Biophysical Characterization of the C-propeptide Trimer from Human Procollagen III Reveals a Tri-lobed Structure*

Received for publication, September 6, 2001, and in revised form, October 25, 2001
Published, JBC Papers in Press, October 29, 2001, DOI 10.1074/jbc.M108611200

Simonetta Bernocco‡, Stéphanie Finet‡, Christine Ebel, Denise Eichenberger‡,
Marlène Mazzorana‡‡, Jean Farjanel‡, and David J. S. Hulmes‡ ‡

From the ‡ Institut de Biologie et Chimie des Protéines, CNRS UMR 5086, Université Claude Bernard Lyon 1, 69367 Lyon cedex 7, the ‡ European Synchrotron Radiation Facility, 38043 Grenoble cedex, and the ‡ Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, 38027 Grenoble cedex 1, France

Procollagen C-propeptide domains direct chain association during intracellular assembly of procollagen molecules. In addition, they control collagen solubility during extracellular proteolytic processing and fibril formation and interact with cell surface receptors and extracellular matrix components involved in feedback inhibition, mineralization, cell growth arrest, and chemotaxis. At present, three-dimensional structural information for the C-propeptides, which would help to understand the underlying molecular mechanisms, is lacking. Here we have carried out a biophysical study of the recombinant C-propeptide trimer from human procollagen III using laser light scattering, analytical ultracentrifugation, and small angle x-ray scattering. The results show that the trimer is an elongated molecule, which by modeling of the x-ray scattering data appears to be cruciform in shape with three large lobes and one minor lobe. We speculate that each of the major lobes corresponds to one of the three component polypeptide chains, which come together in a junction region to connect to the rest of the procollagen molecule.

Fibril-forming collagens (types I, II, III, V, and XI) are synthesized and secreted into the extracellular matrix in precursor form, procollagens (~500 kDa), with large N- and C-terminal propeptide regions (1). The propeptides increase the solubility of the procollagen molecule, thus preventing premature fibril formation inside the cell (2). After secretion, propeptides are cleaved to varying extents by specific procollagen N- and C-proteinases (3–5) and also other proteinases (6, 7), thereby triggering fibril assembly of collagen. Once cleaved, both N- and C-propeptides are thought to control further collagen synthesis by a process of feedback inhibition (8–11). Also, in the extracellular matrix, the N-propeptides are involved in growth factor signaling (12), whereas the C-propeptides have been implicated in interactions with procollagen C-proteinase enhancer (13), mineralization (14–16), integrin receptor binding (17, 18), cell growth arrest (19), and chemotaxis (20).

In addition to their extracellular roles, numerous observations demonstrate the importance of the C-propeptide domain in chaperone-assisted chain association during intracellular assembly of procollagen molecules (21–24). Each procollagen molecule consists of three polypeptide chains encoded by one or more genes, giving rise to homotrimers or heterotrimers, respectively, with specific chain stoichiometries. For example, procollagen III molecules are homotrimers with the chain composition pron1(III)3, whereas procollagen I molecules are normally heterotrimers of the form pron1(I)2pro2(I). The C-propeptides direct association within the rough endoplasmic reticulum to ensure correct chain stoichiometry, which is particularly important in cells producing more than one procollagen type. Once associated into a trimer, and after prolyl hydroxylation, triple helix formation within the collagenous region is initiated at the C terminus and proceeds in a zipper-like manner toward the N terminus (22, 25). The importance of the C-propeptides in chain association has been demonstrated by both naturally occurring (26) and engineered (22, 27) mutations/deletions, which result in failure of or impaired procollagen molecular assembly. Very recently, it has been demonstrated that recombinant type I collagen molecules devoid of N- or C-terminal propeptides can assemble in a yeast expression system to form correctly aligned triple-helices (28), albeit with poor control of chain stoichiometry. Because expression was necessarily carried out at relatively low temperatures, however, the significance of these results for mammalian cells is unclear. Overwhelming evidence indicates that in mammalian cells the C-propeptides are essential for procollagen chain association with the correct stoichiometry.

