The C-proteinase that processes procollagens to fibrillar collagens is identical to the protein previously identified as bone morphogenic protein-1

(astacins/tolloid/decapentaplegic/embryonic patterning)

SHI-WU LI*, ALEKSANDER L. SIERON*, ANDRZEJ FERTALA, YOSHIO HOJIMA, WILLIAM V. ARNOLD, AND DARWIN J. PROCKOP†

Department of Biochemistry and Molecular Biology, and the Jefferson Institute of Molecular Medicine, Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA 19107

Contributed by Darwin J. Prockop, January 16, 1996

ABSTRACT Bone morphogenic protein-1 (BMP-1) was originally identified as one of several BMPs that induced new bone formation when implanted into ectopic sites in rodents. BMP-1, however, differed from other BMPs in that its structure was not similar to transforming growth factor β . Instead, it had a large domain homologous to a metalloendopeptidase isolated from crayfish, an epidermal growth factor-like domain, and three regions of internal sequence homology referred to as CUB domains. Therefore, BMP-1 was a member of the “astacin families” of zinc-requiring endopeptidases. Many astacins have been shown to play critical roles in embryonic hatching, dorsal/ventral patterning, and early developmental decisions. Here, we have obtained amino acid sequences and isolated cDNA clones for procollagen C-proteinase (EC 3.4.24.19), an enzyme that is essential for the processing of procollagens to fibrillar collagens. The results demonstrate that procollagen C-proteinase is identical to BMP-1.

Bone morphogenic protein (BMP) was originally identified by Urist (1) and then by Reddi and Huggins (2) as an activity found in extracts of demineralized bone that induced new bone formation when implanted into ectopic sites in rodents. In 1988, Wozney and coworkers (3) obtained partial amino acid sequences of several components with bone morphogenic activity and used the sequence information to isolate three cDNA clones that they named BMP-1, BMP-2A, and BMP-3. On the basis of their structural similarity, BMP-2A and BMP-3 were identified as members of the transforming growth factor β (TGF- β) superfamily. Subsequently, five additional BMPs that were also similar to the TGF- β superfamily were identified (4, 5). BMP-1, however, was apparently isolated as a complex with the other BMPs because it differed in structure from TGF- β . It contained a large domain homologous to a metalloendopeptidase isolated from crayfish (6); an epidermal growth factor (EGF)-like domain; and three regions of internal sequence homology referred to as CUB domains because they are found in the complement components C1r/C1s, the sea urchin protein Uegf, and BMP-1 (7). The metalloendopeptidase from crayfish is now recognized as one of the simplest members of the “astacin family” of over 17 similar zinc-requiring endopeptidases (7–10). Many astacins contain extended C-terminal domains similar to those in BMP-1, and several of the larger astacins play critical roles in embryo hatching, dorsal/ventral patterning, and early developmental decisions. Also, there is indirect evidence that one or more of the larger astacins act in the same signaling pathway as members of the TGF- β family, perhaps through direct protein–

protein interactions (see ref. 7). Here, we have purified and obtained amino acid sequences of peptide fragments from procollagen C-proteinase (EC 3.4.24.19), an enzyme that is essential for the processing of procollagens to fibrillar collagens (11–14). We have used the amino acid sequences to isolate cDNA clones and found that procollagen C-proteinase is identical in structure to the protein previously identified as BMP-1 (3, 15, 16).

MATERIALS AND METHODS

Purification of Procollagen C-Proteinase. The enzyme was purified to homogeneity from organ cultures of chicken embryo tendons (17). In the final step, the protein was separated by SDS/PAGE. Tryptic peptides from the gel band containing the enzyme (17) were sequenced by the Wistar Protein Microsequencing Facility (Philadelphia, PA). The protein band was electroeluted onto a filter and digested *in situ* with trypsin. The tryptic peptides were separated on a reverse-phase C18 column (Supelco; model LC18DB) eluted with a gradient of 0.1–0.9% trifluoroacetic acid in 70% acetonitrile. Individual peaks from the column were assayed for homogeneity by time-of-flight matrix-assisted laser desorption mass spectrometry (Lasermat; Finnigan-MAT, San Jose, CA), and homogeneous fractions were sequenced by Edman degradation with an automated instrument (ABI Prism 377; Applied Biosystems).

