THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 53, pp. 44536–44545, December 28, 2012

The NC2 Domain of Type IX Collagen Determines the Chain Register of the Triple Helix*[S]

Received for publication, September 6, 2012, and in revised form, October 2, 2012 Published, JBC Papers in Press, November 6, 2012, DOI 10.1074/jbc.M112.417543

Sergei P. Boudko and Hans Peter Bächinger[1]

From the Research Department, Shriners Hospital for Children and the Department of Biochemistry and Molecular Biology, Oregon Health and Science University, Portland, Oregan 97239

Background: The chains of the collagen triple helix are staggered by one residue.
Results: The NC2 domain of type IX collagen establishes a stagger for the adjacent triple helix.
Conclusion: The NC2 domain is a useful tool to build stagger-specific hetero-trimeric triple helices.
Significance: The construction of native-like collagen peptides of any type is now achievable.

Precise mapping and unraveling the mechanism of interaction or degradation of a certain type of collagen triple helix requires the generation of short and stable collagenous fragments. This is a great challenge especially for hetero-trimeric collagens, where chain composition and register (stagger) are important factors. No system has been reported that can be efficiently used to generate a natural collagenous fragment with exact chain composition and desired chain register. The NC2 domain (only 35–50 residues) of FACIT collagens is a potent trimerization domain. In the case of type IX collagen it provides the efficient selection and hetero-trimerization of three distinct chains. The ability of the NC2 domain to determine the chain register of the triple helix is studied. We generated three possible sequence combinations ($\alpha1\alpha1\alpha2$, $\alpha1\alpha2\alpha1$, $\alpha2\alpha1\alpha1$) of a type I collagen fragment (the binding region for the von Willebrand factor A3 domain) attached to the NC2 domain. In addition, two control combinations were produced that constitute homo-trimers of $(\alpha1)_3$ or $(\alpha2)_3$. For the hetero-trimeric constructs, $\alpha1\alpha1\alpha2$ demonstrated a higher melting temperature than the other two. Binding experiments with the von Willebrand factor A3 domain revealed the homo-trimer of $(\alpha1)_3$ as the strongest binding construct, whereas the homo-trimer of $(\alpha2)_3$ showed no binding. For hetero-trimers, $\alpha1\alpha1\alpha2$ was found to be the strongest binding construct. Differences in thermal stability and binding to the A3 domain unambiguously demonstrate that the NC2 domain of type IX collagen determines not only the chain composition but also the chain register of the adjacent triple helix.

Different collagen types are major constituents of the extracellular matrix providing tissue strength, integrity, and a structurally organized and actively instructing environment for living cells in our body. There are 28 different types of collagen that are encoded by 47 genes (1). Differential splicing of some collagens further increases this diversity. Moreover, in humans there are about 20 genes that encode collagen-like proteins. The real complexity of the "collagen world" comes from the supramolecular structures (fibrils, fibers, networks) that are formed by one or more collagens in association with other extracellular components.

Collagens and collagen-like proteins are characterized by a presence of triple helical regions consisting of a repeating sequence *GXY*, where *G* is a glycine residue, and *X* and *Y* are any amino acid but frequently they are proline and 4-hydroxyproline, respectively. A really unique feature of this trimeric motif is a stagger (or chain register) between each chain. Even in homo-trimeric collagen, identical residues within the triple helix are not equivalent as has been observed in crystal structures of collagen peptides. There are always a leading, middle, and trailing chains shifted by one residue. Type I collagen is the most abundant and most studied collagen, which composed of two $\alpha1$ and one $\alpha2$ chains (1). The chain stagger of type I collagen remains elusive, and all three variants, $\alpha1\alpha1\alpha2$ (2), $\alpha1\alpha2\alpha1$ (3, 4), and $\alpha2\alpha1\alpha1$ (5) were suggested.

Our current knowledge of the collagen interactome is generally limited to only a few abundant collagen types. Precise mapping of collagen interaction sites and their binding mechanisms are difficult to address due to exceptional length of the molecule. Generation of short fragments of collagen is a general method, and great progress was achieved in this direction. Collagen peptide toolkits were successfully implemented for type II and III homo-trimeric collagens and used to precisely map and study binding to integrin $\alpha2\beta1$ (6), von Willebrand factor (7), DDR2 (8), gpVI (9), LAIR-1 (10), SPARC (11), and YadA (12). However, this approach is only applicable to homo-trimeric types of collagen. There are a number of hetero-trimeric collagen types, such as I, IV, V, VI, IX, and XI, which have two or even three distinct chains. In this case a problem with chain stoichiometry is additionally complicated by the chain register. For example, instead of forming a single species of hetero-trimeric $\alpha1\alpha2\alpha3$ with a defined stagger, a mixture of three different chains, $\alpha1$, $\alpha2$, and $\alpha3$, can potentially form 27 different varieties (excluding those with misaligned chains); among them 6 will represent differently staggered forms of all three $\alpha1$, $\alpha2$, $\alpha3$ chains. Attempts were made to develop a system with the chain composition and stagger control. One was based on the

* The work was supported by a grant from Shriners Hospitals for Children.
[S] This article contains supplemental Tables S1 and S2 and Figs. S1–S7.
The atomic coordinates and structure factors (code 4GYX) have been deposited in the Protein Data Bank (http://wwpdb.org/).
[1] To whom correspondence may be addressed: Shriners Hospital for Children, Research Dept., 3101 SW Sam Jackson Park Rd., Portland, OR 97239. Tel.: 503-221-3433; Fax: 503-221-3451; E-mail: hpb@shcc.org.

cysteine knot of type III collagen that is located at the very C-terminal end of the triple helical domain. The knot covalently links all three chains, and it was thus hypothesized that it should also lock a specific stagger (13). A simplified artificial version of the knot was designed and implemented using regioselective thiol chemistry (14). Nevertheless, the actual stagger in this system has never been resolved. Moreover, the recent crystal structure of type III collagen fragment with the natural cysteine knot (15) and a new crystal structure reported here raised a question of whether it can determine (lock) a specific stagger. Another approach in achieving the determined stagger is a system of three artificial collagen-like chains bearing compensating charges that allow a specific assembly of hetero-trimeric molecules (16). Surprisingly, the highly preferred stagger was detected using NMR technique (17). Despite successful development of the artificial hetero-trimeric collagen system with the stagger control, its application for natural sequences seems to be impractical due to unpredictable effects of the loaded natural sequence on composition and chain register.

Trimerization is an obligatory step in collagen triple helix formation. For long collagen chains it requires a special trimerization domain that selects type-specific chains and nucleates the triple helix formation in a zipper-like fashion (18). These domains have different folds and are of a non-triple-helical nature. We have recently identified small (35–55 residues) and efficient trimerization domains in collagen types XVIII and XV (19, 20) and XIX and IX (21, 22), which are useful tools for folding and stabilization of collagenous peptides because they represent a natural way to initiate folding and stabilize the triple helix. Among them the non-collagenous domain 2 (NC2)[2] of type IX collagen is unique because it provides the hetero-trimerization of three distinct collagen chains, *i.e.* α1, α2, and α3 (22). We hypothesized here that it also determines the chain register of the adjacent triple helix. To test this hypothesis, a set of different chain compositions based on type I collagen was generated. The collagen sequences comprised a potential binding site for the von Willebrand factor A3 domain (vWFA3). It was speculated previously that only a specific stagger binds vWFA3 (7).