Throughout the fibrillar procollagens, the C-propeptide domain is highly conserved (1, 29), which suggests a common overall three-dimensional structure in the C-propeptide region once the three chains have assembled to form a trimer. Within this conserved framework, a relatively variable and discontinuous sequence of 15 amino acid residues has been identified as being required for type-specific chain selection and trimer formation (30). In addition, the C-propeptide trimer (composed of three chains each of molecular mass ~30 kDa) contains both inter- and intramolecular disulfide bonds, the former involving cysteine residues in the N-terminal region of each chain and the latter involving cysteines in the C-terminal region (31). Beyond this, there exists no three-dimensional structural information for any of the procollagen C-propeptide trimers that might be used to understand the mechanism of chain selection and association. Furthermore, there are no known homologous proteins for which three-dimensional structures are available, molecular modeling by homology is precluded. Recently, a baculovirus system has been developed for the
expression of recombinant human procollagen III C-propeptide in insect cells (32). In this system, disulfide-linked trimers are formed and secreted into the culture medium that appear to have the same secondary structure as native C-propeptide trimers from chick procollagen I. Here we have used a variety of biophysical approaches to study the overall shape of the procollagen III C-propeptide trimer (CPIII)1 in solution. We show that the trimer has an elongated shape, which by small angle x-ray scattering appears to be a cruciform structure with three large lobes and one small lobe, consistent with each large lobe corresponding to one of the three constituent chains.

EXPERIMENTAL PROCEDURES

Cell Culture and Protein Expression—Trichoplusia ni (BTI-TN-5B1-4, High Five) insect cells (Invitrogen) maintained in suspension culture at 1.0 × 107 cells/ml were infected with baculovirus (kind gift from Dr. D. Prokop (32)) and incubated at 27 °C in Express Five serum-free medium (Invitrogen) complemented with 16 mM l-glutamine (Invitrogen), 100 units/ml penicillin, 0.1 mg/ml streptomycin (Sigma), and 0.1% Pluronic® F-68 (Invitrogen). Conditioned medium was then centrifuged at 900 × g for 15 min followed by the addition of enzyme inhibitors to the supernatant to final concentrations 10 mM EDTA, 10 mM 5,5'-dithio-bis(2-nitrobenzoic acid), and 0.5 mM phenylmethylsulfonyl fluoride (all from Sigma). After a second centrifugation for 15 min at 20,000 × g and 4 °C, the supernatant was stored at −80 °C for up to 10 months without significant loss of protein integrity.

Protein Purification—Three chromatographic steps were used to purify recombinant CPIII. Typically 800 ml of conditioned medium was processed. Unless otherwise stated, all procedures were carried out at 4 °C. After the addition of 4-aminobenzamidine dihydrochloride (final concentration, 1 mM) and adjustment of the pH to 7.4 using a stock solution of 1 M Tris-HCl, pH 8.5, medium was centrifuged for 15 min at 10,000 × g to pellet any suspended material. Clarified medium was loaded at 75 m1/h onto a 5 × 2.5-cm column of concanavalin A-Sepharose (Amersham Biosciences, Inc.) pre-equilibrated in buffer A (50 mM Tris-HCl, 300 mM NaCl, 1 mM CaCl2, 1 mM MgCl2, pH 7.4). After extensive washing to remove non-bound material, CPIII was eluted at 10 ml/h using buffer A containing 1 mM methyl-α-D-mannopyranoside (Sigma). The eluate was then diluted with 5 volumes of 50 mM Tris-HCl, pH 8.5, and loaded onto a 10 × 2.6-cm column of DEAE-Sepharose (Amersham Biosciences, Inc.) pre-equilibrated in buffer B (50 mM Tris-HCl, pH 8.5). After extensive washing, bound proteins were eluted with a 500 ml of linear gradient of buffer B containing 50–300 mM NaCl. Fractions of interest were then pooled, 3.4 mM (NH4)2SO4 was added to a final concentration of 1 mM, and the sample was loaded at 10 ml/h and at room temperature on a 5 × 1.6-cm column of butyl-Sepharose (Amersham Biosciences, Inc.) pre-equilibrated in buffer C (50 mM Tris-HCl, 1 mM (NH4)2SO4, pH 8.0). After washing, bound proteins were eluted with 10 ml/h and at room temperature using a linear gradient of buffer C containing 1 to 0 mM (NH4)2SO4. After analysis by SDS-PAGE, fractions of interest were pooled, concentrated up to 10 mg/ml using an UltraFree 15 device (Millipore, 30-kDa cut off), and stored at −80 °C.

Protein Assay—CPIII concentrations were measured by absorbance at 280 nm using a calculated extinction coefficient of 2.12 ml/mg·cm−1, based on the presence of 8 cysteine, 5 tryptophan, and 7 tyrosine residues in the sequence of each polypeptide chain (33). This value was confirmed using a commercial protein assay (Pierce) based on the Bradford method (34) using bovine serum albumin as the protein standard.