Isolation of cDNAs. Total RNA was extracted from normal human skin fibroblasts (RNAeasy; Qiagen, Chatsworth, CA) and reverse transcribed with random primers (First Strand cDNA Synthesis Kit; Pharmacia). The cDNA was amplified by PCR with a pair of primers designed on the basis of the amino acid sequence of two of the peptide fragments (peptides 1 and 6 in Table 1). The primers were B-3 (ATGACTTCGACAG-CATCATGC) and B-4 (CTCCAGATAGTCGTAGGCACA). The PCR product was ³²P-labeled with random primers (Prime-It; Stratagene) and used as a probe to screen a cDNA library prepared from human skin fibroblasts (CRL1262, patient with osteogenesis imperfecta; ATCC) inserted into a λ phage (Zap II; Stratagene).

Expression of cDNAs. To express the cDNAs for procollagen C-proteinase, overlapping clones of the cDNAs were cleaved and ligated to generate full-length cDNAs. The cDNAs were inserted into the expression vector pcDNA3

Abbreviations: BMP, bone morphogenic protein; CUB, three regions of internal sequence homology found in the complement components C1r/C1s, the sea urchin protein Uegf, and BMP-1; pCP-1 and pCP-2, the short and longer full-length clones for human procollagen C-proteinase isolated and characterized here; TGF- β , transforming growth factor β ; EGF, epidermal growth factor; PEG, polyethylene glycol.

*S.-W.L. and A.L.S. contributed equally to this work.

†To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Amino acid sequences from chick pCP and corresponding amino acids encoded by human cDNAs for the shorter BMP-1 initially isolated (3) and one of the longer tolloid-like BMP-1s recently isolated (16)

Peptide	Observed sequence*	Amino acid positions	
		BMP-1/pCP-1	Tolloid-like BMP-1/pCP-2
1	MEPQEVESLGETYDFDSIMHYAR	252–274	252–274
2	NTFSR	275–279	275–279
3	GIFLDT	280–285	280–285
4	EVN KP YFEAGVRSPIGQR	290–302	290–302
5	T LPEPIVSSDSR	401–411	401–411
6	AYDYLEVR	489–496	489–496
7	LWLK	525–528	525–528
8	R C R EVDECSRPNNGGXEOK	547–562	547–562
9	N CD K SGFVLHDMKHDCKEAGSEHR		731–750

*Observed sequences are from tryptic peptides from chick pCP. Sequences above the continuous sequences are different amino acids encoded by human cDNAs (3, 16).

(Invitrogen) and used to prepare stable transfectants (11, 12) of a human tumor cell line (HT-1080 or CCL-121; ATCC) by calcium phosphate precipitation with a commercial kit (Promega). Cells were initially cultured for 24 hr in high-glucose DMEM containing 10% fetal bovine serum (Cellgro; Mediatech, Herndon, VA) on 80-cm² culture dishes with 1×10^6 cells per dish. The cells were transfected by incubation for 18 hr with a calcium phosphate precipitate containing 10 μ g of the linearized plasmid. The cultures were incubated in fresh medium for another 24 hr, passaged after 1:10 dilution in 80-cm² dishes, and then grown under selection with 400 μ g/ml of G418 (GIBCO/BRL) for 12 days. Neomycin-resistant clones were transferred to 12-well microtiter plates, grown to confluency, and cultured in two 24-well plates for an additional 24 hr. Total RNA was extracted from the cell layer of one well (RNAeasy; Qiagen) and used for Northern blot assay after separation by electrophoresis on a 1% agarose gel and transfer to nitrocellulose filters. The filters were probed using a ³²P-labeled cDNA for the shortest cDNA for procollagen C-proteinase (pCP-1).

