion of activated ADAMTS-2. A procollagen III N-proteinase. *J Biol Chem* **278**, 19549-19557
12.   Keene, D. R., Sakai, L. Y., Bächinger, H. P., and Burgeson, R. E. (1987) Type III collagen can be present on banded collagen fibrils regardless of fibril diameter. *J Cell Biol* **105**, 2393-2402
13.   Fleischmajer, R., MacDonald, E. D., Perlish, J. S., Burgeson, R. E., and Fisher, L. W. (1990) Dermal collagen fibrils are hybrids of type I and type III collagen molecules. *J*

Struct Biol **105**, 162-169

14. Okuyama, K., Bächinger, H. P., Mizuno, K., Boudko, S., Engel, J., Berisio, R., and Vitagliano, L. (2009) Re: Microfibrillar structure of type I collagen in situ. *Acta Crystallogr D Biol Crystallogr* **65**, 1007-1008; author reply 1009-1010

15. Liu, X., Wu, H., Byrne, M., Krane, S., and Jaenisch, R. (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proc Natl Acad Sci U S A* **94**, 1852-1856

16. Dalgleish, R. (1997) The human type I collagen mutation database. *Nucleic Acids Res* **25**, 181-187

17. Dalgleish, R. (1998) The Human Collagen Mutation Database 1998. *Nucleic Acids Res* **26**, 253-255

18. Schwarze, U., Schievink, W. I., Petty, E., Jaff, M. R., Babovic-Vuksanovic, D., Cherry, K. J., Pepin, M., and Byers, P. H. (2001) Haploinsufficiency for one COL3A1 allele of type III procollagen results in a phenotype similar to the vascular form of Ehlers-Danlos syndrome, Ehlers-Danlos syndrome type IV. *Am J Hum Genet* **69**, 989-1001

19. Cooper, T. K., Zhong, Q., Krawczyk, M., Tae, H. J., Muller, G. A., Schubert, R., Myers, L. A., Dietz, H. C., Talan, M. I., and Briest, W. (2010) The haploinsufficient Col3a1 mouse as a model for vascular Ehlers-Danlos syndrome. *Vet Pathol* **47**, 1028-1039

20. Smith, L. B., Hadoke, P. W., Dyer, E., Denvir, M. A., Brownstein, D., Miller, E., Nelson, N., Wells, S., Cheeseman, M., and Greenfield, A. (2011) Haploinsufficiency of the murine Col3a1 locus causes aortic dissection: a novel model of the vascular type of Ehlers-Danlos syndrome. *Cardiovasc Res* **90**, 182-190

21. Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **266**, 75-81

22. Brodsky, B., Thiagarajan, G., Madhan, B., and Kar, K. (2008) Triple-helical peptides: an approach to collagen conformation, stability, and self-association. *Biopolymers* **89**, 345-353

23. Bonadio, J., and Byers, P. H. (1985) Subtle structural alterations in the chains of type I procollagen produce osteogenesis imperfecta type II. *Nature* **316**, 363-366

24. Vranka, J. A., Pokidysheva, E., Hayashi, L., Zientek, K., Mizuno, K., Ishikawa, Y., Maddox, K., Tufa, S., Keene, D. R., Klein, R., and Bächinger, H. P. (2010) Prolyl 3-hydroxylase 1 null mice display abnormalities in fibrillar collagen-rich tissues such as tendons, skin, and bones. *J Biol Chem* **285**, 17253-17262

25. Steinmann, B., Royce, P. M., and Superti-Furga, A. (2002) The Ehlers-Danlos syndrome. in *Connective tissue and its heritable disorders: Molecular, genetic and medical aspects* (Royce, P. M., and Steinmann, B. eds.), 2nd Ed., Wiley-Liss, Inc. pp 431-523

26. Boudko, S. P., Sasaki, T., Engel, J., Lerch, T. F., Nix, J., Chapman, M. S., and Bächinger, H. P. (2009) Crystal structure of human collagen XVIII trimerization domain: A novel collagen trimerization Fold. *J Mol Biol* **392**, 787-802

