Structure and Function of Procollagen C-Proteinase (mTolloid) Domains Determined by Protease Digestion, Circular Dichroism, Binding to Procollagen Type I, and Computer Modeling†

Aleksander L. Sieron,*,‡ Anna Tretiakova,§ Bradford A. Jameson,§ Mark L. Segall,|| Sissel Lund-Katz,|| M. Tariq Khan,‡ Shi-wu Li,‡ and Walter Stöcker⊥

Center for Gene Therapy, Cancer Center, and Department of Biochemistry, MCP Hahnemann University, MCP Hahnemann School of Medicine, Philadelphia, Pennsylvania 19102, and Institute of Zoophysiology and Molecular Physiology, University of Münster, Germany

Received October 5, 1999

ABSTRACT: Procollagen C-proteinase-2 (pCP-2, mTld) is derived from the longest splicing variant of the gene encoding bone morphogenetic protein 1 (BMP-1). The variants have identical amino terminal signal peptides, prodomains and astacin-like protease domains. However, they differ in the length of their carboxy terminal part, which in pCP-2 has the composition CUB1, CUB2, EGF-like1, CUB3, EGF-like2, CUB4, CUB5, and C-tail. In the shorter form, pCP-1 (i.e., BMP-1), the sequence ends after the CUB3-domain. Using a combination of mutagenesis and structural approaches, we have investigated the structure and function of subfragments of pCP-2. The full-length latent recombinant enzyme and its N-terminally truncated form lacking the prodomain were tested for their enzymic activity. The intact protein showed only partial processing of procollagen type I, whereas the truncated form expressed enzymic activity indistinguishable from its native counterpart purified from chick embryo tendons. These results clearly demonstrated that the prodomain is required for the latency of the enzyme but not for its correct folding. Limited proteolysis of the recombinant protein with α-chymotrypsin produced four discrete fragments revealing the location of cleavage sites between the repetitive CUB/EGF domains. The results provide evidence that the CUB sequences form independently folded modules that are stabilized by two pairs of internal disulfide bridges. The modules are linked to each other by more flexible, hinge-like peptides. Solid-phase binding assays with isolated CUB domains and immobilized procollagen type I demonstrated that the first three but not the last two CUB domains specifically bound to the substrate. To define putative sites for CUB−CUB or CUB−substrate interactions, we generated molecular models for pCP-2 CUB domains. The models were obtained using as a template the structure of CUB domain in zona pellucida adhesion protein PSP-I/PSP-II from porcine sperm. The predicted conformations for homology models were, subsequently, confirmed by circular dichroism spectroscopy of polypeptide domains isolated following limited proteolysis with α-chymotrypsin.

Procollagen C-proteinase (pCP)[1] also known as bone morphogenetic protein 1 (BMP-1) is a homologue of the *Drosophila* development protein tolloid. The vertebrate enzyme was originally identified as a neutral zinc metalloproteinase that specifically cleaved the large C-propeptides from the three procollagen precursors of the major fibrillar collagens (types I, II, and III) at Ala−Asp or Gly−Asp peptide bonds (*1*−*6*). It was purified as a 100 kDa protein (pCP-2, mTld) from chick embryo tendons (*7*) and subsequently as a 70 kDa protein (pCP-1) from the culture media of mouse fibroblasts (*8*). Amino acid sequencing of these enzymes and subsequent cDNA cloning revealed that the mouse pCP-1 was identical to BMP-1, whereas the chick pCP-2 was identical to human mTld (*9*, *10*). The two forms of the enzyme are products of the same gene as a result of alternative splicing (*9*, *11*). A third splicing variant that codes for an 80 kDa protein (pCP-3) was identified by Northern analysis (*9*). The sequences of three alternative RNA splicing variants of C-proteinase allowed the classification of the enzymes as members of metzincin superfamily of metalloproteinases (*12*, *13*). The zinc-binding protease domain is

† This work was supported in part by Allegheny-Singer Research Institute Grant LTR 03/19/97 to A.L.S. and a grant from the Deutsche Forschungs gemeinschaft (DFG) (Sto 185/3-2 and SFB492) to W.S.

* To whom correspondence should be addressed. Phone: (215) 762-7196. Fax: (215) 762-7408. E-mail: Aleksander.Sieron@drexel.edu.

‡ Center for Gene Therapy.

§ Cancer Center.

|| Department of Biochemistry.

⊥ Institute of Zoophysiology and Molecular Physiology.

[1] Abbreviations: pCP-1, the 70-kDa RNA splicing variant of procollagen C-proteinase that is identical to bone morphogenetic protein-1; pCP-2, the 100-kDa RNA splicing variant of procollagen C-proteinase that is identical to mammalian tolloid; pCP-3/BMP-1/His, the 80-kDa variant of procollagen C-proteinase that ends in a unique histidine-rich sequence; SOG, short gastrulation protein from *Drosophila*; BMP-1, etc., bone morphogenetic protein-1, etc.; Tld, developmental protein tolloid of *Drosophila*; mTld, mammalian tolloid; mTll, protein homologous to mTld; CUB domains, repetitive domains found in complement, sea urchin and BMP-1; PSP-I/PSP-II, heterodimeric zona pellucida binding protein of the porcine sperm; aSFP, bovine acidic seminal fluid protein.

10.1021/bi992312o CCC: $19.00 © 2000 American Chemical Society
Published on Web 03/02/2000

identical in all variants, and resembles the digestive enzyme astacin from crayfish. Therefore, procollagen C-proteinase variants are classified in the astacin family of metalloproteases (*12*−*15*) and the known crystal structure of astacin (*16*, *17*) can be used as structural template for the catalytic site of the C-proteinase. The N-terminal pro-domain identical in all procollagen C-proteinase variants contains sequences found in prepropeptides of latent forms of many astacin-like enzymes. It has been postulated and shown that the activition of astacins from inactive proforms requires proteolytic removal of the prodomains (*14*, *16*, *18*).

The 70-kDa variant of C-proteinase has a C-terminal region with three repetitive CUB domains similar to sequences found in the complement components C1r/C1s, in the sea urchin protein UeGF and in BMP-1. The EGF-like motif is located between the second and third CUB domain (*19*). The 80-kDa variant is identical to the 70-kDa protein except for the insertion of the 121 amino acids rich in His, Ser, and Thr after the third CUB domain (*9*). The 100-kDa variant contains a second EGF motif and two additional CUB domains (*10*).

The alternatively spliced forms of BMP-1 are expressed differently in embryonic and adult tissues (*9*). Transgenic mice homozygous for null alleles of C-proteinase/mTolloid died at birth because of persistent hernia of the gut (*20*). Although they have had essentially normal skeleton, electron micrographs showed the presence of barbed wire-like collagen fibrils probably due to incomplete propeptide processing. This severe but limited phenotype could be explained by the presence of two newly described genes encoding the mammalian tolloid-like proteins 1 and 2 (mTll1 and mTll2) with about 76% identity with mTld (*21*−*24*). mTld and mTll1 were found to be coexpressed in the majority of mouse tissues except heart and brain, where mTld was absent (*22*). These observations provide evidence that mTll1 could partially compensate for the lack of mTld activity in developing mouse embryos. However, presence of the C-propeptides on the surface of collagen fibrils in mtld null mice indicates that the mTll1 putative procollagen C-proteinase-like activity was insufficient.

