Structure and Expression of the cDNA Encoding Human Neutrophil Collagenase

By Prasad Devarajan, Kasim Mookhtiar, Harold Van Wart, and Nancy Berliner

We have isolated and characterized a 2.4-kb cDNA clone encoding human neutrophil collagenase (HNC), a member of the family of matrix metalloproteinases restricted to secondary granules within neutrophils. Partial amino acid sequence was used to deduce oligonucleotide probes. These probes were used to screen a human granulocyte cDNA library derived from messenger RNA (mRNA) from a patient with chronic granulocytic leukemia. Cell-free translation of RNA produced from the cDNA produced a 52-Kd protein that was recognized by anti-HNC antibody. The cDNA clone was sequenced and shown to encode a 467-residue protein whose sequence matched those regions currently known for HNC. The enzyme exhibits 58% homology to human fibroblast collagenase and has the same domain structure. It consists of a 20-residue signal peptide, and an 80-residue propeptide that is lost on autolytic activation by cleavage of an M-L bond. Other regions identified include the autolytic degradation site, the “cysteine switch” residue that is involved in latency and activation, and a putative zinc binding sequence. HNC has six potential N-linked glycosylation sites. The cDNA hybridized to a 3.4-kb mRNA in RNA from a patient with chronic granulocytic leukemia, but not to RNA from uninduced HL60 cells or HL60 cells that had been induced to undergo granulocytic or monocytic maturation with dimethyl sulfoxide or 12-O-tetradecanoylphorbol 13-acetate, respectively. These results parallel those seen with lactoferrin and transcobalamin I, two other secondary granule proteins.

THE TURNOVER OF THE extracellular matrix associated with growth, healing, and a variety of pathologic disorders is believed to be catalyzed by the matrix metalloproteinases (MMP).1 These enzymes are zinc proteinases that are collectively capable of degrading all of the major matrix macromolecules. Important members of this protein family are the interstitial collagenases that initiate the breakdown of type I, II, and III collagens. The interstitial collagenses from human fibroblasts (HFC), which has been cloned2 and extensively studied,3,4 is very similar or identical to the collagenase produced by most other human cells.5,6 In contrast, human neutrophil collagenase (HNC)7 is an immunologically distinct enzyme5 with a different apparent molecular weight,6 glycoprotein processing,7 and collagen8-10 and peptide (unpublished data) specificity. Recently, the determination of the sequence of a limited N-terminal portion of HNC11,12 has confirmed that HFC and HNC are distinct but homologous collagenases. HNC expression is apparently restricted to the neutrophil, where it is a major constituent of secondary granules.12

Our knowledge of the similarities and domain structures of the human MMP gene family has been greatly enhanced by the cloning and comparative sequence analysis of its members.13,16-20 In this study, knowledge of the partial amino acid sequence of HNC21 has been used to isolate a cDNA clone that encodes for the entire protein. An analysis of the cDNA-derived amino acid sequence has allowed a detailed comparison of HNC and HFC and the other MMP. In addition to establishing the domain structure of HNC, homologies in the putative zinc binding region of the catalytic domain and “cysteine-switch” region of the propeptide domain that are thought to be involved in the latency and activation of the zymogen22 have been examined. Comparison of the full sequence of HNC with available experimental data gives the locations of the signal peptidease, autolytic activation, and autolytic degradation sites, as well as information on the glycosylation sites.

In addition, the expression of HNC has been examined in the context of its identity as a neutrophil secondary granule protein. Secondary granule content protein gene expression is thought to be coordinately regulated at the level of transcription, and to be restricted to the stages of neutrophil maturation associated with secondary granule formation.23 HL60 cells are a human leukemia cell line that can undergo limited granulocytic maturation on exposure to dimethyl sulfoxide (DMSO) and limited monocytic maturation on treatment with 12-O-tetradecanoylphorbol 13-acetate (TPA); however, they do not make secondary granules, and do not produce any of the proteins contained within the secondary granule.24 The absence of lactoferrin and transcobalamin I, two of the major content proteins of the secondary granule, has been shown previously to be associated with a complete absence of messenger RNA (mRNA) encoding these proteins.25,26 We have used the HNC cDNA as a probe for HNC mRNA in peripheral blood from patients with chronic granulocytic leukemia (CGL) and in HL60 cells to examine whether HNC expression parallels that of the other secondary granule protein genes.

