*Proc. Natl. Acad. Sci. USA*
Vol. 84, pp. 2228–2232, April 1987
Biochemistry

# Primary structure of human neutrophil elastase

S. SINHA*†, W. WATOREK*‡, S. KARR*, J. GILES*, W. BODE§, AND JAMES TRAVIS*¶

*Department of Biochemistry, University of Georgia, Athens, GA 30602; †Department of Pharmacology, State University of New York, Stony Brook, NY 11794; and §Max-Planck-Institut fur Biochemie, Martinsreid, Federal Republic of Germany

*Communicated by Clarence A. Ryan, December 23, 1986 (received for review October 1, 1986)*

**ABSTRACT** **The complete amino acid sequence of human neutrophil elastase has been determined. The protein consists of 218 amino acid residues, contains two asparagine-linked carbohydrate side chains, and is joined together by four disulfide bonds. Comparison of the sequence to other serine proteinases indicates only moderate homology with porcine pancreatic elastase (43.0%) or neutrophil cathepsin G (37.2%). In particular, many of the residues suggested to play important roles in the mechanism by which the pancreatic elastase functions are significantly changed in the neutrophil enzyme, indicating alternative types of binding with the human proteinase.**

Human neutrophils contain a battery of hydrolytic enzymes that normally are utilized for the degradation of foreign materials ingested as part of the phagocytic process (1). Although it is not completely clear whether such enzymes are also involved in tissue remodeling, this has been suggested as an alternate function (2). It is known that both neutrophil turnover and phagocytosis, themselves, do result in the leakage of enzymes into the extracellular milieu where they may cause extensive damage to connective tissue unless checked by controlling inhibitors (2). Indeed, a popular theory suggests that the development of pulmonary emphysema occurs as a result of insufficient levels of proteinase inhibitors (either locally or plasma-derived) whose primary functions are to inhibit these neutrophil enzymes (3). In particular, there is strong evidence that implicates neutrophil elastase as the proteinase most directly involved in abnormal lung connective tissue turnover. This is based on the fact that individuals devoid in the major controlling inhibitor of this enzyme, plasma $\alpha_1$-proteinase inhibitor, tend to develop obstructive lung disease (familial emphysema) much earlier than those with normal inhibitor levels (4).

We have been interested in the potential function of both neutrophil elastase and cathepsin G in protein turnover. Both enzymes are capable of degrading a wide variety of substrates, including elastin, collagen, and proteoglycan, although it appears that elastase is more efficient in the turnover of such macromolecules (5, 6). Therefore, determination of their primary structures should be of great value not only for understanding the mechanism by which each functions but also for developing specific inhibitors that might be useful in aiding in the control of their activities outside of the cell. This report describes the determination by conventional protein sequencing strategies of the primary structure of neutrophil elastase, a very basic glycoprotein that exists in a series of isoenzyme forms (7, 8). In a separate report (9), a description of the primary structure of cathepsin G has been made by using a combination of recombinant DNA and protein-sequencing technology.

## MATERIALS AND METHODS

**Materials.** Human neutrophil elastase was prepared from the granules of both normal and myeloid leukemia cells by affinity chromatography on Trasylol-Sepharose as described (10). Individual isoenzymes were separated by ion-exchange chromatography (11), and the major component, referred to as $E_4$, was used in most of the studies described below. However, in some experiments mixtures of all four of the isoenzyme forms of elastase were used because early studies (10) had indicated differences in carbohydrate content rather than in amino acid composition. This was verified by sequence analysis of the reduced, carboxamidomethylated protein mixture through 40 Edman cycles without any indication of microheterogeneity in the sequences obtained (11). Indeed, no evidence of microheterogeneity in any of the peptides analyzed could be found. It should also be pointed out that there were no obvious differences in protein structure between elastases isolated from normal or myeloid leukemia cells.

