Crystal Structure of the N-terminal NC4 Domain of Collagen IX, a Zinc Binding Member of the Laminin-Neurexin-Sex Hormone Binding Globulin (LNS) Domain Family*

Received for publication, March 23, 2007, and in revised form, May 22, 2007 Published, JBC Papers in Press, June 6, 2007 DOI 10.1074/jbc.M702514200

Veli-Matti Leppänen¹, Helena Tossavainen¹, Perttu Permi¹, Lari Lehtis₂, Gunilla Rönnholm¹, Adrian Goldman³, Ilkka Kilpeläinen⁴⁵, and Tero Pihlajaama¹¹

From the ¹Program in Structural Biology and Biophysics, Institute of Biotechnology and the ³Laboratory of Organic Chemistry, Department of Chemistry, University of Helsinki, FI-00014 Helsinki, Finland

Collagen IX, located on the surface of collagen fibrils, is crucial for cartilage integrity and stability. The N-terminal NC4 domain of the α1(IX) chain is probably important in this because it interacts with various macromolecules such as proteoglycans and cartilage oligomeric matrix protein. At least 17 distinct collagen polypeptides carry an NC4-like unit near their N terminus, but this report, describing the crystal structure of NC4 at 1.8-Å resolution, represents the first atomic level structure for these domains. The structure is similar to previously characterized laminin-neurexin-sex hormone binding globulin (LNS) structures, dominated by an antiparallel β-sheet sandwich. In addition, a zinc ion was found in a position similar to that of the metal binding site of other LNS domains. A partial backbone NMR assignment of NC4 was obtained and utilized in NMR titration studies to investigate the zinc binding in solution state and to quantitate the affinity of metal binding. The Kₐ of 11.5 mM suggests a regulatory rather than a structural role for zinc in solution. NMR titration with a heparin tetrasaccharide revealed the presence of a secondary binding site for heparin on NC4, showing structural and functional conservation with thrombospondin-1, but a markedly reduced affinity for the ligand. Also the overall arrangement of the N and C termini of NC4 resembles most closely the N-terminal domain of thrombospondin-1, distinguishing the two from the majority of the published LNS structures.

Collagens are a family of extracellular matrix proteins comprised of at least 28 distinct types, all characterized by the presence of one or more elongated triple helical regions. This common motif is supplemented with a variety of non-helical domains to add unique properties for molecular or supramolecular assembly, for tissue-specific targeting, or for interactions with other constituents of the matrix (1, 2). An example of such a supplementary element is a module showing similarity to the N-terminal domains of thrombospondins (TSPNs)² (3). This domain occurs in at least 17 different collagen polypeptides, located N-terminal to the triple helix, but little is known about the detailed structures and physiological functions of these units (2).

One of the TSPNs found in collagen chains is the NC4 domain of collagen IX, formed by about 245 N-terminal residues of the mature α1(IX) polypeptide. In cartilage tissue, collagen IX molecules are covalently bound to the heteropolymeric collagen fibrils that provide the tissue with high tensile strength to resist the swelling pressure caused by proteoglycans. The shorter arm of the triple helix of collagen IX projects away from the fibril body so that the NC4 domain can easily interact with other macromolecules. The temporal and spatial control over the presence or absence of NC4 on collagen IX via alternative splicing (4, 5) suggests that the domain has functional significance. The NC4 domain, as well as three other regions along the collagen IX triple helix, has been shown to interact in vitro with cartilage oligomeric matrix protein (COMP) and with heparin (6, 7), but the significance of these findings in vivo remains obscure. The mutual in vitro interactions of collagen IX, COMP, and matrilin-3 (6, 8, 9), as well as the linkage of the defects in the respective genes to the same chondrodysplasia phenotype, i.e. multiple epiphyseal dysplasia (10–15), suggest that these three multimeric proteins collaborate in the supramolecular assembly of thin cartilage fibrils into a fibrillar network necessary for long term tissue stability (9). Indeed, mice lacking functional collagen IX develop upon aging a degenerative joint disease resembling human osteoarthritis (16). Collagen IX has also been shown to interact with all known collagen receptor integrins on cell surfaces, suggesting that collagen IX, and perhaps other members of the same subfamily of fibril-associated collagens with interrupted triple helices

The abbreviations used are: TSPN, N-terminal domain of thrombospondin; COMP, cartilage oligomeric matrix protein; FACIT, fibril-associated collagen with interrupted triple helices; HSQC, heteronuclear single quantum coherence; LNS, laminin-neurexin-sex hormone binding globulin; MAD, multivavelength anomalous dispersion; SHBG, sex hormone binding globulin; TSPN-1, N-terminal domain of thrombospondin-1; MES, 4-morpholineethanesulfonic acid; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

*This work was supported by Academy of Finland Grants 106411 (to T. P.), 1206228 (to I. K.), 70891 (to V.-M. L.), 1105157 and 1111771 (to A. G.), and 106852 (to P. P.) as well as by funding from the Sigrid Juselius Foundation and Biocentrum Helsinki (to A. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2UUR) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/)

¹ To whom correspondence should be addressed: NMR Laboratory, Inst. of Biotechnology, P. O. Box 65, University of Helsinki, FI-00014 Helsinki, Finland. Tel.: 358-9-19159544; Fax: 358-9-19159541; E-mail:Tero.Pihlajaamaa@helsinki.fi.
Crystal Structure of the Collagen IX NC4 Domain

(FACITs), may serve in the adhesion of the fibril networks to cells (17).

An atomic level structure is currently not available for any of the collagen TSPN units. On the basis of amino acid sequence conservation and the apparent similarities of the secondary structural elements, it was predicted that the TSPN module is homologous with pentraxins and with members of the laminin-neurexin-sex hormone binding globulin (LNS) domain family (18, 19). Three-dimensional structures of LNS modules from the laminin α2 chain (α2LG5) (20), sex hormone binding globulin (SHBG) (21), neurexins 1β (22) and 1α (23), calnexin (24), growth arrest-specific protein Gas6 (25), agrin (26), and the N-terminal domain of thrombospondin-1 (TSPN-1) (27) along with structures of the pentraxins serum amyloid P component (28) and C-reactive protein (29, 30) are now available. All members of the superfAMILY share a β-sandwich fold with a convex and a concave sheet, each comprised of six or seven antiparallel β-strands. This ~200-residue unit is reminiscent of certain types of lectins, containing a jelly roll fold. In the course of evolution, the common LNS structure has been utilized in docking surfaces for various ligands such as glycosaminoglycans, simple carbohydrates, cell surface receptors, and other constituents of the extracellular matrix. In many LNS domains, the interactions appear to cluster to a specific area on the rim of the β-sandwich. In most published LNS structures a divalent cation of apparent functional importance is also located in the same region, whereas in lectins, the binding sites for both the ligands and the cation affecting ligand binding are located on the concave β-sheet (31, 32).

To understand the physiological function of the NC4 domain and its homologues in other collagens, knowledge of the three-dimensional structure is essential. We have previously reported the production and initial characterization of the recombinant NC4 domain of human collagen IX (7). Here we describe the crystal structure of NC4, the first collagen LNS domain solved, and show that NC4 is indeed a metal-coordinating LNS module with binding sites for multiple ligands.