Various complexes demonstrated different thermal stability with the one demonstrating the highest melting temperature. Moreover, the complexes demonstrated different binding affinities to vWFA3. Consequently, the NC2 domain of type IX collagen determines the chain register of the triple helix. The type IX collagen hetero-trimerization domain is a light-weight non-triple-helical fragment that provides a technically simple and natural way for the generation of triple helical collagenous fragments of any type and length with predicted chain composition and stagger.

EXPERIMENTAL PROCEDURES

Peptide Synthesis—The peptides col3HypCC and col3ProCC (Fig. 1) were synthesized on an ABI433A peptide synthesizer with 0.25 mM Fmoc-L-Val-PEG-PS resin, a 4-fold excess of Fmoc amino acids, and *O*-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate as the activating agent. The Fmoc amino acids carried the protection groups Cys-trityl (Trt), Hyp-tert-butyl (*t*-Bu), Gln-Trt , Arg-2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl (Pbf), and Hyp-*t*-Bu. The peptide was cleaved off the resin and deprotected for 4 h at room temperature with 90% trifluoroacetic acid, 5% thioanisole, 3% 1,2-ethanedithiol, and 2% anisole. Subsequently, the peptide was precipitated in cold ether, redissolved in H_2O, and lyophilized. The reduced peptide was then purified by reverse phase HPLC using a C_{18} column (Vydac, Hesperia, CA; 50 × 250 mm, 10–15-μm particle size, 300-Å pores) with an acetonitrile/water gradient and 0.1% trifluoroacetic acid as ion-pairing agent. Finally, the peptide was characterized by electrospray/quadrupole/time-of-flight mass spectrometry (Q-tof micro; Waters Associates) and amino acid analysis.

Peptide Folding, Oxidation, and Purification of Disulfide-linked Trimer—The lyophilized, reduced peptides were dissolved in degassed and N_2-saturated 50 mM sodium acetate buffer, pH 4.5, under N_2 atmosphere and were kept at 4 °C for 24 h to allow triple helix formation before oxidation. Oxidation was carried out by the addition of reduced (10 mM) and oxidized (1 mM) glutathione and exposure to atmospheric O_2. The pH was raised to 8.3 with a saturated solution of Tris. Oxidation was carried out for 5–7 days, and the mass of the peptides was periodically analyzed by liquid chromatography-mass spectrometry. The maximum yield of a covalently linked trimeric form was ~60–70%. To separate the covalently linked trimeric form from other oligomers, the oxidized crude material was dissolved in deionized 8 M urea solution with 0.1% trifluoroacetic acid to prevent disulfide exchange and applied to a sieve column. Trimer-containing fractions were pooled and further purified by reverse phase HPLC using a C_{18} column.

Crystallization and Structure Determination—The purified and lyophilized covalently linked trimeric peptide containing the type III collagen sequence, col3HypCC, was dissolved at a concentration of 10 mg/ml in 5 mM acetic acid. The peptide was crystallized at 22 °C using the hanging drop vapor diffusion method. For crystallization, 2 μl of the peptide solution was mixed with 2 μl of the reservoir solution of 22–28% polyethylene glycol 3350, 0.1 M sodium acetate, pH 5.5. The crystals appeared as plates in a period of 1–5 days and are monoclinic. They belong to space group $P2_1$ with $a = 19.07$ Å, $b = 26.99$ Å, $c = 64.12$ Å, and $\beta = 92.48°$. The crystals were shortly dipped into the reservoir solution with added 10% 1,2-propanediol (served as the cryo-protectant condition) and frozen in liquid nitrogen. Data collection was performed using crystals flash-frozen to 100 K on the NOIR-1 detector system at the Molecular Biology Consortium Beamline 4.2.2 of the Advanced Light Source, Lawrence Berkeley National Laboratory.

The collected images were indexed, integrated, and scaled using iMOSFLM (23) and SCALA from the CCP4 suite (24). The AMORE program (25, 26) was used within the CCP4 suite (24) to find an initial molecular replacement solution. A crystal structure of the VWF binding type III collagen-derived triple helical peptide (PDB ID 4DMT) (2) was used as a search model. A single outstanding solution was generated by AMORE. Iterative cycles of model correction and refinement were performed using COOT (27) and PHENIX (28), respectively. B factors of protein atoms were refined as two TLS groups cover-

[2] The abbreviations used are: NC2, non-collagenous domain 2; vWFA3, von Willebrand factor A3 domain; Fmoc, *N*-(9-fluorenyl)methoxycarbonyl.

ing the N-terminal (less structured) and C-terminal segments. The quality of the models was assessed with program MolProbity (29). Data collection and refinement statistics are summarized in supplemental Table S1. The atomic coordinates and structure factors were deposited to the Protein Data Bank (PDB ID 4GYX). The figures were generated with programs PyMOL and CCP4MG (30).

Cloning, Expression, and Purification of vWFA3—To facilitate expression and purification of von Willebrand factor A3 domain, it was cloned as part of a fusion molecule with a His-tagged thioredoxin with a thrombin cleavage site (HisTrx-thr) to cleave off products later. The plasmid pET23-HisTrx (22) has multiple cloning sites just after the HisTrx-thr gene. The plasmid pQE-9-VWF-A3 (31) containing the cDNA of vWFA3 was a gift from Miguel Cruz. The vWFA3 domain was PCR-amplified using forward and reverse primers (supplemental Table S2) and cloned into the pET23-HisTrx vector using the restriction sites BamHI and SalI. The DNA insert was verified by Sanger dideoxy DNA sequencing.

The recombinant vWFA3 was expressed in the *Escherichia coli* BL21(DE3) host strain (Novagen). Colonies from freshly transformed competent cells were resuspended in 1 liter of 2× TY medium (16 g of Tryptone, 10 g of yeast extract, and 5 g of NaCl per liter), grown to A_{600} ~0.6–0.8, and induced by adding isopropyl 1-thio-β-D-galactopyranoside to a final concentration of 1 mM and expressed at 20 °C for 16–20 h. Cells were chilled at 4 °C, harvested by centrifugation at 3000 × *g* for 15 min, resuspended in 20 ml of 20 mM Tris/HCl buffer, pH 8, and disrupted by ultrasonication. All the insoluble material was removed by centrifugation at 15,000 × *g* for 30 min. In contrast to the previously reported expression of vWFA3 as insoluble inclusion bodies (31), 60–80% of the fusion protein was expressed in soluble form. The soluble fraction was transferred to a beaker, and 30% streptomycin sulfate was slowly added to the intensively mixing solution to a final concentration of 1% and incubated for 15 min. The precipitated DNA fragments and other insoluble materials were removed by centrifugation at 15,000 × *g* for 15 min. To the cleared solution saturated ammonium sulfate was slowly added while mixing up to 35% of saturated solution to salt-out the recombinant protein. The protein was pelleted by centrifugation at 10,000 × *g* for 10 min and solubilized in 50 mM sodium phosphate buffer, pH 7.

After solubilization, the buffer was supplemented with 500 mM NaCl, 20 mM imidazole, and purified by immobilized metal affinity chromatography on HisTrap™ HP column (GE Healthcare) charged with Ni^{2+}. The protein was eluted with 50 mM sodium phosphate buffer, pH 7, supplemented with 500 mM NaCl and 500 mM imidazole.

Thrombin cleavage was performed at room temperature for 24 h with 1 unit/ml recombinant thrombin protease (Baculo-Gold™; BD Biosciences) in 50 mM Tris buffer, pH 8.0, supplemented with 150 mM NaCl. Thrombin cleaved material was run over the nickel-charged column to separate vWFA3 from the His-tagged thioredoxin or uncleaved material.