The similarity of mammalian tolloids to *Drosophila* tolloid establishes a new role for these enzymes in development (*25*, *26*). *Drosophila* tolloid was recently shown to cleave the chordin analogue SOG (short gastruation) and thereby release active DPP (decapentaplegic) from a latent complex (*27*). In *Xenopus*, a tolloid-related proteinase (xolloid) cleaved chordin to release active BMP-4, a homologue of DPP (*28*). Activation of the BMPs, in turn, is an essential step in dorsal-ventral patterning in *Drosophila* and *Xenopus*, as well as in zebrafish (*29*). Proteins with structures homologous to tolloid in both vertebrates and nonvertebrates were shown to be critical in early development as well as in downstream events during subsequent tissue differentiation and remodeling (*14*).

Numerous reports provide evidence that the substrate specificity of complex serine proteinases [for example, tissue-type plasminogen activator (tPA) and thrombin] and of many metalloproteinases is determined by secondary binding sites (exosites) localized in their noncatalytic domains (*30*−*37*). Here, the three variants of procollagen C-proteinase retain identical catalytic domains but differ by the number of CUB/EGF motifs. The highly homologous enzymes mTll1 and mTll2, although, retaining virtually identical protease domains do not compensate the lack of mTld activity. Therefore, we investigated the structure/function relations of the procollagen C-proteinase and subfragments of this enzyme using mutagenesis, limited proteolysis, solid-phase protein−protein binding assays, computer modeling, and circular dichroism spectroscopy.

EXPERIMENTAL PROCEDURES

Synthesis of Recombinant Procollagen C-Proteinase. To synthesize the latent pCP-2, the DNA construct described previously (*10*) was used. To prepare a construct for pCP-2 without the prodomain, the cDNA encoding the full-length C-proteinase (*10*) was cleaved with *Aat*II/*Sph*I. To ensure efficient secretion of recombinant protein, the DNA fragment was ligated into pcDNAII (Invitrogen) vector containing an insert of cDNA encoding the signal peptide for α1 chain of procollagen type I in which an *Aat*II site was created at the 3′-end. The final construct consisted of coding sequences for the COL1A1 signal peptide linked to the complete C-proteinase except for the first 360 bp encoding signal peptide and propeptide. The construct was transferred into *Spe*I/*Not*I sites in pRC RSV expression vector (Invitrogen) and was sequenced to exclude any mutations. The linearized plasmid (*Bgl*II) was used for stable transfection of HT-1080 cells and selection with G418 (400 mg/L) as described previously (*10*). Neomycin-resistant clones were assayed for secretion of truncated C-proteinase (*38*). The culture medium was precipitated overnight at 4 °C with 8% poly(ethylene glycol), and the precipitated proteins were analyzed for presence of recombinant enzyme by Western blotting using a rabbit polyclonal anti-pCP-2/mTld antibody.

Production and Purification of the Recombinant Proteins. Human recombinant latent pCP-2 (*10*) or the variant lacking the propeptide was purified from 48 L of medium collected over 6 days from 80 (1050 cm^2) roller bottles culture of HT-1080 cells (Cell Culture Center, Cellex Biosciences, Inc., Minneapolis, MN). The medium was filtered through a double glass fiber filter (1.6 μm), and NaCl, NaN_3, and Tris-HCl were added to final concentrations 0.15 M, 0.01%, and 50 mM, respectively. The pH was adjusted to 7.5 at room temperature. Filtered medium was fractionated using a 300 kDa cut off spiral cartridge filter (PLTK, Prep/Scale TFF, Millipore), and the filtrate was concentrated to the final volume of about 1000 mL using a 30 kDa cut off spiral cartridge filter (Millipore). The concentrate was loaded onto a green A Dye matrix gel affinity column (*7*). The column was washed with 5 vol of 0.5 M NaCl and the same volume 0.9 M NaCl in 0.05 M Tris-HCl, pH 7.5, buffer to remove a small amount of MMP-2 and MMP-9 that are secreted from HT-1080 cells. The recombinant protein was eluted from the column with 2 M urea and 3 M NaCl in the same buffer. The recovered protein was further purified using a concanavalin A Sepharose column as described elsewhere (*7*). Samples were eluted in buffer containing 1.5 M NaCl, 1 M urea, 0.3 M methyl α-D-manno-pyranoside, 0.01% NaN_3, and 0.1 M Tris-HCl, pH 7.5. The sample was then chromatographed on a gel filtration column (HiLoad Superdex S-200 column; Pharmacia). Fractions containing the recombinant protein identified by Western blot analysis were concentrated to a final concentration 0.3 mg/mL.

Enzymic Activity Assay. The specific enzymic activity assay was performed with minor modifications as described previously (*7*). Briefly, about 0.3 μg of recombinant latent enzyme or 2 units of its active form were incubated for 2 h at 35 °C with about 0.3 μg of ^{14}C-labeled chick procollagen type I in the buffer containing 0.15 M NaCl, 7 mM $CaCl_2$, 0.015% Brij 30, 0.01% NaN_3, and 50 mM Tris-HCl, pH 7.5. The total volume of the reaction mixture was 20 μL. The cleavage was terminated by adding 0.25 volume 5× protein sample buffer (10% glycerol, 0.5% SDS, 25 mM EDTA, and 0.05% bromophenol blue in 0.3 M Tris-HCl, pH 6.8) and boiling the sample for 5 min. Reaction products were separated in 7.5% SDS−polyacrylamide gel and assayed using a phosphor storage imager (STORM 840; Molecular Dynamics). The comparison of relative enzymic activities was based on the amount of free C-propeptide trimer measured with help of ImageQuant program (Molecular Dynamics, CA). The program counts pixels corresponding to the amount of energy emitted by ^{14}C radioisotope to the phosphoimager screen (plate), which subsequently is converted by the "reader" to an image (*39*).

Limited Digestion of Recombinant Procollagen C-Proteinase with α-Chymotrypsin. To define optimal conditions, 200 μg/mL of purified recombinant latent pCP-2 was incubated at 35 °C for 1−15 h with 1−10 μg/mL of α-chymotrypsin-TLCK (specific activity 60 units/mg; Sigma). The reaction was terminated by incubation at room temperature for 1 h with 50 μg/mL of soybean trypsin and chymotrypsin inhibitor (Sigma). Reaction products with or without reduction were analyzed by SDS PAGE in 10% gels stained with colloidal Coomassie Blue G (Sigma).

For N-terminal sequencing, 200 μg/mL of pCP-2 retaining the prodomain was incubated in 80 μL of buffer containing 10 μg/mL α-chymotrypsin for 15 h at 35 °C. The reaction products were separated by electrophoresis in 10% polyacrylamide gel containing SDS and transferred onto PVDF membrane (Bio-Rad). The protease-resistant fragments were recovered from the membrane and assayed by amino acid sequencing by Edman degradation at the Wistar Protein Microsequencing Facility, Philadelphia, PA.

