Cloning of cDNA for Proteinase 3: A Serine Protease, Antibiotic, and Autoantigen from Human Neutrophils

By David Campanelli, Maxine Melchior, Yiping Fu, Munehiro Nakata, Howard Shuman,* Carl Nathan, and Joelle E. Gabay

From the Beatrice and Samuel A. Seaver Laboratory, Division of Hematology-Oncology, Department of Medicine, Cornell University Medical College, New York, New York 10021; and the *Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York, New York 10032

Summary

Closely similar but nonidentical NH2-terminal amino acid sequences have been reported for a protein or proteins in human neutrophils whose bioactivities is/are diverse (as a serine protease, antibiotic, and Wegener’s granulomatosis autoantigen) but that share(s) several features: localization in the azurophil granules, a molecular mass of ~29 kD, reactivity with diisopropylfluorophosphate, and the ability to degrade elastin. We previously purified one such entity, termed p29b. Using a monospecific antibody, we have cloned from human bone marrow a cDNA encoding the complete p29b protein in its mature form, along with pre- and pro-sequences. The predicted amino acid sequence agrees closely with the NH2-terminalsequence obtained previously from purified p29b, as well as with sequences newly obtained from CNBr fragments. The primary structure is highly homologous to elastase, cathepsin G, T cell granzymes, and other serine proteases, and shares both the catalytic triad and substrate binding pocket of elastase. Hybridization of the full-length cDNA with restriction enzyme digests of human genomic DNA revealed only one fragment. This suggests that the closely related species described previously are the same, and can be subsumed by the term used for the first-described activity, proteinase 3. Proteinase 3 is more abundant in neutrophils than elastase and has a similar proteolytic profile and specific activity. Thus, proteinase 3 may share the role previously attributed to neutrophil elastase in tissue damage, and has the potential to function as an antimicrobial agent.

1 Abbreviations used in this paper: DFP, diisopropylfluorophosphate; IPTG, isopropyl-β-D-thiogalactoside; PR-3, proteinase 3.
oxygenucleotide construct to HL-60 cells induced their myeloid differentiation (14). Very recently, a preliminary report by Jenne et al. (15) has revised the NH$_2$-terminal sequence for the Wegener's autoantigen (12), such that it now matches all 20 NH$_2$-terminal residues reported for p29b (3), except for Gln$^{19}$ in place of Glu$^{19}$. Based on PCR analysis, Jenne et al. (15) also proposed a revision of the 5' region of the myeloblastin cDNA, according to which the deduced NH$_2$-terminal 20 residues of myeloblastin would likewise match those of the Wegener's autoantigen and p29b.

We have used a monospecific anti-p29b IgG to clone a cDNA that encodes the entire mature protein. Below, we provide the sequence of this clone, compare its deduced amino acid sequence with the observed sequence of proteolytic fragments of p29b, note the extensive homology of this molecule to cathepsin G, elastase, and the granymes of T cells, and demonstrate that only one human genomic DNA fragment appears to hybridize with the p29b cDNA. Thus, it is likely that the cDNA presented here encodes a single PMN azurophil granule protein that has the potential for elastolysis, microbial killing, and the regulation of myeloid differentiation, and serves as an autoantigen.

### Materials and Methods

**Amino Acid Sequence of CNBr Fragments.** 50 µg of purified p29b (3, 4) was incubated with a 500-fold excess (with respect to methionine content) of CNBr in 0.1 N HCl for 48 h at room temperature in the dark. The samples were subjected to SDS-PAGE. Resolved peptides were transferred to polyvinylidene difluoride membranes (Millipore Continental Water Systems, Bedford, MA) (16) and sequenced directly by automated Edman degradation on a gas-phase sequenator (470; Applied Biosystems, Inc., Foster City, CA).