Two additional purification steps were applied, namely, the hydrophobic chromatography on phenyl-Sepharose and the anion-exchange chromatography on Q-Sepharose. The protein solution was adjusted to 1 M ammonium sulfate, 50 mM sodium phosphate, pH 7, and run through the phenyl-Sepharose column (GE Healthcare). The protein was then eluted with 0.5 M ammonium sulfate. The protein was extensively dialyzed against 50 mM Tris, pH 8, loaded onto the Q-Sepharose column (GE Healthcare) and eluted with a linear gradient of NaCl. The pure protein was eluted at ~100 mM NaCl. The fractions containing the pure proteins were pooled and dialyzed into appropriate buffers. The correct molecular weight was confirmed by mass spectroscopy. Sedimentation equilibrium experiment confirmed the monomeric state of vWFA3 in solution. When necessary, vWFA3 was concentrated using an Amicon Ultra 10K concentrator (Millipore). Protein concentrations were determined by amino acid analysis. The total yield of the purified protein was ~30 mg starting from 1 liter of medium.

Cloning, Expression, and Purification of $(GPP)_7$ and Type III Collagen Construct—A synthetic gene encoding $(GPP)_7$ sequence linked to a fragment encoding NC2 domain of type XIX collagen was made by Integrated DNA Technologies (supplemental Fig. S1). The gene was re-cloned into the pET23-HisTrx expression vector (22) using BamHI and SalI restriction sites, which resulted in a new vector, pET23-HisTrx_gpp7-NC2XIXB. The final construct was expressed as a fusion molecule with a His-tagged thioredoxin and a thrombin cleavage site. In this chimeric construct the collagenous part, $(GPP)_7$, was stabilized by the NC2 trimerization domain of type XIX collagen.

An oligonucleotide 5′-CCAGGCCCGCCAGGTCCGCCAGGTCCGCCGGGTCCGCGCGGCCAGCCAGGTGTTATGGGTTTTCCTGGGCCACCGGGGCCGCCCGGACCG-3′ was used as a template for PCR amplification of a fragment encoding type III collagen vWFA3 binding region flanked by $(GPP)_4$ and $(GPP)_3$ using primers shown in supplemental Table S2. The fragment was cloned into the pET23-HisTrx_gpp7-NC2XIXB vector, where the $(GPP)_7$-encoding portion was replaced with the fragment of interest using ApaI and XmaI cloning sites.

The proteins were expressed and purified using the same strategy as for vWFA3 with a few modifications. Proteins salting out were achieved at 30% of saturated solution of ammonium sulfate. After separation of thrombin cleavage products on the Ni^{2+} column, two ion exchange chromatographies were applied, the anion exchange on Q-Sepharose (GE Healthcare) and the cation exchange on SP-Sepharose (GE Healthcare). The proteins were eluted at ~150 mM NaCl from the Q-Sepharose column equilibrated and operated using 50 mM Tris/HCl, pH 8. The SP-Sepharose column was run using 50 mM HEPES, pH 7, and the elution was achieved with 100 mM NaCl. The total yield of each purified protein was ~7 mg starting from 1 liter of medium.

Cloning of Type I Collagen Complexes—Two synthetic genes encoding human type I collagen fragments of either chain α1 or α2 flanked by $(GPP)_4$ and $(GPP)_3$ and linked to a fragment encoding either α2 or α3 chain of NC2 domain of type IX collagen, respectively, with codons optimized for *E. coli* expression were made by Integrated DNA Technologies (supplemental Fig. S1). The genes were re-cloned into the pET23-HisTrx expression vector (22) using BamHI and SalI restriction sites. The final constructs were expressed as fusion molecules with a His-tagged thioredoxin with a thrombin cleavage site. Due to a strong background (without adding isopropyl 1-thio-β-D-ga-

lactopyranoside) expression of the proteins, a significant amount of degraded material was observed, which was found to be inappropriate for further experiments. The fusion constructs were re-cloned into the pET22b(+) vector (Novagen), which has a tighter expression control, using XbaI and SalI restriction sites. The resulting plasmids, pET22-HisTrx_Iα1-IXα2 and pET22-HisTrx_Iα2-IXα3, were used as vectors to generate four other required constructs, *i.e.* pET22-HisTrx_Iα1-IXα1, pET22-HisTrx_Iα1-IXα3, pET22-HisTrx_Iα2-IXα1, and pET22-HisTrx_Iα2-IXα2. PCR-amplified fragments encoding α1, α2, or α3 chains of type IX collagen NC2 domain were cloned accordingly using the XmaI and SalI restriction sites flanking the NC2 domain sequence of the corresponding vector. As templates for PCR the original synthetic genes and a pET23-His-Trx-NC2IXα1 plasmid (22) were used. Sequences of primers are presented in supplemental Table S2.

Expression, Refolding, and Purification of Type I Collagen Constructs—The recombinant proteins were expressed separately in the *E. coli* BL21(DE3) host strain (Novagen). Colonies from freshly transformed competent cells were resuspended in 2× TY medium, grown to A_{600} ~0.6–0.8 at 37 °C, chilled down to an appropriate temperature, and induced by adding 1 mM isopropyl 1-thio-β-D-galactopyranoside. Whereas the constructs with α1 or α3 sequences of NC2 were prominently expressed at 12 °C within 3–4 days, the constructs with α2 of NC2 required cultivation at 4 °C for 7–10 days to produce comparable amounts of the proteins (supplemental Fig. S2).

To form the required complexes, cells were accordingly combined and subjected to purification and reshuffling. The following nomenclature designates five different complexes: 111, a combination of Iα1-IXα1, Iα1-IXα2, and Iα1-IXα3; 222, a combination of Iα2-IXα1, Iα2-IXα2, and Iα2-IXα3; 112, a combination of Iα1-IXα1, Iα1-IXα2, and Iα2-IXα3; 121, a combination of Iα1-IXα1, Iα2-IXα2, and Iα1-IXα3; 211, a combination of Iα2-IXα1, Iα1-IXα2, and Iα1-IXα3. In other words, numbers 1 and 2 designate α1 or α2 chains of type I collagen, respectively, linked to the α1, α2, and α3 chains of type IX collagen. For each complex equal volumes (0.33 liters) of three different cell cultures were mixed and subjected to initial purification as described for vWFA3w with several modifications. After the first purification on the Ni^{2+} column each complex eluted in 15 ml was diluted 4 times with 100 mM Tris/HCl, pH 8.35. EDTA was added to a 0.5 mM final concentration to chelate any divalent cations that catalyze oxidation. Reduced and oxidized glutathiones were added to final concentrations of 10 and 1 mM, respectively, and pH was adjusted to 8.35 with saturated solution of Tris base. The samples were incubated at 37 °C for 1 h, transferred into beakers at room temperature, treated with 1 μl/ml diisopropylfluorophosphate to inhibit serine proteases, and stirred for 24 h. The samples were extensively dialyzed against the thrombin cleavage buffer and subjected to thrombin cleavage and separation on the Ni^{2+} column as described for vWFA3.

Two additional purification steps were a separation on the heparin-Sepharose (GE Healthcare) column (eluted at ~90 mM NaCl, buffered with 50 mM Tris/HCl, pH 8) and a final polishing on the Q-Sepharose column (eluted at ~50 mM NaCl, buffered with 50 mM Tris/HCl, pH 8). Fig. 4 shows the electrophoresis of the purified proteins. Correct molecular masses and composition identities were confirmed by the mass spectroscopy. Peptide concentrations were determined by amino acid analysis. The final yields of the purified complexes were 4–5 mg starting from 1 liter of medium.

