Collagens comprise a large superfamily of extracellular matrix proteins that play diverse roles in tissue function. The mechanism by which newly synthesized collagen chains recognize each other and assemble into specific triple-helical molecules is a fundamental question that remains unanswered. Emerging evidence suggests a role for the noncollagenous domain (NC1) located at the C-terminal end of each chain. In this study, we have investigated the molecular mechanism underlying chain selection in the assembly of collagen IV. Using surface plasmon resonance, we have determined the kinetics of interaction and assembly of the \( \alpha_{1}(IV) \) and \( \alpha_{2}(IV) \) NC1 domains. We show that the differential affinity of \( \alpha_{2}(IV) \) NC1 domain for dimer formation underlies the driving force in the mechanism of chain discrimination. Given its characteristic domain recognition and affinity for the \( \alpha_{1}(IV) \) NC1 domain, we conclude that the \( \alpha_{2}(IV) \) chain plays a regulatory role in directing chain composition in the assembly of collagen IV. Detailed crystal structure analysis of the \([\alpha_{1}2]\) NC1 hexamer and sequence alignments of the NC1 domains of all six \( \alpha \)-chains from mammalian species revealed the residues involved in the molecular recognition of NC1 domains. We further identified a hypervariable region of 15 residues and a \( \beta \)-hairpin structural motif of 13 residues as two prominent regions which mediate chain selection in the assembly of collagen IV. To our knowledge, this report is the first to combine kinetics and structural data to describe molecular basis for chain selection in the assembly of a collagen molecule.

Collagens comprise a major superfamily of extracellular matrix (ECM) proteins that play a key role in the structural integrity of all tissues. At least 27 different collagen types, consisting of 42 distinct gene products have been identified in vertebrates, underlining their vast diversity in biological functions such as tissue compartmentalization and specialization during the development (1). In the endoplasmic reticulum (ER), the newly synthesized collagen chains assemble into triple-helical molecules with specific chain compositions, which oligomerize to form supramolecular structures including filaments and networks after secretion to the ECM. Some collagens are obligate homotrimers, such as collagen III which comprises three identical \( \alpha_{1}(III) \) chains forming a \( \alpha_{1}(III)_{3} \) procollagen, whereas others form heterotrimers containing at least one different \( \alpha \)-chain; such as collagen I (2). Despite significant sequence identity and a propensity to form triple helices, procollagen chains have an extraordinary ability to discriminate between each other in the ER to form specific collagen types. For example, skin fibroblasts express six highly homologous but genetically distinct fibril-forming procollagen chains that are assembled in a type-specific manner to form type I, III and V collagens. The mechanism by which different collagen chains are selected for assembly is a fundamental question which remains unanswered.

Emerging evidence suggests that the C-terminal propeptide of the procollagen molecules plays a key role in the chain-specific recognition (3-8). Bulleid and colleagues demonstrated that transferring the C-propeptide of \( \text{pro} \alpha_{1}(III) \) chain to the corresponding region of the naturally heterotrimeric \( \text{pro} \alpha_{2}(I) \) molecule was sufficient to
direct the new hybrid molecules to form stable homotrimers (9). They also identified a discontinuous variable sequence motif of 15 residues within the C-propeptide of proα1(III) chain, which when transferred to the corresponding region of proα2(I) chain, enabled those chains to form homotrimers. The 3-dimensional location of the motif awaits high resolution structural analysis, but analytical ultracentrifugation and small angle X-ray scattering studies suggest that chain-specificity is defined by the core of a three-lobed structure of the C-propeptide of procollagen III (10,11).

In the case of network-forming collagens, parallel findings have indicated that the NC1 domains encode chain recognition for molecular assembly (12,13). We have shown that the NC1 domains have an extraordinary capacity to reassemble in vitro to their native hexamer structure (12,13). A recent cell culture study has also shown that the NC1 domain is necessary to initiate and direct triple-helix formation and assembly of collagen IV (8).

Collagen IV provides a unique opportunity to delineate the molecular basis of chain selection in the assembly of triple-helical collagen molecules. It consists of 6 homologous α-chains (α1-α6) which out of 56 possible combinations assembles only into 3 specific trimeric molecules (α1.α1.α2, α3.α4.α5 and α5.α5.α6), reflecting a remarkable specificity (12-15). Importantly, the recently resolved crystal structure of its NC1 domains has provided detailed structural information to decipher 3-dimensional features of the recognition sites (16,17). The capacity of NC1 domains to reassemble in vitro provides the means to study the mechanism of assembly. Furthermore, the primary structures of the six α-chains from several mammalian species are known, providing the means for comparative analysis between sequence variability and chain specificity across species.

In the present study, we have studied the molecular mechanism and the structural basis for chain selection in the assembly of collagen IV. We have utilized a surface plasmon resonance (SPR) biosensor and designed specific mathematical models to determine the kinetics of interaction and assembly of isolated α1(IV) and α2(IV) NC1 domains as the classical model for the collagen IV assembly. We show that the higher affinity of α2(IV) NC1 monomer for dimer formation with the α1(IV) NC1 monomer directs the molecular stoichiometry of chains and assembly of (α1)2α2 trimer. Furthermore, crystal structure analysis of the NC1 hexamer and sequence alignments of all six α-chains across mammalian species identified the molecular recognition sites which include a hypervariable region of 15 residues and a 13-residue long β-hairpin motif involved in the domain swapping mechanism described earlier (16). At the domain interface, the two regions are located adjacent to each other and directly involved in the monomer-monomer interactions. Given its genetic hypervariability across all six α-chains and its critical location at the domain interface, the 15-residue long hypervariable region is thought to play a major role in the early mechanism of chain selection and assembly of collagen IV.

**MATERIALS AND METHODS**

**Purification of NC1 domains.** Frozen Bovine eye lens capsule were purchased from Pel-Freeze Biological (Rogers, AR). The NC1 hexamer was prepared by digestion of lens basement membrane (LBM) using bacterial collagenase, followed by purification on a DE-52 ion-exchange chromatography and a Superdex S-300 gel chromatography column (Amersham Biosciences) as described previously (18). Monomers of α1 and α2 NC1 domains were separated from homo- and heterodimers by chromatofocusing using a Mono-P column (Amersham Biosciences) as described (18). The fractions were analyzed by Western blot analysis using chain-specific monoclonal antibodies against α1(IV) and α2(IV) chains as described (12).

**Acid-dissociation and in vitro reconstitution.** Equal volumes (0.5 mL) of native NC1 hexamer (0.5 mg/mL) were loaded into 0.5 mL slide-A-Lyzer dialysis cassettes (PIERCE, Rockford, IL) with a Mw cut off value of 10,000 and equilibrated overnight with 10 mM KHPO₄ buffers with pH adjusted to 7.4, 5.0 and 4.5. For complete dissociation a sample was equilibrated overnight in 10 mM citrate buffer, pH 3.0. For the reconstitution study, a pH 3.0-dissociated sample was re-equilibrated overnight against 10 mM KHPO₄ buffer, pH 7.4. After equilibrium, a
volume (50 μL) of each sample was analyzed by size-exclusion chromatography using a Superdex 200 (S200 HR 10/30) column (Amersham Biosciences) connected to a HPLC system (ÄKTApurifier, Amersham Biosciences). The column was equilibrated with the same buffers containing 150 mM NaCl and the samples were run at the constant flow rate of 0.5 mL/min. The elution volume of each protein species was compared to that of protein standards as described below. A sample of monomer fractions from both α1 and α2 NC1 domains isolated by chromatofocusing were also brought to equal concentration, dialyzed against 10 mM KHPO₄ buffer (pH 7.4) and analyzed on the gel filtration column. The remaining of the samples after equilibration was used for circular dichroism study as described below.