**Methods.** The primary structure of human neutrophil elastase was determined by using the reduced, carboxamidomethylated protein as starting material (12). The derivatized protein (generally 1.0 μmol) was subjected to separate digestions with (*i*) cyanogen bromide (in 70% formic acid at room temperature in the absence of light for 18 hr), (*ii*) 70% formic acid at 37°C for 72 hr, or (*iii*) 4.0 M iodosobenzoic acid in guanidine·hydrochloride/80% acetic acid at room temperature in the absence of light for 18 hr (13) to obtain fragments of reasonable size for either direct sequencing or for further degradation with either L-1-tosylamido-2-phenylethyl chloromethyl ketone-treated porcine trypsin or L-1-tosylamido-2-lysyl chloromethyl ketone-treated bovine α-chymotrypsin (1:50 enzyme/substrate molar ratio at pH 8.0 and 37°C for 5 hr). Reduced, aminoethylated protein (14) was treated with endoproteinase Lys-C (Boehringer-Mannheim) (15) to obtain peptides derived by cleavage at sites after the modified cysteine residues. Digests were separated by gel filtration on Sephadex G-50 and/or reversed-phase HPLC with a trifluoroacetic acid/water/acetonitrile gradient, with the large fragments being analyzed for purity by $NaDodSO_4$/polyacrylamide gel electrophoresis (16). All peptides were analyzed with a Beckman model 119CL amino acid analyzer, and their sequences were determined either partially or totally by automatic Edman degradation using either the Beckman model 890C sequencer or the Applied Biosystems model 470 gas-phase sequencer, both with programs supplied by the manufacturers. Phenylthiohydantoin derivatives of amino acids were identified either by on-line analysis with the 120A phenylthiohydantoin analyzer (Applied Biosystems, Foster City, CA) or by manual conversion and identification on a Waters HPLC system (17). Glycopeptides were detected by

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

‡Permanent address: Institute of Biochemistry, Wroclaw University, Wroclaw, Poland.
¶To whom reprint requests should be addressed.

analysis of individual fragments for neutral sugar (18), hexosamine (19), or sialic acid content (20).

## RESULTS

The initial strategy planned for the determination of the primary amino acid structure of neutrophil elastase involved cleavage with cyanogen bromide after the three methionine residues, followed by sequencing of the individual fragments. Unfortunately, two of the three methionine residues were found to be present at the beginning of the protein (residues 15 and 37, peptides C-1 and C-2), while cleavage after the third methionine (residue 125) gave only low yields (peptide C-4) and anomalous cleavage after Trp-127 (peptide C-4'). Nevertheless, we were able to sequence a large cyanogen bromide fragment (C-3) through an Asp-Pro sequence at positions 81 and 82 that suggested the design of specific experiments to cleave the reduced carboxamidomethylated protein at this position. Such a digestion yielded two fragments in high yield (70% recovery), and we were able to use these fragments (A-1 and A-2) as well as the cyanogen bromide-derived peptides for determination of nearly all of the primary structure of elastase. Several tryptic, chymotryptic, and endoproteinase Lys-C-derived peptides were analyzed as well as fragments obtained from iodosobenzoic acid digestion. However, for clarity, only those fragments obtained that yielded substantial information as to the primary structure of neutrophil elastase (Fig. 1) or those which gave overlapping peptide sequences are discussed.