Analytical Ultracentrifugation—Sedimentation equilibrium measurements were performed with a Beckman model XLA analytical ultracentrifuge. Samples were in 50 mM sodium phosphate, pH 7, supplemented with 150 mM NaCl (PBS). Protein concentrations varied from 0.7 to 0.9 mg/ml. Absorbance was measured at 240 nm for the collagen peptides or at 280 nm for vWFA3. Runs were carried out at 4 °C in an An60-Ti rotor using 12-mm cells and Epon two-channel centerpieces. Speed used was 20,000 rpm. Data analysis was done using the Sedfit program.

Fluorescence-based Collagen Binding Assay—The measurement of fluorescence spectra was performed on an SLM8000C instrument (SLM Instruments, Inc.) modified by ISS (ISS Inc.) using the software provided by the manufacturer. The excitation wavelength was 270 nm (8-nm bandwidth). Emission scans were obtained over the wavelength range of 300–420 nm (2-nm bandwidth). Stock solutions of col3HypCC (3.49 mM trimer) or col3ProCC (1.32 mM trimer) were added to a solution of recombinant vWFA3 (the final concentration of 198 nM) in 50 mM Tris/HCl, pH 8.0, containing 0.15 M NaCl in a 1 × 1-cm cell in a thermostatted block at 25 °C with constant stirring. The samples were equilibrated for 2 h before emission spectra were acquired. Correction of the spectra was done by subtracting the emission of the collagen peptides alone (this was a very small correction). Maxima of the emission spectra for different peptide concentration were determined by fitting the profiles in the range from 320 to 350 nm using the normal (or Gaussian) distribution function.

Circular Dichroism Spectroscopy—Circular dichroism (CD) spectra were recorded on an AVIV model 202 spectropolarimeter (AVIV Instruments, Inc.) with a thermostatted quartz cell of 1-mm path length. Samples were in 50 mM sodium phosphate, pH 7. Protein concentrations were adjusted to ~0.2 mg/ml. Thermal scanning profiles were recorded at 225 nm to monitor the collagen triple helix transition. The analysis of melting profiles was performed using the hysteresis formalism developed in Ref. 32.

Differential Scanning Calorimetry—The temperature dependence of the partial heat capacity was measured in an N-DSC II differential scanning calorimeter (Calorimetry Science Corp.) with the heating rate of 60 °C/h. Samples were in PBS. Protein concentrations varied from 0.7 to 0.9 mg/ml. Base-line subtraction and calculation of ΔH_{cal} was performed using the software provided by the manufacturer.

Affinity Chromatography—15 mg of vWFA3 was coupled to 1.5 ml of CNBr-activated Sepharose 4B (GE Healthcare) according to the manufacturer's instructions. The resulting vWFA3-Sepharose resin was packed into a 6-cm-long column and equilibrated with 50 mM Tris/HCl, 150 mM NaCl, pH 8. The column was operated at room temperature at the flow rate of 0.5 ml/min. Each collagen peptide of the same concentration (0.5 mg/ml) was injected in a volume of 0.1 ml. The elution profile was monitored at 220 nm.

col3HypCC: GPOGPOGPRGQOGVMGFOGPOGPOGPCCGGV
col3ProCC: GPOGPOGPOGPRGQPGVMGFPGPOGPOGPOGPCCGGV
(GPP)$_7$-NC2XIXB: GSGPPGPPGPPGPPGPPGPPGPPGIPADAVSFEEIKKYINQEVLRIFEERMAVFLSQ
III-NC2XIXB: GSGPPGPPGPPGPPGPRGQPGVMGFPGPPGPPGPPGIPADAVSFEEIKKYINQEVLRIFEERMAVFLSQ
Iα1-IXα1: GSGPPGPPGPPGPPGARGQAGVMGFPGPPGPPGPPGRAPTDQHIKQVCMRVIQEHFAEMAASLKRPDSGAT
Iα1-IXα2: GSGPPGPPGPPGPPGARGQAGVMGFPGPPGPPGPPGRDATDQHIVDVALKMLQEQLAEVAVSAKREALGAV
Iα1-IXα3: GSGPPGPPGPPGPPGARGQAGVMGFPGPPGPPGPPGKEASEQRIRELCGGMISEQIAQLAAHLRKPLAPGSI
Iα2-IXα1: GSGPPGPPGPPGPPGARGEPGNIGFPGPPGPPGPPGRAPTDQHIKQVCMRVIQEHFAEMAASLKRPDSGAT
Iα2-IXα2: GSGPPGPPGPPGPPGARGEPGNIGFPGPPGPPGPPGRDATDQHIVDVALKMLQEQLAEVAVSAKREALGAV
Iα2-IXα3: GSGPPGPPGPPGPPGARGEPGNIGFPGPPGPPGPPGKEASEQRIRELCGGMISEQIAQLAAHLRKPLAPGSI

FIGURE 1. **Sequences of the peptides.** *O* represents 4(*R*)-hydroxyproline. Collagen sequences constituting the von Willebrand factor A3 binding domain are *underlined*.

Surface Plasmon Resonance—The BIAcore X system (BIAcore) was utilized to study the interaction of vWFA3 with different collagen substrates. Collagen peptides were covalently coupled to CM5 sensor chips via amino groups. Specifically, the CM5 surfaces were activated with 0.4 M l-ethyl-3-(3-dimethylaminopropyl)carbodiimide/0.1 M sulfo-*N*-hydroxysuccinimide for 7 min. The flow rate of 5 μl/min was used for all steps. After activation, 5–20 μg/ml of collagen substrate in 10 mM HEPES buffer, pH 6.0, was passed over the surface, producing ~2000–3000 response units (approximate surface density of ~2.5–3.0 ng/mm^2). Unreacted sites on the CM5 surfaces were blocked by treatment with 1 M ethanolamine hydrochloride for 7 min. Binding experiments were performed at 25 °C using 10 mM HEPES buffer, pH 7.5, supplemented with 150 mM NaCl. vWFA3 solutions at different concentrations in a volume of 25 μl were perfused over the sensor chip surfaces to obtain real-time binding data at a flow rate of 5 μl/min. Using this flow rate and surface immobilization conditions, there were no transport-limited effects on binding profiles. Binding was monitored using a laser system that detects changes in refractive index. The signal generated, in arbitrary response units, is proportional to a change in surface mass. The association and dissociation phases were not resolved that indicated a high dissociation constant (usually in μM and higher ranges). Therefore, it was not possible to obtain association and dissociation rates. After vWFA3 injection the response signal quickly reached the steady value and then returned to the buffer background signal when the injection was replaced by the running buffer (supplemental Fig. S7). No regeneration was necessary between consequent injections of vWFA3.

Channel 2 (coupled with the negative control peptide) response was subtracted from channel 1 (coupled with the peptide of interest) to eliminate the contribution from nonspecific binding and surface effects. To compensate for the differences in responses caused by varying peptide coupling efficiency, the ratio of response signal upon binding to response change observed during sensor chip coupling was used. Because masses of vWFA3 (22 kDa) and tested peptides (~21 kDa) are comparable, this ratio approximately reflected the binding ratio of two molecules.

RESULTS

Crystal Structure of Type III Collagen Model Peptide with the Cysteine Knot—The synthetic peptide that contained the minimal vWF binding site GPRGQOGVMGFO flanked by two GPO units on both sides and covalently trimerized within the C-terminally added sequence of type III collagen cysteine knot GPCCGGV (Fig. 1) was crystallized, and its structure was solved by molecular replacement and refined to 1.5 Å resolution (supplemental Table S1). The asymmetric unit contains a single triple-helical peptide with chains labeled A, B, and C that correspond to the leading, middle, and trailing chains, respectively. As has been observed in other collagen peptide crystal structures, the N- and C-terminal ends of each of the three triple helical chains display weaker electron density and appear disordered. In our case the N-terminal residues are particularly disordered and display significantly elevated B-factors (Fig. 2*A*) due to lack of crystal contacts.

