ELSEVIER

FEBS Letters 341 (1994) 54–58

FEBS 13751

FEBS LETTERS

The thrombospondin-like chains of cartilage oligomeric matrix protein are assembled by a five-stranded α-helical bundle between residues 20 and 83

Vladimir P. Efimov, Ariel Lustig, Jürgen Engel*

Department of Biophysical Chemistry, Biozentrum University, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Received 20 January 1994

Abstract

The N-terminal fragment of rat cartilage oligomeric matrix protein (COMP), comprising residues 20–83, was over-expressed in *E. coli* and purified under non-denaturing conditions. The fragment forms pentamers similar to the assembly domain of the native protein. Its five chains can be covalently linked in vitro by oxidation of cysteines 68 and 71. The fragment adopts a predominantly α-helical structure as judged by circular dichroism spectroscopy. On the basis of these findings we propose the model of a five-stranded α-helical bundle for the assembly domain of COMP. The studied sequence is conserved in thrombospondins 3 and 4 thus raising the possibility that these proteins are also pentamers.

Key words: Thrombospondin; α-Helical bundle

1. Introduction

Cartilage oligomeric matrix protein (COMP) is a pentameric glycoprotein synthesized by chondrocytes in cartilage [1,2]. It belongs to the thrombospondin (TSP) gene family [3]. The protein has a bouquet-like shape with five 28 nm-long arms containing a peripherial globular domain, a flexible strand, and a central assembly domain, where five chains are joined together in a cylindrical structure [2]. Five identical subunits are connected by disulfide bonds [1–3].

In addition to COMP, four distinct TSP genes (TSP1, TSP2, TSP3, and TSP4) have been described [4,5]. TSP1 and TSP2 are very similar in domain structure. The prototype human platelet TSP1 [6] is a trimeric protein [7]. Two cysteine residues (Cys-252, Cys-256) are responsible for interchain disulfide linkage of TSP1 [8]. In TSP2 these cysteines are conserved and it can form homotrimers and heterotrimers with TSP1 [9]. It was suggested that assembly of TSP1 involves formation of an α-helical coiled coil structure which is further stabilized by disulfide bonds [8]. TSP3, TSP4, and COMP differ from TSP1 and TSP2 in that they have an additional type 2 (EGF-like) repeat, and lack both the procollagen domain and the type 1 (properdin-like) repeats [3–5]. Oligomeric structures and assembly mechanisms of TSP3, and TSP4 are not known.

Here we report structural characterization of the assembly domain of COMP prepared by expression in *E. coli*.

2. Materials and methods

2.1. Expression of the N-terminal COMP fragment

Bluescript SK vector containing rat COMP cDNA [3] was a kind gift of Dr. P. Antonsson. It was used as a template for PCR amplification of a DNA fragment coding for residues 20–83 of COMP. Oligonucleotides TTT CAT ATG CAG GGC CAG ATC CCG CTG CTG G (introduces an *Nde*I site and the initiator codon ATG befor Gln-20 of COMP) and TTT GGA TCC TTA CAG GCT CAG ACC GGG G (introduces the stop codon TAA after Val-83 of COMP and a *Bam*HI site) were used as primers. The product of PCR was treated with *Nde*I and *Bam*HI restriction enzymes and cloned into the pET-3b vector [10] at *Nde*I–*Bam*HI sites. The resulting plasmid was designated p3b-COMP.

Expression was carried out using the T7 promoter in BL21(DE3) cells as described [10]. Bacteria carrying p3b-COMP were grown at 37°C in 2 × TY medium containing 1% glucose and 200 mg/l ampicilin to an optical density at 600 nm of 0.76 and were then induced by 1 mM IPTG. After 4 h of further incubation the cells were harvested (10,000 × *g*, 10 min) and stored at −70°C.