Binding Assay of Procollagen C-Proteinase-2 (mTld) and Its Isolated CUB/EGF Domains to Substrate, Procollagen I. For these experiments, procollagen I was purified to homogeneity from 17-day old chick embryo tendon fibroblasts as previously reported (*7*). The procollagen I was then covalently linked to an inert bead matrix through N-hydroxysuccinimide on the resin. For the reaction, procollagen I (2.5 μg with specific activity of 14 000 cpm/μg) was incubated with 100 μL of 50% suspension of Affigel-10 (Bio-Rad). After 5 h at 4 °C, unbound procollagen was removed by washing with 2 M urea, 3 M NaCl, 11 mM $CaCl_2$, and 50 mM Tris-HCl buffer, pH 7.5. The resin was blocked with 1% bovine albumin in PBS. Procollagen C-proteinase-2 (mTld) or protease fragments (about 20 μg/mL) were incubated with the resin in 50 μL of 0.15 M NaCl, 11 mM $CaCl_2$, 0.015% Brij30, 0.1% NaN_3, and 50 mM Tris-HCl buffer, pH 7.4, for 2 h at 4 °C. The bound proteins, after washing, were recovered by boiling the isolated resin fraction in 50 μL of buffer with a final concentration of 0.1% SDS and 1% 2-mercaptoethanol. The competition assay was conducted in the presence of soluble procollagen type I at concentration 17 μg/mL.

Alignment and Homology Modeling of CUB Domains. A molecular model of the CUB domains of the pCP-2/mTld was constructed using the three-dimensional structure of the porcine sperm adhesion protein PSP-I/PSP-II as a structural template (GenBank accession numbers UO2626 and UO2627, respectively) (*40*). [Dr. Antonio Romero (CSIC, Madrid, Spain) kindly provided the coordinates of the PSP-I prior to their release in PDB database.] This template was selected based on a FASTA search of the sequence database (*41*, *42*). The CUB domains were aligned with one another and with PSP-I/PSP-II sequences using the ClustalW (Multiple Sequence Alignment) method (*13*, *43*). The sequence for each CUB domain was corrected based on the crystal structure coordinates of PSP-I.

Modeling was performed on Indigo II and Octane (Silicon Graphics) computers using the Sybyl software package (Tripos and Associates, St. Louis, MO) as described previously (*44*) and the Insight II package (Homology, Discover; MSI, San Diego, CA). The gaps shown in Table 1 were inserted using the protein loop search algorithm provided in the Sybyl suite of software. In essence, an appropriate loop was selected from a fragment database based on homology, distance geometry of the anchoring residues present in the template, and the length of the desired loop. All insertions were grafted into existing loops in the template (Table 1). The CUB5 domain contains a 7 amino acid carboxyl terminal extension, which was modeled as short α-helix.

The side chains of the PSP-I protein were consecutively replaced with the relevant CUB domain sequences. Modeling calculations were performed using the Kollman All_Atom force field. Explicit hydrogens were used in all calculations, and the charges were calculated according to the method of Gasteiger-Huckle. Minimizations were performed using the conjugate−gradient method, and the structure was minimized to convergence. Subsequently, four rounds of energy-dependent simulations of molecular motion (dynamics) were performed (100 fs) followed by 50 steps of minimization. The "rocking" motion of alternating rounds of dynamics and minimization was used for approaching a local energy minimum.

CD Spectroscopy of Recombinant Protein. About 1.5 mg of recombinant pCP-2 was incubated with α-chymotrypsin under the conditions described above. The fragments were separated on a size-exclusion column (HiLoad Superdex-200; Pharmacia). Fractions containing the 90 or 40 kDa fragments were concentrated to a final concentration of about 0.3 mg/mL using a YM-30 filtration membrane (Amicon Diaflo Systems). The samples were dialyzed against salt-free 10 mM phosphate buffer (pH 7.5), centrifuged (30 min at 10000*g*) to remove precipitated protein, and analyzed by CD spectrometry using a thermostated quartz cell with a path length of 0.1 cm in JASCO J-600 spectropolarimeter. A thermocouple thermometer (Type T460 ATT; Omega Engineering, Inc) monitored the temperature of the sample. The temperature of the circulating water bath was controlled at 24 °C by a temperature programmer (J-700 for Windows; NESLAB RTE 111/Bath/Circulator) with a microprocessor controller. Calculations of secondary structures were performed using the three structure-prediction programs by Contin (*45*), Estima (*46*), and Selcon (*47*) to obtain average values. The protein concentration used for calculations was

Table 1: Sequence Alignment

```
                 10[a]        20                  30        40        50
                  .           .                   .         .         .
PSP-I     LDYHACGGRLTDDYGTIFTYKGPK---------TECVWTLQVDPKYKLLVSIPTLNLT----CGKE
CUB1 322  -----CGETLQDSTGNFSSPEYPN----GYSAHMHCVWRISVTPGEKIILNFTSLDLYRSRLCWYD
CUB2 435  -----CGGDVKKDYGHIQSPNYPD----DYRPSKVCIWRIQVSEGFHVGLTFQSFEIERHDSCAYD
CUB3 591  -----CGGFLTKLNGSITSPGWPK----EYPPNKNCIWQLVAPTQYRISLQFDFFETEGNDVCKYD
CUB4 747  -----CDHKVTSTSGTITSPNWPD----KYPSKKECTWAISSTRGHRVKLTFMEMDIESQPECAYD
CUB5 860  -----CGGQVRADVKTKDLYSQPQFGDNNYPGGVDCEWVIVAEEGYGVELVFQTFEVEEETDCGYD
          111aaaa22222bbb333333ccccccccccc444444444dddd55555555555eeeeeeee6
    .     ßßßllllßßßßßßlllßßßßßßßllllllllllllßßßßßßßßßßlllllßßßßßßßßßßßlllllllllß


                  60          70          80        90        100       110
                   .          .           .         .         .         .
PSP-I     YVEVLEGA-PGSKSLGKFCE-G-LSILNRGSSGMTVKYKRDSGHPASPYEIIFLRDSQG
CUB1 379  YVEVRDGFWRKAPLRGRFCGSKLPEPIVSTDSRLWVEFRSSSNWVGKGFFAVYEAI
CUB2 492  YLEVRDGHSESSTLIGRYCGYEKPDDIKSTSSRLWLKFVSDGSINKAGFAVNFFKE
CUB3 648  FVEVRSGLTADSKLHGKFCGSEKPEVITSQYNNMRVEFKSDNTVSKKGFKAHFFSD
CUB4 804  TLEVFDGRDRKAPVLGRFCGSKKPEPVLATGSRMFLRFYSDNSVQRKGFQASHATE
CUB5 921  YMELFDGYDSTAPRLGRYCRSGLPEEVYSAGDSVLVKFHSDDTITKKGFHLRYTSTKFQDTLHSRK
          66666ffffffff77777ggggg8888hhhhhhh999999iiiiiii00000000
          ßßßßßlllllllllßßßßßßllßßßßßßßßlllllllßßßßßßßlllllllllßßßßßßßßß

1-0 = ß-strand 1-10

a-i = loops a-i

a) - PSP-I/PSP-II numbering
```

Table 2: CD Data on C−Proteinase (mTolloid)[a]

protein or fragment	α-helix[b] (%)	β-sheet[b] (%)	β-turns and[b] random coil (%)
C-proteinase (mTolloid)	±5	70	23
90 kDa fragment	±5	75	20
40 kDa fragment	±5	60	34
calcd for CUB4 and CUB5[a]	0	59	41

[a] Based on model in Figures 4 and 5. [b] $n = 3$.

determined by the Bradford method (*48*). The inherent error for the measurement was ±5%.