EXPERIMENTAL PROCEDURES

Production of Recombinant NC4 Domain—The recombinant NC4 domain of human collagen IX without the 23-residue signal peptide was produced and purified essentially as described previously (7). The residue numbering used below starts from the first residue of the mature, secreted polypeptide. Instead of ammonium sulfate precipitation, the clarified cell lysate was subjected directly to affinity chromatography on a glutathione-Sepharose column (GE Healthcare) using Factor Xa protease (GE Healthcare) digestion to elute the NC4. After heparin affinity and cation exchange chromatography steps, the sample was further polished by gel filtration on a Superdex 75 column (GE Healthcare), dialyzed into water, and lyophilized. In vitro mutagenesis (Stratagene) was used to create a mutant form of NC4, termed NC4<sup>Δ</sup>23, lacking the first 17 residues of the mature polypeptide and replacing, for technical reasons, Asn<sup>18</sup> and Glu<sup>19</sup> with Asp and Leu, respectively. We also separately made the D190A/D192A double mutant. CD spectropolarimetry and two-dimensional NMR spectroscopy were used to verify that the mutations did not cause abnormal folding of NC4.

Selenomethionine incorporation was achieved by inoculating the Escherichia coli Rosetta<sup>TM</sup>(DE3) cells (Novagen) into LB broth and growing at 37 °C to an absorbance of about 0.7. 1-L-SeMet was then added to 60 μg/ml, expression was induced after 10 min with 1 mM isopropyl 1-thio-β-D-galactopyranoside, and the cells were harvested after a 4-h culture at 37 °C. Incorporation of the stable isotopes<sup>1</sup>N and<sup>13</sup>C for NMR studies was achieved by growing the BL21(DE3) cells to induction cell density in LB broth, collecting the cells by centrifugation, washing by brief resuspension in 0.5× phosphate-buffered saline supplemented with the antibiotic, recollected the cells, and finally resuspending in M9 minimal medium supplemented with 1 g/liter [13C]<sub>2</sub>NH<sub>4</sub>Cl and 2 g/liter D-glucose or D-[13C]<sub>2</sub>glucose. After shaking for 15 min at 37 °C, the induction was carried out as above. For perdeuteration, the M9 medium was prepared in 99% D<sub>2</sub>O. The incorporation of isotopic labels was verified by mass spectrometry using an Ultraflex TOF/TOF instrument (Bruker Daltonics Inc., Bremen, Germany) in a positive ion linear mode using sinapinic acid as a matrix. The instrument was externally calibrated.

Crystallography—Crystals of NC4 were grown in sitting drops over a reservoir solution of 100 mM MES, pH 6.5, 10 mM ZnSO<sub>4</sub>, and 20% monomethyl ether-polyethylene glycol 550. The drops were prepared by mixing 2 μl of the reservoir solution and 2 μl of the protein solution at 20 mg/ml. The crystals belong to space group 14 (a = 81.88 Å, c = 71.43 Å) with one molecule per asymmetric unit and a solvent content of 44%. For data collection at −180 °C, crystals were frozen in liquid nitrogen with the well solution containing 10% (v/v) ethylene glycol.

Multimode anomalous diffraction (MAD) data on a frozen selenomethionine derivative crystal were collected to 1.8 Å using the BM14/2 beamline at European Synchrotron Radiation Facility (Grenoble, France) at three wavelengths (Table 1). The data sets were indexed and scaled with the program X-ray Detector Software (33). The Crystallography NMR software (Ref. 34) was used to find the three expected selenium sites. A fourth, strong anomalous scatterer was found, showing an increasing signal from the remote to the selenium K absorption edge at 1.283 Å. The selenium sites together with the Zn<sup>2+</sup> ion due to its presence in the crystallization buffer and the fact that it has the K absorption edge at 1.283 Å. The selenium sites with the Zn<sup>2+</sup> were used to estimate the experimental MAD phases at 2.0 Å in Crystallography NMR software (Table 1). The electron density map obtained upon solvent flipping in Crystallography NMR software was used for initial model building.

We built an initial model of residues 19–24 and 30–233 using O (35). Using the inflection point data, this model was subjected to iterative rounds of refinement in Crystallography NMR software (34) and manual rebuilding. Refinement was carried out using energy minimization, simulated annealing, and individual temperature factor (B-factor) refinement. Water molecules were added to peaks above 3.5σ in the F<sub>o</sub> − F<sub>c</sub> difference map if they had suitable hydrogen bonding geometry. The final model, with good stereochemistry (Table 1), consists of 210 residues of which nine side chains display alternative conformations. In addition to the Zn<sup>2+</sup>, the solvent model includes...
a sulfate ion, four ethylene glycol, and 209 water molecules. Residues 1–18, 25–29, and 234–245 are not seen in the electron density. PROCHECK (36) was used to assign secondary structure elements and calculate the Ramachandran plot. Of all the non-Gly/non-Pro residues, 86.1% have main chain torsion angles in the most favored regions, and there are no residues in the disallowed regions. Structural comparisons were carried out using DALI (37).

**NMR Spectroscopy** — For the backbone chemical shift assignment of NC4, HNCACB and HN(CO)CACB spectra were acquired at 40 °C on a Varian Unity INOVA spectrometer operating at 600-MHz 1H frequency and equipped with a cryogenically cooled probe head and an actively shielded z-gradient system. The sample contained ~0.5 mM 13C,15N-labeled NC4 in 50 mM NaCl, 20 mM BisTris buffer, pH 6.5 at 40 °C, 92.5% H2O, 7.5% D2O. Spectra were processed with Vnmr 6.1 revision C (Varian Inc., Palo Alto, CA) and analyzed with Sparky 3.106 (38).

To obtain apo-[15N]NC4, the NMR sample was washed free of divalent cations in the presence of 2 mM NaCl and 100 mM EDTA on a spinnable concentrator (Millipore) by concentrating 10-fold followed by a similar wash without added EDTA. Additional washing steps with 20 mM BisTris buffer, pH 6.5, were carried out until the calculated concentration of EDTA was 20 μM or less. This procedure gave a two-dimensional 15N HSQC spectrum identical to that from the size exclusion-based method that we adapted from the procedure used for removal of Ca2+ from calerythrin (39).

For NMR titration with divalent cations as ligands, a Varian Unity INOVA 600 MHz spectrometer equipped with an actively shielded z-axis gradient probe head was used to measure a series of two-dimensional 15N HSQC spectra for apo-[15N]NC4 in the absence or presence of cations. For the Zn2+ titration, a series of spectra was recorded in which the concentration of ZnCl2 was gradually increased. Concentrations corresponding to molar ratios of 0:1, 1:1, 5:1, 10:1, 20:1, 40:1, and 80:1 of Zn2+ to NC4 were used, and a 1H, 15N HSQC spectrum was acquired at each titration point. Average chemical shift changes were calculated with equation 1

\[ \Delta \delta = \left( 6 \Delta \delta_H^2 + 0.17 \times (\Delta \delta_N)^2 \right)^{1/2} \]

for five residues that could be traced with the highest confidence in the Zn2+ titration spectra. These values were plotted as a function of Zn2+ concentration, and curve fitting with nonlinear regression (40) was performed to obtain dissociation constants for the interaction of Ca2+ and Mg2+ with NC4, spectra were measured in the presence of a single cation concentration, corresponding to a molar ratio of 25:1 of the cation to NC4.