**Calculation of molecular mass and Stokes radii.** Molecular mass and stokes radius for each protein was estimated by comparing their peak elution volumes (Vₑ) to a linear standard graph, obtained by plotting the logarithm of Mₑ versus Vₑ of selected protein standards with a known molecular mass and Stokes radius: thyroglobulin (Mₑ 669,000, Stokes radius 85.0 Å), gamma globulin (Mₑ 158,000, Stokes radius 52.2 Å), ovalbumin (Mₑ 44,000, Stokes radius 30.5 Å), and myoglobin (Mₑ 17,000, Stokes radius 20.8) as described (19). The protein standards were run in the same buffers and running conditions as for each study described above.

**Circular Dichroism.** Protein samples from acid-dissociation and in vitro-reconstitution study were diluted with HPLC-grade water to a concentration of 0.15 mg/mL. The CD spectra were recorded for each sample at 25 °C on a Jasco J-810 CD spectropolarimeter using 1 nm steps from 195-250 nm and a 0.1-cm path length. Buffers were scanned three times and their average data were generated. Protein sample was also scanned three times and their average data were calculated and buffer subtracted. Mean residue ellipticity [θ], expressed in deg.cm².dmol⁻¹, was calculated on the basis of mean residue mass (MRW) of 110 per residue and a total number of 240 amino acid residues per NC1 domain according to the equation [θ] = 100 ⋅ θₑₛ/C ⋅ n ⋅ l , where θₑₛ is observed ellipticity in millidegrees, C is the concentration of the protein in mM, n is the number of amino acid residues and l is the length path in cm.

**Surface Plasmon Resonance.** Binding analysis based on surface plasmon resonance (SPR) was performed using a BIAcore 2000 optical biosensor (BIAcore AB, Uppsala, Sweden). Analyte-ligand interactions were measured in resonance units (1 RU = 1 pg/mm² bound protein).

**Ligand immobilization.** All proteins were dialyzed against running buffer, TBST-EDTA (10 mM Tris, 0.15 M NaCl, 0.005% surfactant Tween 20, 3.0 mM EDTA, pH 7.4). Ligands to be immobilized were diluted in immobilization buffer (10 mM Na-Acetate buffer, pH 3.0) to a final concentration of 20 μg/mL and immobilized in equivalent molar ratios. Amine-coupling chemistry was used to immobilize the ligand proteins on carboxymethylated dextran surfaces of CM5 research-grade sensor chips (Biacore AB, Uppsala, Sweden). Prior to immobilization, the sensor chips were washed twice with 1 M ethanolamine (pH 8.0) at 20 μL/min for one minute each, followed by two additional washes with running buffer. The flow cell surfaces were activated by injecting a 1:1 mixture of 0.05 M N-hydroxysuccinimide and 0.2 M 3-(N,N-dimethylamino)propyl-N-ethylcarbodi-imide at a flow rate of 20 μL/min at 25 °C for 10 min. For immobilization, ligands were injected at a flow rate of 10 μL/min. Different CM5 chips with both high density (ca. 5000 RU) and low density (ca. 200 RU) of ligand-surfaces were generated to examine the binding at different ligand-densities. Flow cells 1 and 2 were used to immobilize α1 and α2 NC1 domains, respectively. BSA as negative control was immobilized on flow cell 3 and was used for background subtractions. The fourth flow cell was treated only with activation and blocking solutions and was used as a blank surface for monitoring the refractive indices of buffers and samples. After immobilization, the remaining reactive esters on the carboxymethyl-dextran surfaces were blocked with 1 M ethanolamine (pH 8.0).

**Protein interaction analysis.** All interactions were performed at 25 °C. Analyte samples were diluted in the running buffer and injected in triplicates at different concentrations ranging from 0.25, 0.5, 1.0, 2.0 to 4.0 μM. The flow rate for interaction was set at 20 μL/min with a contact
time for 300 seconds followed by dissociation for an equal time. After each cycle, the surfaces were regenerated by two subsequent injections of 40 μL of 1M Ethanolamine (pH 8.0) followed by equilibration with the running buffer for 10 minutes.

**Data analysis and determination of kinetic constants.** All data processing including scale transformation, background and reference subtractions were performed using BIAevaluation 3.0 (BIAcore AB, Uppsala, Sweden). Data evaluation and calculations based on numerical approaches including global curve fitting to interaction models and calculation of kinetic constants were performed with IGOR Pro (Version 4.01.A, WaveMetrics, Inc., Lake Oswego, OR). Curve fit functions based on specific interaction models for trimers and hexamers were arranged as external operation routines, written and compiled in C, using CodeWarrior IDE (Version 5.1.1.1105, Metrowerks Corporation, Austin, TX) and applied together with the Global Fit Procedure of IGOR Pro, based on a nonlinear least-squares method, utilizing the Levenberg-Marquard algorithm (20). For numerical integration, a fifth order Runge-Kutta-Fehlberg method was applied (20). The applied interaction models assumed a mass transport limited process based on the previously described two compartment model (21). Mathematical descriptions of these models are provided in the attached supplemental data.

The flow cell diffusion layer height ($h_{\text{diff}}$) and the mass transport coefficient ($k_c$) at a given flow rate were obtained according to a previously described approach (22). Global curve-fit calculations were performed with fixed values for the constant parameters $M_w$, $h_{\text{diff}}$ and $k_c$, whereas the global parameters $k_a$, $k_d$ and $R_{\text{max}}$ were defined as free-running parameters to be determined by the global fit routine. However, the $R_{\text{max}}$ value was constrained by upper limit equal to a theoretical surface saturation with the highest $M_w$ complex of the particular interaction model. The local parameter $C_d$ (analyte monomer concentration) was also defined as a free running variable but with upper and lower constraints within the range of total analyte concentration. Curve fit routines were in all cases trimmed to keep the amount of fit variables at minimum.

**Analysis of intermolecular contacts and structural graphics.** The intermolecular contacts between the α-chains were studied using DIMPLOT, which is included within the software package LIGPLOT (23). The analyses were performed in conjunction with HBPLUS (24). Three types of bonds were analyzed: hydrogen bonds, ion pairs (salt bridges), and non-polar bonds. The geometric criteria defining hydrogen bonds and ion pairs, consisting of minimum angles and maximum distances, were as previously defined (25). Non-polar bonds were defined as residues interacting, not with hydrogen bonds or as ion pairs, but rather though hydrophobic and van der Waal interactions. The calculations were performed on a chain pair basis, excluding crystal packing contacts. The LIGPLOT output was modified in Adobe Photoshop to optimize the visual clarity of the residue positions. Structural graphics were performed by The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA, USA, http://www.pymol.org).

Primary sequences of α1-α6 chains. The complete amino acid sequences of NC1 domains of α1-α6 chains of mammalian collagen IV were obtained from the NCBI Entrez (http://www.ncbi.nlm.nih.gov/).