To assay the clones for C-proteinase activity, positive clones were grown to confluence in 175-cm² flasks in DMEM containing 10% fetal bovine serum and then for 24 hr in 15 ml of serum-free medium. The medium was precipitated for 18 hr at 4°C with ¼ vol of 25% polyethylene glycol (PEG; 8000 Da) in 0.4 M NaCl and 0.1% NaN₃ in 0.1 M Tris-HCl buffer (pH 7.4). The precipitate was isolated by centrifugation at $15,000 \times g$ for 1 hr and partially solubilized in 20 μ l of reaction buffer (0.1 M NaCl/7 mM CaCl₂/0.015% Brij 35/0.01% NaN₃ in 25 mM Tris-HCl buffer, pH 7.5). For the C-proteinase assay, the samples were incubated for 3 hr at 35°C in a total volume of 25 μ l of reaction buffer containing 1 μ g of ¹⁴C-labeled type I procollagen purified from chicken embryo fibroblasts (17). The reaction was terminated by adding ¼ vol of 11.5% SDS, 10% glycerol, and 0.05% bromophenol blue in 0.125 M Tris-HCl buffer (pH 6.8). The products were separated on an SDS/7.5% polyacrylamide gel (MiniProtein apparatus; Bio-Rad), and the gel was assayed on a phosphor storage imager (Molecular Dynamics).

RESULTS

Amino Acid Sequences of Procollagen C-Proteinase. Procollagen C-proteinase from chicken embryos was purified to homogeneity (17), and tryptic peptides of the enzyme were sequenced. Nine sequences of different peptides were ob-

tained (see Table 1). With minor conservative substitutions attributable to the species difference, eight of the peptides contained sequences found in the protein initially identified as human BMP-1 (3). The ninth peptide had a sequence of 20 amino acids found in the C-terminal domain of one of the longer forms of BMP-1 recently identified in mouse (15) and human (16) tissues.

cDNA Clones for the Enzyme. The amino acid sequences of the enzyme (Table 1) were used to design primers for reverse transcription-PCR. As predicted from the amino acid sequence of BMP-1 (3), the PCR product was 592 bp. The PCR product was then used to screen a cDNA library prepared from cultured human skin fibroblasts. Five positive clones ranging in size from 0.16 to 2.9 kb were obtained (Fig. 1). Sequencing of the clones indicated that they encoded all the amino acid sequences found in the nine peptides from the chicken procollagen C-proteinase (Table 1) and in the previously published sequences of BMP-1 (3, 15, 16). Analysis of overlapping sequences of the clones indicated that the cDNAs coded for two proteins of different lengths, one of 730 aa (pCP-1) and a second of 986 aa (pCP-2). The first 702 codons for the two proteins were identical. Beginning with the codon for amino acid 703, however, pCP-2 had a new sequence that coded for a second EGF-like domain and a fourth and fifth CUB domain. Therefore, the structure of pCP-2 (Fig. 1) was the same as one of the tolloid-like longer forms of BMP-1 isolated from mouse (15) and human (16) tissues that arise from alternative splicing of the gene (16). The only difference in the sequences obtained here from the reported human sequences (16) was an -AAC-codon for asparagine in place of a -GAC-codon for aspartate at amino acid position 24. The results were consistent with the conclusion that the two mRNAs arose by alternative splicing of RNA transcripts from the same gene (16).

Expression of the cDNAs. Full-length cDNAs for pCP-1 and pCP-2 were stably transfected into a human tumor cell line (HT-1080) that we had used previously for expression of recombinant procollagens (11, 12). Neomycin-resistant clones were screened for expression of the cDNAs as mRNAs. As indicated in Fig. 2, two clones, each transfected with pCP-1 or pCP-2, had high levels of corresponding mRNAs.