MATERIALS AND METHODS

Materials. HNC was isolated from buffy coats and [3H]aceteylated rat tendon collagen was prepared as described previously.27,28 The Immuno-Blot assay kit, goat antirabbit IgG antibody conjugated to alkaline phosphatase, biotinylated mouse monoclonal antibodies, and avidin-alkaline phosphatase were purchased from Bio-Rad Laboratories, Richmond, CA. P-chloromercuribenzoate (PCMB) was purchased from Sigma (St Louis, MO). DNA sequencing was performed on an automatic DNA sequencing machine at the Biotechnology Resource Center of the University of Florida. The Immuno-Blot assay kit, goat antirabbit IgG antibody conjugated to alkaline phosphatase, biotinylated mouse monoclonal antibodies, and avidin-alkaline phosphatase were purchased from Bio-Rad Laboratories, Richmond, CA. P-chloromercuribenzoate (PCMB) was purchased from Sigma (St Louis, MO). DNA sequencing was performed on an automatic DNA sequencing machine at the Biotechnology Resource Center of the University of Florida.

From the Hematology Section, Department of Internal Medicine, Yale Medical School, New Haven, CT; and the Department of Chemistry and Institute of Molecular Biophysics, Florida State University, Tallahassee.

Submitted October 18, 1990; accepted February 8, 1991.

P.D. is supported by NICHD Grant No. HD-22297 from the National Institutes of Health and by a Small Research Grant from the American Cancer Society.

Address reprint requests to Nancy Berliner, MD, Hematology Section, Yale Medical School WWB 423, 333 Cedar St, New Haven, CT 06510.

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. section 1754 solely to indicate this fact.

© 1991 by The American Society of Hematology.
formed using the Sequenase kit (US Biochemicals, Cleveland, OH); probes were radiolabeled by nick translation using the kit from Boehringer-Manheim (Indianapolis, IN); restriction enzymes and T4 polynucleotide kinase were purchased from New England Biolabs (Beverly, MA); in vitro translations were performed using rabbit reticulocyte lysates purchased from Promega (Madison, WI); polymerase chain reaction (PCR) kit was from Perkin-Elmer-Cetus (Norwalk, CT).

Protein sequencing. A sample (100 μg) of pro-HNC (58 Kd) was incubated at 50°C for 1 hour in assay buffer. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the sample showed that this treatment produced a mixture of intact 58-Kd HNC and a 27-Kd fragment. The remaining intact HNC was removed by chromatography over Sepharose-CH-Pro-Leu-Gly-NH2 and the sample electroblotted onto polyvinylidene difluoride and subjected to sequential N-terminal sequence analysis. The analysis was performed with an Applied Biosystems (Foster City, CA) 470A protein sequencer equipped with a 120A on-line PTH-analyzer.

Preparation of oligonucleotide probes. Two synthetic oligonucleotides of 50 bp were prepared corresponding to published partial amino acid sequence of the HNC propeptide. A computer program, BIGPROBE (Arizona Board of Regents), was used to design “guessmers” based on codon usage, dinucleotide frequency, and potential probe self-complementarity. The predicted probes contained the following sequences: AAGGAGAAGAACACCAA-ATGGTGACCGGCAAGCCCAATGAGGAGACCCTGGAGAAGATCTACCA, corresponding to the amino acid sequence of residues 25 to 41; and GGTTGACCGGCAAGCCCAATGAGGAGACCCTGGAGAAGATCTACCA, corresponding to sequences from 72 to 88. The probes were 5’ end-labeled using T4 polynucleotide kinase.

Screening of cDNA library. The construction of the CGL cDNA library in Lambda-Zap II (Stratagene, LaJolla, CA) has been described previously. The library was propagated in XL1-Blue cells (Stratagene), and plated at 50,000 plaques/plate. Quadruplicate nitrocellulose filters were lifted, and hybridized overnight at 45°C. Hybridization was performed in duplicate to each of the two oligonucleotides in 6X SSC/0.05% Na pyrophosphate for 15 minutes each at 27°C and 45°C. Positive plaques were identified by autoradiography, and purified by secondary and tertiary screening, which yielded three clones that hybridized to both probes were identified. These clones were isolated as pBluescript phagemids, and analyzed by restriction enzyme digestion. The clone containing the largest insert (2.4 kb) was sequenced completely by the dideoxy method.