2.2. Purification of the recombinant protein

Pellet of bacteria from 1 l of culture was resuspended in 20 ml of TE buffer (20 mM Tris-HCl, pH 8.0, 1 mM EDTA) containing 0.1 mg/ml of lysozyme and incubated at 25°C for 30 min. Bacteria were lysed by sonication and centrifuged at 15,000 × *g* at 4°C for 15 min to remove insoluble material. These conditions were also used in subsequent centrifugations. 2 ml of streptomicin sulphate solution (30% w/v) were mixed with the supernatant. After 15 min of incubation on ice the precipitate was removed by centrifugation. Ammonium sulphate was

*Corresponding author. Fax: (41) (61) 267 2189.
E-mail: engel@urz.unibas.ch

Abbreviations: COMP, cartilage oligomeric matrix protein; TSP, thrombospondin; PCR, polymerase chain reaction; IPTG, isopropyl-β-D-thiogalactopyranoside; EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; CD, circular dichroism.

SSDI 0014-5793(94)00136-J

added to the supernatant to 36% saturation. After 1 h of incubation on ice the precipitate was collected by centrifugation, resuspended in 2 ml of TE buffer, and applied to a 10 ml column of hydroxylapatite (Bio-Rad; DNA grade) which had been equilibrated with 10 mM sodium phosphate, pH 7.6. The column was washed with the equilibrating buffer and the flow-through protein fraction containing mainly the recombinant protein was collected. This procedure yielded about 30 mg of 95% pure protein.

2.3. Protein chemistry and electrophoresis

Protein concentration was determined by the method of Waddell based on the difference between spectrophotometric absorptions at 215 and 225 nm [11].

Tricine SDS-PAGE system of Schägger and von Jagow [12] was used to analyze bacterial lysates and to estimate molar mass of the recombinant protein.

Disulfide bond formation was performed as follows. The protein was completely reduced by incubation with 100 × molar excess of DTT for 30 min at 37°C and precipitated by 50% ammonium sulfate. The pellet was resuspended in 0.2 M Tris-HCl, pH 8.8, 0.2 m NaCl, 1 mM EDTA to a protein concentration of 1.5 mg/ml. The protein was oxidized at room temperature by addition of oxidized and reduced glutathione to final concentrations of 10 mM and 2 mM, respectively. To monitor the progress of disulfide bond formation aliquots were withdrawn at various times and free cysteines were blocked with 100 mM iodoacetamide for 2 min at 25°C. Samples were analyzed by 15% SDS-PAGE [13] under non-reducing conditions.

Native electrophoresis was conducted in 12% PAGE using the Laemmli system [13] but SDS and reducing agents were omited from all solutions. Native gels were photopolymerized with 0.0004% riboflavin and 0.001% ammonium persulfate.

Gels were stained with Coomassie blue G250 according to [12].

2.4. Biophysical studies

Prior to each experiment samples were dialyzed against the appropriate buffer. Reduced protein was studied in the presence of 1 mM DTT and 0.1 mM EDTA.

Analytical ultracentrifugation measurements were performed on a Beckmann XLA centrifuge equipped with an AN-60Ti rotor. The protein was used at a concentration of 0.2 mg/ml in 50 mM sodium phosphate, pH 7.6. The absorbance was measured at 230 nm. In sedimentation equilibrium experiments three runs at 17,000, 20,000 and 24,000 rpm were performed and calculated molar masses were averaged. An apparent specific volume of 0.73 ml/g was used in calculations.

Circular dichroism spectra between 185 and 250 nm were recorded on a Jasco J720 spectropolarimeter at a protein concentration of 0.2 mg/ml in 10 mM sodium phosphate, pH 7.6, 25°C.

Negative staining with uranyl acetate and glycerol spray/rotary shadowing with platinum were performed as described [14].

2.5. Computational methods

α-Helical content was calculated from CD spectra according to [15] using the programm PROVEC3 written by C.H. Robert at the Biocenter of Basel University.

A homology search was performed in the GenEMBL database with the COMP sequence from Gln-20 to Val-83 using the program TFASTA (W.R. Pearson, GCG Programs) based on the algorithm of Pearson and Lipman [16].

The probability of coiled coil formation was calculated using the programm PEPCOIL (P. Rice, EGCG Programs) based on the algorithm of Lupas et al. [17]. A window of 28 residues was used.