27. Wirz, J. A., Boudko, S. P., Lerch, T. F., Chapman, M. S., and Bächinger, H. P. (2011) Crystal structure of the human collagen XV trimerization domain: a potent trimerizing unit common to multiplexin collagens. *Matrix Biol* **30**, 9-15

28. Boudko, S. P., Engel, J., and Bächinger, H. P. (2008) Trimerization and triple helix stabilization of the collagen XIX NC2 domain. *J Biol Chem* **283**, 34345-34351

29. Boudko, S. P., Engel, J., and Bächinger, H. P. (2012) The crucial role of trimerization domains in collagen folding. *Int J Biochem Cell Biol* **44**, 21-32

11

30. Mizuno, K., Boudko, S. P., Engel, J., and Bächinger, H. P. (2010) Kinetic hysteresis in collagen folding. *Biophys J* **98**, 3004-3014

31. Boudko, S. P., and Bächinger, H. P. (2012) The NC2 domain of type IX collagen determines the chain register of the triple helix. *J Biol Chem* **287**, 44536-44545

32. Richards, A. J., Lloyd, J. C., Ward, P. N., De Paepe, A., Narcisi, P., and Pope, F. M. (1991) Characterisation of a glycine to valine substitution at amino acid position 910 of the triple helical region of type III collagen in a patient with Ehlers-Danlos syndrome type IV. *J Med Genet* **28**, 458-463

33. Smith, L. T., Schwarze, U., Goldstein, J., and Byers, P. H. (1997) Mutations in the COL3A1 gene result in the Ehlers-Danlos syndrome type IV and alterations in the size and distribution of the major collagen fibrils of the dermis. *J Invest Dermatol* **108**, 241-247

34. Morris, N. P., Keene, D. R., Glanville, R. W., Bentz, H., and Burgeson, R. E. (1986) The tissue form of type VII collagen is an antiparallel dimer. *J Biol Chem* **261**, 5638-5644

35. Bächinger, H. P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization. *Eur J Biochem* **106**, 619-632

36. Harbury, P. B., Kim, P. S., and Alber, T. (1994) Crystal structure of an isoleucine-zipper trimer. *Nature* **371**, 80-83

37. Boudko, S. P., Engel, J., Okuyama, K., Mizuno, K., Bächinger, H. P., and Schumacher, M. A. (2008) Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *J Biol Chem* **283**, 32580-32589

38. Aebi, U., and Baschong, W. (2006) Glycrol spraying/Low-angle rotary shadowing. in *Cell Biology* (Celis, J. E. ed.), 3rd Ed., Academic Press. pp 241-246

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

40. Okuyama, K., Xu, X., Iguchi, M., and Noguchi, K. (2006) Revision of collagen molecular structure. *Biopolymers* **84**, 181-191

41. Dölz, R., Engel, J., and Kühn, K. (1988) Folding of collagen IV. *Eur J Biochem* **178**, 357-366

42. Bächinger, H. P., and Engel, J. (2005) The thermodynamics and kinetics of collagen folding. in *Protein Folding Handbook* (Buchner, J., and Kiefhaber, T. eds.), Wiley, New York, USA. pp 1059-1110

43. Boudko, S., Frank, S., Kammerer, R. A., Stetefeld, J., Schulthess, T., Landwehr, R., Lustig, A., Bächinger, H. P., and Engel, J. (2002) Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. *J Mol Biol* **317**, 459-470

44. Schmid, F. X. (2001) Prolyl isomerases. *Adv Protein Chem* **59**, 243-282

45. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J Mol Biol* **279**, 449-460

46. Yoshizumi, A., Fletcher, J. M., Yu, Z., Persikov, A. V., Bartlett, G. J., Boyle, A. L., Vincent, T. L., Woolfson, D. N., and Brodsky, B. (2011) Designed coiled coils promote folding of a recombinant bacterial collagen. *J Biol Chem* **286**, 17512-17520

47. Yu, Z., Mirochnitchenko, O., Xu, C., Yoshizumi, A., Brodsky, B., and Inouye, M. (2010) Noncollagenous region of the streptococcal collagen-like protein is a trimerization

domain that supports refolding of adjacent homologous and heterologous collagenous domains. *Protein Sci* **19**, 775-785