Electrophoresis and Western Blotting—SDS-PAGE (10% acrylamide for samples reduced with dithiothreitol, 6% acrylamide in non-reducing conditions) was carried out according to Laemmli (35) using reagents from Bio-Rad. Western blotting and immuno-labeling on polyvinylidene difluoride membranes were done according to Towbin et al. (36). The three monoclonal antibodies used for the immuno-labeling (kind gifts from Dr. E. Burchardt (37)) were produced against single chains of CPIII expressed in Escherichia coli. Their epitopes are as follows: 483D34 (amino acids 1–207), 48B14 (amino acids 207–245) (numbering follows the TrEMBL CPIII sequence, GenBank™ accession number Q15112). Detection was with a commercial anti-mouse secondary antibody (Dako) coupled with alkaline phosphatase followed by color development using an alkaline phosphatase conjugate substrate kit (Bio-Rad).

Deglycosylation—Deglycosylation in native conditions was carried out in 50 mM Hepes, 100 mM NaCl, pH 7.4, using CPIII at a concentration of 0.1 mg/ml and N-glycosidase F (Roche Molecular Biochemicals) at 0.1 unit/μg substrate and incubating for 4 h at 37 °C. Deglycosylation of denatured CPIII was done on protein previously heated for 10 min at 100 °C in the presence of 1% SDS then diluted to 0.1% SDS with buffer containing 0.15% Nonidet P-40 and applying the same procedure as used for native conditions.

SAXS—Samples were analyzed in Buffer D (20 mM Hepes, 150 mM NaCl, pH 7.4) by both static and dynamic light scattering (38) using a Malvern 4700 spectrometer and 7132 256-channel correlator with a 40-nm He-Ne laser (Siemens). After analysis, solutions were centrifuged at 4 °C for 15 min at 15,000 × g, then supernatants were transferred to 10-mm-diameter sample cells and examined at 25 °C. Samples were analyzed in the concentration range of 1.8–4.2 mg/ml. For static light scattering, samples were analyzed in the angular range 30–130°, and the molecular mass of CPIII was calculated using a Zimm plot (38). Rayleigh ratios were determined with reference to a toluene standard, and a value of 0.182 ml/g was assumed for the refractive index increment. For dynamic light scattering, samples were analyzed at 90°, and correlation curves were analyzed using the CONTIN program provided by the manufacturer. Rayleigh ratio (2s, where s = 1–Dcorrected/D0) and diffusion coefficients (experimentally measured s and scorrected/D0) were related to the frictional factor f and hydrodynamic diameter D0 by the relation D0 = RTNf = RTf(1−D0). Here D0 is the radius of gyration, T is the absolute temperature, N is Avogadro’s number, and f the solvent viscosity.

Analytical Ultracentrifugation—Sedimentation velocity experiments were performed using a Beckman XL-A analytical ultracentrifuge and an AN-60 Ti rotor (Beckman Instruments). The experiments were carried out at 25 °C in Buffer D. Three samples of 400 μl at protein concentrations of 0.24, 0.44, and 0.81 mg/ml were loaded into 1.2-cm path cells and centrifuged at 42,000 rpm. Scans were recorded at 278 nm every 5 min using a 0.003-mm radial spacing. Sedimentation profiles were analyzed by different methods using time derivative analysis or direct modeling of boundary profiles in terms of one non-interacting component (dcdf + and Svedberg from J. Philo, Sedfit from Ref. 39). Sedfit takes advantage of a radial and time-independent noise subtraction procedure (40). These procedures allow the evaluation of both sedimentation (s) and diffusion (D) coefficients, from which the molar mass is derived using the Svedberg equation: M = sTsD1−p. We estimated the partial specific volume of the protein to be 0.721 ml/g (assuming a high molar mass of 18,000 Da for a single polypeptide chain), the solvent density ρ to be 1.004 g/ml, and the solvent viscosity η to be 0.908 mPa·s, at 25 °C using Sednterp software (V1.01; developed by D. B. Haynes, T. Laue, and J. Philo) for the calculation of the corrected S20,w and D20,w values.