Protein sequence analysis. Sequence analysis and molecular weight calculations were performed with the GAP (Gap wt 3.0; Gap length 0.1) and PEPTIDESORT programs of the sequence analysis software package of Genetics Computer Group (GCG) Inc (Madison, WI) using a VAX computer.

In vitro translation of HNC protein. The HNC cDNA plasmid was linearized with Xho I, and capped mRNA synthesized using T3 RNA polymerase. After quantification by spectrophotometry, 250 ng of RNA was translated in a 25 μL rabbit reticulocyte lysate reaction containing 10S-methionine. After preincubation with non-immune rabbit serum, translated products were immunoprecipitated with rabbit anti-HNC antibody, and translation products before and after precipitation were visualized after fractionation on SDS-PAGE and enhanced fluorography of the dried gel. In vitro translation was also performed using unlabeled methionine for use in Western blots.

**Electrophoresis and immunoblots.** SDS-PAGE experiments were performed as described by Laemmli. Western blots were performed using rabbit anti-HNC as the primary antibody and goat antirabbit IgG alkaline phosphatase conjugate as the second antibody, as described in the BioRad Immuno-Blot assay kit. To block binding to cross-reacting proteins within the reticulocyte lysate, anti-HNC antibody was preadsorbed with an aliquot of the lysate and filtered before binding to the nitrocellulose.

**Assay of collagenase activity.** Samples of in vitro translated protein (100 μL) were diluted to 1 mL with 50 mmol/L Tricine, 0.2 mol/L NaCl, 10 mmol/L CaCl2, 50 μmol/L ZnCl2, pH 7.5, and concentrated/dialyzed versus this same buffer with a Bio-Molecular Dynamics Micro-ProDiCon apparatus to final volume of 100 μL. The samples were then diluted 50-fold with assay buffer (same as above, except containing 0.05% Brij-35) and a 5-μL aliquot was activated by treatment with 100 μmol/L PCMB for 30 minutes. The activated sample was added to 75 μL of a 5 μmol/L solution of [3H]acetylated rat type 1 collagen. The reaction mixture was incubated at 30°C for 5 hours and the hydrolysis of the collagen analyzed by SDS-PAGE.

**Northern blot analysis.** Fifteen micrograms of total RNA was electrophoresed in a 1% agarose/formaldehyde denaturing gel, blotted onto nitrocellulose, and hybridized overnight at 42°C to the 32P-labeled nick translated probes. The blot was washed twice in 0.1% SSC/0.1% SDS for 30 minutes at 55°C and autoradiographed.

**Determination of S′ mRNA sequence by PCR.** Single strand synthesis of 10 μg of CGL RNA was performed using AMV reverse transcriptase with random hexamer priming. The resultant cDNA was tailcd with oligo dC. One fifth of the cDNA was used for a PCR reaction using 100 pmol/L of an antisense oligonucleotide based on the S′ sequence of the cDNA and oligo dG. PCR products were ligated into plasmid vectors and sequenced.

**RESULTS**

**Isolation of cDNA clone.** The entire sequence of the 2.4-kb cDNA clone was determined (Fig 1). Identity of the clone was confirmed by matching DNA sequence to known amino acid sequence. The cDNA encodes a protein of 467 amino acids, with a 20-residue leader sequence and an 80-residue propeptide. The mature protein, as predicted by the amino terminal sequence of the mature enzyme begins at amino acid residue 101. The cDNA contains a 900 nucleotide 3′ untranslated region with a single AATAAA polyadenylation site. While this manuscript was in preparation, Hasty et al reported the isolation of a 2.2-kb cDNA clone for HNC. The sequence reported here differs by only two nucleotides in the coding sequence of the protein. The clone reported here otherwise differs only in the length of the 3′ untranslated region of the cDNA.

**Predicted protein sequence of HNC and comparison with HFC.** A comparison of the sequences of prepro-HNC and prepro-HFC shows that they are homologous collagenases with many similar features (Fig 2). The two preproenzymes consist of 467 and 468 residues with predicted molecular weights of 53,443 and 53,860, respectively. Alignment of the two sequences shows 58% identity and 73% chemical similarity between the two enzymes.