48. Bächinger, H. P., Morris, N. P., and Davis, J. M. (1993) Thermal stability and folding of the collagen triple helix and the effects of mutations in osteogenesis imperfecta on the triple helix of type I collagen. *Am J Med Genet* **45**, 152-162

49. Raghunath, M., Bruckner, P., and Steinmann, B. (1994) Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. *J Mol Biol* **236**, 940-949

50. Boudko, S. P., Zientek, K. D., Vance, J., Hacker, J. L., Engel, J., and Bächinger, H. P. (2010) The NC2 domain of collagen IX provides chain selection and heterotrimerization. *J Biol Chem* **285**, 23721-23731

FOOTNOTES

[1] In this study, we number the type III collagen sequence starting from the first residue of the major triple helical domain. Thus the residue numbers are 167 smaller than in the translated molecule.

[2] Personal communication by Dr. Yoshihiro Ishikawa at Shriners Hospitals for Children, Portland Research Center

Table 1 The apparent initial first order rates of folding

	rate (sec^{-1}) (4 °C)	ratio to WT
WT	$4.0 \pm 0.07 \times 10^{-4}$	1
G1018A	$1.7 \pm 0.03 \times 10^{-4}$	0.43
G1018V	$7.6 \pm 0.03 \times 10^{-5}$	0.19
bovine skin QF*	3.7×10^{-4}	

* calculated using the data from Bächinger et al. (35)

Table 2 The apparent rates of folding calculated using the model

	rate (sec^{-1})	k_m/k_n
First 7,500 sec		
k_n	8.1×10^{-4}	1
k_m G-to-A	7.3×10^{-4}	0.90
k_m G-to-V	1.8×10^{-4}	0.22
24,000 sec		
k_n	9.2×10^{-4}	1
k_m G-to-A	5.8×10^{-4}	0.63
k_m G-to-V	1.2×10^{-4}	0.13
four k_m		
k_n	9.4×10^{-4}	1
k_m G910A	5.6×10^{-4}	0.59
k_m G1018A	5.6×10^{-4}	0.60
k_m G910V	8.8×10^{-5}	0.09
k_m G1018V	1.3×10^{-4}	0.14

14

FIGURE LEGENDS

FIGURE 1. **The sequence of the designed human type III collagen C-terminal quarter fragment with (GPP)$_7$ and the type XIX collagen NC2 domain.**
252 (775-1026) residues of the human type III collagen C-terminal end were used for the experiments schematically shown in (A). The Pro1026 is the last tripeptide unit before the cystine knot sequence -Gly-Pro-Cys-Cys-Gly-Gly- in the natural sequence. In the designed fragment the cystine knot is replaced by the (Gly-Pro-Pro)$_7$-NC2 domain of type XIX collagen. (GlyProPro)$_7$ was attached in order to align the three staggered chains properly. The positions of the mutated residue G910 and G1018 are indicated in bold. The complete sequence of the expressed quarter fragment is shown in (B).

FIGURE 2. **Rotary shadowing images of the QF.** The purified QF protein (0.1 mg/ml) was dialyzed against 100 mM ammonium bicarbonate at 4 °C, sprayed onto mica, and shadowed with platinum (34). The images were viewed by transmission electron microscopy. A representative field is shown in (A). Only molecules with defined, extended tails were measured. The observed average length of the strings is 73.8 nm (SD 6.9 nm). Statistical analysis was performed and a standard error of the mean value was calculated. The length distribution is shown in (B).

FIGURE 3. Thermal transition curves of the QFs. The samples were run in 20 mM Tris/HCl buffer, pH 7.0, containing 200 mM NaCl and 0.05% Tween 20. Samples were equilibrated at 4 °C for more than two days and heated with a rate of 10 °C/h. The CD signal at 223 nm was monitored. At 35 °C the heating was stopper for 10 min in order to achieve complete unfolding of the QF. The temperature was then lowered at a rate of 10 °C/h. The lines are colored as follows, WT (black), G910A (red), and G910V (blue) shown in (A), WT (black), G1018A (green), and G1018V (violet) shown in (B). The arrows indicate the direction of the temperature change.