Small Angle X-ray Scattering—For SAXS, CPIII samples in buffer D were analyzed at 20 °C on beamline ID2 (42) at the European Synchrotron Radiation Facility. Gridded samples were spread on a 4 to 10 mg/ml/ml dose range 4–10 mg/ml were placed in a quartz capillary (GLAS, 2-mm diameter, 10-μm thickness) mounted in a thermostatted flow-through cell. Scattering was measured using a two-dimensional detector, either an x-ray Image Intensifier FReLoN CCD camera (at 2.5 m from the sample) or a multiview proportional gas-filled detector. Gridded samples were spread on a photodiode dose range 4-10 mg/ml were spread on a 4 to 10 mg/ml dose range 4–10 mg/ml were placed in a quartz capillary (GLAS, 2-mm diameter, 10-μm thickness) mounted in a thermostatted flow-through cell. Scattering was measured using a two-dimensional detector, either an x-ray Image Intensifier FReLoN CCD camera (at 2.5 m from the sample) or a multiview proportional gas-filled detector (at 1 m) using x-rays of wavelength λ = 0.995 Å. Data were averaged from individual exposures of 500 ms (CCD detector, 1024 × 1024 pixels) or 600 s (gas-filled detector, 512 × 512 pixels). Two-dimensional data reduction consisted of normalization for detector response, exposure time and sample transmission, absolute intensity calibration, azimuthal integration, and background subtraction from buffer alone to obtain the normalized scattered intensity I as a function of Q or s, where Q = 2πs (Q being the scattering angle). Data from the CCD detector in the Q range 0.01–0.15 Å−1 were merged with data from the gas detector in the Q range 0.15–0.3 Å−1. No concentration dependence in the scattering curves was observed at the concentrations used.

SAXS data were analyzed using Guinier plots (43) to determine the radius of gyration Rg as well as the apparent molecular mass of CPIII. The latter was obtained by extrapolation to zero angle with reference to a bovine serum standard. The program GNOM (44) was used to determine the distance distribution function p(ri) after eliminating data for Q < 0.0272 Å−1 to suppress subsidiary maxima at large distances (45). For modeling of the structure, three programs were used: SASHA (46), DAMMIN (47), and DALI_GA (45, 48). Model building using spherical harmonics (SASHA) was carried out up to a maximum harmonic order of 4, corresponding to 19 independent pa-

1 The abbreviations used are: CPIII, C-terminal propeptide region of the human type III procollagen molecule; SAXS, small angle x-ray scattering.
Structure of the Procollagen C-propeptide Trimer

RESULTS

Protein Production and Characterization—Optimal production of recombinant CPIII was found when insect cells were infected with baculovirus for 45 h with a multiplicity of infection (expressed as number of viral particles per cell) = 3. As previously reported (32), the presence in the medium of the ~90-kDa disulfide-linked C-propeptide trimmer was confirmed by SDS-PAGE in reducing and non-reducing conditions (not shown). In reducing conditions, two bands (I and II) were systematically observed with apparent molecular masses of 34 and 32 kDa, respectively. Both bands were identified by immunoblotting using monoclonal antibody 48B14, specific for the central region of the human procollagen III C-propeptide (not shown). The relative intensities of these bands varied as a function of multiplicity of infection with the upper band (band I), predominating at a multiplicity of infection = 3. In these conditions CPIII production, representing 30% of total protein in the medium, was about 20 mg/liter.

Recombinant CPIII was purified using a three-step procedure (Fig. 1). As previously reported (32), initial purification was by concanavalin-A affinity chromatography. Subsequent cation exchange chromatography at low pH led to considerable losses due to proteolysis; hence, this step was replaced by DEAE anion-exchange at pH 8.5. A final purification step by hydrophobic interaction on butyl-Sepharose resulted in essentially pure CPIII that was enriched in band I in the early part of the elution gradient (see Fig. 1). The total yield of purified CPIII from 1 liter of conditioned medium was 6.4 mg.