The titration spectra with heparin oligosaccharides were acquired at 600-MHz 1H frequency (40) in 33 mM NaCl, 20 mM BisTris buffer, pH 6.5 at 40 °C, 92.5% H2O, 7.5% D2O. For titration with the heparin tetrasaccharide (Dextra Laboratories, Reading, UK) concentration ratios of 0:1, 3:1, 6:1, 15:1, 24:1, and 36:1 were used, and an averaged dissociation constant was determined as above using data from five residues.

To investigate the interaction of naturally occurring retinoids with NC4, solid all-trans-retinoic acid or vitamin D3 (Sigma-Aldrich) was incubated for at least 24 h with [15N]NC4 at 20 °C in the presence of Zn2+ or EDTA (41). Insoluble retinoid was removed by centrifugation, and a two-dimensional 15N HSQC spectrum was measured as above. Alternatively, concentrated stock solution of retinoid prepared in deuterated dimethyl sulfoxide was added in excess to a [15N]NC4 NMR sample, keeping the amount of Me2SO below 1%, and a two-dimensional 15N HSQC spectrum was measured (42).

**RESULTS**

Structure Determination — The three-dimensional crystal structure of NC4 was solved by collecting MAD data from a crystal of selenomethionine-substituted protein. In addition to the three expected selenium sites, a Zn2+ cation was found and used in phasing. The asymmetric unit encompasses a single NC4 domain. The model was refined at 1.8-Å resolution to a crystallographic R factor of 18.8% and an Rfree of 21.4%. Data collection and refinement statistics are summarized in Table 1. Analysis using PROCHECK (36) shows that all non-glycine and non-proline residues are in the most favored or additionally allowed regions of the Ramachandran plot. Except for residues 1–18, 25–29, and 234–245, the main chain is well defined by the electron density.

Structure of NC4 — The NC4 structure has a globular lectin-like architecture and contains 14 β-strands, four small α-helices, and two 3_10 helix-like turns (Fig. 1). The most prominent feature is two antiparallel six-stranded β-sheets sandwiched together. The first β-sheet is composed of strands β1, β5, β10, β11, β12, and β14, and the solvent-exposed face is convex (the convex sheet). The second β-sheet, approximately parallel with the first one, correspondingly has a concave solvent-exposed face (the concave sheet) and is composed of antiparallel strands β4, β6, β7, β8, β9, and β13. The loops connecting β6 to β7 and β13 to β14 rise up and form a depression that is orthogonal to the direction of the strands and runs the length of the sheet (Fig. 1).

Although the C-terminal portion of the structure is covered by the lectin-like β-sandwich with relatively short connecting loops, the N-terminal part possesses larger excursions. The first β-strand, on the convex side, is preceded by 32 residues, including Cys21 that forms a disulfide bridge with Cys219 at the end of β14. Despite this linkage to the rest of the structure, only the residues immediately adjacent to Cys21 (residues 19–24) are visible in the electron density of the first 29 residues (Fig. 1B). The first β-strand is followed by a long insertion bearing two short α-helices, α1 and α2, and a β-hairpin loop (strands β2 and β3). The insertion closes one edge of the β-sandwich (Fig. 1A). From here, the chain continues to the other β-sheet, contributing a short strand, β4, that is directly followed by the third α-helix, α3. The last α-helix, α4, is part of the connecting loop between strands β5 and β6. Following the last strand, β14, there is a C-terminal extension of 15 residues (residues 219–233) on the convex side. In addition to Cys219, Cys229 also forms a disulfide bridge (Cys175–Cys229) in this area and links the C terminus to the convex sheet.

A strong anomalous scatterer was found when the MAD phases were estimated. The anomalous signal (Fig. 2) for this scatterer was highest in the selenium K edge inflection point data set, and because 10 mM ZnSO4 was required for crystallization...
Crystal Structure of the Collagen IX NC4 Domain

**TABLE 1**

Summary of X-ray data collection and refinement statistics

| Data collection | \( \lambda_{\text{peak}} \) | \( \lambda_{\text{remote}} \) | \( \lambda_{\text{inflection point}} \) |
|----------------|-----------------|-----------------|-----------------|
| Wavelength (Å) | 0.9796          | 0.8856          | 0.9797          |
| Resolution range (Å)* | 20–1.8 (1.9–1.8) | 20–1.8 (1.9–1.8) | 20–1.8 (1.9–1.8) |

| Number of reflections | Total | 164,652 | 165,303 | 164,402 |
|-----------------------|-------|---------|---------|---------|
| Unique                | 21,920 | 21,920  | 21,918  |
| Completeness (%)*     | 99.9 (100.0) | 99.9 (100.0) | 99.9 (100.0) |
| \( I / I_{\text{c}} \) | 18.3 (4.1) | 19.5 (4.4) | 20.5 (5.1) |
| \( R_{\text{sym}} \) (%)* | 5.6 (43.5) | 5.3 (40.0) | 4.9 (33.5) |
| Number of selenium sites | 3     | 3       | 3       |

| Overall figure of merit for MAD phasing at 2.0-Å resolution |
|------------------------------------------------------------|
| Acentric | 0.48 |
| Centric | 0.49 |

| Refinement | \( \lambda_{\text{inflection point}} \) |
|------------|-----------------|
| Resolution range (Å) | 20–1.8 |
| Reflections | 21,918 |
| \( R_{\text{work}} \) (%) | 18.8 |
| \( R_{\text{free}} \) (%)* | 21.4 |

| Number of atoms |
|-----------------|
| Protein | 1,582 |
| \( \text{Zn}^{2+} \) | 1 |
| Solvent | 235 |

| Average B-factor (Å²) |
|----------------------|
| Protein atoms | 28.6 |
| \( \text{Zn}^{2+} \) | 25.8 |
| Solvent atoms | 44.0 |

| r.m.s. deviation from ideal values |
|-----------------------------------|
| Bond lengths (Å) | 0.007 |
| Angles (°) | 1.5 |

* Values in parentheses correspond to the highest resolution shell.
* The \( R_{\text{free}} \) was calculated with 5% of the data omitted from structure refinement.
* Root mean square.

FIGURE 1. The NC4 structure. **A**, a schematic ribbon diagram of the NC4 structure approximately perpendicular to the \( \beta \)-sheets. The secondary structure elements are labeled where applicable. \( \beta \)-Strands are numbered from 1 to 14. Disulfide bridges are highlighted in yellow. The \( \text{Zn}^{2+} \) ion is shown as a green sphere. **B**, the same as in A, but the molecule has been rotated –90° around the vertical axis. The figures were prepared using MOLSCRIPT (43) and RASTER3D (44).

zation, we assigned the anomalous scatterer as \( \text{Zn}^{2+} \). Crystals could not be obtained in the presence of \( \text{Ca}^{2+} \) or \( \text{Mg}^{2+} \). The \( \text{Zn}^{2+} \) binding site is located on an edge of the \( \beta \)-sandwich (Fig. 1) and is formed by the \( \beta 6–\beta 7 \) and \( \beta 12–\beta 13 \) loops. The \( \text{Zn}^{2+} \) is coordinated by the side chains of Asp\(^{190} \) and Asp\(^{192} \), a water molecule, and His\(^{230} \) from a symmetry-related NC4 molecule (Fig. 2). The interactions with the aspartates are monodentate, resulting in a favorable tetragonal coordination of the \( \text{Zn}^{2+} \) (46). The zinc-ligand distances range from 2.03 to 2.18 Å (average, 2.11 Å). Overall the coordinating amino acid residues and the metal-ligand distances are typical for \( \text{Zn}^{2+} \) and thus support our assignment of this anomalous scatterer as \( \text{Zn}^{2+} \). The side chain of Ser\(^{113} \) forms a hydrogen bond with the \( \text{Zn}^{2+} \)-bound water molecule. Interestingly one more aspartate, Asp\(^{112} \), is also close to the \( \text{Zn}^{2+} \) and could interact with it by a simple side chain rotation. The binding affinity of \( \text{Zn}^{2+} \) as well as the effect of \( \text{Zn}^{2+} \) on the interaction of NC4 with heparin were analyzed by NMR and affinity chromatography, respectively (see below).