**cDNA Cloning.** A Agt11 bone human marrow cDNA library (Clontech, Palo Alto, CA) was plated on Escherichia coli Y1090 (17). Nitrocellulose filter lift saturated with isopropyl-β-D-thiogalactoside (IPTG; Sigma Chemical Co., St. Louis, MO) were dried, blocked with 5% nonfat dry milk and 0.05% Tween 20 in PBS, rinsed, probed with a 1:500 dilution of monospecific rabbit anti-p29b IgG (4) that had been preabsorbed with E. coli Y1090, rinsed, and treated with goat anti-rabbit IgG conjugated with alkaline phosphatase (Boehringer Mannheim Biochemicals). Plates positive by a colorimetric phosphatase reaction were purified to homogeneity, as monitored by rescreening with the antibodies. Plate cultures were lysed in chloroform, clarified by centrifugation, and rid of bacterial DNA with DNase (Pharmacia Fine Chemicals, Piscataway, NJ). Phage were collected by ultracentrifugation (89,500 g, 2 h, 4°C), purified on spin columns of DEAE cellulose (3-Prime 3-Prime, Westchester, PA) (18), and the DNA was isolated according to the manufacturer's instructions. The cDNA insert was recovered by restriction with EcoRI (Boehringer Mannheim Biochemicals) and electrophoresis in low melt/low gelling media in 5%agarose (Bio-Rad Laboratories, Richmond, CA) as visualized with long-wavelength UV light. The appropriate gel slice was melted and extracted with phenol and ether, and the DNA was precipitated with ethanol. pBluescript SK+ (Stratagene, La Jolla, CA) was digested with EcoRI and dephosphorylated (calf intestinal alkaline phosphatase; Boehringer Mannheim Biochemicals) for insertion of the p29b cDNA with T4 DNA ligase (Stratagene). The resultant plasmid was used to transform competent E. coli XL1 Blue (Stratagene) (19). Recombinants were screened by blue/white color (β-galactosidase) selection on luria broth agar with ampicillin, 20 mM IPTG, and 80 µg/ml 5-bromo-4-chloro-3-indoly-β-D-galactoside. After transformation of DH5 cells (20), the recombinant plasmids were amplified and purified (21). cDNA was sequenced by the dideoxynucleotide chain termination method using double-stranded plasmid DNA as a template (22) with a kit from U.S. Biochemical Corp. (Cleveland, OH) and deoxyadenosine 5'-[32P]Triphosphate (New England Nuclear, Boston, MA). After initial use of universal primers, sequencing was continued independently on both strands with sequentially constructed 20-25-mer oligonucleotides so as to obtain sequences in regions overlapping by 20-40 bp. Results were analyzed by DNASIS and PROSIS software (Hitachi America, Ltd., San Bruno, CA). Alignment to other proteases was scored according to Lipman and Pearson (23) with the program FASTP. Structural predictions were obtained as described (24).

**Southern Hybridization.** Human placental genomic DNA (10 µg; Clontech) was digested to completion with EcoRI, HindIII, BamHI, PstI, or BglII (Boehringer Mannheim Biochemicals). The digests were transferred onto a nylon membrane (Schleicher & Schuell, Inc., Keene, NH) and hybridized with p29b cDNA [32P]labeled by random priming (25). Prehybridization (2 h in 5 x Denhardt's solution) and hybridization (18 h in 10% dextran sulfate) were performed at 42°C (high stringency) or 32°C (low stringency) in 50% (vol/vol) formamide in 6x SSC/0.5% SDS with 100 µg/ml sonicated salmon sperm DNA. Filters were washed twice for 15 min at room temperature with 6 x SSC/0.5% SDS, twice at 37°C and once at 65°C with 1 x SSC/0.5% SDS (high stringency), or twice at room temperature with 2 x SSC/0.1% SDS and twice at 37°C with 0.25 x SSC/0.1% SDS (low stringency).