As indicated in Fig. 3, medium from clones expressing either pCP-1 or pCP-2 contained enzymatic activity that specifically cleaved type I procollagen into the expected products, i.e., pN α 1(I) chains, pN α 2(I) chains, and a disulfide-linked C-propeptide trimer (11, 12, 17). There was no detectable activity in the medium from untransfected cells. PAGE of the samples

FIG. 1. Sequenced peptides from procollagen C-proteinase, the encoded structures of pCP-1 and pCP-2, and the isolated cDNA clones. The first diagram shows peptides sequenced from chicken pCP (see Table 1). The second and third diagrams show domains encoded for BMP-1 (pCP-1) and a longer tolloid-like BMP-1 (pCP-2). The fourth diagram shows the cDNAs isolated here. Scales are shown at the bottom. The cDNA sequence closest to the 3' end extended to 3560 bp. As discussed in text, the astacin-like domain contains a signature zinc-binding sequence. The EGF-like domains probably bind Ca^{2+} , which is also required for enzymic activity of pCP (see ref. 17).

after reduction demonstrated that the C1 subunit from the pro α 1(I) chain and the C2 subunit from the pro α 2(I) chain were of the expected size and obtained in the expected ratio of 2:1 (not shown). Complete cleavage of type I procollagen to pN α 1(I) chains, pN α 2(I) chains, and the C-propeptide was obtained with medium from a clone transfected with pCP-2 and partially fractionated and concentrated by membrane filtration (Fig. 3, lane 6).

DISCUSSION

The results here demonstrate that procollagen C-proteinase (17) is the same protein as BMP-1 (3, 15, 16). Since BMP-1 was

FIG. 2. Northern blot assay of total RNA from HT-1080 cells. The filter was probed with a ^{32}P -labeled clone of pCP-1 (nucleotides 837-2487) that was labeled by random primer extension with ^{32}P to a specific activity of 4×10^8 cpm/ μg . The filter was exposed to x-ray film at -70°C for 6 hr. Lanes 1 and 2, two clones transfected with pCP-1; lane 3, clone transfected with vector (pcDNA-3) without cDNA insert; lanes 4 and 5, two clones transfected with pCP-2; and lane 6, RNA from untransfected host cells (HT-1080).

first identified (3), its function was closely linked both to the TGF- β -like activities of other BMPs (4) and to the endopeptidase activities of large members of the astacin family (7). In *Drosophila*, mutations in the astacin-like protein tolloid produce a patterning phenotype of dorsal/ventral patterning similar but less severe than the phenotype produced by mutations in decapentaplegic, a protein in the TGF- β family (18–20). Therefore, tolloid and decapentaplegic may be in the same signaling pathway (7). In *Xenopus*, the astacin-like protein UVS.2 is located on the hatching gland of embryos and breaks down membranes so that embryos can emerge (8, 21). In fish, the small astacin-like proteins choriolysin H and L play a similar role in embryo hatching (22, 23). In sea urchin, the astacin-like proteins BP-10 and SpAN have a role in differentiation of endodermal linkages and embryonic patterning (24, 25). In hydra, the astacin-like protein HMP-1 plays a role in patterning and the formation of tentacle cells, apparently by degradation of extracellular matrix proteins (26). In mice, two astacin-like proteins (meprin A and B) are found on the brush borders of renal and intestinal cells and apparently have a role in degradation of peptides such as parathyroid hormone and of extracellular matrix proteins (7, 27). All astacin-like proteins have a characteristic 18-aa sequence that begins with a signature zinc-binding sequence of HEXXH found in most metalloproteases (7–10). The repetitive C-terminal domains found on the larger astacins like BMP-1 suggest that the proteins may have been assembled as modules during evolution to provide proteases that differ in specificity and function much as the proposed modular assembly that led from trypsin to the complex system of coagulation enzymes in mammals (28).

Procollagen C-proteinase is a biosynthetic enzyme that is involved in the processing of several procollagens to collagens