NMR Chemical Shift Assignment—

The dispersion for the majority of the cross peaks in the \( ^{1} \text{H}, ^{15} \text{N} \) HSQC spectrum of NC4 is good. Unfortunately in the middle of the spectrum there is a cluster of peaks with many of the peaks overlapped and some of very high intensity. This indicates that part of the NC4 main chain contains highly flexible parts. A total of 208 backbone cross-peaks could be distinguished in the spectrum. Backbone chemical shift assignment of NC4 was pursued with three-dimensional HNCACB and HN(CO)CACB spectra (47, 48). We acquired this pair of spectra from three different samples, a \( ^{2} \text{H}, ^{13} \text{C}, ^{15} \text{N} \) triple-labeled sample at pH 4 and 7.5 as well as a \( ^{13} \text{C}, ^{15} \text{N} \) double-labeled sample at pH 6.5. The latter...
NMR Titration of NC4 with Zn$^{2+}$. A collection of two-dimensional $^{15}$N HSQC spectra of NC4 in increasing concentrations of ZnCl$_2$ (see ‘Experimental Procedures’) was obtained, and each spectrum was assigned a distinct color and overlayed on top of the spectrum of NC4 measured in the absence of divalent cations (red). A, a close-up of the overlaid spectra showing Zn$^{2+}$-dependent linear transition of the cross-peak representing the backbone amide bond of Ser$^{113}$ B, five cross-peaks from the spectral overlay were selected for plotting the calculated average chemical shift changes. After curve fitting a $K_d$ of $11.5 \pm 1.1$ mM was obtained for the binding of Zn$^{2+}$.

NMR Titration with Heparin Oligosaccharides Reveals a Low Affinity Binding Site—NMR titrations of $^{15}$N-labeled NC4 with heparin di-, tetra-, or hexasaccharides or longer heparin fragments were performed. When overlaying the titration spectra, linear movements were seen for less than 10% of the cross-peaks upon titration with heparin di- or tetrasaccharide (not shown). Most of these chemical shift perturbations were in one part of the NC4 molecule formed by loops $\beta3$–$\beta4$ and $\beta13$–$\beta14$, strand $\beta5$, and the neighboring helix $\alpha4$, suggesting the presence of a previously unidentified, low affinity heparin binding site on NC4. Analysis of the perturbation data obtained with a tetrasaccharide gave a $K_d$ of 1.0 mM (Fig. 4). The primary heparin binding site of NC4 is located within the first eight residues of the domain (7); these are currently of undetermined...
structure and without a backbone chemical shift assignment. The cross-peaks originating from these residues may be among the few unassigned peaks demonstrating chemical shift perturbation upon titration with the tetrasaccharide, or they may be undetectable under the experimental conditions used. Notably among the assigned residues Val10 is located closest to the N terminus and shows clear perturbation upon heparin addition (data not shown). In addition, when the tetrasaccharide titration was performed at pH 5, large perturbations were seen for four unassigned cross-peaks reaching saturation at a 10-fold lower ligand concentration (data not shown). Titrations with heparin hexasaccharide or larger fragments resulted in precipitation of NC4 and an associated reduction of the spectral intensity and quality even at subequimolar amounts of the ligands. The same effect was seen with highly sulfated heparin analogues such as sucrose octasulfate and myo-inositol hexasulfate (not shown).

NC4 Lacking the N Terminus Has Only Residual Heparin Affinity—To verify our initial mapping of the heparin binding site of NC4 to the N terminus of the domain (7), we created mutant NC4_{Ndel} lacking the first 17 residues of the domain. Analytical heparin affinity chromatography showed that N-terminally truncated NC4 could not bind to the column at a physiological salt concentration. When bound to the column in a buffer containing no added NaCl, NC4_{Ndel} eluted at 0.13 M salt compared with the 0.31 M salt required for elution of wild-type NC4 (Fig. 5). The D190A/D192A double mutant, which cannot bind Zn^{2+} bound to the heparin column as tightly as wild type, and the presence of excess EDTA had no significant effect on the elution profile of wild-type NC4 (data not shown). This indicates that a Zn^{2+} cation bound to NC4 is unlikely to affect its interaction with a glycosaminoglycan in vivo.

Comparison of the Structural Model of NC4 with Other Structures—Searching the Protein Data Bank using DALI (37) revealed that NC4 has structural similarities to the large number of lectins and lectin-like domains in the concanavalin A-like lectin and glucanase superfamily (Structural Classification of Proteins; Ref. 49). Despite the modest sequence identity (≤16%), all of the structurally homologous domains share the central lectin fold feature: a pair of antiparallel β-sheets curved to form a concave and a convex side. NC4 is structurally closest to the 23.5-kDa TSPN-1 (Ref. 27; Fig. 6A) and can be superimposed with a root mean square deviation of 2.8 Å for 187 Cα atoms. In addition to TSPN-1 and several glycolytic enzymes, the top hits include serum amyloid P component (28), neurexin 1β (22), calnexin (24), laminin G-like domain (LG5; Ref. 20), agrin G3 domain (26), Gas6 (25), and SHBG (21). These multifunctional protein domains, including TSPN-1, are involved in heparin, steroid ligand, and protein-protein interactions rather than in binding simple carbohydrates. They are referred to collectively as LNS domains (22, 26, 32). Because NC4 may function in both protein-protein interactions and heparin binding (7), we include NC4 as a new member of the LNS domain family.

The significant structural similarity between NC4 and the other LNS domains is mainly in the β-sandwich strands and the short connecting loops and hairpin turns. The N and C termini in LNS domains are located on one side of the β-sandwich, and this is where the structures differ the most. TSPN-1 is most
Crystal Structure of the Collagen IX NC4 Domain

FIGURE 6. Structural comparison of NC4 with other LNS domains. A, stereoview of superposition of the NC4 (green) and the TPSN-1 domain (yellow) (27) coiled-coil diagrams. The NC4 aspartate side chains coordinating the Zn$^{2+}$ are shown as ball-and-stick models, and the Zn$^{2+}$ ion is shown as a green sphere. The N and C termini are labeled with corresponding coloring. B, superposition of LNS domains laminin LG5 (red), SHBG (blue), agrin G3 (pink), and NC4 (green) with similar metal binding sites. C, close-up of B. The side chains involved in metal ion binding are shown as ball-and-stick models. For clarity, only the SHBG and NC4 ligands coordinating the Zn$^{2+}$ ions and Ser$^{113}$ in NC4 are labeled. Also for clarity, the neurexin 1$_{\alpha}$LNS#2 with a Ca$^{2+}$ ion in a similar position to that in laminin LG5 and agrin G3 domains is not shown. The $\beta$-strands are labeled using the NC4 numbering as in Fig. 1A.