Both collagenase zymogens are known to activate autolytically in response to various stimuli and also to undergo an autolytic degradation that results in loss of collagenase activity. In an earlier study, the sequence of a 35-
DNA made from CGL RNA.

Fig. 1. Nucleotide sequence of the DNA for HNC. Brackets indicate PCR obtained by PCR.
Fig 2. Comparison of the amino acid sequences of HNC and HFC predicted from cDNA clones. Sections of the HNC sequence determined by amino acid sequencing are shown above the predicted sequence. The alignment of the two collagenase sequences necessitates gaps and these residue numbers for each enzyme are shown at the right. Potential N-linked glycosylation sites are indicated with an asterisk, while vertical arrows indicate the (a) signal peptidase, (b) autolytic activation, and (c) autolytic degradation processing sites. The cysteine-switch and putative zinc binding regions are underlined.

Residue region of HNC near the border between the propeptide and catalytic domains was determined. This sequence, shown above residues 85 to 119 of the cDNA-derived sequence in Fig 2, is confirmed in the sequence of the cloned gene. The N-terminus of autolytically activated pro-HNC is indicating that activation occurs by hydrolysis of the M-L bond. Pro-HFC activated by hydrolysis of one of the three bonds in the Q-F-V sequence. These autolytic activation loci mark the border between the propeptide and catalytic domains of these collagenases.

The first 15 residues of the 27-Kd autolytic degradation fragment of pro-HNC have been previously sequenced. That sequence is shown above the cDNA-derived sequence of HNC in Fig 2, where the identified residues match completely with residues 263 to 277. Thus, the autolytic degradation of pro-HNC arises via hydrolysis of the G-L bond. By comparison, the analogous event in pro-HFC occurs downstream at the P-L bond. This degradation site marks the approximate border between the catalytic and hemopexin-like domains. Thus, HFC and HNC have the same domain structure and similar, but not identical, proteolytic processing sites.

The basis for latency in all of the MMP zymogens is believed to be a complex between a “cysteine-switch” residue in the propeptide domain and the zinc atom in the active site. Pro-HNC has the requisite cysteine-switch residue that lies in a characteristic and highly conserved region of amino acids that corresponds to residues 89 to 97 that are underlined in Fig 2. It also has the necessary zinc binding site that conforms to the VIY-A-A-H-E-LIIIF-G-H-SIA-LIM-G consensus sequence, at residues 214 to 224 (underlined in Fig 2).

One of the major differences between HNC and HFC is the carbohydrate processing of the protein chains. HNC is thought to be glycosylated at three sites. By contrast, HFC is often unglycosylated, and when glycosylated, it is probably at a single site. Reflective of this difference, pro-HNC contains six potential N-linked glycosylation sites, while pro-HFC has only two.

Analysis of in vitro translation products from cDNA-directed RNA. In vitro translation of RNA synthesized from the protein chain.
cDNA yielded a major protein product of approximately 52 Kd with a minor product at approximately 43 Kd, both of which were immunoprecipitated with rabbit anti-HNC antibody (Fig 3A). A third band of intermediate size was seen in the reaction containing no RNA template, and is not precipitated with antibody; this probably results from translation of residual RNA in the reticulocyte lysate. The 52-Kd product corresponds to the predicted size for the full-length unglycosylated HNC protein; the smaller band is of a size appropriate for the autocatalytically activated enzyme. Minor products of slightly higher molecular weight may correspond to partially glycosylated products, which may occur at low frequency in in vitro translation reactions. These results were confirmed by Western blot analysis, which identified a broad band corresponding to proteins of 52 Kd and slightly larger; this again suggests that the in vitro translated protein underwent partial glycosylation (Fig 3B).

The products of the in vitro translation were shown to have full collagenolytic activity (Fig 4). A 1:800 dilution of the product of translation in the reticulocyte lysate fully degraded radiolabeled collagen within 5 hours.