3. Results

3.1. Expression and purification of the N-terminal COMP domain

The expression vector p3b-COMP, constructed as described in section 2, codes for the 64 amino acids of the rat COMP sequence preceded by an initiator methionine (M_w = 7.18 kDa). The COMP sequence starts from Gln-20 immediately after a putative signal peptide cleavage site and terminates after Val-83 before the first type 2 repeat [3]. Confirming our expression vector by sequenc-

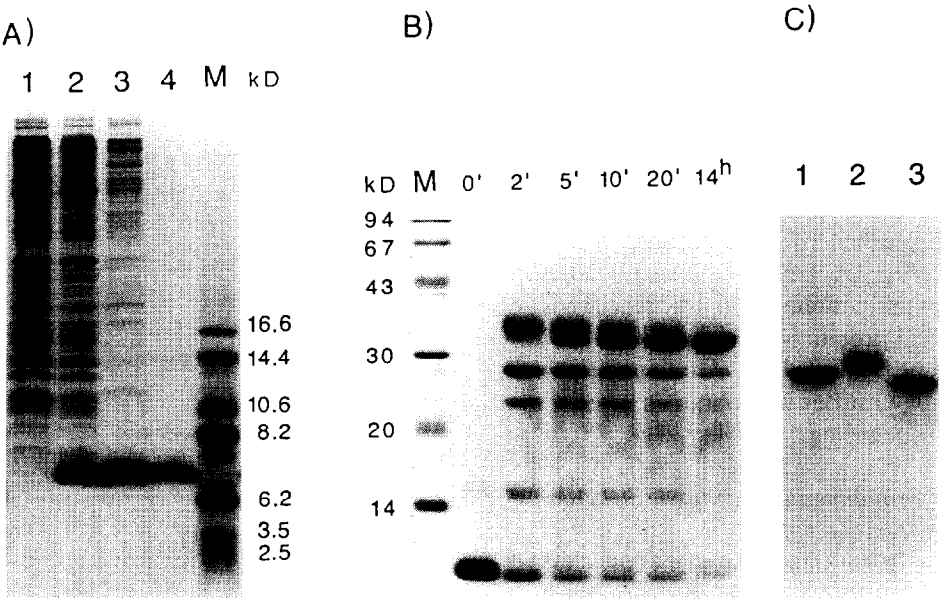

Fig. 1. Expression, purification and electrophoretic analysis of the COMP assembly domain. (A) Extracts from induced bacterial cultures (soluble proteins) containing pET-3b vector (lane 1) and p3b-COMP (lane 2); recombinant protein after ammonium sulfate precipitation (lane 3) and after chromatography on hydroxylapatite (6 μg, lane 4). (B) Kinetics of disulfide bond formation determined as described in section 2. 12 μg of recombinant protein were analyzed on 15% SDS-PAGE under non-reducing condition at the times indicated after starting of oxidation. (C) Native electrophoresis of the recombinant protein. Completely oxidized protein (lane 1); protein (0.4 mg/ml) was reduced by 20 mM DTT (30 min, 37°C) and blocked for 2 min with 100 mM iodoacetamid (lane 2) or iodoacetic acid (lane 3). 4 μg of protein were applied in each line. Lanes M contain molecular weight markers.

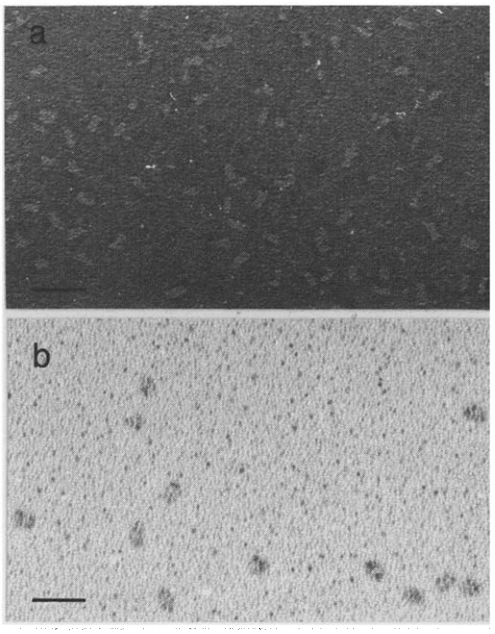

Fig. 2. Electron micrographs of the recombinant COMP fragment obtained after negative staining with uranyl acetate (a) and rotary shadowing with carbon/platinum (b). Bars = 30 nm.

ing we found descrepancies with the published COMP sequence [3]. Namely, the sequence between nucleotides 160–170 was CGACAGCAGGT rather then CGACACAGGGT. Sequencing of the original COMP clone using our PCR primers revealed the same sequence. This correction changes both His-53 and Arg-54 into Gln.