similar to NC4 because they both have similar structures in the N-terminal excursion between strands $\beta$1 and $\beta$4 and in the C terminus (NC4 numbering; Figs. 1 and 6A) as well as in the $\beta$-sandwich region. However, the first 30 residues seem to be unique for NC4. TSPN-1 and NC4 share the overall topology of the N-terminal excursion including the $\beta_{2}$$-$$\beta_{3}$ hairpin loop and the helical ($\alpha_{1}$; Fig. 1) segment immediately after strand $\beta$1. In calnexin (24), the N terminus has equivalents to $\beta$1 and to the $\beta_{2}$$-$$\beta_{3}$ hairpin loop of NC4 but has a much longer insertion between strands $\beta$1 and $\beta$2. Interestingly in all other LNS domains this area is mainly occupied by C-terminal residues, and the $\beta_{2}$$-$$\beta_{3}$ hairpin loop is where the N and C termini meet. In these other LNS domains, NC4 $\beta$14 is followed by a hairpin turn, and the NC4 strand $\beta$1 is replaced with the last strand in their convex or concave (SHBG) sheet of the $\beta$-sandwich. This is also the case for LNS domain pairs both in Gas6 (25) and laminin (LG4–LG5; Ref. 50). Not surprisingly, the only counterpart to the C-terminal disulfide bridge of NC4 is in TSPN-1 where the only TSPN-1 disulfide bridge links the C terminus to the $\beta$11–$\beta$12 turn, analogous to the Cys$^{175}$$-$$\$Cys$^{229}$ bond of NC4 (Fig. 1). In general, the LNS structures are more homologous on the side opposite to the N and C termini. In the connecting loops, the most notable differences are seen in the $\beta_{4}$$-$$\beta_{5}$, $\beta_{6}$$-$$\beta_{7}$, $\beta_{12}$$-$$\beta_{13}$, and $\beta_{13}$$-$$\beta_{14}$ loops (NC4 numbering). The $\beta_{6}$$-$$\beta_{7}$ and $\beta_{12}$$-$$\beta_{13}$ loops are involved in metal binding in many LNS domains (see below and Fig. 6, B and C).

NC4 and the metal ion-binding LNS domains have a similar metal binding site. The Ca$^{2+}$ ions in the laminin LG5 and agrin G3 domains have octahedral coordination with four protein ligands (20, 26). The tetragonal coordination of Zn$^{2+}$ in NC4 consists of three protein ligands, two from a single molecule and a third one from a symmetry-related molecule of the crystal. In NC4, the long connecting loop between $\beta$-strands 12 and 13 provides two aspartate side chains, and the hairpin loop between $\beta$-strands 6 and 7 is hydrogen-bonded to the Zn$^{2+}$ via a water molecule and the Ser$^{113}$ side chain (NC4 strand numbering; Figs. 1, 2, and 6C). In laminin LG5 and agrin G3 domains both of the corresponding loops contribute an aspartate side chain to the Ca$^{2+}$ coordination (Fig. 6C). The laminin LG5 and agrin G3 $\beta_{12}$$-$$\beta_{13}$ loops and $\beta_{8}$$-$$\beta_{9}$ hairpin turns complete the coordination by one main chain carbonyl oxygen each. In SHBG, the Zn$^{2+}$ ion affecting ligand binding is coordinated by histidines both from the long connecting loop $\beta_{12}$$-$$\beta_{13}$ and the hairpin turn $\beta_{8}$$-$$\beta_{9}$ and by aspartate from the $\beta_{6}$$-$$\beta_{7}$ hairpin loop (51). In a very recent report, the second LNS/LG domain of neurexin 1$\alpha$ (1$\alpha$LNS#2; Ref. 23) was shown to have a Ca$^{2+}$ ion in a similar position and with analogous coordination to laminin LG5 and agrin G3 domains. However, the neurexin 1$\alpha$LNS#2 has only three protein ligands due to the lack of counterpart for the aspartate in the $\beta_{12}$$-$$\beta_{13}$ loop. The metal ions are up to 13 Å apart when the structures are superimposed (Fig. 6, B and C).

The Central Cavity of NC4 Does Not Incorporate Retinoids—SHBG is known to accommodate various hydrophobic compounds, i.e. steroid hormones, at the central cavity between the $\beta$-sheets (51). Because COMP, functionally coupled to collagen IX, has also been suggested to serve as a store for hydrophobic molecules (41), the naturally occurring retinoids, we studied the ability of NC4 to incorporate such compounds. The experiments were carried out in the presence of Zn$^{2+}$ or EDTA in case the cation binding to NC4 should serve a gate-keeping function. Incubations of NC4 with an excess of solid all-trans-retinoic acid or vitamin D$_{3}$, or with concentrated stock solu-
tions of the two, did not result in specific perturbations of the chemical shifts in two-dimensional HSQC spectrum of NC4 (data not shown). The negative result seems to rule out the possibility that NC4 and COMP, in addition to their similar tissue distribution, mutual interactions, and shared involvement in the pathogenesis of multiple epiphyseal dysplasia, would also share a function as a store for retinoids suggested for COMP.

Sequence Alignment of the FACIT Collagen LNS Domains—A sequence alignment for the collagen LNS units has been reported (52), but it contained data from multiple species and did not include all appropriate FACIT sequences. To facilitate the evaluation of the conserved properties of the FACIT LNS domains, we performed an alignment of the amino acid sequence of NC4 with sequences of other LNS modules of human collagens with ClustalW (Fig. 7).

The TSPN-1 sequence was included as the most homologous published structure, and the laminin/H92512LG5 was included as a representative of other thoroughly characterized LNS modules. The laminin/H92511 chain of collagen XI was included to represent the LNS modules of collagens V, XI, XV, XVIII, XXIV, and XXVII, which are not members of the FACIT subfamily. The alignment shows that the -strands aligning with the strands 5-7, 10, 11, 13, and 14 of NC4 present the highest similarity of residues at corresponding positions. The central thirds of the sequences present the best alignment, whereas more gaps are inserted at the N- and C-terminal thirds. As already noted above based on structural superimpositions, the overall arrangement of the N and C termini of NC4 is homologous with that in a recent model of the TSPN-1 (27). In contrast, other LNS modules of human collagens with ClustalW (Fig. 7). The TSPN-1 sequence was included as the most homologous published structure, and the laminin α2LG5 was included as a representative of other thoroughly characterized LNS modules. The α1 chain of collagen XI was included to represent the LNS modules of collagens V, XI, XV, XVIII, XXIV, and XXVII, which are not members of the FACIT subfamily. The alignment shows that the -strands aligning with the strands 5-7, 10, 11, 13, and 14 of NC4 present the highest similarity of residues at corresponding positions. The central thirds of the sequences present the best alignment, whereas more gaps are inserted at the N- and C-terminal thirds. As already noted above based on structural superimpositions, the overall arrangement of the N and C termini of NC4 is homologous with that in a recent model of the TSPN-1 (27). In contrast, other