**RESULTS**

**Structural features and stability of NC1 hexamer.** We and others have recently reported the atomic structure of the NC1 hexamer (16,17). The structure provides a foundation on which to elucidate the mechanism of molecular assembly and identify the recognition sites for interaction. Thus we briefly present key features relevant to this topic.

Each NC1 hexamer consists of two identical hetero-trimers with a molecular stoichiometry of two α1 and one α2 NC1 domains. The heterotrimers form a virtually flat surface of about 4400 Å², through which they interact end-to-end to form a compact ellipsoid-shaped molecule (Fig. 1A). The hexamer is stabilized through extensive hydrophobic and hydrophilic interactions without a need for disulfide bonds, whereas the trimers are formed through a 3-dimensional domain swapping mechanism (16). In these types of interactions, a β-hairpin from each chain is swapped into a four-stranded anti-parallel β-sheet from the flanking chain to form a six-stranded anti-parallel β-sheet
structure, providing a stable inter-subunit contact (Fig. 1B and C). Each NC1 subunit possesses two sites through which it interacts with the neighboring subunits; one site covers the β-hairpin and the other includes the docking site for the domain swapping (Fig. 1C). Since the α2 NC1 domain forms two different heterodimers with the neighboring α1 NC1 domains, each α1 NC1 domain has been indicated with different color to distinguish them (Fig. 1B).

To establish a deeper understanding about the intermolecular forces involved in chain selection and stability of NC1 domain interactions within the hexamer, we performed a series of acid dissociation and reassembly studies, circular dichroism at different pH, and structural analysis of the X-ray crystal structure of the hexamer as described below.

**Acid dissociation and reassembly of NC1 domains.** Gradual and pH-dependent dissociation of the native NC1 hexamer and its reassembly was examined by analytical gel filtration. The native hexamer was prepared from bovine lens capsule basement membrane (LBM), since its collagen fraction consists almost entirely of α1(IV) and α2(IV) chains with lowest degree of covalently crosslinked NC1 domain (26).

As shown in figure 2A, the native hexamer migrated as a single protein complex at pH 7.4 with a relative molecular mass of 85±4 kDa (calculated Stokes’ radius of 43±2 Å). This molecular migration is significantly slower than what is expected from the calculated molecular mass of a globular hexameric molecule consisting of six NC1 domains (6 × 25 = 150 kDa). We propose that this is most likely due to the domain swapping interactions, which result in a highly compact and volumetrically smaller molecular shape – hence resulting in slower migration in a gel chromatography column. Nonetheless, from the calculated Stokes radius, a theoretical molecular diameter of 86±4 Å can be derived, which is in good agreement with the molecular dimensions of the hexamer (95 x 70 Å) reported by X-ray crystallography (17).

At pH 5.0 the hexamer was still intact, however, at pH 4.5 it began to dissociate as revealed by appearance of a minor shoulder peak which could correspond to a trimer peak (Figure 2A). At pH 3.0, the hexamer is almost completely dissociated into monomers migrating as a major peak with a relative molecular mass of 24±1 kDa (calculated Stokes’ radius of 27±1 Å). A minor fraction of the hexamer did not dissociated into monomers at pH 3.0 and migrated as a dimer with a relative molecular mass of 46±2 kDa (Stokes’ radius 35±2 Å). This minor fraction (ca. 20%) corresponds to covalently crosslinked dimers (27).

Finally, when a sample of the dissociated hexamer was returned to neutral pH by overnight dialysis, the hexamer configuration was regained, indicating that a mixture of α1(IV) and α2(IV) NC1 monomers can be reassembled to a hexamer, as it was reported earlier (12).

By circular dichroism, we further investigated the pH-dependent conformational change of the NC1 domains during acid-dissociation of the hexamer and reassembly in neutral pH. The CD spectra of the native hexamer at pH 7.4 in the far-UV range (195-250 nm) was compared to samples at pH 5.0, pH 3.0 and to that of a reassociated hexamer in pH 7.4. The CD spectra for all samples demonstrated a negative band peaking at 199 nm and a positive band with a broad maximum at 225-235 nm (Figure 2B). The negative peak at 199 nm is likely to derive from the unordered random coil structure of the NC1 domains, which in the crystal structure comprise about 35% in α1 and 41% in α2 NC1 domain. In the crystal structure, the β-strands constitute about 53% of the hexamer structure. However, the lack of characteristic β-sheet CD signals is most probably due to the significant contribution of the CD spectra from the unordered portion of the structure, as mentioned above, and the abundance of aromatic residues and disulfide chromophores which can make a strong contribution in the CD spectrum in the range of 215-235 nm (28,29). In fact, the notion that the positive band at 225-235 nm disappeared in the denatured sample at pH 3.0 indicates that this positive band is most likely to arise from the aromatic residues in the native structure of the hexamer.

Nonetheless, based on the apparent difference in the CD spectra of the acid-dissociated sample at pH 3.0 and the spectra of other samples, it could be concluded that a complete dissociation of the hexamer occurs at pH 3.0 due to changes in the secondary structure. This conformational change seems to be reversible as the CD spectra for the
refolded sample at pH 7.4 aligned with that of native structure (Fig. 2B). This is also in agreement with our gel filtration data, showing that the native hexamer configuration can be recovered once the pH is changed to 7.4 (Fig. 2A).

To correlate the pH-dependent dissociation study with the structure, we further identified the intermolecular parameters that define the stability of trimers and hexamers. Table 1 summarizes the number of polar interactions (hydrogen bonds and ion pairs) and non-polar contacts (hydrophobic and van der Waals) between the NC1 domains within a trimer (side-by-side interactions) and between the trimers (end-to-end interactions). From the table, one can conclude that the total number of non-polar interactions both within a trimer (175 total) and between the trimers (110 total) are the dominating type of interactions. However, the number of ion pairs between the trimers (13 total) is significantly higher than that within a trimer (4 total), indicating a greater sensitivity to low pH for the trimer-trimer interaction. As illustrated in figure 2C, there are 6 conserved ion pairs (Arg77-Glu176) between the chains A/D, B/F, and C/E at the trimer interface. At the surface of each trimer, there is also a conserved glutamate (Glu41 in A/B chains and Glu38 in C chain) that form ion pair or pH-sensitive hydrogen bond with proton-donors from the opposite trimer (Fig. 2C). The loss of these ion pairs upon acid treatment will impact inter- and intramolecular stability, resulting in rapid dissociation of the hexamer into trimers. Indeed, our acid-dissociation studies clearly showed that the hexamer peak separated into two peaks, of which the smaller one most likely corresponds to a trimer (Fig. 2A, pH 4.5). The greater number of pH-sensitive ion pairs at the trimer surface also supports the gel filtration data indicating that hexamer dissociation occurs first at the trimer interface followed by dissociation of each trimer into dimers and monomers.

As for side-by-side interactions (Fig. 2C) within each trimer (e.g. ABC), we found four ion pairs, out of which three locate at the B-C interface (Glu113-Lys230, Lys57-Asp125, Arg41-Glu153), and one at the A-C interface (Glu153-Lys42). At the A-C interface, there is also an aspartate (Asp70) that forms hydrogen bond which upon protonation will dissociate.

In the hexamer there is additional glutamate and aspartate residues that do not directly participate in intermolecular interactions, but upon protonation in acidic environment are likely to reduce the hydrogen-bonding potential, aside from disrupting ion pairs, resulting in intramolecular destabilization and intermolecular dissociation.