FIGURE 4. Trypsin digestion of the QFs. (A) The QF sample solutions were incubated at 15, 20, or 33 °C for 0, 5, or 10 min as indicated. Trypsin (8 µg/ml at final concentration) was added and incubated for 5 or 10 min. After the 12% SDS-PAGE the proteins were blotted onto a membrane and subjected to Edman degradation. Blots stained with GelCode Blue are shown. (B) Trypsin-resistant peptides identified by the N-terminal peptide sequencing are indicated by arrows with the residue number. Lys and Arg sites are indicated as vertical lines above and below the horizontal line, respectively. The potential cleavage sites for trypsin are as follows; Lys 810, 815, 864, 893, 927, 939, 983, and Arg 789, 801, 825, 845, 857, 867, 897, 915, 924, 933, 942, 972, 999, 1002, 1005. Mutation sites are marked with big arrows.

FIGURE 5 Refolding kinetics of the QF by CD. The QF proteins were unfolded at 33 °C for 10 min. Refolding at 4°C was monitored at 223 nm by CD. The traces are for WT (black), G910A (red), G910V (blue), G1018A (green), and G1018V (violet). Shown are the long-term experiments (A) and the first 1,000 sec periods (B) with the linear fits used to calculate the initial first order rates in Table 1.

FIGURE 6 Refolding kinetics of the QF analyzed by trypsin resistance of folded fragments. QFs of WT, G910A, G910V, G1018A and G1018V proteins were unfolded at 33 °C for 10 min. The refolding of the triple helix was monitored by trypsin cleavage resistance of the folded helical portions. The unfolded collagen is readily cleaved by trypsin, whereas the refolded triple helical domain is resistant. The time points indicated are when the trypsin solution was added to the sample. After the 2 min. incubation on ice, the enzyme was inactivated. (A) The proteins were analyzed by 12% SDS-PAGE, followed by GelCode Blue staining. The time point is indicated in min unless specified as hour (h). St: starting material, E: trypsin, +E: trypsin treated sample without heat denaturation, M: protein marker, trypsin shows two bands indicated as arrows in WT panel. The molecular weight of the globular protein markers is indicated in kDa. (B) The gel of WT was scanned and the relative abundance of several bands was plotted. The bands are as indicated on the first gel (WT). The relative abundance of the e band (fully folded chains) of all

15

peptides was quantitated and are shown in (C).

FIGURE 7 Simple model to fit the refolding curves. (A) The C-terminal QF of type III collagen is trimerized by $(GPP)_7$- type XIX collagen NC2 domain (XIX NC2). After the heat treatment at 33 °C, the type III collagen part is unfolded but $(GPP)_7$ keeps three chains at the correct stagger. Refolding starts at the C-terminus of QF. For the peptides with G910 mutation the refolding trace is about the same as for WT until the residue G910 is encountered. At the mutated site, the renucleation of the triple helix is necessary to restart the triple helix formation. The mutated peptides with the Val mutations take longer time to renucleate than those with the Ala mutations. For the G1018 peptides the initial refolding is affected from the very beginning. Once the triple helix is renucleated after the G1018 site the zipper-like folding proceeds with the same rate. (B) The model used for the fitting is illustrated. Details are described in the text. All the reaction steps are assumed as the first order. The folding rate for the helix fragment without a mutation is defined as k_n. The renucleation rate (without propagation of helices) at the site of a mutation is defined as k_m. The vertical line in the middle of QF indicates the point of a half conversion. The sites of mutation are indicated as v. (C) Experimental and fitted curves. The experimental curves are the same as in Figure 5A. A single parameter $k_{m,A}$ was used for the both Ala mutants or $k_{m,V}$ for the Val mutants. The experimental data are plotted as circles, WT (black), G910A (red), G910V (blue), G1018A (green), and G1018V (violet). The fitted curves are indicated as lines with the corresponding color. (D) All k_m parameters were fitted individually for each mutant. The same colors are used as in (C).