RESULTS

Characterization of Recombinant Latent pCP-2 and pCP-2 Lacking the Prodomain. DNA constructs coding for pCP-2 and truncated pCP-2 were stably expressed in HT-1080 cells. Recombinant proteins secreted into the media were purified from high-expressing clones. Both recombinant proteins were purified to homogeneity as assayed by Western blotting and Coomassie Blue G staining of polyacrylamide gels. Figure 1A shows a purification result of the intact C-proteinase/mTld. Identical purity was achieved for the truncated form of the protein (not shown).

The native pCP-1 (BMP-1) and pCP-2 (mTld) were originally purified as active enzymes (*7*, *8*), and no latent forms were identified; therefore, we tested the full-length pCP-2 and its amino terminally truncated form lacking the propeptide for enzymic activity. The form lacking the prodomain had specific activity on procollagen I, the same as pCP-2 purified from chick embryo tendons (Figure 1B, lanes 2 and 4). The full-length, unprocessed protein, however, had only about one-tenth the specific activity of both, the truncated recombinant pCP-2 and the native processed enzyme (Figure 1B, lane 3). The residual activity of the full-length form could be due to the presence of some processed mature form of the enzyme. These results also indicate that the propeptide is not essential for correct folding of the protein in human fibrosarcoma cells (HT-1080).

Defining the CUB/EGF Domains by Limited Proteolysis. The domain structure of the CUB/EGF homologous sequences in pCP-2/mTld was defined by limited digestion with α-chymotrypsin. As indicated in Figure 2, limited proteolysis of the 120 kDa protein generated fragments of 90, 70, 60, and 40 kDa. The same fragments were obtained when the digestion was carried out at 35 °C for 15 h with 1−10 μg/mL of α-chymotrypsin, or for 1−15 h with 10 μg/mL of

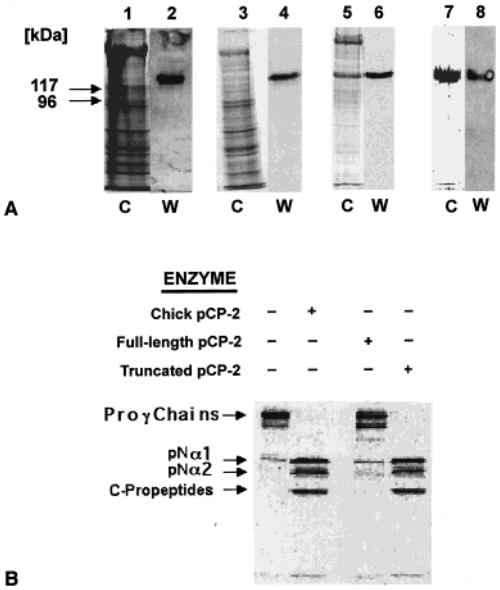

FIGURE 1: Purification of recombinant pCP-2/mTld and enzymic activity assay on procollagen I. (A) The recombinant enzyme from concentrated medium (lanes 1 and 2) was eluted from a green A Dye matrix gel affinity column with 3 M NaCl, 2 M urea, and 0.05 M Tris-HCl, pH 7.5 (lanes 3 and 4). Subsequently, the recombinant enzyme was recovered from a concanavalin A Sepharose column (lanes 5 and 6) under conditions described previously (*7*). The concentrated sample was finally purified to homogeneity by size exclusion chromatography on HiLoad Superdex S-200 column (lanes 7 to 8). C and W indicate Coomassie stained gel and Western blot, respectively. (B) Assays for 2 h at 35 °C using chick procollagen I as a substrate and 2 units (about 0.02 μg) chick pCP-2, 0.3 μg of recombinant pCP-2 and 2 units (about 0.02 μg) of recombinant truncated pCP-2. The substrate (0.3 μg in 20 μL) was labeled with ^{14}C-amino acids and assayed on a SDS−polyacrylamide gel with a phosphorimager. Symbols: Chick pCP-2, 2 units of C-proteinase (Tolloid) from chick embryos; full-length pCP-2, about 0.3 μg of recombinant C-proteinase (mTolloid); Truncated pCP-2, 2 units of recombinant C-proteinase (mTolloid) prepared with construct lacking sequences for the N-terminal propeptide.

α-chymotrypsin (not shown). Therefore, the protease-resistant fragments probably define separately folded domains. Electrophoretic analysis of the intact enzyme and the digestion products with and without reduction yielded the same patterns except that the reduced samples showed lower mobility (not shown), probably due to the disruption of the compact globular structure maintained by intramolecular disulfide bridges. This compact, native like structure was not altered by SDS alone or by treatment with chymotrypsin. Faster migration of nonreduced polypeptides when compared to reduced sample might imply that the disulfide bonds stabilized these globular entities.

N-Terminal sequencing of the fragments eluted from the gel demonstrated that the 90 kDa fragment lacked the propeptide and protease domain (right panel, Figure 2). The 70 kDa fragment contained the complete C-terminal region except the first CUB domain. The 60 kDa fragment lacked the first two CUB domains. Some preparations of the 60 kDa fragment also contained a small amount of an overlapping fragment that contained sequences spanning CUB2 to CUB4. The 40 kDa fragment contained just the last two CUB domains. Treatment of the full-length pCP-2 from chicken with N-glycosidase causes a 10 kDa shift on SDS electrophoresis (data not shown), indicating N-linked glycosylation. A search for potential sites for N-linked glycosylation using prosite yielded four sites, which were located in the catalytic domain (one site), the CUB1 domain (two sites), and the CUB3-domain (one site). Hence, the N-glycosylation sites seem to be distributed all over the protein. Therefore, the sizes of the N-terminally truncated chymotryptic fragments lacking the catalytic domain can probably deduced without major deviation from the SDS−gels.

The fragments were isolated without reduction; therefore, the results indicate that disulfide bonds do not link the separate domains and imply that these cysteines are fundamental to maintain the intra-domain structure.

Demonstration That Three First but Not Last Two CUB/EGF Domains Specifically Bind to Procollagen Type I. The presence of the exosites outside the catalytic domain was reported for numerous proteinases (*30*−*37*). The results obtained here from an enzymic activity assay of the truncated