**Residues 1–50.** Amino-terminal sequence of the reduced carboxamidomethylated protein and that of peptides C-1,

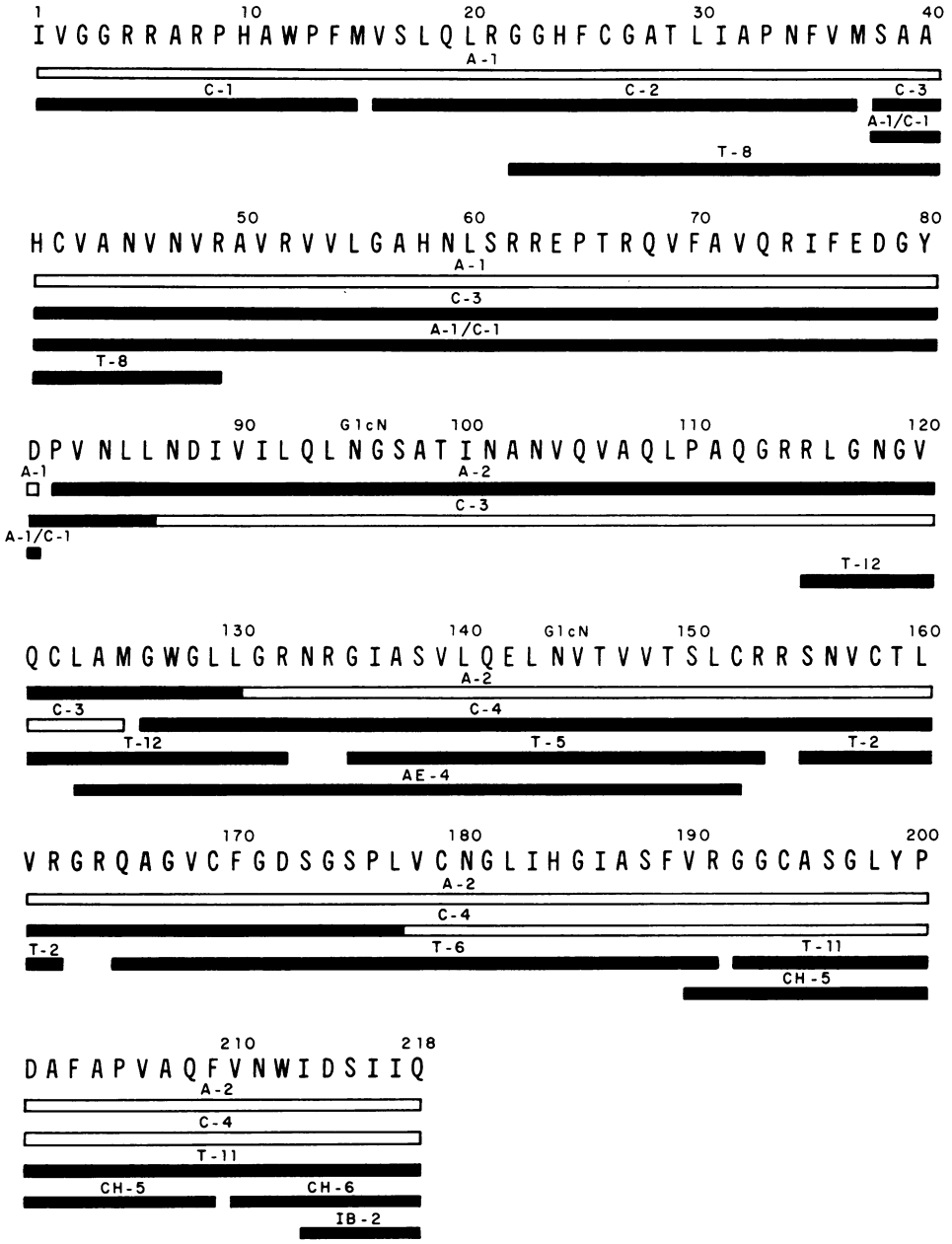


FIG. 1. Amino acid sequence analysis of human neutrophil elastase. Peptide fragments used for structural analysis are shown by bars and letters, designated by the cleavage method used as follows: A, cleavage by formic acid; C, cleavage by cyanogen bromide; IB, cleavage by iodosobenzoic acid; C, chymotrypsin; T, trypsin; and AE, endoproteinase Lys C digestion of reduced, aminoethylated protein. Sequences established by automated Edman degradation of peptides are shown with filled bars. GlcN, positions of glycosylation. The COOH-terminal sequence Iln-Gln was confirmed by the order of release of glutamine and isoleucine by carboxypeptidase A. The recommended one-letter notation for amino acids (21) is used.

C-2, and C-3 gave data that unambiguously positioned these residues. Previously, we had identified (11) a glycine residue at position 38, and this was reexamined and found to be serine as reported by others (22).

**Residues 51–86.** Cleavage of peptide A-1 with cyanogen bromide gave three fragments as expected. One of these, A-1/C-1, was sequenced to completion through 44 cycles with ready identification of Asp-18 in the last cycle. Analysis of peptide C-3 through 49 Edman cycles yielded data that confirmed the structure of peptide A-1/C-1 and also overlapped the Asp-Pro sequence at positions 81 and 82 for an additional five residues to Leu-86.

**Residues 87–130.** Sequence analysis of peptide A-2 through 49 residues were readily obtained, and the position of this peptide in the protein was confirmed by its identity with the last five residues that we were able to identify during the sequencing of peptide C-3. The data obtained was unambiguous, with the exception of residue 95, where no major residue could be identified by HPLC. Analysis of both peptides C-3 and A-2 had indicated that these were glycopeptides, suggesting that a carbohydrate-linked asparagine residue was in this position. This was established in a more conclusive manner by x-ray crystallographic analysis (23).

The last few residues that we were able to establish for the partial structure of peptide A-2 overlapped with both peptide T-12, which was completely sequenced, and the amino-terminal sequences of peptides C-4 and AE-4, the latter being obtained from an endoproteinase Lys C digest of the reduced, aminoethylated protein. It should be pointed out that endoproteinase Lys C digestion of reduced, carboxamidomethylated elastase was negative, suggesting the absence of lysine residues from the native protein, a fact that was confirmed during total sequence analysis.