To determine the nature of bands I and II seen by SDS-PAGE in reducing conditions, purified CPIII was probed by immunoblotting using monoclonal antibodies specific for the N-terminal 30 residues (antibody 48D34) and C-terminal 39 residues (antibody 48D19) of human CPIII (37). Both bands I and II were recognized by both antibodies (not shown), indicating that the difference in apparent molecular mass was not due to partial proteolytic degradation. A further possibility was differences in post-translational modifications, in particular, N-linked glycosylation. The observed molecular mass difference (~2 kDa) corresponds to the presence of a single high mannose glycan Man8GlcNAc2, as commonly found in glycoproteins produced in insect cells (41). In addition, there is a single asparagine linked N-glycosylation site in each of the three identical chains of CPIII. Recombinant CPIII, both native and denatured, was thus subjected to treatment by N-glycosidase F, then analyzed by SDS-PAGE in reducing and non-reducing conditions. As shown in Fig. 2, N-glycosidase F treatment of native CPIII led to the appearance of four bands migrating in the region of 100 kDa by SDS-PAGE in non-reducing conditions as well as a relative increase in the intensity of band II in reducing conditions. When exposed to N-glycosidase F after denaturation of CPIII, deglycosylation was complete, resulting in a single fast migrating band in non-reducing conditions and the total disappearance of band I in reducing conditions. We conclude that the four bands observed in non-reducing conditions correspond to CPIII trimers glycosylated at a single site on 0, 1, 2, or 3 chains. Thus, the recombinant CPIII used in these experiments consists of a mixture of mostly fully glycosylated (i.e. on all three chains) and partially glycosylated (i.e. on two of three chains) trimers.

Light Scattering—Analysis of CPIII by static light scattering (not shown) revealed no significant angular dependence in scattering intensity, indicating the absence of large aggregates in the concentration range studied (2–4 mg/ml). Extrapolation to zero concentration and zero angle using a Zimm plot gave a molecular mass of 103 ± 8 kDa, consistent with the mass calculated from the amino acid sequence of 88 kDa (Table I). We conclude that CPIII behaves as a freely soluble trimer (composed of three polypeptide chains) in solution.

Dynamic light scattering of CPIII at 4 mg/ml gave an apparent diffusion coefficient (Table I) corresponding to a hydrodynamic radius 1.35 times greater than for a normally hydrated spherical protein of mass 88 kDa. This indicated that CPIII has a highly elongated shape equivalent to a prolate (cigar-like) or oblate (disc-like) ellipsoid with axial ratios 6.6 or 7.5, respectively.

Analytical Centrifugation—Sedimentation velocity profiles were nicely modeled by considering one single non-interacting...
species, giving similar results whatever the data treatment procedure and the sample concentration. One example is shown in Fig. 3. Sedimentation and diffusion coefficients are reported in Table I. The diffusion coefficient is similar but slightly larger than that measured by dynamic light scattering, a probable consequence of the slight heterogeneity in the glycosylation of the sample (see above). The molecular mass (90 kDa) calculated from the sedimentation and diffusion coefficients obtained by analytical ultracentrifugation and dynamic light scattering, respectively, agreed closely with that based on the known amino acid sequence (88 kDa).

**Small Angle X-ray Scattering**—Measurement of the x-ray scattering intensity as a function of the scattering angle gives information on the overall shape of proteins in solution (45, 47, 48). Scattering data were obtained for CPIII in the concentration range 4–10 mg/ml at two different sample-detector distances and then corrected for background and merged to produce the curve shown in Fig. 4. Guinier analysis of the low angle region (not shown) gave a linear profile corresponding to a radius of gyration of 33.4 Å, equivalent to a sphere of diameter 86.2 Å. Extrapolation to the zero angle gave an apparent molecular mass of 100 kDa, consistent with data obtained by light scattering and analytical ultracentrifugation (Table I). Determination of the distance distribution function \( p(r) \) using the program GNOM revealed an asymmetric curve (Fig. 5), with a tail extending to \(-115 \) Å, corresponding to the maximum interatomic spacing within the structure. These values compare with a diameter of 58.6 Å for a spherical (unhydrated) particle based on the known molecular weight (Table I). Thus standard SAXS analysis confirmed the molecular mass as well as the elongated shape of the protein indicated by light scattering and analytical centrifugation.

To obtain the low resolution structure of CPIII, model structures were fitted to the full angular range of the SAXS data using the complementary approaches of spherical harmonic shape restoration (46) and dummy atom modeling (45, 47, 48). Using spherical harmonics, the program SASHA produced a cruciform structure (Fig. 6a) with three long arms of equal length and a fourth short arm. Each of the arms was \(-20 \) Å in thickness. The maximum length of the structure (115 Å) was consistent with that determined using GNOM. In orthogonal views, the structure was flat, with a width of \(-50 \) Å. Agreement between observed and calculated SAXS data was excellent (Fig. 4).