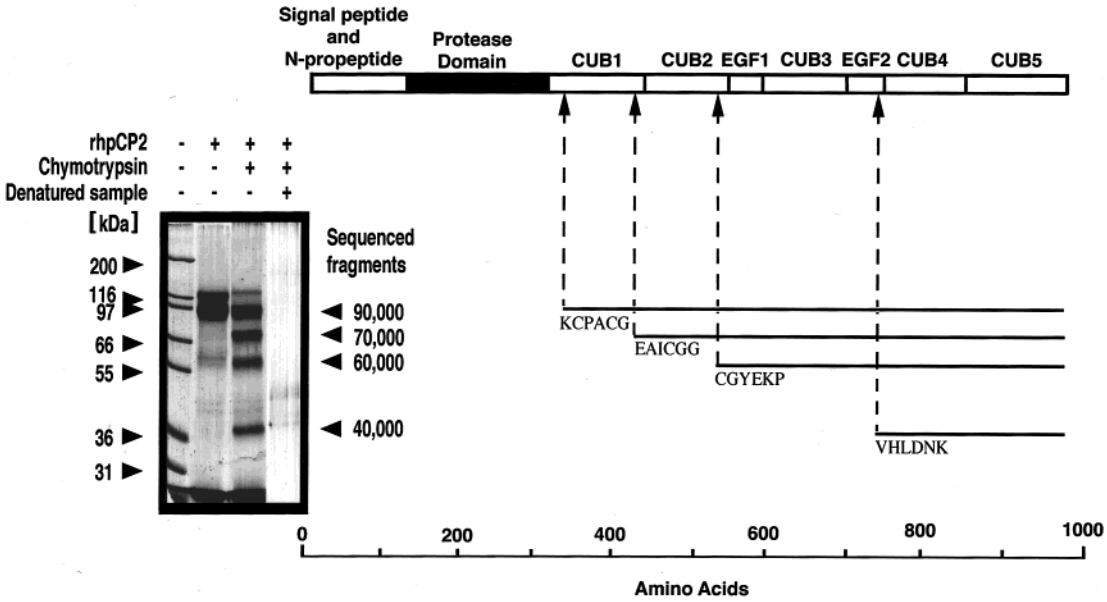

FIGURE 2: Defining CUB/EGF domains by limited digestion with α-chymotrypsin. (Left) SDS−PAGE of recombinant pCP-2/mTld (200 μg/mL) digested for 15 h at 35 °C with α-chymotrypsin (10 μg/mL). Gel was stained with colloidal Coomassie Blue G. (Right) N-Terminal sequences and sizes of fragments obtained from gel in the left panel.

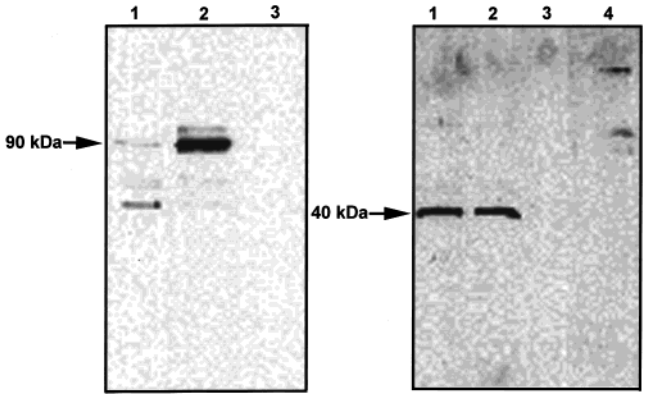

FIGURE 3: Binding of recombinant C-proteinase (mTolloid) and its fragments to procollagen I. Human procollagen I was covalently bound to an insoluble resin through a functional N-hydroxysuccinimide group on the resin (AffiGel-10; Bio-Rad). Recombinant C-proteinase fragments (20 μg/mL) were then incubated with the resin for 5 h at 4 °C. The bound proteins were eluted with SDS/2-mercaptoethanol, separated on an SDS−polyacrylamide gel, and detected by Western blotting with polyclonal antibodies to recombinant C-proteinase. (Left) Binding of the 90 kDa fragment. Lane 1: Binding to resin that did not contain procollagen I. Lane 2: Test sample with resin containing procollagen I. Lane 3: Competitive assay with 17 μg/mL procollagen I. (Right) Assay similar to LEFT PANEL with 40 kDa fragment of C-proteinase. Lanes 1: Control sample of supernatant from incubation with resin not containing procollagen I immobilized on the resin. Lane 2: Control sample of supernatant from incubation with procollagen I. Lanes 3 and 4: Assays with the 40 kDa fragment as with 90 kDa polypeptide in lanes 1 and 2 of the left panel.

recombinant human pCP-2 and on isolation without reduction of single polypeptides containing different numbers of CUB/EGF motifs provide evidence that the protease resistant fragments had native conformation. Therefore, we tested the possibility that the isolated polypeptides may bind specifically to a substrate, procollagen type I. As shown in lane 2 of the left panel of Figure 3, the 90 kDa fragment specifically bound to procollagen I, immobilized on the resin. The interaction of the 90 kDa pCP-2 fragment with the resin-bound-procollagen was competitively inhibited by the presence of substrate (17 μg/mL) in the incubation mixture (lane 3). The last two lanes in the right panel of Figure 3 are identical to the two first lanes in the left panel in this figure except that the 40 kDa fragment was assayed for binding to the procollagen. These results demonstrate that the 40 kDa fragment lacking the first three CUB domains does not bind to type I procollagen. The peptide was recovered in the incubation buffer from both reactions, the control without procollagen (lane 1 in the right panel of Figure 3), and from the reaction with procollagen immobilized on the resin (lane 2 in the right panel of Figure 3). Therefore, the specific binding site(s) (exosite) for procollagen I seem to be located within the first three CUB domains of pCP-2. The data support a concept that the CUB1-3 and the EGF1 domains are important for substrate (procollagen I) binding.

Alignment and Homology Modeling of CUB Domains. Structural information on pCP-2 is still not available. Hence, we used sequence alignment and homology modeling, in an attempt to elucidate the structural and functional properties of single CUB domains. A database search for proteins homologous to the CUB domains of the procollagen C-proteinase yielded members of the astacin family of zinc endopeptidases (*13*, *14*), complement proteins such as C1s

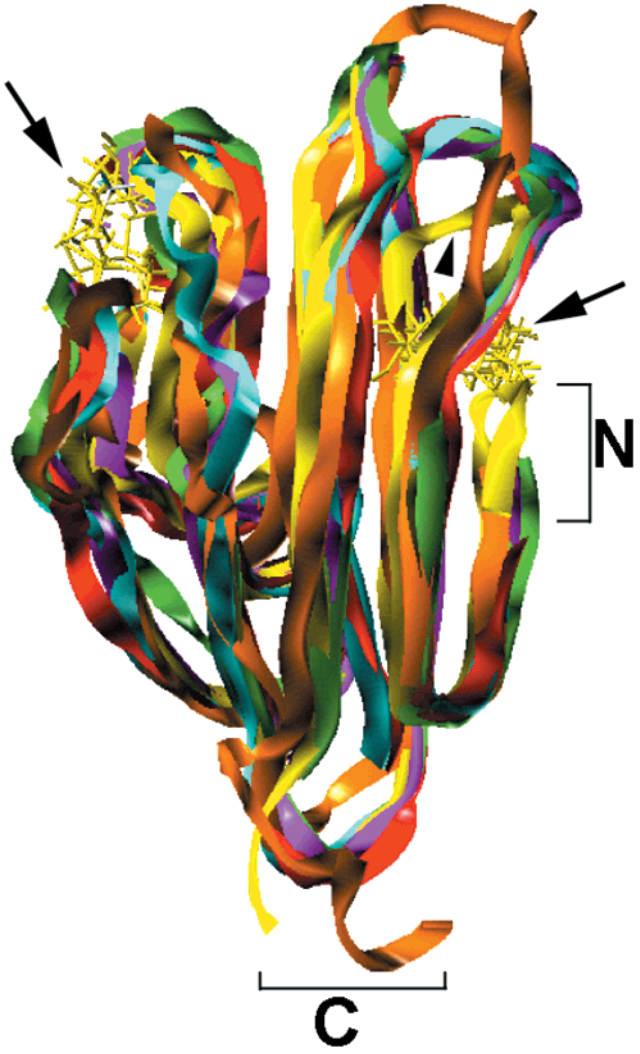

FIGURE 4: The energy minimized molecular models of the five CUB domains of human pCP-2/mTld prepared by DISCOVER. Models were calculated on the basis of crystal structure coordinates of the porcine PSP-I (kindly provided by Dr. Antonio Romero). CUB1, CUB2, CUB3, CUB4, CUB5, and PSP-1 are shown in cyan, green, magenta, red, orange, and yellow, respectively. The N-termini are located on the right-hand side (N) and the C-termini at the bottom (C).

and C1r, and a series of mammalian proteins involved in sperm adhesion (e.g., refs *39* and *40*). The proteins with the closest similarity to CUB domains whose three-dimensional structures have been solved included the porcine zona pellucida binding heterodimeric PSP-I/PSP-II protein (*40*) and bovine acidic seminal fluid protein (aSFP) (*49*).