The final model consists of residues 1–28 of chain A, residues 1–28 of chain B, and residues 1–26 of chain C. Whereas a strong electron density was observed for the central region of the peptide (Fig. 2*B*) that correlated with the lowest B-factors, the unambiguous tracing of the C-terminal region was limited. Although we built one disulfide bond between Cys-27(A) and Cys-28(B) (A1-B2 connectivity), which has the best geometry and best fits the density (Fig. 2*C*), alternative disulfide bond connections were possible, *i.e.* A2-B1 or A2-C1. None of these connections allowed the complete structure of the cysteine knot to be built, including all three disulfide bonds, in accord with the diffraction data and geometry constrains. The best connectivity, A1-B2, corresponds to the one previously reported for the crystal structure of the natural type III collagen fragment Gly-991–Gly-1032 (15). Despite the same connectivity between A1 and B2 cysteines, the main chain geometry of these peptides within the cysteine knot region significantly deviates (Fig. 2*D*). It is highly probable that the cysteine knot can adopt several sets of disulfide bond connections and thus cannot lock the specific chain stagger within an adjacent triple helix.

Hydroxyproline to Proline Substitution within the Binding Region of Type III Collagen and Its Affinity to vWFA3—The central 4-hydroxyproline (O) residue within the GPRGQOGVMGFO sequence was reported to be crucial for the binding to vWFA3 as its substitution to the alanine residue abolished the binding (7). Because the recombinant production of collagenous peptides with the post-translational modification of proline to hydroxyproline residues in milligram quantities is rather challenging, especially using the bacterial system, we were interested whether non-modified proline residues would not sufficiently reduce the binding. Thus two synthetic peptides col3HypCC and col3ProCC (Fig. 1) with hydroxypro-

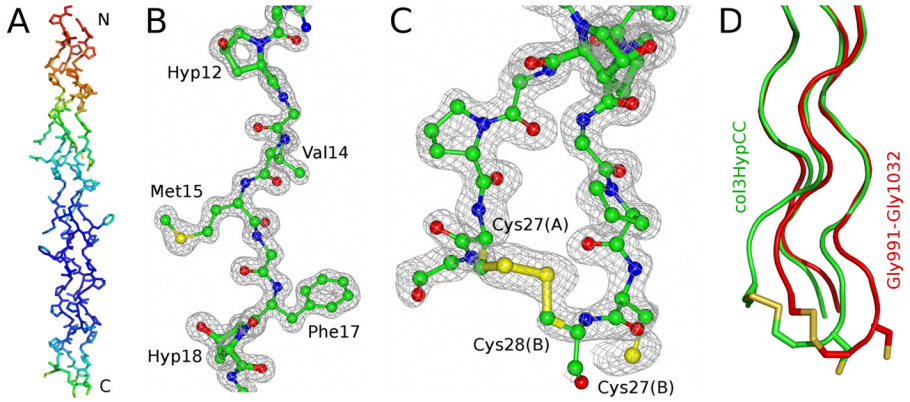

FIGURE 2. **Crystal structure of col3HypCC.** *A*, the overall triple helix is shown for the col3HypCC structure and colored according to B-factor, with the gradation of *blue* to *red* reflecting increasing B-factors (*i.e. blue* represents low B-factors, and *red* represents high B-factors). *B*, 1.5-Å resolution $2F_o - F_c$ map of the col3HypCC structure shows the central region. The map is contoured at 1.5 σ and shown as a *gray mesh*. The structure is shown as *balls and sticks* and colored according to atom type with carbon, nitrogen, oxygen, and sulfur shown in *yellow*, *blue*, *red*, and *green*, respectively. *C*, $2F_o - F_c$ map contoured at 1.0 σ shows the density for the cysteine knot region. Clear density is only observed for the A1-B2 connection, which is consistent with the previously published structure of type III collagen fragment Gly-991–Gly-1032 (15). *D*, shown is cysteine knot region superimposition of col3HypCC and Gly-991–Gly-1032 (15). Coordinates of main chain atoms of the last two tripeptide units preceding the cysteine knot region were used for the least squares fitting to align the triple helices. A significant variation was observed within the cysteine knot region.

lines and prolines, respectively, extended with the sequence of the cysteine knot of type III collagen were produced in the covalently trimerized form. The cysteine knot forms interchain disulfide bonds at the end of the triple helix, which substantially stabilizes the triple helix and prevents chain dissociation in diluted solutions (33, 34). To compensate for the destabilizing effect of the hydroxyproline to proline substitution on the stability of the triple helix, the sequence of col3ProCC has two additional GPO units (one on each side of the chain). The resulting peptides remain in their triple helical structure upon heating up to 30 °C (thermal stability profiles for both peptides are shown in supplemental Fig. S3). Fluorescence of a single Trp residue within the vWFA3 domain was found to be sensitive to the collagen peptide binding, and the blue shift in the spectrum was observed upon the complex formation. To explore the binding affinity the fluorescence spectra of vWFA3 at a fixed concentration in the presence of varying amounts of peptides were recorded at 20 °C, and a peak position was determined as a function of peptide concentration. The peak position shifted from 336.5 to 332.2 nm. The curves were globally fitted, and the dissociation constants were determined to be 1.2 and 2.9 μM for col3HypCC and col3ProCC, respectively (Fig. 3). Although the Hyp to Pro substitution within the vWFA3 binding site decreases the binding affinity, the change is not detrimental. Therefore, all other peptides used in this study were produced in the bacterial system that lacks the 4-hydroxylation of the proline residues.

Production of Homo- and Hetero-trimeric Collagen Complexes with Controlled Composition—Type III collagen is a homo-trimer that consists of three α1 chains. A host guest peptide including the short sequence of type III collagen as well as an empty host peptide $(GPP)_7$ were expressed as fusion proteins with the NC2 homo-trimerization domain of type XIX collagen (21) (Fig. 1). Thermal stability of collagenous domains in these proteins was analyzed by the CD spectroscopy (supplemental Fig. S4). Both proteins demonstrated unfolding transitions at temperature ranges higher than used for binding measurements.

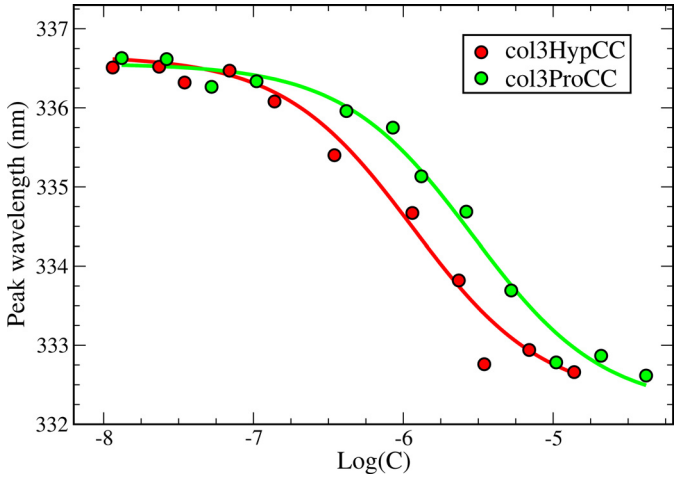

FIGURE 3. **Binding affinities of Hyp or Pro-containing type III collagen peptides to vWFA3.** Fluorescence peak shift of vWFA3 as a function of col3HypCC (*red*) and col3ProCC (*green*) concentration is shown.