16

Figure 1

(A)
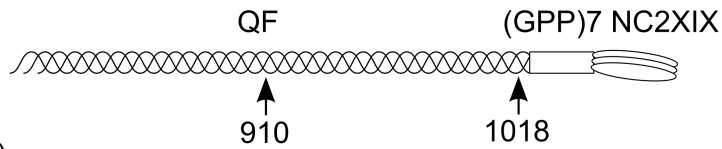
(B)

```
GS

GAPGPLGIAGITGARGLAGPPGMPGPRGSPGPQGVKGESGKP
775
GANGLSGERGPPGPQGLPGLAGTAGEPGRDGNPGSDGLPGRD

GSPGGKGDRGENGSPGAPGAPGHPGPPGPVGPAGKSGDRGES

GPAGPAGAPGPAGSRGAPGPQGPRGDKGETGERGAAGIKGHR
          910
GFPGNPGAPGSPGPAGQQGAIGSPGPAGPRGPVGPSGPPGKD

GTSGHPGPIGPPGPRGNRGERGSEGSPGHPGQPGPPGPPGAP
                         1018      1026
GPPGPPGPPGPPGPPGPP                   (GPP)7

GIPADAVSFEEIKKYINQEVLRIFEERMAVFLSQ   (NC2XIX)
```
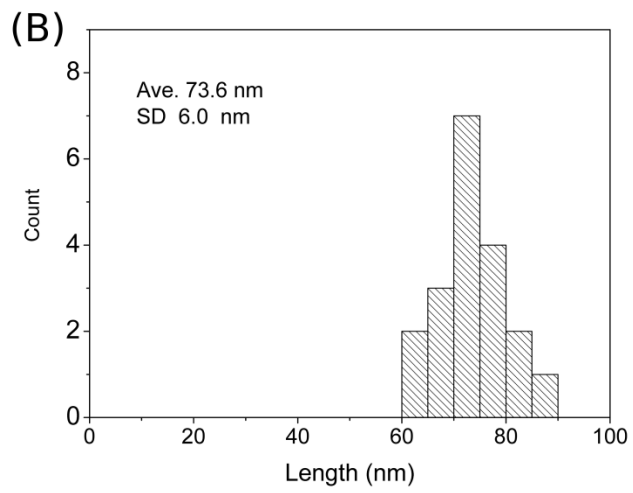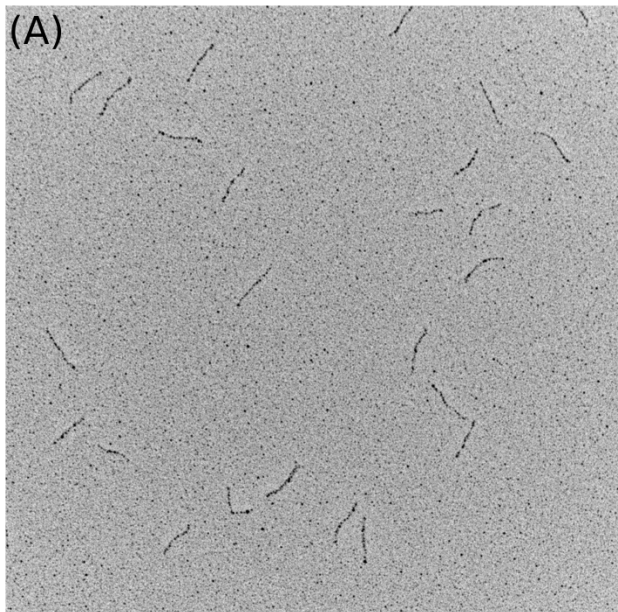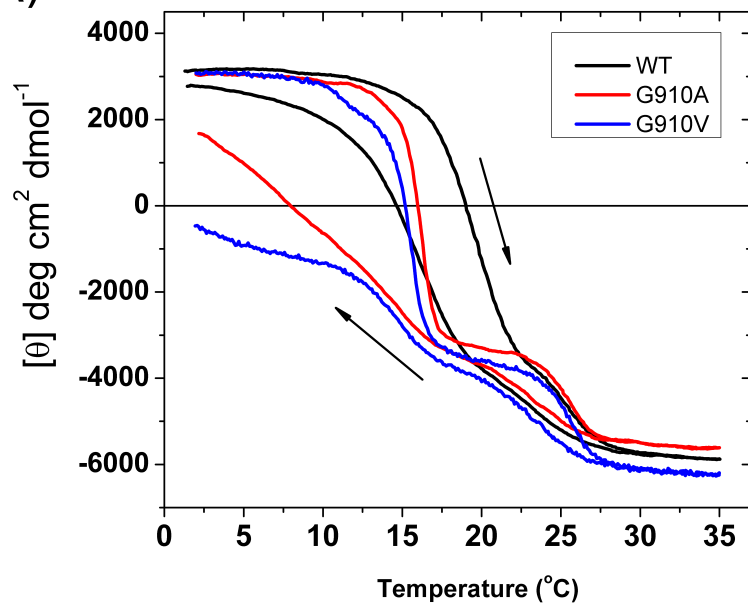
17