We next examined the ability of α1 and α2 NC1 domains to form homohexamers. Monomers of α1 and α2 NC1 domains were separated from each other by the means of chromatofocusing, which simultaneously dissociated and resolved the hexamer into monomer fractions of α1 (fractions 4-6) and α2 (fractions 17-20) NC1 domains by pH gradient (Fig. 3A). These fractions were dialyzed against TBST-EDTA buffer (pH 7.4) at a concentration of 0.5 mg/mL overnight and were analyzed by gel filtration. As shown in figure 3B, the α1(IV) NC1 domain migrated as both monomer and hexamer with almost equal ratio (55% monomer and 45% hexamer), whereas the α2(IV) NC1 domain migrated almost entirely (96% of total protein) as monomer (Fig. 3B). In the case of α1 NC1 domain, the ratio between the hexamer and monomer peaks was concentration-dependent and could be shifted towards more hexamer by increasing the total concentration (data not shown). This indicated a concentration-dependent equilibrium shift.

Thus in agreement with the previous observations (12), it was concluded that a hexamer configuration could be obtained by either a mixture of isolated α1 and α2 NC1 domains or by α1 NC1 domain alone (Fig. 3C). With these analyses, we further established that the isolated NC1 domains used in the present study were properly refolded after acid-treatment and had the capacity to reassemble.

The inability of the α2 NC1 domain to form homohexamer, even at higher concentrations, strongly indicated that the molecular stoichiometry of chains is regulated by the specific interactions of certain combinations of NC1 domains. This together with the finding that within a trimer, the number of H-bonds and ion pairs at each binding site varies significantly between α2 and α1 NC1 domains (Table 1) further raised the question as to which factors regulate the stoichiometry and selection of chains in the assembly of collagen IV. To address this question, we carried out a
systematic analysis of the kinetics of interaction between the individual NC1 domains, utilizing surface plasmon resonance (SPR) technology.

Kinetics of interaction by surface plasmon resonance biosensor. In order to investigate the affinity between the NC1 domains, we systematically examined the kinetics of interaction of isolated α1 and α2 NC1 domains using a SPR biosensor. To ensure that the domains were in monomeric form when immobilized, the isolated NC1 domains were immobilized at pH 3.0 using the amine-coupling chemistry (see Materials and Methods). The choice of amine-coupling chemistry for immobilization at low pH was an essential step in our experimental procedure and excluded a potential ligand dimerization or complex formation prior to or during immobilization. Other immobilization methods, such as capturing by immobilized antibodies or streptavidin, were inappropriate since the NC1 domain could not be captured at low pH using these methods.

Binding activity and specificity was verified by injecting a sample of 50 μM of α1 NC1 domain on a surface of α2 NC1 domain immobilized at high surface density (~1500 RU). BSA was immobilized at similar surface density as negative control. As shown in figure 4A, interaction of α1 NC1 domain to immobilized α2 NC1 domain resulted in a strong and specific binding response as compared to the BSA control surface.

We next examined the kinetics of interactions by injecting a series of different concentrations (0.25, 0.5, 1.0, 2.0 and 4.0 μM) of isolated NC1 domains over parallel surfaces of α1 and α2 NC1 domains immobilized at a surface density of 200 RU. Sensorgrams were recorded for both homotypic (α1-α1 and α2-α2) and heterotypic (α1-α2 and α2-α1) interactions. Analysis of sensorgrams revealed specific and concentration-dependent binding for all four combinations of interactions, however, the binding responses as well as the association and dissociation profiles varied significantly between each set of interaction (Fig. 4B-E). The highest binding response was obtained when the α1 NC1 domain was injected over an immobilized α2 surface (Fig. 4B), whereas the lowest binding response was found for α1-α1 interaction (Fig. 4E). We further sought to systematically analyze the sensorgrams by mathematical models available in the BIAsolution software; however, none of the provided models returned satisfactory curve fit (supplementary Fig. 1), except for the α2-α2 interaction which was best described by a simple 1:1 interaction with conformational change (Fig. 4D).

To explore new kinetic models that could best describe the mechanism of interaction, we developed several specific mathematical models, two of which were a heterotrimer model and a homohexamer model. The models were designed taking the in vitro reconstitution studies (Fig. 2 and 3) and the crystal structure of the NC1 hexamer in consideration (16). Since we found that the number of contact sites at each monomer-monomer interface varied significantly (Table 1), the curve-fit routines for the heterotrimer model were designed to distinguish between the kinetics of these two side-by-side interactions, herein referred to as site 1 and site 2 (Fig. 1 C). Schematic diagram and description of the kinetic models are presented in the supplementary data (supplementary Fig. 2 and kinetic models).

By the use of global curve fit analysis, the mathematical models were successfully fitted to the recorded sensorgrams and the kinetic rate constants were calculated (Table 2). The heterotrimer model (supplementary Fig. 2A) was perfectly fitted to the experimental data generated from α1-α2 interaction (Fig. 4B). As we had predicted, the α1-α2 interaction revealed a clear difference in kinetics of interactions through sites 1 and 2. This difference in affinity was reflected in association (kₐ) and dissociation (kₐ) rate constants for each site. The association rate between α1 and α2 NC1 domains was found to be 7 times faster through site 1 than through site 2; however, dissociation rate through site 2 was 60 times slower (Table 2). This clearly indicated that the α2 NC1 domain formed a more stable complex with α1 NC1 domain through site 2, resulting in 8-fold lower equilibrium affinity constant, Kᵢ (Table 2).

Interestingly, the sensorgrams generated for the α2-α1 interaction in which α1 was the ligand (Fig. 4C) fitted best to a simplified version of the
heterotrimer model. In this simplified version, the last interaction step describing the closure of the trimer was excluded (supplementary Fig. 2A), indicating that a stable heterotrimer of (α2)2α1 is less likely to exist. This observation is also consistent with the fact that no heterotrimer with the stoichiometry of two α2 chains and one α1 chain has ever been reported.

The sensorgrams deriving from the α1-α1 interactions (Fig. 4E) were best fitted to the homohexamer model (supplementary 2B), supporting the experimental data that the isolated α1 NC1 monomers can assemble to form homohexamer in solution (Fig. 3B). Other models that were tested on these recordings did not return statistically convincing fit-results (supplementary Fig. 1e-g), indicating that the hexamer model was an appropriate model to describe the α1-α1 interaction. As indicated by the profile of the sensorgrams (Fig. 4E), relatively rapid association and dissociation rate constants were obtained for the formation of the α1-homodimer (Table 2). Formation of the α1-homotrimer was relatively slow resulting in an equilibrium affinity constant of 13 μM. The fastest association rate constant was calculated for the formation of the hexamer, resulting in a relatively lower equilibrium dissociation constant of 0.4 μM.

As shown in figure 4F, change in the distribution of different forms of α1 NC1 domain is dependent on total protein concentration. The combination of rate constants calculated for this homogeneous system indicated a strong equilibrium shift towards the hexamer form as the total protein concentration increases. The relative distribution of different NC1 forms indicated that at a given total protein concentration, for example 1 mg/ml, about 60% of the NC1 domains are in the hexamer form, 20% in the dimer form whereas each of the trimer and monomer forms comprise 10% of the total concentration.