Type I collagen is a hetero-trimer that consists of two α1 and one α2 chains. Therefore, there are three possible arrangements of chains, where the α2 chain can be in the leading, middle, or trailing position. To cover all possibilities, three hetero-trimeric complexes were produced, *i.e.* 112, 121, and 211, containing the type IX collagen NC2 hetero-trimerization domain. In these abbreviations the number means the chain of type I collagen, whereas the position determines the chain of the type IX collagen NC2 domain. For example, 112 is the combination of Iα1-IXα1, Iα1-IXα2, and Iα2-IXα3 (see Fig. 1 for sequences). Note that this abbreviation may not necessarily reflect the α2 chain absolute position in the chain register; it only reflects the relative position as linked to the NC2 domain. Two homo-trimeric combinations were also produced for controls, namely 111 and 222, which was composed solely of either α1 or α2 chains, respectively. All six chains (last sequences in Fig. 1) required for these complexes were expressed separately and then mixed accordingly and purified as described under "Experimental Procedures." The purity and composition of the complexes were analyzed under non- and reducing conditions

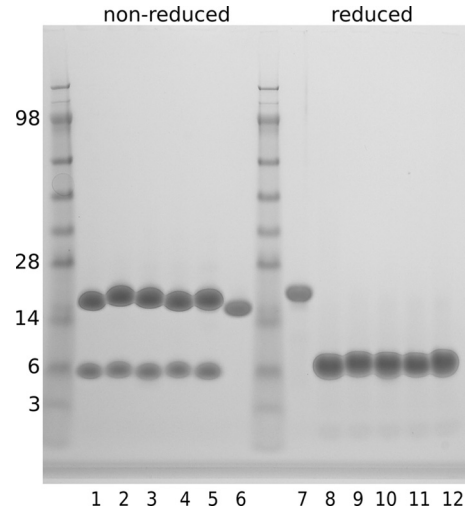

FIGURE 4. **The purity of recombinant proteins.** 4–12% NuPAGE Coomassie-stained gel of the purified proteins 111, 222, 112, 121, 211, and vWFA3 is shown. *Lanes 1–5* and *8–12* show non-reduced and reduced 111, 222, 112, 121, 211, respectively. *Lanes 6* and *7* are non-reduced and reduced vWFA3.

on the SDS-PAGE (Fig. 4). Two bands were observed under the non-reducing conditions, with the higher band representing the disulfide-bonded α1 and α3 chains of NC2. Fig. 4 also demonstrates the purity of the recombinantly produced vWFA3 domain. The reduction of the intrachain disulfide bond of vWFA3 remarkably decreases its migration rate. Sedimentation equilibrium experiments confirmed the trimeric nature of the collagen complexes with no signs of dissociation into individual subunits or higher oligomeric aggregation (supplemental Fig. S5).

Thermal Stability of Hetero-trimeric Complexes Analyzed by CD and Differential Scanning Calorimetry—The thermal denaturation profiles of the collagen triple helix within the hetero-trimeric complexes were assessed using the CD spectroscopy (Fig. 5*A*) and differential scanning calorimetry (Fig. 5*B*). The NC2 domain retained its native trimeric structure within the used temperature intervals (22). Both methods demonstrated the noticeably higher melting temperature for the 112 complex as compared with 121 and 211 (Fig. 5). The detailed analysis of the melting profiles from CD experiments that takes into account the kinetic hysteresis phenomenon was performed (supplemental Fig. S6) in accord with Ref. 32. The thermodynamic parameters derived from this analysis are presented in Table 1. The equilibrium melting temperature of 112 was 1.5–1.9 °C higher than of 121 or 211. The enthalpy gain for 112 was also remarkable. These differences clearly point that the NC2 domain plays a role in chain register of the adjacent chains. Because type I collagen sequences are separated from the NC2 domain by three GPP units (Fig. 1), any interaction between type I and type IX collagen sequences must be excluded.

Collagen Peptides-vWFA3 Binding Assessed by the Affinity Chromatography—Equal amounts of the collagen complexes were sequentially injected onto the vWFA3-coupled column, and the elution profiles were recorded. The column equilibration and the loading and elution of proteins were performed using the same buffer. The elution profiles are shown in Fig. 6*A*. The strongest retention was observed for the type III collagen

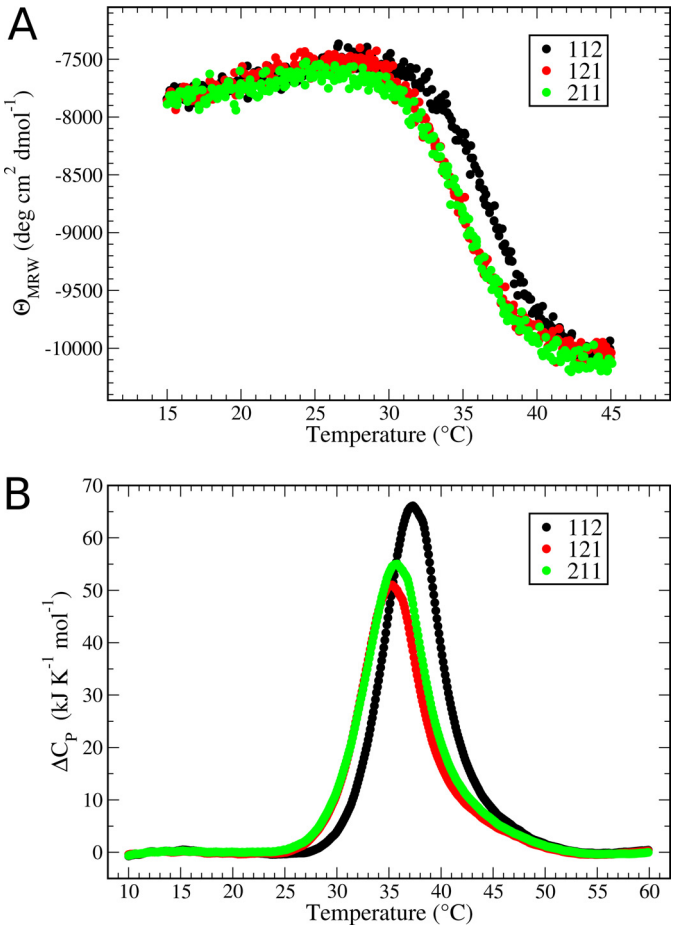

FIGURE 5. **Thermal unfolding profiles of hetero-trimeric constructs.** *A*, circular dichroism transitions of the 112, 121, and 211 complexes in PBS buffer were monitored at 225 nm with the scanning rate of 1 °C/min. *B*, shown is differential scanning calorimetry of 112, 121, and 211 in PBS using the scanning rate of 1 °C/min.

TABLE 1
Thermodynamic parameters of the complexes

Parameters	112	121	211	111	222
ΔH^0 (kJ mol^{-1})	−340.7	−327.4	−326.5	−345.3	−329.0
ΔS^0 (J mol^{-1} K^{-1})	−1102.6	−1066.4	−1062.2	−1121.1	−1067.9
T_m (°C)	35.8	33.9	34.3	34.8	34.9

construct (111.III), which was effectively eluted only with 1 M NaCl (data not shown). The small peak between 1 and 1.5 ml reflects non-interacting impurities in the sample of 111.III. Type I collagen constructs demonstrated various retentions ranging from the strongest one for the 111 homo-trimeric construct to the non-interacting 222. Whereas 112 showed slightly stronger binding than 211, both had the remarkably longer retention than 121 among the hetero-trimeric complexes.