**Residues 131–181.** The position of peptide C-4 was readily established by its overlap with the sequence obtained for peptide A-2. C-4 was sequenced through 51 residues; with the exception of the first 2 residues, the sequence of C-4 was identical with that of peptide C-4′, which was found in large quantities in cyanogen bromide digests. Both peptides gave structural data past the putative reactive serine residue to leucine-177, with the exception of a blank residue at position 144. X-ray crystallographic data was again used to establish that an asparagine residue linked to a carbohydrate side chain was present in this position (23). Confirmation of the data obtained in this long sequence run was obtained by determination of the complete structures of the endoproteinase Lys C-derived peptide AE-4 and the tryptic peptides T-12, T-5, T-2, and T-6.

**Residues 182–192.** The structure of peptide T-6 gave the sequence from residue 165 to residue 191 and overlapped with that of the last 13 residues that we were able to obtain for peptide C-4. A chymotryptic digest of peptide C-4′ provided fragments Ch-5 and Ch-6, while a tryptic digest yielded peptide T-11. Fragment Ch-5 overlapped with the last two residues of T-6 and, while this is not rigorous proof, its position can be established by two facts. First, we used a chymotryptic digest, and the position at which we suggest cleavage had occurred would be after Phe-189. Second, x-ray crystal structure analysis indicates that no other amino acids can be inserted into this position to explain the possible presence of other residues (23). Therefore, the position of peptide Ch-5 would appear to be correct.

**Residues 193–218.** The structure of peptide Ch-5 was determined completely and represented residues 193–203. It overlapped with the structure of peptide T-11, which also was sequenced in its entirety and represented the carboxyl-terminal sequence of the protein. This was confirmed by sequence analysis of both peptides Ch-6 and IB-2, each of which contained not only part of the sequence of T-11 but also ended in a glutamine residue. The presence of glutamine and traces of isoleucine at the carboxyl terminus of neutrophil elastase was also noted after extensive treatment of the reduced, carboxymethylated protein with carboxypeptidase A (data not shown).

## DISCUSSION

To date, the term "elastase" has been reserved for those enzymes that are capable of solubilizing peptide fragments from the insoluble, highly cross-linked protein referred to as elastin. In fact, virtually all of the studies reported on this class of serine proteinase have involved an examination of the porcine enzyme (24), with only limited studies on other elastolytic proteinases having been made. Thus, it is highly intriguing to make a comparison of such a well-studied enzyme with that from human neutrophils. In actuality, both are very different types of enzymes, with the porcine enzyme (*i*) being extracellularly secreted as a zymogen, (*ii*) having a specificity primarily toward proteolysis after alanine residues, (*iii*) being sensitive to increased ionic strength, and (*iv*) being synthesized as a single nonglycosylated entity (24). In contrast, human neutrophil elastase is primarily a lysosomal enzyme, being electrostatically bound (presumably through its large number of arginine residues) to an insoluble sulfated polysaccharide matrix. It prefers bulkier side chains in the $P_1$ position, such as valine residues (25), and this is probably the major reason it degrades elastin at a much slower rate than does the porcine enzyme (26), the substrate having a much larger number of alanine vs. valine residues in its amino acid composition (27). Neutrophil elastase also differs from its pancreatic counterpart in that (*i*) it shows increased activity at higher ionic strength (28); (*ii*) it is synthesized as a series of isoenzymes, each containing different amounts of carbohydrate (10); and (*iii*) it is affected by the presence of hydrophobic groups, being inactivated by fatty acids and activated by fatty alcohols (29). Finally, it is intriguing that the esterolytic activity of neutrophil elastase is significantly enhanced when it is bound to the natural proteinase inhibitor $\alpha_2$-macroglobulin (30), whereas the pancreatic enzyme is unaffected. This suggests that the neutrophil enzyme is able to undergo major conformational changes to more easily accommodate low molecular weight substrates.

The sequence analysis of human leukocyte elastase shows that it is a single-chain polypeptide of 218 amino acids with four intramolecular disulfide bonds linking eight half-cystine residues; these residues have been established by x-ray crystallography (23) to be linked as follows: Cys-26 to Cys-42, Cys-122 to Cys-179, Cys-132 to Cys-158, Cys-169 to Cys-194.