Extracts from induced *E.coli* BL21(DE3) cells harboring the expression vector p3b-COMP and empty pET-3b vector as a control were analysed by SDS-PAGE (Fig. 1A, lanes 2,3). Induction by IPTG yielded an abundant product which migrated at the expected position and which was absent in controls. The protein was completely soluble and no degradation products were detected.

The protein was purified by precipitation with 36% ammonium sulfate (Fig. 1A, lane 4) followed by elution from a hydroxylapatite column (Fig. 1A, lane 5). The identity of the purified protein was confirmed by protein sequencing and amino acid analysis. A unique sequence of NH_2-Met-Gln-Gly-Gln-Ile- was obtained after 5 cycles of N-terminal sequencing.

3.2. *Disulfide bond formation and oligomeric structure of the recombinant protein*

The recombinant N-terminal fragment of COMP bears two cysteines (68 and 71). Based on the similarities with the structure of TSP1 these cysteines have been suggested as candidates for holding the pentamer together via disulfide bonds [3]. When the recombinant protein was analyzed on SDS-PAGE under non-reducing conditions immediately after hydroxylapatite purification, only monomers, dimers and a small amount of trimers were detected (not shown). However, as is demonstrated in Fig. 1B, covalently linked pentamers were readily obtained by oxidation of the protein with glutathione. The accumulation of the final pentameric structure was preceeded by the appearence of five intermediate products. According to their electrophoretic mobilities four of them could correspond to monomers, dimers, trimers and tetramers. The fifth intermediate product had a slightly lower mobility than the final structure. This intermediate is most likely a pentamer with four disulfide bonds, while in the final pentameric structure there are five disulfide bonds.

Analysis of fully oxidized and reduced proteins on native PAGE (Fig. 1C) revealed predominantly single bands with only minute differences in mobilities apparently caused by the groups which have been used to block free cysteines. This finding implies that both reduced and oxidized proteins have the same pentameric structure. Analytical ultracentrifugation was used to confirm this conclusion. Sedimentation velocity experiments at 56,000 rpm yielded sedimentation coefficients (20°C, water) of 2.54 S both for the reduced and oxidized proteins. Sedimentation equilibrium experiments yielded a M_w of 36 ± 2 kDa for oxidized protein and 33 ± 2 kDa for reduced protein. These values are consistent with a pentameric structure of the recombinant polypeptide (expected M_w = 36 kDa).

3.3. *Electron microscopy and circular dichroism studies*

The structural similarity between the COMP fragment prepared by expression in *E. coli* and the corresponding segment in the intact native protein was proven by elec-

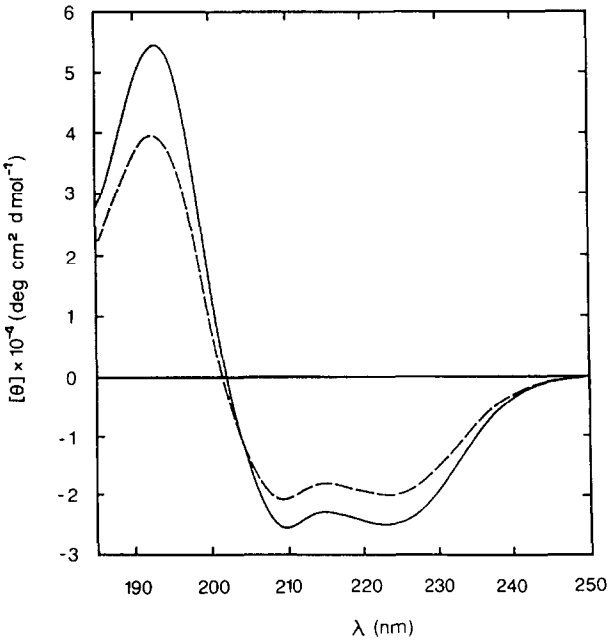

Fig. 3. Circular dichroism spectra of oxidized (solid line) and reduced (dashed line) recombinant protein.