Collagen Peptides-vWFA3 Binding Measured by Surface Plasmon Resonance—The collagen complexes were immobilized on the sensorchip surfaces, and vWFA3 was injected in the concentration range of 0.0002–2.08 mM to monitor the change in the surface plasmon resonance signal in real time, which reflected an amount of bound material. After sample injection, a running buffer was passed over the surfaces to monitor the dissociation. The association and dissociation kinetics were not resolved, and the total response signal was used as a measure of

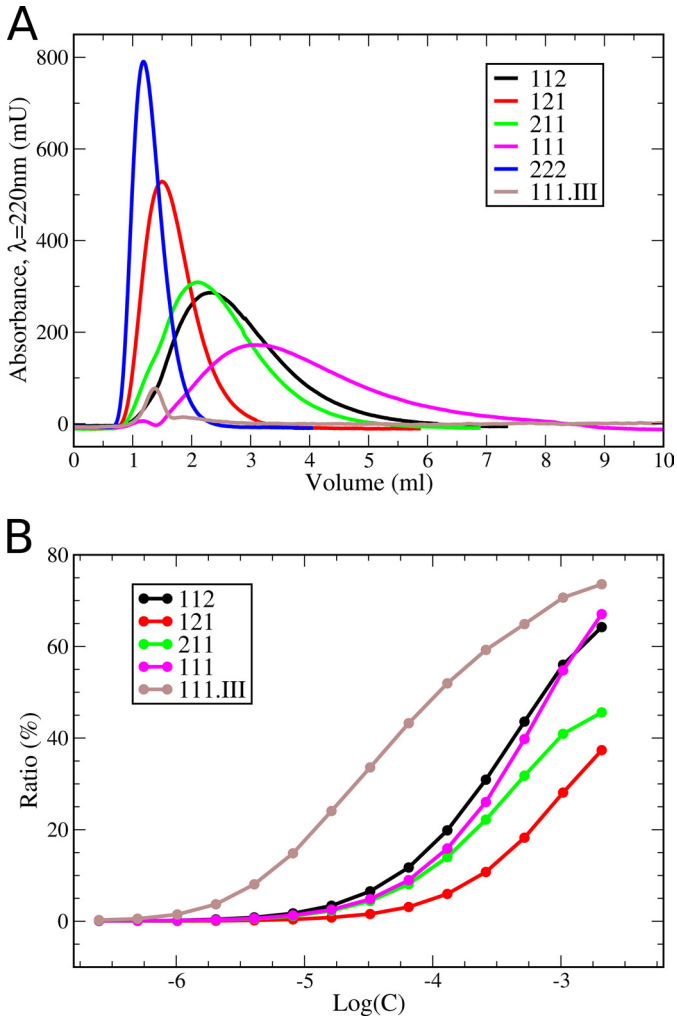

FIGURE 6. **Binding of collagen peptides to vWFA3.** *A*, shown is vWFA3-Sepharose affinity chromatography. Elution profiles of type III collagen model peptide (111.III) and various chain combinations of type I collagen model (111, 222, 112, 121, and 211) are shown. *B*, surface plasmon resonance binding data is shown. The vWFA3 domain binding to the surface-coupled collagen model peptides is shown.

the equilibrium binding. The reference channels were coupled with $(GPP)_7$-NC2XIXB or 222-NC2IX for type III or type I collagen complexes, respectively, in order to subtract nonspecific binding and bulk refractive index effects (supplemental Fig. S7).

The equilibrium binding curves of vWFA3 on the surfaces with the collagen complexes are shown in Fig. 6*B*. For scaling purposes the data are given as ratio between the actual response signals of ligand binding (vWFA3) to the overall change in the coupling response signal of different substrates (collagens). The type III collagen fragment (111.III) demonstrated the highest binding affinity. The type I collagen variants 111 and 112 showed similar affinity. The 121 construct revealed the weakest binding potential among the hetero-trimeric constructs, which is in accord with the affinity chromatography.

DISCUSSION

The only method reported for the stagger-controlled assembly of collagen triple helical peptides with the natural sequences (14) is based on several assumptions. First, the type III collagen cysteine knot sequence GPCCGG specifically locks a certain chain stagger within the triple helix. Second, the simplified version of this sequence GCGG/GCCGG/GPCGG for leading/middle/trailing chains and the regioselective connection between cysteines mimics the natural knot. Neither of these assumptions has been experimentally proven. Moreover, no systematic study has been done to demonstrate any preference of one over any other stagger possibility in any kind of assays. The crystal structures of the natural fragment of the type III collagen (15) and of col3HypCC (this study) did not demonstrate a specific structure of the cysteine knot, which might be a consequence of alternative disulfide connections and/or multiple conformations. Therefore, the use of the knot for the stagger control remains questionable. It is worth mentioning here that type III collagen from a natural source might still have the specific connectivity within the cysteine knot as its structure formation is the result of a set of events such as the C-propeptide trimerization and folding and cleavage of the latter. Moreover, the telopeptide sequences, which extend the GPCCGG sequence, might influence the knot structure.

In this study we explored the stagger control potential of the NC2 hetero-trimerization domain of type IX collagen. Whereas the cysteine knot formation requires the preceding formation of the triple helix (34) or regioselective chemistry (14), the NC2 domain is a potent hetero-trimerization domain that selects and brings three distinct chains together, forms a stable hetero-trimeric complex, and thus initiates the triple helix formation (22). The NC2 domain is an autonomously folding unit, and its assembly should precede the tripe helix formation in order to dictate it. For this reason the stability of the adjacent triple helical sequences was optimized by varying the number of flanking GPP units (data not shown). In other words the ability of collagenous sequences to compete with the NC2 domain for the trimer formation was eliminated, whereas its ability to form the trimer when trimerized by the NC2 domain remained. To demonstrate the stagger control of the NC2 domain, a systematic analysis of all possible stagger variants of type I collagen fragment was performed. Because type I collagen has two α1 and one α2 chains per trimer, we permuted the α2 chain to all three chains of the NC2 domain and supplemented α1 chains accordingly.

Whereas the fusion of type I collagen sequences with the hetero-trimerization domain of type IX collagen was performed to clarify the role of the NC2 domain in stagger determination, all other homo-trimeric proteins were stabilized either by the type III collagen cysteine knot or the NC2 domain of type XIX collagen to ensure the triple helix stability at diluted solutions.

Two alternative methods were used to demonstrate discriminative binding of homo- and hetero-trimeric complexes of type I collagen, *i.e.* affinity chromatography with the immobilized substrate (vWFA3) and surface plasmon resonance with the immobilized ligands (collagens). Whereas the affinity chromatography was controlled by association/dissociation kinetics, the surface plasmon resonance data reflected only the equilibrium binding thermodynamics. This combination provided the comprehensive and complementary data. None of the hetero-trimeric combinations of type I collagen was able to approach the efficiency observed with the type III sequence.

This is somewhat unexpected from recent speculations based on the crystal structure of the complex between type III collagen peptide and the vWFA3 domain (2), where the authors anticipated an identical interface for a specific type I collagen stagger. On the other hand this is in agreement with the previous observation of much lower affinity of type I collagen to vWFA3 compared with type III (35). Our finding supports this experimental observation. Moreover, the ability of the homotrimeric combination 111 of type I collagen to retain longer on the vWFA3-coupled Sepharose than any hetero-trimeric form raise a question of whether the structure-based speculations are appropriate. Among the hetero-trimeric variants of type I collagen, 112 is the strongest binding composition to vWFA3, and 211 has a slightly weaker binding potential. The 121 construct is the least interacting variant. The overall order of interacting potential for the collagen variants is 111.III $\gg$ 111 $>$ 112 $>$ 211 $>$ 121 $>$ 222. Given the strongest binding to vWFA3 among the hetero-trimeric variants and the highest melting temperature, the 112 composition represents the most likely stagger of type I collagen. The lack of structural information on the type IX collagen NC2 domain and its interface with the collagenous portion does not allow us to conclude about the absolute stagger of type I collagen. Nevertheless, the most likely stagger of type I collagen can be practically reproduced by placing the α2 chain of type I collagen in front of the α3 chain of the NC2 domain of type IX collagen and the α1 chains in front of α1 and α2, respectively.