There are two positions at which the protein is N-glycosylated (Asn-95 and Asn-144), both being indirectly deduced from the lack of an unidentifiable phenylthiohydantoin-conjugated amino acid at these positions during sequencing of appropriate peptides and independently confirmed from crystallographic data (23). Since we could find no evidence for any microheterogeneity during protein sequence analysis, we conclude that the isoenzyme character of neutrophil elastase preparations detected by acid gel electrophoresis is due to minor differences in carbohydrate content, as originally suggested earlier in this group (10).

In Fig. 2 a comparison is made of the sequence of human neutrophil elastase with porcine pancreatic elastase (31), optimized for maximum homology. In general, the sequence of the leukocyte enzyme parallels that of the porcine enzyme with ≈43% identical residues being noted. However, some important deviations that clearly would reflect major differences in the shapes of the two proteins can readily be noted, including a major 10-amino acid deletion in the "methionine loop" (32) between residues Cys-158 and Lys-169 of the porcine enzyme. Several small deletions relative to pancreatic elastase—between Gly-23 and His-28, Tyr-82 and Asp-86, Gly-176 and Val-180, Cys-194 and Asn-197, and Gly-198

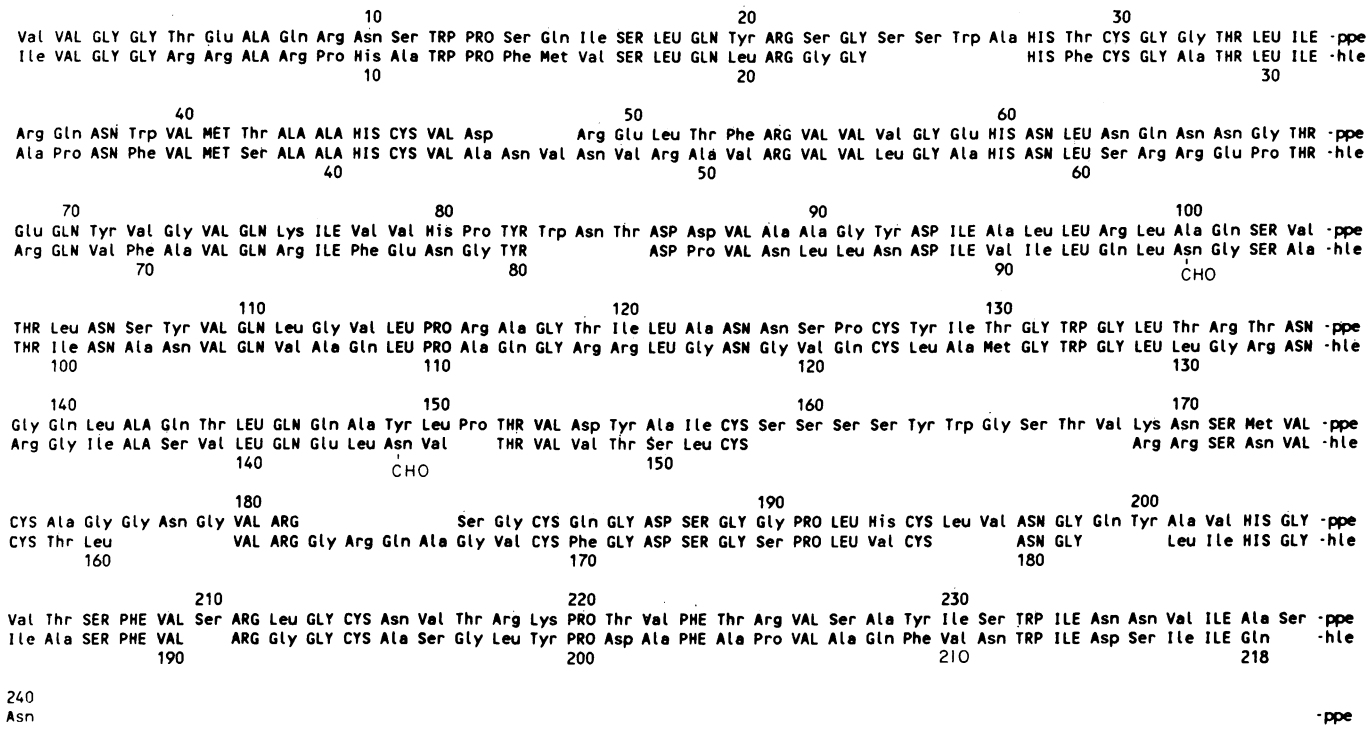


FIG. 2. Comparative amino acid sequence of porcine pancreatic and human neutrophil elastase. Identical residues are capitalized and gaps positioned to give the most realistic homology.

and Ala-201—and two point deletions at the positions of Pro-151 and Ser-210 are also necessary to obtain maximum homology. There are also two insertion positions—a 2-amino acid insertion between Asp-48 and Arg-49 of the porcine enzyme and a 4-amino acid insertion between Arg-181 and Ser-182.