**Northern blot analysis of HNC expression.** The full-length cDNA was used as a probe of total RNA from a patient with CGL, and yielded a single mRNA band corresponding to a molecular weight of approximately 3.4 kb. Because the 5' end of the cDNA clone did not encode a stop codon preceding the initiator methionine, we used PCR to determine whether the discrepancy in the size of the cDNA and the mRNA species detected in CGL represented additional 5' untranslated sequences. Sub-cloned PCR fragments from the reaction using the 5' antisense oligo from HNC and tailing with oligo-dG extended the 5' sequence 30 bases and introduced a stop codon 5' of the initiator methionine (depicted in brackets in Fig 1). This result suggests that the remainder of the mRNA seen on Northern is composed of additional 3' untranslated sequences.

We have also examined HNC expression in total RNA from uninduced HL60 cells and HL60 cells induced with DMSO and TPA. As has been previously noted for lactoferrin and transcobalamin I,2 HU0 cells do not express HNC mRNA at any stage of maturation (Fig 5).

**DISCUSSION**

We report the cloning of a full length cDNA for human neutrophil collagenase. The identified clone contains a 2.4-kb insert that includes the entire coding region for the
HNC protein as well as both 5' and 3' untranslated regions. The mRNA species that hybridizes to the HNC probe, however, is approximately 3.4 kb. The large difference in size between the cDNA and its corresponding mRNA species suggests that the mRNA has a very long untranslated region at either the 3' or the 5' end. We used the PCR to determine the 5' extent of the mRNA, and have confirmed that the majority of the untranslated region lies at the 3' end of the gene. The functional significance of this long 3' untranslated region is unclear.

We have compared the sequence and structural features of HNC with that of HFC and have shown strong similarities between the two proteins in areas of functional importance, including the cysteine-switch region, zinc binding site, and autocatalytic sites. One relatively surprising finding is the strong homology between the signal peptides for the two enzymes. One might have predicted a greater difference in this portion of the protein, because the neutrophil enzyme is stored within granules and the fibroblast enzyme is constitutively secreted.

One important difference between HNC and HFC is the degree of carbohydrate processing undergone by the two enzymes. The fibroblast enzyme is unglycosylated or minimally glycosylated, whereas the neutrophil enzyme is usually glycosylated at at least three sites. The reason for this difference in carbohydrate processing is unknown. It may be related to the fact that HFC is secreted while HNC is targeted to specific granules for intracellular storage. Our data showing abundant activity in protein synthesized in vitro in a cell-free system confirms that the glycosylation is not necessary for collagenase activity. These studies also confirm that protein synthesized in a cell-free system probably folds properly, because its function is preserved. This finding suggests that in vitro translation should be a viable system in which to study the structure-function characteristics of mutagenized HNC proteins.

HNC, by virtue of its localization within the neutrophil secondary granule, is also of interest as a marker of granulocyte differentiation. To evaluate its expression in this context, we have used the probe to study mRNA from CGL cells and the human HL60 leukemic cell line. As previously noted, HL60 cells can be induced to limited granulocytic maturation in response to DMSO, however, they do not acquire secondary granules and do not produce detectable amounts of secondary granule content proteins. Previous studies have shown that HL60 cells do not express lactoferrin or transcobalamin at any phase of induction. The results reported here show that collagenase mRNA, like that of the other two secondary granule proteins, is present in CGL cells, but not in uninduced or induced HL60 cells. This finding is further evidence for the hypothesis that neutrophil secondary granule protein gene expression is coordinately controlled at the level of mRNA transcription. Furthermore, the absence of expression of all of these genes in HL60 cells induced to undergo phenotypic maturation to granulocytes suggests that part of the malignant phenotype expressed by HL60 cells may include a defect in a common regulatory factor controlling expression of all of the secondary granule protein genes.

ACKNOWLEDGMENT

The authors are indebted to Josan Chung and Marianna Diószegi for expert technical assistance.

REFERENCES


40. Birkedal-Hansen B, Moore WG, Taylor RE, Bliow AS, Birkedal-Hansen H: Monoclonal antibodies to human fibroblast...
procollagenase: Inhibition of enzymatic activity, affinity purification of the enzyme, and evidence for clustering of epitopes in the NH2-terminal end of the activated enzyme. Biochemistry 27:6751, 1988


Structure and expression of the cDNA encoding human neutrophil collagenase

P Devarajan, K Mookhtiar, H Van Wart and N Berliner