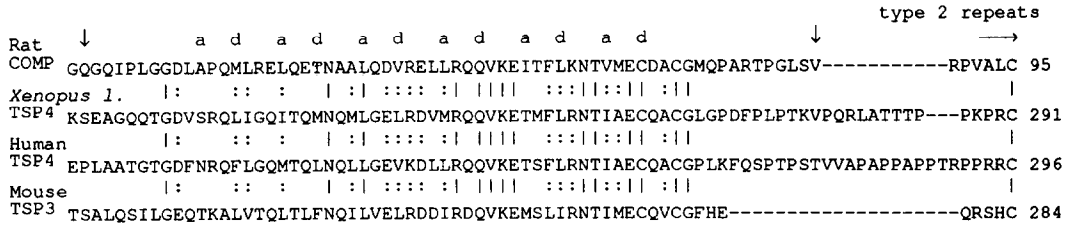

Fig. 4. Homology between pentamerizing domain of COMP and the regions before the first type 2 repeat of TSP3 and TSP4. The strictly conserved residues are indicated by vertical lines and the similar residues by dots. The first and last residues of COMP recombinant fragment are marked by vertical arrows. Possible a and d positions of heptad repeats are indicated.

tron microscopy. As shown in Fig. 2a, negative staining of the recombinant protein reveals cylindrical structures of 9.4 ± 1.2 nm in length and 3.7 ± 0.8 nm in width. Dimensions of 7.7 ± 0.6 nm in length and 3.3 ± 0.4 nm in diameter have been reported for the assembly domain of COMP isolated from the Swarm rat chondrosarcoma [2]. An assymetric shape of the molecules can also be noted after rotary shadowing (Fig. 2b).

Both oxidized and reduced recombinant proteins have CD spectra characteristic for an α-helical structure (Fig. 3). An α-helical content of 77% was predicted from the CD spectrum of the oxydized protein using the variable selection method [15].

4. Discussion

The N-terminal fragment of COMP comprising residues 20–83 was expressed in the cytoplasm of *E. coli* BL21 (DE3) to a high level and in a soluble form. No degradation products were detected. These data serve as indirect evidence that the fragment adopts a compact stable conformation. The structural identity of the recombinant protein with the assembly domain of COMP is proven by its ability to form covalently linked pentamers and by its electron microscopic appearence. The necessity to use in vitro oxidation to introduce disulfide bonds into the recombinant protein is not unexpected as proteins are usually expressed in the cytoplasm of *E. coli* in a reduced form [18–20]. Under the conditions employed, pentamerization of the COMP assembly domain does not depend significantly on disulfide bond formation, as judged from native electrophoresis and sedimentation equilibrium experiments. However, a somewhat lower absolute value of ellipticity and molar mass estimated by sedimentation equilibrium for the reduced protein could indicate partial dissociation.

The 77% α-helical content calculated from the CD spectrum corresponds to 50 residues in the 65 residue peptide. Such an α-helical region would have a length of 7.5 nm, which could account for most of the length of the 9.4 nm-long cylindrical structure seen by electron microscopy. Disulfide bond formation between Cys-68 and Cys-71 is only consistent with a parallel arrangement of five chains. Therefore, we suggest that the assembly domain of COMP consists mainly of five parallel α-helices assembled in a bundle. The occurence of α-helical bundles and their similarity with coiled coil structure has been described [21]. We analysed the COMP sequence using the algorithm of Lupas et al. [16] which predicts coiled coils by comparing a protein sequence with sequences in a database of known coiled coil domains. Only residues 32–65 were predicted to form a coiled coil (with a probability of 0.99). The crystal structure of a bundle of four parallel α-helices has been reported recently [22]. To our knowledge parallel five-stranded α-helical bundles have not yet been found.