PSP-I/PSP-II is a heterodimer of subunits sharing about 45% sequence identity (*40*). Here we used the PSP-I subunit for alignment and modeling of the CUB domains of the human pCP-2 (Figure 4). PSP-I comprises 109 amino acid residues arranged in two five-stranded β-pleated sheets (strands β1 through β10), each containing three parallel and two antiparallel strands. The β-strands are interconnected by nine loops termed LA through LI (*40*). Two disulfide bridges conserved in each CUB domain link loop LA to strand β4 and loop LA to loop LG, respectively.

An alignment of the CUB1 to five domains of human pCP-2 and the PSP-I monomer is possible if five insertions are introduced (Table 1). According to the underlying

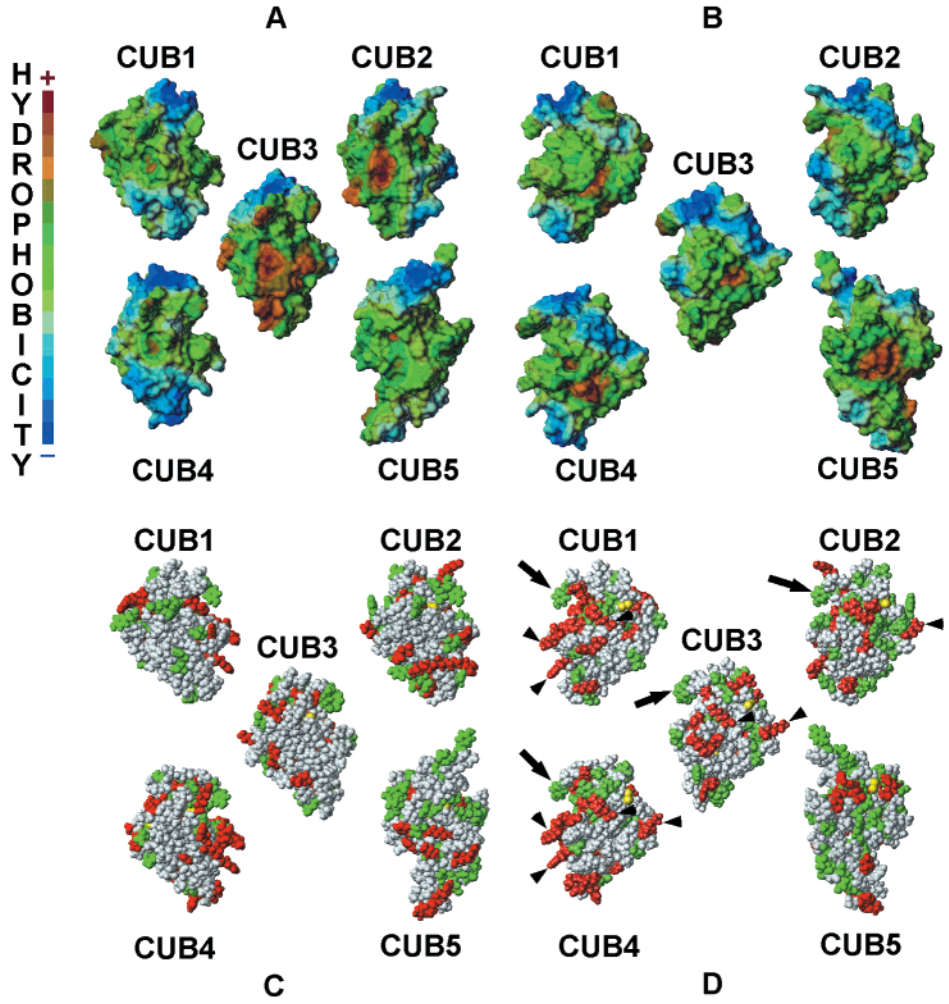

FIGURE 5: Surface images (Connolly dot surface, rolling sphere of 1.4 Å, prepared by INSIGHT II) of two opposite views of the five CUB domains of human pCP-2/mTld. (A and C) Focus on β-strand 6 (i.e., Asp378 to Arg383 in CUB1); (bottom) focus on β-strand 5 (i.e., Glu358 to Phe364 in CUB1). (A and B) Corresponding electrostatic properties of the same images prepared by MOLCAD, Tripos. (C and D) Positively charged surface residues in red, negatively charged residues in green, cysteines in yellow. Arrowheads point to conserved positively charged surface residues and the arrows indicate conserved negatively charged surface residues.

topology of the corresponding three-dimensional structures of PSP-I, these insertions are located in loops LC, LE, LF, and LG (Table 1). Thus, these insertions do not affect the overall folding of the protein (Figure 4).

The five individual CUB domains share between 30 and 42% identical residues. Most notably, the positions of the four conserved cysteines are identical between the PSP-I protein and all five CUB modules. The percent identities between PSP-I and CUB domains range from 17% for CUB5 domain to 27% for CUB3 domain.

According to the homology models of all CUB domains and in agreement with the PSP-I structure, the four cysteine residues of each CUB domain form two intradomain disulfide bridges (Figure 4), which is supported by biochemical data. If, on the contrary, the disulfides would link the CUB domains to each other, the nonreduced electrophoretic pattern after α-chymotrypsin digestion should contain fewer bands. Similarly, we did not detect by SDS−PAGE fragments other than 90 or 40 kDa in corresponding fractions after size-exclusion chromatography.

Although the models display a high degree of overall similarity there are some significant differences. Loop LC, for example, is much larger in C-proteinase CUB1 to four domains than in PSP-I, and it is extra large in CUB5 domain (indicated by an arrowhead in Figure 4). Another difference is seen in CUB1 domain, where, the four-residue insertion in loop LE contains two basic residues, whereas in CUB4, there are two acidic residues and four acidic residues are found in the corresponding region of CUB5 (Table 1).

The calculated Connolly dot surfaces revealed striking differences in hydrophobicity and charge distribution of their opposite views (Figure 5, panels A and B). The relatively large hydrophobic patch in the CUB3 domain is formed by a series of residues underlined in Table 1 (intense brown area in the lower half of CUB3 in Figure 5A). Part of this patch is also present in CUB2 but not in the other CUB domains (residues underlined in Table 1).

The corresponding opposite surfaces are less variable in CUB1, CUB2, CUB3, and CUB4 (Figure 5B). They all accommodate a relatively small and shallow hydrophobic cleft that is least apolar in the case of CUB2. In the same region of CUB5, the shallow cleft is replaced by a pronounced hydrophobic rim (Figure 5B). In all CUB domains, the β-strand-connecting loops pose strong hydrophilic properties, indicating that they could contribute to an external surface of the procollagen C-proteinase as a whole.