FIG. 3. Assays of C-proteinase activity in medium from transfected cells. Medium proteins from positive clones were fractionated either by PEG precipitation or by membrane filtration. For PEG precipitation, 2×10^7 cells in 175-cm² flasks were incubated for 24 hr in serum-free DMEM, and the medium (15 ml) was precipitated as described in text. The precipitated proteins were partially solubilized in 20 μ l of reaction buffer. For fractionation by membrane filtration, 30 ml of the medium was passed through a filter with a high molecular weight cut-off (Amicon; model XM300). A second filter (50 kDa Ultrafree 15; Millipore) was used to concentrate the flowthrough $\approx$ 500-fold and to transfer the sample to the reaction buffer. A 20- μ l portion of each sample was incubated at 35°C for 3 hr with 5 μ l of reaction buffer containing 1 μ g of ¹⁴C-labeled type I procollagen purified from chicken embryo fibroblasts (17). The reaction products were separated on an SDS/7.5% polyacrylamide gel without reduction, and an image was generated on a phosphor storage imager. ¹⁴C-labeled type I procollagen was incubated without additions (lane 1); with 10 units (17) of purified chicken procollagen C-proteinase (lane 2); with proteins precipitated by PEG from 15 ml of medium from a clone transfected with pCP-1 (lane 3); with proteins precipitated by PEG from 15 ml of medium from a clone transfected with pCP-2 (lane 4); with proteins precipitated by PEG from 15 ml of medium from nontransfected host cells (HT-1080) (lane 5); and with proteins partially fractionated and concentrated by membrane filtration from 10 ml of medium from a clone transfected with pCP-2 (lane 6).

that spontaneously self-assemble into fibrils (13, 14). The enzyme cleaves -Ala-Asp- or -Gly-Asp- bonds in type I, type II, and type III procollagens. Without cleavage of the C-propeptides, the procollagens remain as partially processed precursors that are soluble to 0.5–1.0 mg/ml and are not incorporated into fibrils (13). Nonspecific processing of the C-propeptide of type I procollagen is occasionally observed in tissue culture (29, 30) but not *in vivo* (13, 14). No genetic defects in the enzyme have been detected in surveys of hundreds of patients with genetic diseases of connective tissues (14), and transgenic mice without fibrils of type I collagen (31, 32) or type II collagen (33) developed lethal phenotypes. In addition, the C-proteinase was recently shown to process a precursor form of lysyl oxidase to the active enzyme *in vitro*, an observation indicating that it may have a role in the synthesis of the covalent crosslinks that are essential for the normal tensile strength of fibers of both collagen and elastin (M. V. Panchenko and H. M. Kagan, personal communication). Therefore, the enzyme plays an essential role in the biological function of the structural proteins that largely define the size, shape, and strength of most complex organisms. The effects of BMP-1 on bone morphogenesis may or may not be explained by its C-proteinase activity. If the C-proteinase activity is the bone morphogenic principle, alternative splicing of RNA transcripts (11) may provide C-proteinases that specifically cleave different procollagens or latent forms of growth factors such as TGF- β (see ref. 7). It is possible, however, that the effects of BMP-1 on bone morphogenesis may be explained by additional properties of the protein that have not yet been defined.

We are grateful to Katherine Sahm for assistance in carrying out the experiments described here. This work was supported in part by a from the Lucille P. Markey Charitable Trust and National Institutes of Health Grants AR39740 and AR38188.