Changes in several important residues in addition to those introduced by the deletions and insertions detailed above would appear to explain some of the differences in the properties of the two enzymes, especially with regard to substrate specificity. For instance, although the key residues believed to be involved in peptide bond hydrolysis (His-41, Asp-88, and Ser-173; His-57, Asp-102, and Ser-195 in bovine chymotrypsinogen) are strictly conserved (33), some of the residues defining the specificity site are changed. As shown in Fig. 2, Ser-182 and Thr-221 of the porcine enzyme (residues 189 and 226 in chymotrypsinogen), which are near the bottom of the pocket (34), are replaced by Gly-163 and Asp-201, respectively. This latter replacement seems very unusual in that it positions an acidic residue in a primary position during substrate binding. In addition, Thr-29 (residue 41 in chymotrypsinogen), which has been suggested to be part of the $S_3$ and $S_4$ binding sites for this enzyme (34), is replaced by Phe-25 in the human enzyme, in agreement with the normally conserved residue in most serine proteases. However, Val-209 (residue 216 in chymotrypsinogen), which is near the top of the pocket in pancreatic elastase, remains unchanged as Val-190 in the neutrophil enzyme. Significantly, Gln-185 (residue 192 in chymotrypsinogen), which is also believed to be part of the $S_1$ and $S_3$ binding sites in pancreatic elastase, is replaced by the much bulkier and hydrophobic Phe-170 in the neutrophil enzyme.

Of some interest is the fact that human neutrophil elastase has two sites to which carbohydrate is bound. Preliminary data from a detailed analysis of both the type and structure of such side chains indicates that the composition reflects that of a complex carbohydrate, with the individual isoenzymes having differing compositions. It is also possible, however, that the two chains are different, one representing a terminally phosphorylated high-mannose type in order to mark the enzyme for insertion into the lysosomal-like granule matrix of the neutrophil (35), and the other being involved in receptor recognition either for the native enzyme or for enzyme–inhibitor complexes (36).

A comparison of the amino acid sequence of human neutrophil elastase with other mammalian serine proteinases using the FASTP protein database search (37) reveals some interesting homologies (Table 1). Of particular interest is the fact that there are several proteinases with identity close to that of pancreatic elastase, including porcine pancreatic kallikrein and cathepsin G (9). Clearly, it would be of significant interest to compare the crystal structures of these proteinases, when they become available, with both neutrophil and pancreatic elastase to further establish the mechanisms by which the latter pair act as efficient elastolytic enzymes.

In all of our studies, we could find no evidence for zymogen forms of neutrophil elastase, although the nucleotide sequence for a cDNA clone coding for cathepsin G (9), a protease found in the same granules as elastase, would strongly suggest that these enzymes are synthesized as

Table 1. Comparative sequence homology of human neutrophil elastase with other proteinases*

| Enzyme | Source | % identity |
|---|---|---|
| Pancreatic elastase | Porcine | 43.0 |
| Pancreatic kallikrein | Porcine | 39.2 |
| Cathepsin G | Human | 37.2 |
| Collagenolytic protease | Fiddler Crab | 35.4 |
| Preproelastase II | Rat | 34.9 |
| Trypsinogen | Dogfish | 34.8 |
| Plasminogen | Human | 34.0 |
| Factor D | Human | 32.9 |
| Chymotrypsinogen A | Bovine | 31.6 |
| Chymotrypsinogen B | Bovine | 31.2 |

*Comparative sequence analysis utilized the FASTP program (37).

inactive zymogens that are then processed to the active enzyme prior to neutrophil maturity. The isolation and sequencing of similar clones for neutrophil elastase could conceivably identify such putative precursors.