Two and three positively charged clusters are conserved in CUB2, CUB3, CUB4, and in CUB1 and CUB4, respec-

tively (indicated by arrowheads in Figure 5D), which have no corresponding counterparts in CUB5. On the "backside" (the opposite side), there is only a single, negatively charged, cluster, common for CUB1−4, but which is absent from CUB5 (arrows in Figure 5, panels C and D). These features in general provide a rationale for the specific electrostatic properties in the corresponding surface regions of different CUB domains that might be involved in homophilic CUB−CUB domain interactions as well as may contribute to substrate specificity of the enzyme variants. The modeling demonstrates both the structural similarity on one hand and the distinct differences in patches on the surface of CUB1−3 that are candidate sites for intermolecular or intramolecular contacts during enzyme substrate interactions.

CD Assays of Conformation. To verify the computer models of the CUB domains, we examined the secondary structure of full-length pCP-2, 90 kDa, and 40 kDa fragments by CD spectroscopy. The conformation of the 90 kDa (75% of the β-sheet, and 20% of the random coil) was virtually indistinguishable from the intact protein (70 and 23%, respectively) (Table 2). The 40 kDa fragment, which contained only the last two CUB domains had a 15 ± 5% lower content of β-strands. Since the peptide lacked the two EGF-like sequences, one might anticipate that these sequences could contribute to the higher content of β-strands in the full-length pCP-2 and in the 90 kDa fragment of pCP-2. The CD analyses strongly support the modeling data in saying that the C-terminal domains in pCP-2 are almost exclusively arranged in β-strand and turn structures. Most importantly, the data also indicate that the individual C-terminal subfragments are regularly folded.

DISCUSSION

On the basis of the sequence similarity, it has been postulated previously that the large prodomains in pCP variants restrain their enzymic activity (*12*, *14*, *50*). Presence of a putative cleavage site at the transition point between the propeptide and the mature polypeptide chain for a furin-like protease indicates one possibility for activation of the zymogen along the secretory pathway. In our study, no difference has been detected in the procollagen I processing by the native pCP-2 and the recombinant form of the enzyme lacking the entire prodomain up to the putative furin cleavage site, indicating that the prodomain is not critical for correct folding of the recombinant enzyme.

A modular organization is typical for proteins of the extracellular matrix. Previously, extensive efforts have been undertaken to define the structures and functions of modules in C-terminal regions of matrix metalloproteinases such as gelatinases, collagenases, and stromelysin (*51*). Substrate interaction sites outside the catalytic domain have been identified or their presence has been suggested in numerous complex enzymes such as collagenase 3 (*34*), meprin A (*35*), tissue type plasminogen activator (*36*), thrombin (*30*−*33*), and gelatinase A (*37*). The differences in the number of carboxy terminal domains in pCP variants indicate that these parts of the protein might contain exosites that regulate protein targeting and substrate specificity. Our results showed that the first three CUB domains are critical for binding to procollagen I because the last two CUB domains separate did not bind to procollagen I. On the other hand, the pCP-1/BMP-1, which is lacking the last two CUB domains, still processed fibrillar procollagens; however, the K_m value for procollagen I was about 20-fold lower when compared to pCP-2 (compare 7 and 52). The kinetics for the hydrolysis of different fibrillar procollagens by pCP-1/BMP-1 have not been reported so far.

The CUB domain-type sequences are particularly common in complement proteins and in developmentally regulated proteins (*53*). Their role has been proposed in providing sites of protein−protein interactions. Such proteins as aSFP, PSP-I/PSP-II, and the procollagen C-proteinase enhancer protein consist entirely or almost entirely of CUB domains (*40*, *49*, *54*). Others, such as tolloids, contain CUB domains linked to other functional domains. In the complement complex C1, the CUB domains of proteases C1r and C1s are the sites of interaction with C1q, a collagen-like structure. In particular, limited proteolysis of C1s and C1r yielded fragments containing CUB domains and EGF-like domains that retained their capability to bind their cognate target structures (*55*). This supports our observation that fragments of pCP-2 derived by limited proteolysis might likewise be structurally and functionally intact and able to interact with their target substrate procollagen I. The two EGF-domains in pCP-2 are presumably of considerable importance for the structure and function of pCP-2. The fragment that did not bind to procollagen I was devoid of any EGF-like domain. This is in accordance with the data of Thielens et al. (*55*) on C1s and C1r, who did not observe target recognition for separated EGF-motifs. The presence of calcium ions appeared crucial for the protein−protein interactions between C1s and C1r, which is believed to involve specifically the EGF-like motifs.

Recently reported structures for the CUB domains from aSFP and PSP-I/PSP-II (*40*, *49*) somewhat altered the previous theoretical model for CUB domains proposed by Bork and Beckmann (*56*), who had suggested a β-barrel immunoglobulin-like fold. Instead, the CUB-fold resembles a jelly roll type structure (*40*). Our homology models of pCP-2 CUB domains have been verified with respect to their secondary structures composition by CD spectrometry of the proteolytically excised individual domains from recombinant protease. The CUB domains in tolloids are not redundantly attached as repetitive structural motifs. Of special note is that the CUB4 and CUB5 domains differ structurally from the other CUB domains. Although the peptide comprising CUB4 and CUB5 domains did not bind to procollagen I, the domains might provide regulatory functions. There is no information as to how the CUB domains may interact with each other, with the catalytic domain or with substrates. Our model, however, presents several sites for specific interactions most notably in CUB2 and CUB3. Especially, the well-distinguished hydrophobic areas in these domains may be involved in interaction with the cognate substrate, procollagen I, as suggested by the results obtained from binding of α-chymotrypsin generated fragments of pCP-2 to immobilized procollagen I.

The information about surface properties based on the computer model will be an aid in choosing residues for site directed mutagenesis in order to identify further the structural and functional properties of possible sites for protein−protein interactions and of potential exosites. More work is also required to clarify the role of the last two CUB domains

present in the embryonic variant of procollagen C-proteinase-2/mTolloid.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Antonio Romero (CSIC, Madrid, Spain) for generously providing the coordinates for PSP-I/PSP-II molecular structure.