FIGURE 7. Sequence alignment for the LNS modules of FACIT collagens. The amino acid sequences of the LNS modules of FACITs were aligned with ClustalW. Secondary structure predictions were obtained by PSIPRED, and the predicted secondary structure elements showing a confidence value of at least 4 (scale from 0 to 9) were used to further adjust the alignment manually to match the experimentally determined secondary structure elements of NC4. The conservation of residue types is indicated by colored vertical bars for acidic (red), basic (blue), uncharged polar (pink), hydrophobic or bulky aromatic (gray), and cysteine (yellow) side chains. The sequence of NC4 is numbered from the mature N terminus omitting the signal peptide. The -strands (numbered) and the -helices of NC4, as determined by PROCHECK, are indicated with green and blue arrows, respectively. The sequences of the laminin α2LG5 domain (Protein Data Bank code 1DYK) and the TSPN-1 (Protein Data Bank code 1Z78) are included as homologues with a known crystal structure, and their -strands and -helices are indicated with horizontal green and blue bars, respectively. nA1 = the α1 chain of collagen n (where n denotes the collagen type number).
LNS modules of proteins outside the collagen superfamily, as represented by the laminin a2LG5 sequence in the alignment (Fig. 7), appear to be devoid of a region corresponding to the N-terminal third of the NC4 domain. At the same time, these proteins, with the exception of TSPN-1, lack the C-terminal cysteine of NC4 and its conserved surroundings. In a2LG5 the cysteine closest to the C terminus is used to link it to the β-strand corresponding to the β14 of NC4. Close homologues of a2LG5 thus adopt different N- and C-terminal folds compared with the LNS modules of the FACITs and the TSPN-1. In fact, the Cys21–Cys219 disulfide bond that is used to stabilize the N terminus of NC4 seems to be present in all FACIT collagens with the exception of collagen XXI but absent in all published LNS models and in LNS modules of other collagens (Fig. 7). Although the secondary structure elements of the N-terminal third of NC4 appear to have counterparts in other FACITs, the numerous gaps in the alignment at this region suggest that the N termini of FACITs are likely to show specific differences in their folding.

**DISCUSSION**

The fibril-associated collagen IX and other FACIT collagens are proposed to mediate the interaction of a collagen fibril with neighboring fibrils and with other extracellular matrix components. In cartilage, collagen IX covers collagen fibrils so that the NC4 and COL3 domains project away from the fibril body. Several interactions have been suggested, especially for the NC4 domain, reflecting its position and its putative homology to multifunctional LNS domains involved in glycan, proteoglycan, protein, and steroid ligand binding. The functional significance of LNS modules found at the N terminus of numerous collagens is nevertheless mostly a mystery.

To study the NC4 structure-function relationship at the atomic level, we have solved the crystal structure at high resolution. Furthermore having assigned NMR chemical shifts for the majority of residues, we used NMR titration to study the binding of various ligands. The NC4 structure is dominated by two, curved β-sheets sandwiched together (Fig. 1) revealing the homology to lectins, glycolytic enzymes, and LNS domains. Also the structure shows interesting similarity to metal ion-dependent LNS domains. TSPN-1 is structurally most similar to NC4 due to the additional structural similarity in the N- and C-terminal extensions. However, closer examination reveals differences at the atomic level that dictate how these domains gain function. The first ~30 residues, bearing the major heparin binding site (7), are unique for NC4. Unfortunately these N-terminal residues were largely unresolved both in the crystal structure and NMR chemical shift assignments. To our surprise, a Zn²⁺ was found in the NC4 crystal structure. The Zn²⁺ is coordinated by two aspartates from the β12-β13 loop, a water molecule, and a histidine from a symmetry-related molecule. The favorable tetragonal coordination, indicated by the strong anomalous signal and high occupancy, is thus accomplished by a crystal contact suggesting different coordination in solution.

NMR titration of NC4 with Zn²⁺ confirmed the location and showed that the $K_d$ was 11.5 mM. Thus, it seems that under the normal physiological submillimolar levels of Zn²⁺ in tissues, the metal ion in NC4 is not structural. The metal binding may instead provide an important means for regulating the high affinity binding of various ligands, including COMP. Alternatively the Zn²⁺ binding might become significant only under conditions of stress or injury where elevated Zn²⁺ levels can occur or at specific sites within tissue presenting elevated Zn²⁺ levels. Another possibility is that the affinity of the NC4 domain for Zn²⁺ is higher in intact collagen IX.

Many of the LNS domains showing homology to NC4 either are known to have metal ion-dependent ligand binding activities or their crystal structure contains at least one metal ion, usually Ca²⁺. The Ca²⁺ ion in the Gas6 G domain pair interface (25) as well as the one in the calnexin convex sheet (24) is structural rather than functional. On the other hand, the Ca²⁺ ions in the laminin LG5, the agrin G3, and the neurexin 1α_LNS#2 domains are strictly required for α-dystroglycan binding (20, 23, 26, 53). In SHBG, a Ca²⁺ in a position similar to that in Gas6 and a Zn²⁺ affecting estradiol binding have been identified (21, 51). The SHBG Zn²⁺ and the laminin LG5, agrin G3, and neurexin 1α_LNS#2 Ca²⁺ ions occupy a similar position on one edge of the β-sandwich. In NC4, the Zn²⁺ binds to the same edge of the β-sandwich (Fig. 1), and superimposition of the NC4, laminin LG5, agrin G3, SHBG (Fig. 6, B and C), and neurexin 1α_LNS#2 domains reveals not only the overall homology but also common features in the metal binding architecture. The striking similarity is that the same connecting loops are involved in metal ion binding and that these metal ions, except for the Zn²⁺ in NC4, have been shown to affect ligand binding (23, 51, 53–55).

The biological function of the NC4 Zn²⁺ ion remains elusive. Our studies by NMR and affinity chromatography demonstrate that Zn²⁺ has no effect on the heparin affinity of NC4. However, the NC4 Zn²⁺ data in this study and the recent agrin G3 (26) and neurexin 1α_LNS#2 (23) Ca²⁺ binding data together confirm that low affinity metal ions are commonly found on the LNS domain β-sandwich edges, the common ligand interaction regions. These ions can be crucial for ligand binding despite the low affinity of their interaction per se (23, 26). Furthermore incomplete coordination of the metal has been reported also in other LNS domains (20, 50, 51). In the laminin a2LG5 crystals, a buffer-derived SO₄²⁻ or an Asp side chain from a symmetry-related molecule completes the coordination, apparently mimicking a functional group of the ligand, α-dystroglycan (20, 50). Thus, the completion of the metal coordination in NC4 by a histidine of a symmetry-related molecule suggests that a protein ligand may coordinate directly to the Zn²⁺ of NC4 in vivo. Conformational changes triggered upon metal coordination would then increase specificity. A similar mechanism occurs for example in the metal ion-dependent adhesion site of integrins (56) where a metal bridge is a general and critical feature of I domain-ligand interactions. On the other hand, a transient NC4-ligand complex may serve as a high affinity binding site for the Zn²⁺ ion. In solution, the Zn²⁺ of NC4 is apparently coordinated by the aspartates, the Ser113-linked water, and an additional water molecule. This water is likely to be labile, like the coordinating waters on Zn²⁺ and could be easily replaced by a group of a ligand. Furthermore the completion of the metal coordination via a water molecule may be important for the efficiency of the ligand binding activity of NC4. In inte-
Crystal Structure of the Collagen IX NC4 Domain

grips and in numerous metalloenzymes one or two water molecules coordinating to the metal modify its electrophilicity, thus affecting the reactivity (56, 58).