In view of the unusual nature of the structure produced by SASHA, alternative methods were used to model the SAXS data using structures built from dummy atoms in regular three-dimensional array. The simulated annealing approach using DAMMIN (47) gave rise to a variety of different multilobed structures with moderate fits to the high angle region of the data (not shown). In contrast, searching conformational space with the genetic algorithm DALAI_GA (45, 48) resulted in consistent structural features and excellent agreement be-

---

**TABLE I**

| Method                          | Molecular mass | Structural parameters | Diameter of equivalent sphere |
|---------------------------------|----------------|-----------------------|-------------------------------|
| Calculated\(^a\)                | 88.1 kDa       | \( D_{20,\text{w}} = 4.8 \pm 0.3 \) | 58.6 Å\(^b\)                |
| Static light scattering         | 103 ± 8        | \( D_{20,\text{w}} = 5.0 \pm 0.05 \) | 89 ± 6                       |
| Dynamic light scattering        | 90 ± 6         | \( D_{20,\text{w}} = 5.4 \pm 0.4 \) | 80 ± 6                       |
| Analytical ultracentrifugation  | 81 ± 6         | \( \bar{R}_g = 33.4 \pm 0.3 \) | 86.2 ± 0.4                   |
| Small angle x-ray scattering    | 100 ± 10       |                       |                               |

\(^a\) Assuming one high mannose glycan Man\(_6\)GlcNAc\(_2\) per chain (41).

\(^b\) Unhydrated.

\(^c\) Assuming hydration of 0.3 g of water/g of protein.

\(^d\) Mass calculated using \( s_{20,\text{w}} \) from analytical ultracentrifugation.

---

**Fig. 3. Analytical ultracentrifugation of CPIII.** Sedimentation profiles of CPIII at 0.44 mg/ml, centrifuged at 42,000 rpm for up to 3.5 h. The residuals in the upper panel indicate the difference between modeled and experimental absorbances (Abs), which are displayed as lines and dots, respectively, on the lower panel. The analysis was performed on 21 profiles, considering one homogeneous ideal non-interacting species in the program Sedfit.

**Fig. 4. Small angle x-ray scattering of CPIII.** Radially averaged x-ray scattering intensity corrected for detector response and after background subtraction as a function of scattering angle \( 2 \theta \), expressed as \( Q = 4 \sin \theta / \lambda \). Also shown is the best-fit curve obtained using the modeling program SASHA, which is indistinguishable from those obtained with DALAI_GA (not shown).
between observed and calculated SAXS curves (Fig. 4). A gallery of structures produced by DALAI-GA is shown in Fig. 6b. Although the structures show clear variability, all have in common the presence of three major lobes, often arranged at right angles to each other and confined to the same plane, with a maximum dimension of ~120 Å. These structures show a striking similarity to that produced by SASHA (Fig. 6a). In both cases, no constraints were imposed on the modeling either in terms of symmetry or subunit composition.

DISCUSSION

Here we have used the expression system described previously (32) to produce large amounts of recombinant CPIII to investigate its overall shape in solution. The purification procedure was modified to avoid proteolytic losses in low pH conditions, and a further chromatography step was added resulting in essentially pure protein. Recombinant CPIII formed disulfide-linked trimers, as previously demonstrated (32). Unlike the previous report, however, we show here that recombinant CPIII is heterogeneous as a result of partial N-glycosylation. Different variants resulting from the presence or absence of N-glycosylation at a single site on each of the three chains could be partially separated by hydrophobic interaction chromatography as has been demonstrated for other glycoproteins produced in a baculovirus system (51). We found recombinant CPIII to be almost fully N-glycosylated, with a minor variant glycosylated on two of its three chains.

By static light scattering, analytical centrifugation/dynamic light scattering, and SAXS, recombinant CPIII was found to have the expected molecular mass of about 90 kDa at concentrations up to several mg/ml. Thus, there was no evidence of aggregation in the conditions used, consistent with the known function of the C-propeptide domain in increasing the solubility of the procollagen molecule compared with pN-collagen (procollagen minus the C-propeptide domain) by at least 100-fold (52, 53). The value of the diffusion coefficient also gave information on the shape of CPIII in solution. It showed clearly that CPIII is a relatively elongated molecule, with a hydrodynamic radius equivalent to that of a prolate or oblate ellipsoid of axial ratio of at least 4. This was further supported by the small angle x-ray scattering data, where Guinier analysis indicated a radius of gyration some 47% greater than that expected for a sphere with the same molecular mass. Finally, calculation of the distance distribution function from the small angle x-ray scattering data, which shows the distribution of all interatomic distances in the structure, revealed a tail with a maximum dimension of 115 Å, again showing CPIII to be highly non-spherical.