A major hindrance in establishing a full picture of the collagen interactome is very limited availability of different collagen types and inability to control the chain composition and the correct stagger for short collagen fragments for precise mapping of binding sites. The development of collagen toolkits for hetero-trimeric collagens is a big challenge that has not been addressed so far. Toolkits for hetero-trimeric collagens are attractive and highly demanded tools for studying angiogenesis, hemostasis, autoimmunity, tumor growth and inhibition, infection, cell signaling, muscular dystrophies, myopathies, Alzheimer disease, etc. (1). Hetero-trimeric collagens are involved in interactions with pluripotent cell types in our body and thus can be used for development of xeno-free scaffolds for amplification of stem cells. The ability of the type IX collagen NC2 domain to control the chain register in the triple helix can now be practically exploited for the generation of hetero-trimeric collagen fragments of interest with a controlled stagger and development of the hetero-trimeric collagen toolkits.

Acknowledgments—We are grateful to Prof. Miguel Cruz for providing cDNA encoding vWFA3, Dr. Keith Zientek for the mass spectroscopic analysis, Dawn Jennings for the peptide synthesis, Richard Sinclair for the DNA sequencing and the amino acid analysis, and Prof. Michael S. Chapman and Dr. Jay Nix for the x-ray data collection.

REFERENCES

1. Bächinger, H. P., Mizuno, K., Vranka, J. A., and Boudko, S. P. (2010) in *Comprehensive Natural Products II. Chemistry and Biology.* pp. 469–530 Elsevier, Ltd., Oxford, UK
2. Brondijk, T. H., Bihan, D., Farndale, R. W., and Huizinga, E. G. (2012) Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 5253–5258
3. Hofmann, H., Fietzek, P. P., and Kühn, K. (1978) The role of polar and hydrophobic interactions for the molecular packing of type I collagen. A three-dimensional evaluation of the amino acid sequence. *J. Mol. Biol.* **125,** 137–165
4. Orgel, J. P., Irving, T. C., Miller, A., and Wess, T. J. (2006) Microfibrillar structure of type I collagen *in situ*. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 9001–9005
5. Bender, E., Silver, F. H., Hayashi, K., and Trelstad, R. L. (1982) Type I collagen segment long spacing banding patterns. Evidence that the α2 chain is in the reference or A position. *J. Biol. Chem.* **257,** 9653–9657
6. Raynal, N., Hamaia, S. W., Siljander, P. R., Maddox, B., Peachey, A. R., Fernandez, R., Foley, L. J., Slatter, D. A., Jarvis, G. E., and Farndale, R. W. (2006) Use of synthetic peptides to locate novel integrin $\alpha2\beta1$ binding motifs in human collagen III. *J. Biol. Chem.* **281,** 3821–3831
7. Lisman, T., Raynal, N., Groeneveld, D., Maddox, B., Peachey, A. R., Huizinga, E. G., de Groot, P. G., and Farndale, R. W. (2006) A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* **108,** 3753–3756
8. Konitsiotis, A. D., Raynal, N., Bihan, D., Hohenester, E., Farndale, R. W., and Leitinger, B. (2008) Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* **283,** 6861–6868
9. Jarvis, G. E., Raynal, N., Langford, J. P., Onley, D. J., Andrews, A., Smethurst, P. A., and Farndale, R. W. (2008) Identification of a major GpVI binding locus in human type III collagen. *Blood* **111,** 4986–4996
10. Lebbink, R. J., Raynal, N., de Ruiter, T., Bihan, D. G., Farndale, R. W., and Meyaard, L. (2009) Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. *Matrix Biol.* **28,** 202–210
11. Giudici, C., Raynal, N., Wiedemann, H., Cabral, W. A., Marini, J. C., Timpl, R., Bächinger, H. P., Farndale, R. W., Sasaki, T., and Tenni, R. (2008) Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. *J. Biol. Chem.* **283,** 19551–19560
12. Leo, J. C., Elovaara, H., Bihan, D., Pugh, N., Kilpinen, S. K., Raynal, N., Skurnik, M., Farndale, R. W., and Goldman, A. (2010) First analysis of a bacterial collagen-binding protein with collagen Toolkits. Promiscuous binding of YadA to collagens may explain how YadA interferes with host processes. *Infect. Immun.* **78,** 3226–3236
13. Bruckner, P., Bächinger, H. P., Timpl, R., and Engel, J. (1978) Three conformationally distinct domains in the amino-terminal segment of type III procollagen and its rapid triple helix leads to and comes from coil transition. *Eur. J. Biochem.* **90,** 595–603
14. Ottl, J., Battistuta, R., Pieper, M., Tschesche, H., Bode, W., Kühn, K., and Moroder, L. (1996) Design and synthesis of heterotrimeric collagen peptides with a built-in cystine-knot. Models for collagen catabolism by matrix-metalloproteases. *FEBS Lett.* **398,** 31–36
15. Boudko, S. P., Engel, J., Okuyama, K., Mizuno, K., Bächinger, H. P., and Schumacher, M. A. (2008) Crystal structure of human type III collagen Gly-991–Gly-1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *J. Biol. Chem.* **283,** 32580–32589
16. Gauba, V., and Hartgerink, J. D. (2007) Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions. *J. Am. Chem. Soc.* **129,** 2683–2690
17. Fallas, J. A., Gauba, V., and Hartgerink, J. D. (2009) Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J. Biol. Chem.* **284,** 26851–26859
18. 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
19. 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
20. 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. *Ma-*

trix Biol. **30,** 9–15
21. 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
22. 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
23. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM. A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 271–281
24. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242
25. Navaza, J. (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1367–1372
26. Trapani, S., and Navaza, J. (2008) AMoRe. Classical and modern. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 11–16
27. Emsley, P., and Cowtan, K. (2004) Coot. Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
28. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX. A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221
29. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity. All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21
30. McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. (2011) Presenting your structures. The CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 386–394
31. Cruz, M. A., Yuan, H., Lee, J. R., Wise, R. J., and Handin, R. I. (1995) Interaction of the von Willebrand factor (vWF) with collagen. Localization of the primary collagen-binding site by analysis of recombinant vWF A domain polypeptides. *J. Biol. Chem.* **270,** 19668
32. Mizuno, K., Boudko, S. P., Engel, J., and Bächinger, H. P. (2010) Kinetic hysteresis in collagen folding. *Biophys. J.* **98,** 3004–3014
33. Mechling, D. E., and Bächinger, H. P. (2000) The collagen-like peptide (GER)15GPCCG forms pH-dependent covalently linked triple helical trimers. *J. Biol. Chem.* **275,** 14532–14536
34. Boudko, S. P., and Engel, J. (2004) Structure formation in the C terminus of type III collagen guides disulfide cross-linking. *J. Mol. Biol.* **335,** 1289–1297
35. Li, F., Moake, J. L., and McIntire, L. V. (2002) Characterization of von Willebrand factor interaction with collagens in real time using surface plasmon resonance. *Ann. Biomed. Eng.* **30,** 1107–1116