**Results**

**Cloning.** A human bone marrow cDNA expression library in λgt11 was screened with monospecific anti-p29b IgG (4). One of the positive clones, containing an insert of 1,014 bp, was purified, subcloned in pBluescript SK+ , and sequenced (Fig. 1). The open reading frame of 762 bp encodes a 254-amino acid polypeptide, including a 26-residue NH$_2$-terminal peptide and a 228-residue mature protein. The latter corresponds to p29b, based on the following evidence. First, the deduced amino acid sequence for residues 1-20 (numbering for the mature protein) matches exactly the sequence obtained from the NH$_2$ terminus of the purified protein (Glu$^{19}$ was originally observed [3], but Gln$^{19}$ was detected in the CNBr peptide sequenced in the present study and conforms to the cDNA). Second, two peptides from CNBr cleavage of purified p29b provided the sequence of 55 residues, and a [3H]DFP-binding tryptic fragment of p29b reported by Wilde et al. (26) provided another 22 residues; of these 77 residues, 70 were identical to the deduced sequence. Third, the deduced amino acid composition corresponded closely to that determined for purified p29b (not shown).

**Features of the Predicted Structure.** The 5' end of the cDNA encodes a hydrophobic stretch (residues -22 to -9) (Fig. 2) typical of a signal sequence followed by a consensus signal peptidase cleavage site (Ala-X-Ala) (27) at -5 to -3 (Fig. 1). A 24-residue pre-signal peptide (Arg$^{-16}$ to Ala$^{-1}$) 50% homologous to the pre-sequence in the elastase cDNA (28) is followed by a two-residue pro-peptide (Ala$^{-2}$,Glu$^{-1}$), as
in other leukocyte lysosomal proteases (28-30). However, the cDNA insert lacks an ATG translation initiation codon. The 3' end of the cDNA contains a TGA termination codon, as well as a polyadenylation signal (AAATAAA) and a poly(A) tail, suggesting that the clone encompasses the full-length sequence for the mature protein.

Mature p29b is predicted to consist of 228 amino acids with a Mr of 25,000 for the polypeptide backbone. Two potential N-linked glycosylation sites (Asn-X-Ser/Thr) are present, and may account for the apparent isoforms (3, 4).

Eight Cys residues (five clustered in the COOH-terminal third of the molecule) suggest that there may be four intramolecular disulfide bonds, as in elastase (32). His^4, Asp^41, and Ser^176 correspond to the active site triad typical of serine proteases (Fig. 1).

After a hydrophobic region (−22 to −9) corresponding to the presumed signal sequence, the NH2-terminal half of p29b contains six regions averaging 17 residues each, in which there is a marked, periodic alternation of hydrophilicity and hydrophobicity (Fig. 2). The mature protein is predicted by the Chou-Fasman algorithm (33) to contain three short stretches of α-helix totalling 8.3%, and 10 β-sheets totalling 40%. Remarkably, the β-sheets are all predominantly hydrophobic. With 27 basic and 15 acidic amino acids in the mature protein (Fig. 2, bottom), p29b has an overall calculated pl of 7.9. Regions of the molecule predicted to be surface exposed (34) correspond to residues 7-12, 63-72, 83-88, 113-118, 165-168; these 32 residues include four acidic and eight basic amino acids.

Fig. 3 compares the p29b sequence with that of other serine proteases, aligned to maximize homology. p29b exhibits the highest homology with elastase (54%). There is also considerable homology with the T cell enzymes human lymphocyte protease (36%) and granzyme B (33%), and with cathepsin G (35%), rat mast cell protease II (34%), and chymotrypsin (30%). Residues 9-16 of p29b align with the Cys residues of elastase; five align with corresponding residues in the other proteases listed. One of the potential N-linked glycosylation sites in p29b (Asn^70) aligns with a corresponding site in elastase; both enzymes have two such sites. The members of the catalytic triad (His, Asp, Ser) are located at homologous positions in all the aligned sequences, flanked by well-conserved peptides. The residues at −6, 15-17, and 28 relative to the active site serine are thought to determine the