REFERENCES

1. Lapiere C. M., Lenaers A., and Kohn L. D. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 3054−3058.
2. Prockop D. J., Berg R. A., Kivirikko K. I., and Uitto J. (1976) in *Biochemistry of collagen* (Ramachandran, G. N., and Reddi, A. H., Eds.) pp 163−273, Plenum Publishing Corporation, New York.
3. Fessler J. H., and Fessler L. I. (1978) *Annu. Rev. Biochem. 47*, 129−162.
4. Leung M. K., Fessler L. I., Greenberg D. B., and Fessler J. H. (1979) *J. Biol. Chem. 254*, 224−232.
5. Prockop D. J., Kivirikko K. I., Tuderman L., and Guzman N. A. (1979) *N. Engl. J. Med. 301*, 12−23, 75−85.
6. Prockop, D. J., and Kivirikko, K. I. (1984) *N. Engl. J. Med. 311*, 376−386.
7. Hojima, Y., van der Rest, M., and Prockop, D. J. (1985) *J. Biol. Chem. 260*, 15996−16003.
8. Kessler, E., and Adar, R. (1989) *Eur. J. Biochem. 186*, 115−121.
9. Takahara, K., Lyons, G. E., and Greenspan, D. S. (1994) *J. Biol. Chem. 269*, 32572−32578.
10. Li, S.-W., Sieron, A. L., Fertala, A., Hojima, Y., Arnold, W. V., and Prockop, D. J. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 4127−5130.
11. Fukagawa, M., Suzuki, N., Hogan, B. L. M., and Jones, C. M. (1994) *Dev. Biol. 163*, 175−183.
12. Rawlings N. D., and Barrett A. J. (1995) *Methods Enzymol. 248*, 183−228.
13. Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Rüth, F. X., McKay, D. B., and Bode, W. (1995) *Protein Sci. 4*, 823−840.
14. Bond, J. S., and Beynon, R. J. (1995) *Protein Sci. 4*, 1247−1261.
15. Stöcker, W., Gomis-Rüth, F. X., Bode, W., and Zwilling, R. (1993) *Eur. J. Biochem. 214*, 215−231.
16. Bode, W., Gomis-Rüth, F. X., Huber, R., Zwilling, R., Stöcker, W. (1992) *Nature 358*, 164−167.
17. Grams, F., Dive, V., Yiotakis, A., Yiallouros, I., Vassiliou, S., Zwilling, R., Bode, W., and Stöcker, W. (1996) *Nat. Struct. Biol. 3*, 671−675.
18. Johnson, G. D., and Bond, J. S. (1997) *J. Biol. Chem. 272*, 28126−28132.
19. Kessler, E., Takahara, K., Biniaminov, L., Brusel, M., and Greenspan, D. S. (1996) *Science 271*, 360−362.
20. Suzuki, N., Labosky, P. A., Furuta, Y., Hargett, L., Dunn, R., Fogo, A. B., Takahara, K., Peters, D. M., Greenspan, D. S., and Hogan, B. L. (1996) *Development 122*, 3587−3595.
21. Takahara, K., Brevard, R., Hoffman, G. G., Suzuki, N., and Greenspan, D. S. (1996) *Genomics 34*, 157−165.
22. Clark, T. G., Conway, S. J., Scott, I. C., Labosky, P. A., Winnier, G., Bundy, J., Hogan, B. L., and Greenspan, D. S. (1999) *Development 126*, 2631−2642.
23. Clark, T. G., Conway, S. J., Scott, I. C., Labosky, P. A., Winnier, G., Bundy, J., Hogan, B. L., and Greenspan, D. S. (1999) *Development 126*, 2631−2642 (accession no. AF059516).
24. Scott I. C., Blitz I. L., Pappano W. N., Imamura Y., Clark T. G., Steiglitz B. M., Thomas C. L., Maas A. S., Takahara K., Cho K. W. Y., and Greenspan, D. S. (1999) *J. Dev. Biol. 213*, 283−300.
25. Padgett, R. W., Wozney, J. M., and Gelbart, W. M. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2905−2909.
26. Shimell, M. J., Ferguson, E. L., Childs, S. R., and O'Connor, M. B. (1991) *Cell 67*, 469−481.
27. Marques, G., Musacchio, M., Simell, M. J., Wunnenberg-Stapleton, K., Cho, K. W. Y., and O'Connor M. (1997) *Cell 91*, 417−426.
28. Piccolo, S., Agius, E., Lu, B., Goodman, S., Dale, L., and De Robertis, E. M. (1997) *Cell 91*, 407−416.
29. Blader, P., Rastegar, S., Fisher, N., and Strähle, U. (1997) *Science 278*, 1937−1940.
30. DiMaio, J., Ni, F., Gibbs, B., and Konishi, Y. (1991) *FEBS Lett. 282*, 47−52.
31. Qiu, X., Yin, M., Padamanabhan, K. P., Krstenansky, J. L., and Tulinsky, A. (1993) *J. Biol. Chem. 272*, 20318−20326.
32. van de Locht, A., Lambda, D., Bauer, M., Huber, R., Friedrich, T., Kröger, B., Höffken, W., and Bode, W. (1995) *EMBO J. 14*, 5149−5157.
33. Fethiere, J., Tsuda, Y., Coulombe, R., Konishi, Y., and Cygler, M. (1996) *Protein Sci. 5*, 1174−1183.
34. Knäuper, V., Cowell, S., Smith, B., Opez-Otin, C., O'Shea, M., Morris, H., Zardi, L., and Murphy, G. (1997) *J. Biol. Chem. 272*, 7608−7616.
35. Tsukuba, T., and Bond, J. S. (1998) *J. Biol. Chem. 273*, 35260−35267.
36. Ke, S. H., Tachias, K., Lambda, D., Bode, W., and Madison, E. L. (1997) *J. Biol. Chem. 272*, 1811−1816.
37. Overall, C. M., King, A. E., Sam, D. K., Ong, A. D., Lau, T. T. Y., Wallon, U. M., DeClerck, Y. A., and Atherstone, J. (1999) *J. Biol. Chem. 274*, 4421−4429.
38. Fertala, A., Sieron, A. L., Ganguly, A., Li, S.-W., Ala-Kokko, L., Anumula, K. R., and Prockop, D. J. (1994) *Biochem. J. 298*, 31−37.
39. Kanekal, S., Sahai, A., Jones, R. E., and Brown, D. (1995) *J. Pharmacol. Toxicol. Methods 33*, 171−178.
40. Varela, P. F., Romero, A., Sanz, L., Romao, M. J., Töpfer-Petersen, E., and Calvete, J. (1997) *J. Mol. Biol. 274*, 635−649.
41. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444−2448.
42. Needleman, S. B., and Wünsch, C. D. (1970) *J. Mol. Biol. 48*, 443−453.
43. Smith, T. F., and Waterman, M. S. (1981) *J. Mol. Biol. 147*, 195−197.
44. Jameson B. A. (1989) *Nature 341*, 465−466..
45. Provencher, S. W., and Glockner, J. (1981) *Biochemistry 20*, 33−37.
46. Yang, J. T., Wu, C. S., and Martinez, H. M. (1986) *Methods Enzymol. 130*, 208−269.
47. Sreerama, N., and Woody, R. W. (1993) *Anal. Biochem. 72*, 248−254.
48. Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
49. Romao, M. J., Kolln, I., Dias, J. M., Carvalho, A. L., Romero, A., Varela, P. F., Sanz, L., Töpfer-Petersen, E., and Calvete, J. (1997) *J. Mol. Biol. 274*, 650−660.
50. Childs, S. R., and O'Connor, M. B. (1994) *Dev. Biol. 162*, 209−220.
51. Murphy, G., and Knäuper, V. (1997) *Matrix Biol. 15*, 511−518.
52. Adar, R., Kessler, E., and Goldberg, B. (1986) *Collagen Relat. Res. 6*, 267−277.
53. Bork, P. (1992) *Curr. Opin. Struct. Biol. 2*, 413−421.
54. Takahara, K., Kessler, E., Biniaminov, L., Brusel, M., Eddy, R. L., Jani-Sait, S., Shows, T. B., and Greenspan, D. (1994) *J. Biol. Chem. 269*, 26280−26285.
55. Thielens, N. M., Enrie, K., Lacroix, M., Jaquinod, M., Hernandez, J.-F., Esser, A. F., and Arlaud, G. J. (1999) *J. Biol. Chem. 274*, 9149−9159.
56. Bork, P., and Beckmann, G. (1993) *J. Mol. Biol. 231*, 539−545.

BI992312O