1. Urist, M. R. (1965) *Science* **150**, 893–899.
2. Reddi, A. H. & Huggins, C. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1601–1605.
3. Wozney, J. H., Rosen, V., Celeste, A. J., Mitscock, L. M., Whitters, M. J., Kriz, R. W., Hewick, R. M. & Wang, E. A. (1988) *Science* **242**, 1528–1534.
4. Lyons, K. M., Jones, C. M. & Hogan, B. L. M. (1991) *Trends Genet.* **7**, 408–412.
5. Kingsley, D. M. (1994) *Trends Genet.* **10**, 16–21.
6. Titani, K., Torff, H.-J., Hormel, S., Kumar, S., Walsh, K. A., Rödl, J., Neurath, H. & Zwilling, R. (1987) *Biochemistry* **26**, 222–226.
7. Bond, J. S. & Beynon, R. J. (1995) *Protein Sci.* **4**, 1247–1261.
8. Dumermuth, E. & Sterchi, E. E. (1991) *J. Biochem.* **266**, 21381–21385.
9. Gomis-Ruth, F. X., Grams, F., Yiallourous, I., Nar, H., Kusthardt, U., Zwilling, R., Bode, W. & Stocker, W. (1994) *J. Biol. Chem.* **269**, 17111–17117.
10. Stocker, W., Gomis-Ruth, F. X., Bode, W. & Zwilling, R. (1993) *Eur. J. Biochem.* **217**, 215–231.
11. Fertala, A., Sieron, A. L., Ganguly, A., Li, S.-W., Ala-Kokko, L., Anumula, K. R. & Prockop, D. J. (1994) *Biochem. J.* **298**, 31–37.
12. Fertala, A., Sieron, A. L., Hojima, Y., Ganguly, A. & Prockop, D. J. (1994) *J. Biol. Chem.* **269**, 11584–11589.
13. Prockop, D. J. & Hulmes, D. J. S. (1994) in *Extracellular Matrix Assembly and Structure*, eds. Yurchenco, P. D., Birk, D. E. & Mecham, R. P. (Academic, New York), pp. 47–90.
14. Prockop, D. J. & Kivirikko, K. I. (1995) *Annu. Rev. Biochem.* **64**, 403–434.
15. Fukagawa, M., Suzuki, N. P., Hogan, L. M. & Jones, C. M. (1994) *Dev. Biol.* **163**, 175–183.
16. Takahara, K., Lyons, G. E. & Greenspan, D. S. (1994) *J. Biol. Chem.* **269**, 32572–32578.
17. Hojima, Y., van der Rest, M. & Prockop, D. J. (1985) *J. Biol. Chem.* **260**, 15996–16003.
18. Shimell, M. J., Ferguson, E. L., Childs, S. R. & O'Connor, M. B. (1991) *Cell* **67**, 469–481.
19. Jurgens, G., Wieschaus, E., Nusslein-Volhard, C. & Kluding, H. (1984) *Roux's Arch. Dev. Biol.* **193**, 283–295.
20. Padgett, R. W., Wozney, J. M. & Gelbart, W. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2905–2909.
21. Sato, S. M. & Sargent, T. D. (1990) *Dev. Biol.* **137**, 135–141.
22. Yasumasu, S., Katow, S., Umino, Y., Iuchi, I. & Yamagami, K. (1989) *Biochem. Biophys. Res. Commun.* **162**, 58–63.
23. Lee, K. S., Yasumasu, S., Nomura, K. & Iuchi, I. (1994) *FEBS Lett.* **339**, 281–284.
24. Lepage, T., Ghiglione, C. & Gache, C. (1992) *Development (Cambridge, U.K.)* **114**, 147–164.
25. Reynolds, S. D., Augerer, I. M., Palis, J., Nasir, A. & Angerer, R. C. (1992) *Development (Cambridge, U.K.)* **114**, 769–786.
26. Yan, L., Pollock, G. H., Nagase, H. & Sarras, M. P. (1995) *Development (Cambridge, U.K.)* **121**, 1591–1602.
27. Gorbea, C. M., Marchand, P., Jiang, W., Copeland, N. P., Gilberg, D. J., Jenkins, N. A. & Bond, J. S. (1993) *J. Biol. Chem.* **268**, 21035–21043.
28. Patthy, L. (1985) *Cell* **41**, 657–663.
29. Bateman, F., Pillow, J. J., Mascara, T., Medvedec, S., Ramshaw, J. A. M. & Cole, W. G. (1987) *Biochem. J.* **245**, 677–682.
30. Lee, S.-T., Kessler, E. & Greenspan, D. S. (1990) *J. Biol. Chem.* **265**, 21992–21996.
31. Harbers, K., Kuehn, M., Delius, H. & Jaenisch, K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1504–1508.
32. Kratochwil, K., Ghaffari-Tabrizi, N., Holzinger, I. & Habers, K. (1993) *Dev. Dyn.* **198**, 273–283.
33. Li, S.-W., Prockop, D. J., Helminen, H., Fässler, R., Lapveteläinen, T., Kiraly, K., Peltari, A., Arokoski, J., Lui, H., Arita, M. & Khillan, J. S. (1995) *Genes Dev.* **9**, 2821–2830.