Attempts to fit the sensorgrams for α2-α2 interaction (Fig. 4D) to other models such as analyte dimer or homohexamer models failed (supplementary Fig. 1), supporting the previous notions that the α2 NC1 monomer is unable to form trimers or hexamers (Fig. 3B). Instead, a satisfactory curve fit was achieved using a simple 1:1 interaction model with conformational change, indicating that the α2 NC1 domain could at least form a homodimer.

From these kinetic studies, we concluded that the most stable complex is the naturally occurring heterotrimer (α1)2α2. The α2 NC1 domain showed a differential affinity for dimer formation with α1 NC1 domain, indicating that the initial assembly of the heterotrimer may preferentially be directed by the heterodimer formation through the highest affinity binding site (site 2 of α2 NC1 domain). Our analysis further indicated that formation of a closed heterotrimer with two α2 and one α1 NC1 domain was unfavorable perhaps due to the instability of the complex which is most likely caused by the steric hindrance mentioned in the Discussion. Perhaps therefore, such heterotrimer has not been found in vivo or in vitro. The α1 NC1 domain could form stable homodimer, trimer and hexamer in solution, further supporting the current (Fig. 3A) and previous observation (12).

When comparing the overall binding responses for different set of interactions, the α1 as ligand basically resulted in relatively lower R_max values as expected (Fig. 4 and Table 2). The reason for this relatively lower binding signal could be related to the self-interaction property of α1 NC1 domain in solution, resulting in lower concentrations of free monomers for interaction (Fig. 4F). On the other hand, the R_max value for the (α1)α2 heterotrimer (94±1 RU) was approximately twice as much as that for the α2α2 homodimer (45±5 RU), which is in good agreement with the relative values expected for these interactions (Fig. 4B and 4D, respectively). Disregard the discrepancy in R_max values, application of different kinetic models on identical interaction steps in different interaction setups returned similar kinetic rate constants (Table 2), indicating the consistency of the applied kinetic models.

Structural basis for difference in binding affinity between site 1 and site 2. Since our kinetic analysis revealed that the α2 NC1 domain displays different affinities for interaction through site 1 and site 2, we further sought to find structural basis for the differences. To do so, we did a comprehensive analysis of the primary, secondary and tertiary structures of the NC1 domains of both α1 and α2 chains from bovine, human and mouse.
As a fourth criterion in discrimination between site 1 and site 2, we also identified and mapped the amino acid residues involved in direct contacts between the NC1 domains within a trimer (marked with asterisks in Fig. 5A).

Based on sequence identity, each NC1 domain can be divided into two highly similar half-domains; an N-terminal half (upper panel in Fig. 5A) followed by a C-terminal half domain (lower panel), as noted previously (30). At the 3D-level, the N-terminal half covers contact residues that involve the \( \beta \)-hairpin domain swapping and constitute site 1 for interaction while the C-terminal half covers contact residues that represent site 2. Both NC1 domains share a high degree (64\%) of sequence identity (shown in gray), whereas \( \sim 34\% \) of the residues are specific for \( \alpha 1 \) (shown in red) and \( \alpha 2 \) NC1 domains (shown in blue). As we reported earlier (16), the only region with significant secondary structural difference was found at the C-terminal end of the NC1 domains, where sequence variation have resulted in formation of a \( 3_{10} \) helix in the \( \alpha 1 \) NC1 domain as compared to an extended \( \beta \)-strand conformation in the \( \alpha 2 \) NC1 domain (Fig. 5B). Differences in the 3-dimensional structure of \( \alpha 1 \) and \( \alpha 2 \) NC1 domains were calculated as the root-mean-square deviation (RMSD) between the main chains of both domains. From the RMSD values, it was concluded that the overall conformation of the \( \alpha 1 \) and \( \alpha 2 \) NC1 domains are also strikingly similar, except for three short variable regions where the conformational deviation exceeds a distance of 4Å (Fig. 5A). Two of these variable regions, VR1 with 33\% and VR2 with 42\% sequence identity, are located in the N-terminal half domain while the third one, VR3, is located near the C-terminal end of the molecule, where the lowest sequence identity (25\%) and highest conformational deviation (6Å) is found (Fig. 5A).

While the VR1 and VR2 regions showed minimal number of contact residues, this number was particularly high within two other regions; the \( \beta \)-hairpin region in the N-terminal half domain and the VR3 region located in the C-terminal half domain. The \( \beta \)-hairpin motif, which also comprises the domain swapping part of the molecule, is a 13 residues long motif, out of which 12 are involved in monomer-monomer interaction. However, despite some sequence variations at this region, its secondary and tertiary structures remain conserved in both \( \alpha 1 \) and \( \alpha 2 \) NC1 domains (Fig. 5A). On the other hand, the VR3 was distinguished as the only region that: i) contains high sequence variability between both NC1 domains, ii), is the only site of difference in the secondary structure, and iii) covers a larger number of variable residues involved in direct interdomain interactions. Thus, VR3 was recognized as a unique region which most likely plays a key role in domain recognition.

We further investigated whether the VR3 could be involved in the higher affinity binding of \( \alpha 2 \) NC1 domain through site 2 (Table 2). To this end, we identified and characterized all the interdomain contact sites within this region and the binding partners on the other domains. Using the LIGPLOT analysis software (23) designed to calculate protein-ligand interactions, we identified all the hydrogen bonds and non-polar contacts at this region. Figure 6 illustrates the 2-dimensional presentation of all contact sites for the corresponding VR3 at the \( \alpha 1A-\alpha 1B, \alpha 1B-\alpha 2C \) and \( \alpha 2C-\alpha 1A \) interfaces. Interestingly, we found that the number of hydrogen bonds and hydrophobic contacts for \( \alpha 1A-\alpha 1B \) and \( \alpha 2C-\alpha 1A \) interfaces were similar, whereas the \( \alpha 1B-\alpha 2C \) interface showed a far more complex pattern of interactions both in respect to hydrogen bonds and hydrophobic contacts. This analysis clearly showed that the \( \alpha 2 \) NC1 domain interacts considerably stronger to one of the \( \alpha 1 \) NC1 domains (the \( \alpha 1B \)), supporting our kinetic findings that the \( \alpha 2 \) NC1 domain has an asymmetric affinity for interaction through each binding site. Of particular interest is the side chain of a lysine residue (Lys56) on the \( \alpha 1B \) NC1 domain which in addition to non-polar bonds also form two H-bonds with Glu196 and Ile194 and a salt bridge with Asp121 on the \( \alpha 2C \) NC1 domain (Fig. 6, B-C interface).

The impact of the VR1 and VR2 on the stability of dimer formation was also investigated; however, none of these regions were located at the domain interfaces (Fig. 5B) and had minimal number of residues involved in interaction. Thus, our structural analyses clearly distinguished the VR3 of \( \alpha 2 \) NC1 domain as a critical site which confers higher stability for dimer formation. Hence, in agreement with our kinetic data, we
assigned the contact residues on the C-terminal half domain to the site 2 which confers higher affinity for interaction, and designated the site 1 to the contact residues within the N-terminal half domain (Fig. 1 and supplementary Fig. 2A).