The SAXS data were used to fit a low resolution model for the three-dimensional structure of recombinant CPIII. Using the program SASHA (46), which builds up the shape as a sum of spherical harmonics, the best-fit model with maximum harmonic order 4 was obtained. Structure determination using SAXS data is limited by the relatively low information content that results from spherical averaging due to scattering from molecules in all possible orientations in solution. As a result of this, different structures can theoretically give rise to the same scattering curves. This problem does not arise with SASHA (46), however, as long the number of independent parameters used to describe the structure (in this case 19) is no more than about 1.5 times the number of Shannon channels in the data (in
The C-terminal region of the procollagen molecule is depicted as a rectangle, has a length of ~40 Å (61).

This conclusion is supported by the results of the alternative modeling program used, DALAI GA (45, 48), which builds models from dummy atoms by searching all possible conformations using a genetic algorithm. Despite the stochastic nature of the second approach, the results from several independent simulations were in remarkable agreement both with each other and with the results of the SASHA program. In each case, a tri-lobe structure was observed, often with a fourth minor lobe and with all lobes in the same plane. This agreement between the results of the two modeling programs lends further support to the conclusion that this is the correct structure. The fact that additional information or symmetry constraints were imposed during the model fitting whereas the best-fit model is so readily interpretable in terms of the known subunit composition (see below) is an additional vindication of the structure.

The structure resulting from the SAXS data is readily interpretable in terms of the chain composition of the CPIII trimer. Because the subunits are identical, it seems reasonable to assign each of the three large lobes seen in the models to each of the three polypeptide chains. The small minor lobe would then correspond to a junction region involving all three chains that links up to the rest of the procollagen molecule (Fig. 7). In the absence of high atomic resolution data, these structural assignments are provisional. Nevertheless, the proposed structure fits well with what is known about the positions of inter- and intra-chain disulfide bonds (31, 32). Of the eight cysteines present in each procollagen III C-propeptide subunit, those nearest the N terminus (residues 43, 49, 66, and 75) are involved in interchain disulfide bonding, whereas those nearest the C terminus (residues 83, 153, 198, and 245) are required for intrachain disulfide bonding. The size of the putative junction region in the structure (Fig. 7) is consistent with the positions of all interchain disulfide bonds, whereas all intrachain disulfide bonds would be localized to the major lobes. Furthermore, because intrachain disulfide bonding occurs between cysteines 83 and 245 and between cysteines 153 and 198, it can be speculated that each major lobe is stabilized by the Cys-198 pair, whereas each polypeptide chain is folded back on itself so that the Cys-83–Cys-245 bond is near the junction region. This interpretation is inspired by recent ab initio modeling studies on the procollagen I C-propeptide trimer (54). In addition, the model requires that the previously identified (30) discontinuous molecular recognition sequence involved in type-specific chain association (residues 122–133 and 142–144) be within the junction region. This model would explain why frameshift (26, 55–57) or deletion (27) mutations at the C termini of the proα1(I) or proα2(I) chain C-propeptide domains or point mutations affecting formation of intrachain disulfide bonds (26, 58) should prevent or impede trimer assembly because both the C terminus and the “large loop” bond (equivalent to Cys-83–Cys-245 in the procollagen III C-propeptide) would be located in the junction region. A correctly folded C-terminal region stabilized by intrachain disulfide bonding might be an essential requirement for presentation of the molecular recognition sequence.

Finally, the model for the procollagen III C-propeptide domain bears some resemblance to the known structures of the N-terminal domains of collagens VII, XII, and XIV (see Ref. 59). These collagens, which are unrelated to the fibrillar collagens, are characterized by large non-helical N-terminal domains that form trimers with a cruciform-like structure, as visualized by electron microscopy after rotary shadowing. It should be noted, however, that these trimers are considerably larger (~500 kDa) than the procollagen C-propeptides (~90 kDa). Rotary shadowing of procollagen C-propeptides indicates a globular shape of an ~110 Å diameter (60), consistent with the structure reported here although with insufficient resolution to observe finer detail.

Acknowledgments—We thank Dr. D. Prokop for the CPIII baculovirus construct, Dr. E. Burchardt for the anti-CPIII monoclonal antibodies, Dr. T. Narayan for help with small angle x-ray scattering, Dr. K. Beck and Dr. B. Font for discussions, K. Bertoni for technical assistance, and C. Van Herrewege for help with the illustrations.