NC4 binds to COMP in the presence of Zn\(^{2+}\). Initially COMP was shown to bind type I and type II collagens in a Zn\(^{2+}\)-dependent manner (59). We, and others, have shown that COMP binds to collagen IX at several locations, including the NC4 domain (6, 7, 60). The high affinity binding is Zn\(^{2+}\)-dependent, and the interaction is via the COMP C-terminal domain. The regulatory effect of Zn\(^{2+}\) on collagen interaction is apparently caused by binding of the metal to the COMP C-terminal domain, resulting in a change of the conformation of COMP (6). Unfortunately the atomic level basis of this putative Zn\(^{2+}\) dependence is poorly understood, and thus it remains to be seen whether a Zn\(^{2+}\) ion, possibly in the binding site identified here, has a direct role in the interaction of NC4 with COMP. Considering other possible ligands of NC4, \(\alpha\)-dystroglycan is an unlikely candidate as it does not occur in the extracellular matrix, \textit{i.e.} the tissue presenting the highest abundance of collagen IX in its long form. As described under “Results” the ability of NC4 to store hydrophobic compounds in the central cavity, similar to SHBG (51), was studied by NMR titration. NC4, unlike functionally coupled COMP, seems not to bind naturally occurring retinoids in the presence or absence of Zn\(^{2+}\). Furthermore TSPN-1 can bind to two cell surface integrins via sites that correspond to the Zn\(^{2+}\) binding loops \(\beta 6\)–\(\beta 7\) and \(\beta 12\)–\(\beta 13\) in NC4 (61, 62). It will be interesting to determine whether these loops, and the Zn\(^{2+}\) site formed by them, can serve as a docking site for chondrocyte integrins.

The FACIT subfamily currently contains collagens IX, XII, XIV, XVI, XIX, XX, XXI, and XXII (2). Within this group, the N terminus of the molecule is formed by an LNS motif in the case of collagens IX, XVI, and XIX, whereas the rest of the members contain additional motifs, namely von Willebrand factor A-domains and fibronectin type III repeats, located N-terminal to the LNS module. The sequence alignment of FACITs performed here (Fig. 7) shows that they are likely to present highly similar folds. The differences appear to be confined mainly to loop regions, whereas the overall arrangement of the domain is probably conserved. The apparently flexible regions corresponding to the loop between Cys\(^{21}\) and the \(\beta 1\) strand of NC4 show interesting division into long and short forms in FACITs, spanning 10–12 residues in most FACITs but containing an extra 10 amino acids in collagens XVI, XIX, and XXI (Fig. 7). As we could not trace this loop in NC4, the significance of this size difference remains unclear.

The sequence alignment (Fig. 7) also shows that the residues coordinating the Zn\(^{2+}\) in NC4 show limited conservation in the other FACITs. This does not indicate, however, that they could not bind divalent cations because the metal identity, location, and coordinating groups, often involving backbone carbonyl oxygen atoms, are quite variable in the published LNS structures. The lengths of the loops participating in the Zn\(^{2+}\) coordination in NC4 are apparently well conserved within the FACIT subfamily, suggesting that metal coordination may be a common characteristic. This can only be confirmed by experimental studies, however. Metal binding is apparently a property of the homologous proline/arginine-rich protein domain of the non-FACIT collagen XI (52).

Many of the LNS modules are capable of binding the highly sulfated glycosaminoglycan heparin. We have shown previously that NC4 also interacts with heparin with a \(K_d\) of 0.6 \(\mu\)M and mapped the major binding site to the extreme N terminus of the domain (7). The recent crystallographic model of TSPN-1 (27) included a synthetic pentameric heparin as a ligand contacted by the side chains of Arg\(^{29}\), Arg\(^{42}\), and Arg\(^{77}\) of TSPN-1. From our sequence alignment (Fig. 7), it is clear that these residues have positively charged counterparts in NC4 at corresponding locations, \textit{i.e.} Arg\(^{53}\), Lys\(^{65}\), and Arg\(^{95}\). The NMR studies performed here suggest that this region of NC4 does indeed show functional conservation of the heparin binding. This site has an estimated \(K_d\) of 1 mM for heparin; this is several orders of magnitude higher than the \(K_d\) of 71–80 nM reported for the TSPN-1 (63). Closer examination of the crystal structure of NC4, compared with that of TSPN-1 (27), offers some possible explanations for the dramatic differences in the affinities of the two sites. In NC4, the side chains of Lys\(^{65}\) and Arg\(^{95}\) are not well exposed, and the side chain of Arg\(^{53}\) is oriented away from these two residues so that a proper contact surface for heparin is not formed (data not shown). Of these three basic residues of NC4, Arg\(^{95}\) does not have a backbone NMR assignment, Lys\(^{65}\) shows negligible chemical shift perturbation upon heparin titration, and the cross-peak representing Arg\(^{95}\) disappears upon the first titration step. We, therefore, had to calculate the \(K_d\) by monitoring the adjacent residues in the sequence. The surroundings of Lys\(^{65}\) and Arg\(^{95}\) in the electron density map of NC4 have no suspicious density that could originate from a sulfate ion. A SO\(_4^{2-}\) is found at the region containing the positively charged residues His\(^{218}\), Arg\(^{223}\), Arg\(^{225}\), and Arg\(^{226}\) of NC4, but no evidence of heparin binding to this site is seen upon NMR titration. Furthermore NC4\(_{N\text{del}}\) lacking the positively charged N terminus does not bind heparin at physiological salt concentration (Fig. 5), thus confirming our earlier mapping of the binding site. Consequently despite the apparent conservation of the heparin binding basic residues of TSPN-1 in NC4, a completely different basic stretch is responsible for binding of the polyanionic glycan in NC4. This offers yet another reminder of the importance of experimental verification when making predictions on the structure-function relationships of proteins on the basis of sequence homology and existing homologous structures.

Within the FACIT subfamily, collagens XXI and XXII also contain a cluster of basic residues at their N terminus (Fig. 7), but their activity in heparin binding is unknown. On the other hand, most FACITs apparently have positively charged side chains at the location corresponding to the TSPN-1 heparin binding site (Fig. 7), but their heparin binding has not been studied experimentally either. A very recent report (64) describes the binding of heparan sulfate to the LNS module of collagen XI \(\alpha 1\) chain, the coordinating groups originating from the strand corresponding to the \(\beta 12\) of NC4. This region on the surface of the NC4 structure does contain the positive side chains of Arg\(^{166}\), Arg\(^{177}\), and Lys\(^{184}\), but they are 15–20 Å apart and do not show any chemical shift perturbation upon the heparin tetrasaccharide titration. Sequence comparison (Fig. 7)
suggests that the other FACITs also are unlikely to have a heparin binding site in this region.

CONCLUSION

The crystallographic model of the NC4 domain presented here is the first experimentally determined structure available for a collagen LNS module and provides an atomic level basis for understanding the structure-function relationship of NC4. The NMR chemical shift assignment of the backbone amide bonds of NC4 performed here enables detailed analyses of the molecular interactions of NC4 at atomic resolution as exemplified by the results of the Zn$^{2+}$ and heparin tetrasaccharide titrations. The NC4 crystal structure revealed a Zn$^{2+}$ in a position similar to that of other metal ion-dependent LNS domains. NMR titration with Zn$^{2+}$ mapped the same binding site in solution but also revealed its nature of low affinity. Ligand binding in the other metal ion-dependent LNS domains is regulated by low affinity metal binding sites. Similarly the NC4 metal ion may control ligand binding suggesting that this common ligand interaction region is important for NC4 as well. The N-terminal deletion mutant confirmed the major heparin binding site, and NMR titration with a heparin tetrasaccharide revealed a secondary site homologous to TSPN-1. The Zn$^{2+}$ ion is not crucial for heparin binding, supporting its putative role in modulating NC4 multifunctionality. Finally the NC4 structure reveals new insights into the general properties of the LNS family of extracellular protein domains, and it will serve as a model in the detailed characterization of the respective domains of the other FACIT collagens.