Identification of recognition motifs for chain selection. The fact that out of 56 potential chain combinations, only three (α1α1α2, α3α4α5 and α5α5α6) have been found to form trimeric molecules in vivo indicate a high degree for chain selection and underscores an evolutionary conserved mechanism for specific chain selection in the assembly of collagen IV molecules. To further identify recognition motifs involved in chain selection, we aligned the amino acid sequences of the NC1 domains of all six mammalian collagen IV α-chains from human, chimpanze, bovine and mouse (supplementary Fig. 3). As expected, the alignment falls into two subtypes; one includes the α1-like chains (α1, α3 and α5 chains), and the other includes the α2-like chains (α2, α4 and α6 chains). We next superimposed the contact residues identified from the crystal structure of [(α1)2α2] hexamer on the sequence alignment (supplementary Fig. 3). However, here we further distinguished between invariable contact residues (marked with asterisks) and variable ones (asterisks highlighted in black) which are critical for specificity. In doing so, we were able to identify 34 variable contact residues for α1-like chains and 33 for α2-like chains (supplementary Fig. 3). Once again, the VR3 was clearly distinguished as the region with highest sequence variability (87%) but also having a large number of variable contact residues. In addition to the VR3, the domain swapping region also stood out as a region with some sequence variability (47%) across the chains, but also comprising a high degree of variable contact residues.

We further identified which residues in the neighboring domains are the binding partners for the contact residues within the VR3 and β-hairpin motif. This showed that the contact residues of the β-hairpin motif make extensive contacts with 14 residues located on the site 2 of neighboring NC1 domain (supplementary Fig. 3, red lines). Eight out of these 14 residues are variable through all chains and across the species, indicating that in addition to the VR3 region, the β-hairpin motif also plays a selective role in the mechanism of assembly. Similar analysis for the VR3 region revealed even higher degree of sequence variability (8 out of 12 residues) for their binding partners the neighboring domain (supplementary Fig. 3). Thus, it was concluded that the initial mechanism for chain selection must be defined by two recognition motifs on each NC1 monomer; the β-hairpin motif and the VR3 hypervariable region (Fig. 7). The contact residues within both recognition motifs cluster at the monomer-monomer interface and through extensive chain-specific interactions stabilize the interaction (Fig. 7). However, some of these variable contact residues may not be essential for specificity.

DISCUSSION

A fundamental aspect of the tissue-specific assembly of ECM-networks is defined by mechanisms through which different collagen chains discriminate their natural binding partners. Since these molecules are highly homologous, molecular recognition and affinity of interaction are two important factors by which specific chains and their molecular stoichiometry within a particular collagen type is determined. In the present study of collagen IV, the analytical gelfiltration results clearly demonstrated that α1 NC1 domain is capable of self-assembly and hexamer formation. This was also confirmed by SPR analysis, showing that the sensorgrams for α1-α1 interactions (Fig. 4E) could be successfully fitted to the homohexamer model. However, the self-assembly of α1 chain to a trimeric molecule must be less efficient compared to the assembly of the naturally occurring (α1)2α2 heterotrimer, underscoring the prominent role of the α2 chain for the assembly.

The inability of α2 NC1 domain to form stable homotrimer or (α2)2α1 heterotrimer is an important underlying factor which explains why α2 chain occurs only once in the (α1)2α2 heterotrimer in vivo. Interestingly, this inability of self-interaction has been previously predicted using computer modelling of a hypothetical homotrimer of α2 NC1 domain (16) which showed serious steric clashing by the relatively larger equatorial dimension of the α2 NC1 domain at the triple helical junction makes it virtually
impossible for the α2 NC1 domains to assemble into a homotrimer. The relatively narrower width of this particular region in the α1 NC1 domain, however, seems to accommodate such homotrimer formation as three α1 NC1 domains could successfully be packed into a hypothetical homotrimer (16). Our kinetic analyses of homotypic and heterotypic interactions of both domains support these notions and provide experimental evidence that the naturally occurring heterotrimer (α1)2.α2 is the most stable trimer combination and therefore found in vivo.

The finding that the α1B-α2C interface has a dominating number of hydrogen bonds and hydrophobic contacts is in perfect agreement with our SPR analysis, indicating that the α2 NC1 domain displays a higher affinity for interaction to α1 NC1 domain through its site 2 (Table 2). This also implies that out of the two possible heterodimers, α1A-α2C and α1B-α2C, the latter is more stable and thus more likely to be formed prior to formation of the heterotrimer by interaction of another α1(IV) chain.

Similar to α2(IV) chain, α4(IV) and α6(IV) chains also occur only once in the α3.α4.α5 and α5.α5.α6 trimers, respectively. Based on sequence similarity of the NC1 domains, both α4(IV) and α6(IV) belong to the α2-like subtype, and therefore, it is likely that the NC1 domains of these chains also play a similar role in the assembly of their corresponding trimeric molecules as the α2 NC1 domain plays in the assembly of (α1)2.α2 trimer.

Thus, based on our kinetic and extensive structural analysis of NC1 domains it is reasonable to conclude that the molecular recognition motifs for chain selection in the assembly of collagen IV resides within site 1 and site 2 of each NC1 domain. Site 1 includes the β-hairpin domain swapping motif and locates to the N-terminal half, and the site 2 includes the hypervariable region 3 (VR3) which resides at the C-terminal half of the NC1 domain (Fig. 1 and 7).

The present work provides new strategic and intellectual framework to study the mechanism of assembly of other ECM macromolecules with known structure. More interestingly, as recent findings have established a potential inhibitory effect for the NC1 domains in angiogenesis and tumor growth (31,32), we hope that the knowledge gained in this study provide the means to tailor-design specific inhibitors to target angiogenesis in a tissue-specific manner.
REFERENCES

FOOTNOTES

*This work was supported by grants from the National Institutes of Health (R-37DK18381 and PO1DKO65123, to B.G.H.). We thank Dr. Ray Mernaugh at the Molecular Recognition Unit, Vanderbilt University for his excellent support with the BIAcore instrument. The BIAcore instrument at the Molecular Recognition Unit was supported in part by a Cancer Center Support Grant from the National Institutes of Health (CA68485).

1A type IV collagen molecule is a triple-helical molecule composed of three specific collagen IV α-chains or subunits named α1 through α6.

2Abbreviations are: NC1, noncollagenous domain of type IV collagen; ECM, extracellular matrix; ER, endoplasmic reticulum; SPR, surface plasmon resonance; RU, response unit; LBM, lens capsule basement membrane; root-mean-square deviation, RMSD.

FIGURE LEGENDS

**Fig. 1.** Domain organization and configuration of the NC1 domains in hexamer, trimer and monomer forms. **A**, The hexamer is a tightly packed ellipsoid-shaped molecule composed of two identical trimers: an ABC-trimer shown in surface rendered (left side) and a DEF-trimer represented by the secondary structures (right side). Each trimer consists of two α1 NC1 domains (shown in red and green) and one α2 NC1 subunit (shown in blue). The triple-helical collagenous domain (CD) emerging from each trimer is shown for orientation. **B**, left panel illustrates the surface-view of a trimer formed by side-by-side interactions through the domain swapping mechanism. Positions of the AB-, BC- and CA-domain swapping interfaces are indicated on the right panel. **C**, Left panel shows the topology of a NC1 domain (α2C) with two sites for interactions; a domain swapping β-hairpin (site 1) and the docking site for the domain swapping (site 2). Right panel shows higher magnification of the three-dimensional domain swapping between the α2C and the α1A NC1 domains. The β-hairpin structure is stabilized through disulfide bonds (shown in yellow).