Figure 1. Nucleotide and deduced amino acid sequences of p29b (PR-3). The nucleotide sequence is numbered from the first base of the cDNA insert after the EcoR1 linker site. The deduced amino acid sequence (single-letter code) is numbered from the NH2-terminal residue of the mature protein (+1) (3). Regions for which amino acid sequence have been determined directly are underlined. These include residues 1-20, as determined from amino acid sequencing): 29, C for I; 31, G for P; 35, H for L/V; 37, S for A/V; 44, H for I; 130, W for P; 147, N for R.
tein termed myeloblastin was also evident from its cDNA terminal sequences. High homology with the predicted pro-
Discussion

their subcellular localization, relative abundance, molecular mass, enzymatic activity, ability to bind DFP, and NH2-terminal sequences. High homology with the predicted protein termed myeloblastin was also evident from its cDNA sequence. However, apparent discrepancies in their partial amino acid sequences have left unclear the precise interrelat-
edness of these entities, which were isolated by independent techniques according to distinct bioactivities. By obtaining what to our knowledge is the first cDNA sequence encoding the complete mature form of any one of these species, we have been able to perform Southern blotting experiments that strongly suggest the existence of a single hybridizing gene, even at low stringency. It is highly likely that PR-3, p29b, the predominant cytoplasmic autoantigen of Wegener's granulomatosis, and myeloblastin are all encoded by this gene. We propose that the gene product be called proteinase 3 (PR-3), a term with historical precedence (7).

The predicted amino acid sequence of PR-3 differs in 7 of 92 residues from the sequences determined from the purified protein or its fragments. Six of the seven discrepant positions are highly conserved among serine proteases; in each case, the conserved amino acid is that predicted by the cDNA. The discrepancies may arise from ambiguities in amino acid sequencing. Alternatively, PR-3 may be polymorphic. The nonconservative nature of the substitutions militates against this; nonetheless, we plan to test for polymorphism by PCR techniques. Finally, we cannot exclude the remote possibility that an additional gene(s) is linked in tandem to PR3 without intervening sites for the restriction enzymes used in the Southern blotting experiments. Genomic cloning will resolve this issue.

The PR-3 cDNA encodes a precursor with signal (pre) and propeptides typical of serine proteases. The hydrophobic prepeptide of 24 or more residues (the cDNA may be incomplete at the 5' end) may target PR-3 to the endoplasmic retic-
Figure 4. Southern blot of human genomic DNA with p29b (PR-3) cDNA as a probe. Human placental genomic DNA was digested to completion with five restriction enzymes and subjected to electrophoresis in 0.7% agarose. (A) The gel was stained with ethidium bromide and photographed under short-wavelength UV light to show the extent of digestion of the genomic DNA and the migration of size markers (left lane) derived from phage λ DNA digested with HindIII (23, 9.6, 6.6, 4.4, 2.3, 2.0 kb). (B) The digests were transferred to a nylon membrane, hybridized with [32P]-labeled p29b (PR-3) cDNA under high stringency, and washed as described in Materials and Methods, then autoradiographed with an intensifying screen for 4 d at -70°C.

We thank A. Sievertsen and A. Marra for help and suggestions, and D. Romeo for communicating results before publication.

This work was supported by National Institutes of Health grants AI-23807, CA-45218, and CA-43610.

Address correspondence to J. Gabay, Box 57, Cornell University Medical College, 1300 York Avenue, New York, NY 10021.

Received for publication 18 September 1990.

References
1. Odeberg, H., and I. Olsson. 1975. Antibacterial activity of cationic proteins from human granulocytes. J. Clin. Invest. 56:1118.

2. Elsbach, P., and J. Weiss. 1988. Phagocytic cells: oxygen-independent antimicrobial systems. In Inflammation: Basic Principles and Clinical Correlates. J.I. Gallin, I.M. Goldstein, and
Cloning of Proteinase 3

Using mini-prep plasmid DNA for sequencing double-stranded templates with Sequenase™. Biotechniques. 6:544.

22. Chen, E.Y., and P.H. Seeburg. 1985. Supercoiling: a fast and simple method for sequencing plasmid DNA. DNA (NY). 4:165.

23. Lipman, D.J., and W.R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science (Wash. DC). 22:1435.

24. Devereux, J., P. Haebeli, and O. Smithies. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387.

25. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Second edition, vol. 3. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. A1-A13.

26. Wilde, C.G., J.L. Snable, J.E. Griffith, and R.W. Scott. 1990. Characterization of two azurophil granule proteases with active-site homology to neutrophil elastase. J. Biol. Chem. 265:2038.

27. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. 14:4683.

28. Takahashi, H., T. Nukiwa, K. Yoshimura, C.D. Quick, D.J. States, M.D. Holmes, J. Whang-Peng, T. Kutz, and R.G. Crystal. 1988. Structure of the human neutrophil elastase gene. J. Biol. Chem. 263:14739.

29. Salvesen, G., D. Farley, J. Shuman, A. Przybyla, C. Reilly, and J. Travis. 1987. Molecular cloning of human cathepsin G: Structural similarity to mast cell and cytotoxic T lymphocyte proteinases. Biochemistry. 26:2289.

30. Benfey, P.N., F.H. Yin, and P. Leder. 1987. Cloning of the mast cell proteast, RMCP II: evidence for cell-specific expression and a multi-gene family. J. Biol. Chem. 262:5377.

31. Dayhoff, M.O., W.C. Barker, and L.T. Hunt. 1983. Establishing homologies in protein sequences. Methods Enzymol. 91:524.

32. Navia, M.A., B.M. McKeever, J.P. Springer, T.Y. Lin, H.R. Williams, E.M. Fluder, C.P. Dorn, and K. Hoogstein. 1989. Structure of human neutrophil elastase in complex with a peptide chloromethyl ketone inhibitor at 1.84-A resolution. Proc. Natl. Acad. Sci. USA. 86:7.

33. Chou, P.Y., and G.D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. 47:45.

34. Emin, E.A., J.V. Hughes, D.S. Perlow, and J. Boger. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol. 55:836.

35. Jenne, D.E., C. Ray, J.-A. Haeberli, B.-Y. Chao, P. Gross, and J. Tschopp. 1988. Identification and sequencing of cDNA clones encoding the granule-associated serine proteases granzymes D, E, and F of cytolytic T lymphocytes. Proc. Natl. Acad. Sci. USA. 85:4814.

36. Kraut, J. 1977. Serine proteases: structure and mechanism of catalysis. Annu. Rev. Biochem. 46:331.

37. Frank, R.W., R. Gennaro, K. Schneider, M. Przybylski, and D. Romeo. 1990. Amino acid sequences of two proline-rich bacteriocins. Antimicrobial peptides of bovine neutrophils. J. Biol. Chem. In press.

38. Bangalore, N., J. Travis, V.C. Onunka, J. Pohl, and W.M. Shafer. 1990. Identification of the primary antimicrobial domains in human neutrophil cathepsin G. J. Biol. Chem. 265:13584.

39. Chung, L.P., S. Keshav, and S. Gordon. 1988. Cloning the human lysozyme cDNA: Inverted Alu repeat in the mRNA and in situ hybridization for macrophages and Paneth cells. Proc. Natl. Acad. Sci. USA. 85:6227.

40. Ganz, T., M.E. Selsted, D. Szklarek, S.L. Harwig, K. Daher, D.F. Bainton, and R.I. Lehrer. 1985. Defensins. Natural pep-
tide antibiotics of human neutrophils. J. Clin. Invest. 76:1427.
41. Gray, P.W., G. Flaggs, S.R. Leong, R.J. Gumina, J. Weiss, C.E. Ooi, and P. Elsbach. 1989. Cloning of the cDNA of a human neutrophil bactericidal protein. J. Biol. Chem. 264:9505.
42. Kyte, J., and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105.
43. Lobe, C.G., B.B. Finlay, W. Paranchych, V.H. Paetkau, and R.C. Bleackley. 1986. Novel serine proteases encoded by two cytotoxic T lymphocyte-specific genes. Science (Wash. DC). 232:858.
44. Schmid, J., and C. Weissmann. 1987. Induction of mRNA for a serine protease and a β-thromboglobulin-like protein in mitogen-stimulated human leukocytes. J. Immunol. 139:250.
45. Dayhoff, M.O. 1978. Atlas of Protein Sequences and Structure. Vol. 5, suppl. 3. National Biomedical Research Foundation. Washington, D.C. 73–93.