Acknowledgments—We thank Harri Heikkinen for preparation and analyses of mutated NC4 and Anne Hakonen for expert technical assistance. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation resources, and we thank Dr. Gavin Fox for assistance in using beamline BM14. We also thank Dr. Tomi Airenne for help with the data collection.

REFERENCES

1. Myllyharju, J., and Kivirikko, K. I. (2001) Ann. Med. 33, 7–21
2. Ricard-Blum, S., and Ruggiero, F. (2005) Pathol. Biol. 53, 430–442
3. Swiderski, R. E., and Solursh, M. (1992)
4. Savontaus, M., Ihanamäki, T., Perala, M., Metsäranta, M., Sandberg-Lall, M., and Vuorio, E. (1998) Histochem. Cell Biol. 110, 149–159
5. Sasaki, T., Nyavey, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
6. Sasaki, T., Knyazev, P. G., Bork, P., and Reich, J. G. (1998) J. Mol. Biol. 279, 725–730
7. Shagir, J. D., Bergeron, J. I., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., and Cygler, M. (2001) Mol. Cell 6, 633–644
8. Sasaki, T., Nyavey, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
9. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) Matrix Biol. 251, 2286–22905
10. Beckmann, G., Hanke, J., Bork, P., and Reich, J. G. (1998) J. Mol. Biol. 275, 725–730
11. Hohenester, E., Tisi, D., and Reich, J. G. (1999) Mol. Cell 4, 783–792
12. Grishkovskaya, I., Avvakumov, G. V., Sklenar, G., Dales, D., Hammond, G. L., and Muller, Y. A. (2000) EMBO J. 19, 504–512
13. Rucenko, G., Nguyen, T., Chelliah, Y., and Seidenhofer, J. (1999) Cell 99, 93–101
14. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) J. Biol. Chem. 281, 2286–22905
15. Shagir, J. D., Bergeron, J. I., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., and Cygler, M. (2001) Mol. Cell 6, 633–644
16. Sasaki, T., Nyavey, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
17. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) Matrix Biol. 251, 2286–22905
18. Simons, P. D., and Muller, Y. A. (2001) Trends Biochem. Sci. 26, 363–368
19. Sasaki, T., Knyazev, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
20. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) J. Biol. Chem. 281, 2286–22905
21. Shagir, J. D., Bergeron, J. I., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., and Cygler, M. (2001) Mol. Cell 6, 633–644
22. Sasaki, T., Nyavey, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
23. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) J. Biol. Chem. 281, 2286–22905
24. Schrag, J. D., Bergeron, J. I., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., and Cygler, M. (2001) Mol. Cell 6, 633–644
25. Sasaki, T., Nyavey, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
26. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) J. Biol. Chem. 281, 2286–22905
27. Shagir, J. D., Bergeron, J. I., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., and Cygler, M. (2001) Mol. Cell 6, 633–644
28. Sasaki, T., Nyavey, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
29. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) J. Biol. Chem. 281, 2286–22905
30. Schrag, J. D., Bergeron, J. I., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., and Cygler, M. (2001) Mol. Cell 6, 633–644
31. Sasaki, T., Knyazev, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
32. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) J. Biol. Chem. 281, 2286–22905
33. Shagir, J. D., Bergeron, J. I., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., and Cygler, M. (2001) Mol. Cell 6, 633–644
34. Sasaki, T., Knyazev, P. G., Cheburkin, Y., Göhring, W., Tisi, D., Ullrich, A., and Hohenester, E. (2002) J. Biol. Chem. 277, 44164–44170
35. Scheckler, L. R., Henry, L., Sugita, S., and Rude, G. N. (2006) J. Biol. Chem. 281, 2286–22905
Crystal Structure of the Collagen IX NC4 Domain

Permi, P., and Kilpeläinen, I. (2006) Protein Sci. 15, 1760–1768
41. Guo, Y., Bozic, D., Malashkevich, V. N., Kammerer, R. A., Schulthess, T., and Engel, I. (1998) EMBO J. 17, 5265–5272
42. Krishnan, V. V., Sukumar, M., Gierasch, L. M., and Cosman, M. (2000) Biochemistry 39, 9119–9129
43. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
44. Merritt, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524
45. Esnouf, R. M. (1997) J. Mol. Graph. Model. 15, 132–134
46. Harding, M. M. (2001) Acta Crystallogr. Sect. D Biol. Crystallogr. 57, 401–411
47. Wittekind, M., and Müller, L. (1993) J. Magn. Reson. B101, 201–205
48. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Prog. Nucl. Magn. Reson. Spectrosc. 34, 93–158
49. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) J. Mol. Biol. 247, 536–540
50. Tisi, D., Talts, J. F., Timpl, R., and Hohenester, E. (2000) EMBO J. 19, 1432–1440
51. Avvakumov, G. V., Muller, Y. A., and Hammond, G. L. (2000) J. Biol. Chem. 275, 25920–25925
52. Fallahi, A., Kroll, B., Warner, L. R., Oxford, R. J., Irwin, K. M., Mercer, L. M., Shadle, S. E., and Oxford, J. T. (2005) Protein Sci. 14, 1526–1537
53. Sugita, S., Saito, F., Tang, J., Satz, J., Campbell, K., and Südhof, T. C. (2001) J. Cell Biol. 154, 435–445
54. Campanelli, I. T., Gayer, G. G., and Scheller, R. H. (1996) Development 122, 1663–1672
55. Talts, J. F., Andac, Z., Göhring, W., Brancaccio, A., and Timpl, R. (1999) EMBO J. 18, 863–870
56. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J., and Liddington, R. C. (2000) Cell 101, 47–56
57. Berg, J. M., and Shi, Y. (1996) Science 271, 1081–1085
58. Coleman, J. E. (1998) Curr. Opin. Chem. Biol. 2, 222–234
59. Rosenberg, K., Olsson, H., Mörgelin, M., and Heinegård, D. (1998) J. Biol. Chem. 273, 20397–20403
60. Holden, P., Meadows, R. S., Chapman, K. L., Grant, M. E., Kadler, K. E., and Briggs, M. D. (2001) J. Biol. Chem. 276, 6046–6055
61. Li, Z., Calzada, M. J., Sipes, J. M., Cashel, J. A., Krutzsch, H. C., Annis, D. S., Mosher, D. F., and Roberts, D. D. (2002) J. Cell Biol. 157, 509–519
62. Calzada, M. J., Sipes, J. M., Krutzsch, H. C., Yurchenco, P. D., Annis, D. S., Mosher, D. F., and Roberts, D. D. (2003) J. Biol. Chem. 278, 40679–40687
63. Yabkowitz, R., Lowe, J. B., and Dixit, V. M. (1989) J. Biol. Chem. 264, 10888–10896
64. Warner, L. R., Brown, R. J., Yingst, S. M. C., and Oxford, J. T. (2006) J. Biol. Chem. 281, 39507–39516