**Fig. 2.** Acid dissociation and reconstitution of NC1 hexamer. **A**, Gel chromatography data showing the migration of native hexamer (H) at different pH and its gradual dissociation into dimers (D) and monomers (M). At pH 4.5, a shoulder peak (asterisk), likely corresponding to a heterotrimer peak, is distinguished from the hexamer peak. The mixture of NC1 monomers can be reassembled back into the hexamer configuration once the pH is brought back to neutral (pH 7.4 refolded). Arrows indicate elution positions of the standard proteins: gamma globulin (158 kDa), ovalbumin (44 kDa) and myoglobin (17 kDa). **B**, Circular dichroism spectra of the native (– – –) hexamer at pH 7.4, pH 5.0 ( – – –), pH 3.0 (•••••••••) and refolded (– · – · –) NC1 hexamer. Arrows point to negative (199 nm) and positive (225-235 nm) bands, where the CD spectra of the pH 3.0 sample most deviating from the other spectra. **C**, At the trimer-trimer (T-T) interface, there are three conserved arginine residues (R77, white) and three conserved glutamates (E176, yellow) forming a total of 12 ion pairs with their counter part residues on the other trimer. The glutamate residue E41 in B chain also forms an ion pair with a lysine residue on the opposite trimer, while the other two glutamates (E41 in A chain and E38 in C chain) form H-bonds with proton-donors from the opposite trimer. At the monomer-monomer (M-M) interfaces, there are 3 acidic residues at the B-C interface (E113, D125, E153) and one at the A-C interface (E153) forming a total of 4 ion pairs. At the A-C interface, there is also an aspartate (D70) involved in acid-sensitive H-bond formation.

**Fig. 3.** Reassembly of isolated NC1 monomers. **A**, Protein fractions from chromatofocusing separation of α1 and α2 NC1 domains analyzed by SDS-PAGE (upper gel) stained with Coomassie Brilliant Blue...
CBB). The same fractions were also analyzed by immunoblotting using monoclonal antibodies against α1 (Mab1) or α2 (Mab2) NC1 domains. B, Fractions containing isolated monomers of α1 (fractions 4-6) and α2 NC1 domains (fractions 17-20) were used in reconstitution studies. In contrast to α2 NC1 domain, isolated α1 NC1 monomers can form homohexamer once brought into neutral pH. C, Schematic representations of α1 and α2 NC1 domains showing that hexamers can be formed either by two heterotrimers of α1α1α2 (as found in vivo), or from two homotrimers of α1 (found only in vitro), but not by α2 NC1 domain alone.

Fig. 4. Surface plasmon resonance sensorgrams. A, Specific binding responses of α1 NC1 domain (50 μM) injected over a high density surface (1500 RU) of α2 NC1 domain and the negative control protein (BSA) immobilized at equal surface density on a parallel lane. Solid line (in red) represents the corrected response units (cRU) obtained after subtraction of the nonspecific signal recorded from the BSA surface (dotted line) from the initial response (dashed line). For kinetic analysis, isolated monomers of α1 and α2 NC1 domains were immobilized at lower surface densities (ca. 200 RU). B, Sensorgrams for interactions of α1 NC1 domain as analyte and α2 NC1 domain as ligand fitted with a trimer model. C, α2 NC1 domain as analyte and α1 as ligand fitted with an open trimer model. D, Sensorgrams for α2-α2 interaction fitted with a simple 1:1 interaction with conformational change, and E, α1-α1 interactions fitted with a hexamer model. In each experiment (B-E), five different analyte concentrations ranging from 0.25 to 4.0 μM are shown. The sensorgrams, shown in black, are the average of three injections per concentration. The global curve fits are shown in red. The goodness of curve fit is shown as residual plots above each sensorgram. The average chi-square value (χ²) for each curve fit is also indicated. F, Compiled equilibrium graphs for α1-α1 interaction in E showing the relative changes in distribution of monomer (M), dimer (D), trimer (T) and hexamer (H) forms as a function of total α1 concentration in solution. The range of analyte concentrations (0.25-0.4 μM) used in the kinetic analysis is shown as yellow bar.

Fig. 5. Comparison of the primary, secondary, tertiary and quaternary structures of α1 and α2 NC1 domains from bovine (Bov), human (Hum) and mouse (Mou). The sequence alignment shows identical residues (highlighted in gray), α1-specific residues (red) and α2-specific residues (blue). The N-terminal half-domain (upper alignment line) consists of ten β-strands (β1-10) and a short helix (g1). The C-terminal half-domain (lower alignment line) also consists of ten β-strands (β1'-β10') and a short helix (g1'). The major secondary structural difference is shown as a short 3₁₀ helix (g2') in the α1 NC1 domain and a short β-strand (βp') in the α2 NC1 domain. Color-coded RMSD-values represent the calculated root-mean-square deviations in Ångstrom (Å) between the main chains of α1 and α2 NC1 domains. The color-scale varies from 0 Å (yellow) to 6 Å (red, where the highest difference in tertiary structure was calculated). Variable regions (VR1, VR2 and VR3) are boxed. Asterisks show the residues directly involved in inter-domain contacts. B, Surface-rendered representations of NC1 trimers showing the location of variable regions. The chain-specific residues in α1A (red), α1B (green) and α2C (blue) NC1 domains and invariant residues (gray) are shown. Major structural differences (insets) at the VR3 region of α1 chain (3₁₀ helix) and α2 chain (β-strand) are shown.

Fig. 6. Structural comparison between the VR3 regions located at the interfaces between the α1A-α1B, α1A-α2C and α1B-α2C NC1 domains. The position of lysine residue (K56) in α1A and α1B NC1 domain, and their equivalent arginine residue (R55) on the α2 NC1 domain are shown. Two-dimensional representations of residues making direct contact at each interface are shown below each interface. Hydrogen bonds (lines in red), their distances in Ångstrom and non-polar contacts (lines in light green) are shown. At the α1B-α2C interface, the side chain of K56 from α1B chain makes two hydrogen bonds, one salt bridge and several hydrophobic contacts with residues located at the VR3 region of α2C chain.
The number of hydrogen bonds, salt bridges and hydrophobic contacts for each VR3 region are indicated below each column.