Figure 2

(A)

(B)

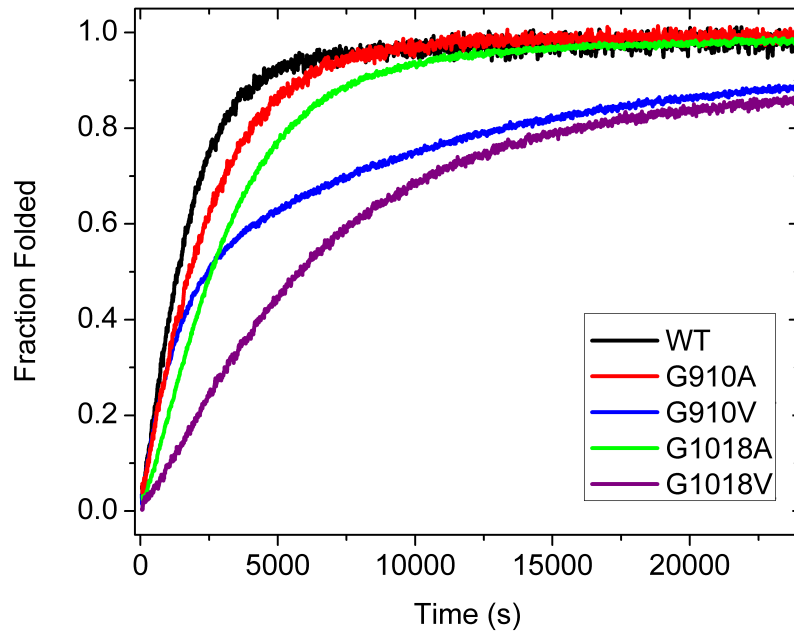Ave. 73.6 nm
SD 6.0 nm

Count

8

6

4

2

0

0   20   40   60   80   100

Length (nm)

18

Figure 3

(A)

(B)

Figure 4

Figure 5

(A)