REFERENCES

1. Kadler, K. E. (1995) in Protein Profile (Sheterline, P., ed) Vol. 2, pp. 491–619, Academic Press, Inc., London

2. Prokop, D. J., and Hulmes, D. J. S. (1994) in Extracellular Matrix Assembly and Structure (Yurchenco, P. D., Birk, D. E., and Mehad, R. P., ed), pp 47–90, Academic Press, Inc, San Diego

3. Kessler, E., Takahara, K., Biniaminov, L., Brusel, M., and Greenspan, D. S. (1996) Science 271, 360–362

4. Prokop, D. J., Sieron, A. L., and Li, S. W. (1998) Matrix Biol. 16, 399–408

5. Moschovici, L., Bernocco, S., Font, B., Rivkin, H., Eichenberger, D., Chejanovsky, N., Hulmes, D. J. S., and Kessler, E. (2001) Eur. J. Biochem. 285, 2991–2996

6. Imamura, Y., Steiglitz, B. M., and Greenspan, D. S. (1998) J. Biol. Chem. 273, 27511–27517

7. Kofford, M. W., Schwartz, L. B., Schecter, N. M., Yager, D. R., Diegelmann, R., and Graham, M. F. (1997) J. Biol. Chem. 272, 7127–7131

8. Katayama, K., Seyer, J. M., Raghow, R., and Kang, A. H. (1991) Biochemistry 30, 7997–7104

9. Wu, C. H., Walton, C. M., and Wu, G. Y. (1991) J. Biol. Chem. 266, 2983–2987

10. Nakata, K., Miyamoto, S., Bernocco, S., Tanaka, M., Utani, A., Krebsbach, P., Rhodes, C., and Yamada, Y. (1996) Ann. N. Y. Acad. Sci. 785, 307–308

11. Mizuno, M., Fujisawa, R., and Kobaki, Y. (2000) FEBS Lett. 479, 120–126

12. Zhu, Y., Oganessian, A., Keene, D. R., and Sandell, L. J. (1999) J. Cell Biol. 144, 1099–1100

13. Kessler, E., and Adar, R. (1989) Eur. J. Biochem. 185, 115–121

14. Kirsch, T., and Pfaffle, M. (1992) FEBS Lett. 310, 143–147

15. Lee, E. R., Smith, C. E., and Poole, A. R. (1996) J. Histochem. Cytochem. 44, 423–433

16. Choglad, A. A., Purdom, I. F., and Hulmes, D. J. S. (1993) J. Biol. Chem. 268, 6107–6114

17. Davies, D., Tuckwell, D. S., Calderwood, D. A., Weston, S. A., Takigawa, M., and Humphries, M. J. (1997) J. Biochem. 246, 574–579

18. Bhattacharyya-Pakrasi, M., Dickes, S. K., and Santos, S. A. (1998) Matrix Biol. 17, 223–232

19. Rushin, J. A., Schmitz, S., Gunn-Moore, F., Sherman, D., Pappas, C. A., Ritchie, J. M., and Haynes, L. W. (1999) J. Neurochem. 73, 1816–1827

20. Palmieri, D., Camardella, L., Ulivi, V., Gusaico, G., and Manduca, P. (2000) J. Biol. Chem. 275, 32658–32663

21. Cheesler, S. D., and Byers, P. H. (1993) J. Biol. Chem. 268, 18226–18233

22. Mclauglin, S. H., and Bulleid, N. J. (1998) Matrix Biol. 17, 369–377

23. Lamande, S. R., and Bateman, J. F. (1999) Semin. Cell Biol. 10, 455–464

24. Bottomley, M. J., Batten, M. R., Lamb, R. A., and Bulleid, N. J. (2001) Curr.
Biophysical Characterization of the C-propeptide Trimer from Human Procollagen III Reveals a Tri-lobed Structure
Simonetta Bernocco, Stéphanie Finet, Christine Ebel, Denise Eichenberger, Marlène Mazzorana, Jean Farjanel and David J. S. Hulmes

J. Biol. Chem. 2001, 276:48930-48936.
doi: 10.1074/jbc.M108611200 originally published online October 29, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M108611200

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

This article cites 56 references, 14 of which can be accessed free at http://www.jbc.org/content/276/52/48930.full.html#ref-list-1