**Fig. 7.** Chain-recognition motifs and the critical contact residues involved in chain selection and assembly of collagen IV triple-helical molecule. Invariant (gray) and variant contact residues (white) are shown for both α2 chain (blue) and α1 chain (red). Specific residues involved in direct contact between the VR3 region of α1 chain and their binding partners on the α2 chain are indicated by arrows. In the heterodimer molecule, the variant contact residues that confer specificity are tightly clustered at the dimer interface. The collagenous domains are shown to indicate the orientation of each chain.
**TABLE 1.** Comparison of binding parameters stabilizing the monomer-monomer interactions within a trimer (side-by-side interactions) and between trimers (end-to-end interactions). The number of hydrogen bonds (H-bonds), salt bridges (ion pairs) and non-polar interactions (NP-bonds) are given for each interdomain interface. Total bonds are shown for each type of interaction within as well as between trimers and also for interactions between each pair of subunits. The ratio of non-polar to polar (H-bonds and ion pairs) is also shown to provide a scalar representative of the prevalent interaction at the chain pair level.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Subunits</th>
<th>H-bonds</th>
<th>Ion pairs</th>
<th>NP-bonds</th>
<th>Total</th>
<th>Ratio (NP/P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within trimers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(side-by-side)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1A-α1B</td>
<td>17</td>
<td>0</td>
<td>50</td>
<td>67</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>α1A-α2C</td>
<td>17</td>
<td>1</td>
<td>60</td>
<td>78</td>
<td></td>
<td>3.3</td>
</tr>
<tr>
<td>α1B-α2C</td>
<td>21</td>
<td>3</td>
<td>65</td>
<td>89</td>
<td></td>
<td>2.7</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>4</td>
<td>175</td>
<td>234</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>Between trimers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(end-to-end)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α1A-α1D</td>
<td>14</td>
<td>4</td>
<td>19</td>
<td>37</td>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td>α1A-α1E</td>
<td>4</td>
<td>0</td>
<td>12</td>
<td>16</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>α1B-α1D</td>
<td>4</td>
<td>0</td>
<td>12</td>
<td>16</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>α1B-α2F</td>
<td>12</td>
<td>5</td>
<td>27</td>
<td>44</td>
<td></td>
<td>1.6</td>
</tr>
<tr>
<td>α2C-α1E</td>
<td>10</td>
<td>4</td>
<td>28</td>
<td>42</td>
<td></td>
<td>2.0</td>
</tr>
<tr>
<td>α2C-α2F</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>13</td>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>13</td>
<td>110</td>
<td>162</td>
<td></td>
<td>1.9</td>
</tr>
</tbody>
</table>
**TABLE 2.** Kinetic rate constants for interaction of α1 and α2 NC1 domains. The data is collected from four sets of experiments in which both α1 and α2 NC1 domains were injected over two separate surfaces of immobilized α1 or α2 NC1 domains. Each interaction represents formation of a particular set of dimer, trimer or hexamer as indicated. Application of a specific kinetic model: heterotrimer, heterotrimer (limited, homohexamer or simple 1:1 molecular interaction model with conformational change (1:1 conf-change), are indicated. The association \( k_a \) and dissociation \( k_d \) rate constants and the equilibrium dissociation constant \( K_D \) for each reaction are tabulated. A, B and C correspond to α1A, α1B and α2C, respectively. The reaction A-B-C ↔ ABC indicates the transition from an open trimer (complex 3, 4 and 5) to the closed trimer (complex 6) as described in supplementary Fig. 2A. ¹C and ²C correspond to interaction through site 1 or site 2 of α2 NC1 domain, respectively. The reaction C-C ↔ C₂ indicates transition in conformational change. The second \( k_a \) values labeled with * has the unit \( s^{-1} \). The \( K_D \) for these reactions (within parenthesis) is therefore unit-less. The goodness of the fit, as indicated by the average chi-square value (Av. \( \chi^2 \)), and the calculated \( R_{max} \) value for each set of interaction are indicated.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic Model</th>
<th>Av. ( \chi^2 )</th>
<th>( R_{max} ) (RU)</th>
<th>( k_a ) (M⁻¹s⁻¹)</th>
<th>( k_d \times 10^4 ) (s⁻¹)</th>
<th>( K_D ) (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A + C ↔ AC</td>
<td>Heterotrimer</td>
<td>0.25</td>
<td>94 ± 1</td>
<td>6,244 ± 140</td>
<td>505 ± 5</td>
<td>8.1</td>
</tr>
<tr>
<td>B + C ↔ BC</td>
<td></td>
<td></td>
<td></td>
<td>867 ± 13</td>
<td>8 ± 1</td>
<td>1.0</td>
</tr>
<tr>
<td>A + B ↔ AB</td>
<td></td>
<td></td>
<td></td>
<td>28,034 ± 1,870</td>
<td>1,236 ± 125</td>
<td>4.4 (0.4)</td>
</tr>
<tr>
<td>A-B-C ↔ ABC</td>
<td></td>
<td></td>
<td></td>
<td>115* ± 2</td>
<td>52 ± 1</td>
<td></td>
</tr>
<tr>
<td>¹C + A ↔ ¹CA</td>
<td>Heterotrimer (limited)</td>
<td>0.27</td>
<td>40 ± 1</td>
<td>6,335 ± 264</td>
<td>600 ± 15</td>
<td>9.5</td>
</tr>
<tr>
<td>²C + A ↔ ²CA</td>
<td></td>
<td></td>
<td></td>
<td>819 ± 30</td>
<td>10 ± 1</td>
<td>1.2</td>
</tr>
<tr>
<td>C + C ↔ C₂</td>
<td></td>
<td></td>
<td></td>
<td>79,628 ± 44,500</td>
<td>11,778 ± 737</td>
<td>15</td>
</tr>
<tr>
<td>C₂ + A ↔ C₂-A</td>
<td></td>
<td></td>
<td></td>
<td>8264 ± 360</td>
<td>46 ± 1</td>
<td>0.6</td>
</tr>
<tr>
<td>A + A ↔ A₂</td>
<td>Homohexamer</td>
<td>0.17</td>
<td>28 ± 1</td>
<td>28,063 ± 27,000</td>
<td>1,180 ± 30</td>
<td>4.2</td>
</tr>
<tr>
<td>A₂ + A ↔ A₃</td>
<td></td>
<td></td>
<td></td>
<td>131 ± 125</td>
<td>17 ± 1</td>
<td>13.0</td>
</tr>
<tr>
<td>A₃ + A₃ ↔ A₆</td>
<td></td>
<td></td>
<td></td>
<td>187,320 ± 65,900</td>
<td>825 ± 581</td>
<td>0.4</td>
</tr>
<tr>
<td>C + C ↔ C-C</td>
<td>1:1 conf-change</td>
<td>0.26</td>
<td>45 ± 5</td>
<td>6,247 ± 1,130</td>
<td>1,223 ± 35</td>
<td>20</td>
</tr>
<tr>
<td>C-C ↔ C₂</td>
<td></td>
<td></td>
<td></td>
<td>5,179* ± 70</td>
<td>2,941 ± 37</td>
<td>(11.1)</td>
</tr>
</tbody>
</table>

17
Fig. 1 (Khoshnoodi et al.)
Fig. 2 (Khoshnoodi et al.)
A

Protein fractions from chromatofocusing

4-6  7-9  10-16  17  18-20

CBB

Mab1

Mab2

B

Reconstitution study of α1 and α2 NC1 domains

Native Hexamer

H M

Fractons 4-6

α1 NC1

Fractons 17-20

α2 NC1

C

α1  α2

α1α2α1 (α1)3 (α2)3

α1α2α1  (α1)6 (α2)6

Fig. 3 (Khoshnoodi et al.)
Fig. 4 (Khoshnoodi et al.)
<table>
<thead>
<tr>
<th>Configuration</th>
<th>H-bonds</th>
<th>NP-bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>α1A–α1B</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>α1A–α2C</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>α1B–α2C</td>
<td>9</td>
<td>28</td>
</tr>
</tbody>
</table>

Fig. 6 (Khoshnoodi et al.)
Fig. 7 (Khoshnoodi et al.)
Mechanism of chain selection in the assembly of collagen IV: A prominent role for the α2 chain
Jamshid Khoshnoodi, Kristmundur Sigmundsson, Jean-Philippe Cartailler, Olga Bondar, Munirathinam Sundaramoorthy and Billy G. Hudson

J. Biol. Chem. published online December 22, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M506555200

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/2005/12/23/M506555200.DC1