(B)
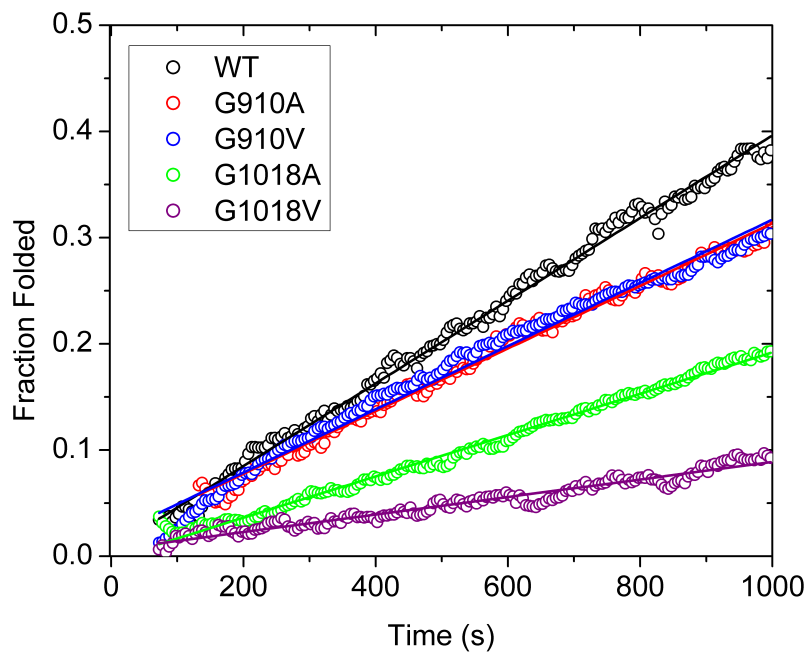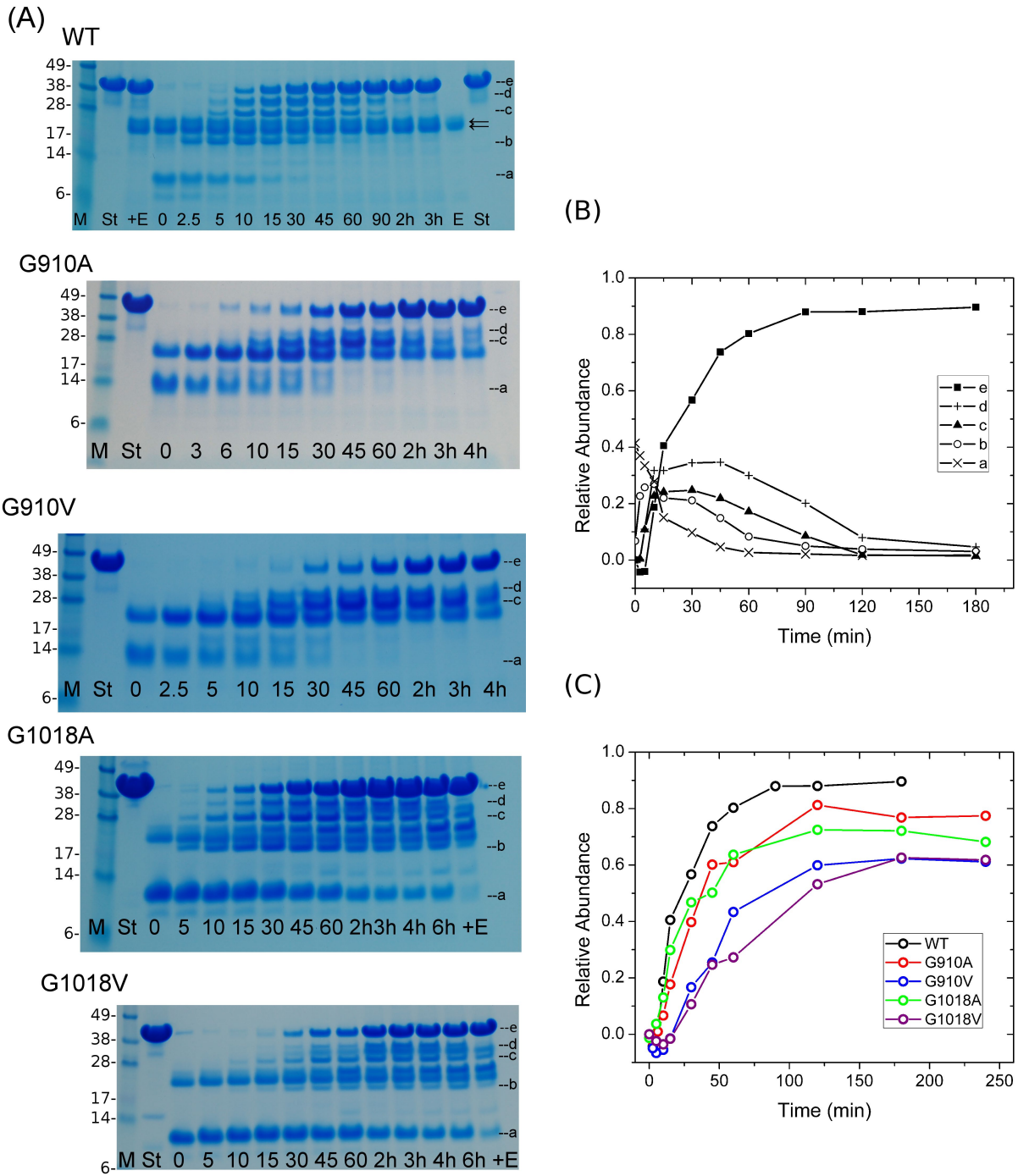
21

Figure 6

Figure 7