Our data suggest that the COMP sequence between Gln-20 and Val-83 functions as a pentamerizing tool. A database search revealed that this sequence shares a conserved stretch of 46 residues (from Gly-27 to Gly-72) with mouse TSP3 [23,24], *Xenopus laevis* TSP4 [25] and human TSP4 [4] (Fig. 4). The conservative motif in all these proteins preceeds the first type 2 repeat. In mouse TSP3 gene this sequence corresponds exactly to one exon [24]. As shown in Fig. 4 the COMP sequence used in this work, as well as homologous regions of TSP3 and TSP4, have a heptad repeat of hydrophobic residues which is commonly believed to indicate an α-helical coiled coil structure [21]. Although the algorithm of Lupas et al. does not predict coiled coils in TSP3 and TSP4 it still yields the highest values of probability of coiled coil formation in the regions of homology with COMP residues 32–65 (0.1 for TSP4 and 0.3 for TSP3). These findings point to regions in TSP3 and TSP4 which might control their oligomerization and raise the possibility that these proteins are pentamers.

Acknowledgements: The authors are most grateful to Dr. P. Antonsson for providing the COMP cDNA clone. We thank Dr. P. Jenö and T. Mini for the oligonucleotide synthesis, amino acid analysis and protein sequencing. We thank C. Fauser and T. Schulthess for their excellent technical assistance. V.P.E. was supported by EMBO fellowship for East European molecular biologists.

References

[1] Hedbom, E., Antonsson, P., Hjerpe, A., Aeschlimann, D., Paulsson, M., Rosa-Pimentel, E., Sommarin, Y., Wendel, M., Oldberg, Å. and Heinegård, D. (1992) J. Biol. Chem. 267, 6132–6136.

[2] Mörgelin, M., Heinegård, D., Engel, J. and Paulsson, M. (1992) J. Biol. Chem. 267, 6137–6141.
[3] Oldberg, Å., Antonsson, P., Lindblom, K. and Heinegård, D. (1992) J. Biol. Chem. 267, 22346–22350.
[4] Lawler, J., Duquette, M., Urry, L., McHenry, K. and Smith, T.F. (1993) J. Mol. Evol. 36, 509–516.
[5] Bornstein, P. (1992) FASEB J. 6, 3290–3299.
[6] Lawler, J. and Hynes, R.O. (1986) J. Cell Biol. 103, 1635–1648.
[7] Lawler, J. (1986) Blood 67, 1197–1209.
[8] Sottile, J., Selegue, J. and Mosher, D.F. (1991) Biochemistry 30, 6556–6562.
[9] O'Rourke, K.M., Laherty, C.D. and Dixit, V.M. (1992) J. Biol. Chem. 267, 24921–24924.
[10] Studier, F.W., Rosenberg, A.N., Dunn, A.H. and Dubendorff, J.W. (1990) Meth. Enzymol. 185, 60–89.
[11] Wolf, P. (1983) Anal. Biochem. 129, 145–155.
[12] Schägger, H. and von Jagoww, G. (1987) Anal. Biochem. 166, 368–379.
[13] Laemmli, U.K. (1970) Nature 227, 680–685.
[14] Engel, J. and Furthmayr, H. (1987) Meth. Enzymol. 145, 3–78.
[15] Manavalan, P. and Johnson, W.C. (1987) Anal. Biochem. 167, 76–85.
[16] Lupas, A., Van Dyke, M. and Stock, J. (1991) Science 252, 1162–1164.
[17] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[18] Pollit, S. and Zalkin, H. (1983) J. Bacteriol. 153, 27–32.
[19] Derman, A.I. and Beckwith, J. (1991) J. Bacteriol. 173, 7719–7722.
[20] Derman, A.I., Prinz, W.A., Belin, D. and Beckwith, J. (1993) Science 262, 1744–1747.
[21] Cohen, C. and Parry, D.A.D. (1990) Proteins Struct. Fuct. Genet., 1–15.
[22] Harbury, P.B., Zhang, T., Kim, P.S. and Alber, T. (1993) Science 262, 1401–1407.
[23] Vos, L.H., Devarayalu, S., de Vries, Y. and Bornstein, P. (1992) J. Biol. Chem. 267, 12192–12196.
[24] Bornstein, P., Devarayalu, S., Edelhoff, S. and Disteche, C.M. (1993) Genomics 15, 607–613.
[25] Lawler, J., Duquette, M., Whittaker, C.A., Adams, J.C., McHenry, K. and DeSimone, D.W. (1993) J. Cell Biol. 120, 1059–1067.