This research was supported in part by grants from the National Heart, Lung, and Blood Institutes and The Council For Tobacco Research-USA (to J.T.) and by a Biomedical Research Support Grant from the National Institutes of Health (to S.S.). The authors wish to thank Dr. Guy Salvesen and Dr. James Powers for helpful discussions during the course of this work and Dr. John Wunderlich and Mr. Thomas Fischer for the gas-phase sequence analyses. We would especially like to show our appreciation to Dr. Jean Hester, M. D. Anderson Hospital, Houston, TX, who provided the neutrophils required for this study.

1. Dewald, B., Rinder-Ludwig, R., Bretz, U. & Baggiolini, M. (1975) *J. Exp. Med.* **141,** 709–723.
2. Travis, J. & Salvesen, G. S. (1983) *Annu. Rev. Biochem.* **52,** 655–709.
3. Matheson, N. R., Janoff, A. & Travis, J. (1982) *Mol. Cell. Biochem.* **45,** 65–71.
4. Laurell, C. B. & Eriksson, S. (1963) *Scand. J. Clin. Lab. Invest.* **15,** 32–40.
5. Roughley, P. (1977) *Biochem. J.* **167,** 639–647.
6. Roughley, P. & Barrett, A. J. (1977) *Biochem. J.* **167,** 629–638.
7. Janoff, A. & Scherer, J. (1968) *J. Exp. Med.* **128,** 1137–1151.
8. Janoff, A. (1973) *Lab. Invest.* **29,** 458–464.
9. Salvesen, G. S., Farley, D., Shuman, J., Przybyla, A., Reilly, C. & Travis, J. (1987) *Biochemistry*, in press.
10. Baugh, R. J. & Travis, J. (1976) *Biochemistry* **15,** 836–843.
11. Travis, J., Giles, P. J., Porcelli, L., Reilly, C., Baugh, R. & Powers, J. (1980) in *Protein Degradation in Health and Disease*, eds. Evered, D. & Whelan, J. (Excerpta, Amsterdam), Vol. 75, pp. 51–68.
12. Crestfield, A. M., Moore, S. & Stein, W. H. (1963) *J. Biol. Chem.* **238,** 622–627.
13. Mahoney, W. C. & Hermodson, M. A. (1979) *Biochemistry* **18,** 3810–3816.
14. Raftery, M. A. & Cole, R. D. (1966) *J. Biol. Chem.* **241,** 3457–3461.
15. Jekel, P. A., Weyer, W. J. & Beintema, J. J. (1983) *Anal. Biochem.* **134,** 347–354.
16. Wyckof, M., Rodbard, D. & Chrambach, A. (1977) *Anal. Biochem.* **78,** 459–482.
17. Glajch, J. L., Gluckman, J. C., Charikofsky, J. G., Minor, J. M. & Kirkland, J. J. (1985) *J. Chromatogr.* **318,** 23–29.
18. Dubois, M., Giles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956) *Anal. Chem.* **28,** 350–356.
19. Gatt, R. & Berman, E. R. (1966) *Anal. Biochem.* **15,** 167–171.
20. Warren, L. (1959) *J. Biol. Chem.* **234,** 1971–1977.
21. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem.* **5,** 151–153.
22. Heck, L., Darby, W., Hunter, F., Bhown, A., Miller, E. & Bennett, C. (1985) *Anal. Biochem.* **149,** 153–160.
23. Bode, W., Huber, R., Meyer, E., Travis, J. & Neumann, S. (1986) *EMBO J.* **51,** 2453–2458.
24. Bieth, J. (1978) *Front. Matrix Biol.* **6,** 1–82.
25. Blow, A. M. J. (1977) *Biochem. J.* **161,** 13–16.
26. Reilly, C. F. & Travis, J. (1980) *Biochim. Biophys. Acta* **621,** 147–153.
27. Keller, S. & Mandl, I. (1972) in *Pulmonary Emphysema and Proteolysis*, ed. Mittman, C. (Academic, New York), pp. 251–259.
28. Starkey, P. M. & Barrett, A. J. (1976) *Biochem. J.* **155,** 255–264.
29. Ashe, B. M. & Zimmerman, M. (1977) *Biochem. Biophys. Res. Commun.* **75,** 194–199.
30. Twumasi, D. Y., Liener, I. E., Galdstone, M. & Levytska, V. (1977) *Nature (London)* **267,** 61–63.
31. Shotton, D. M. & Hartley, B. S. (1970) *Nature (London)* **225,** 802–806.
32. Hartley, B. S. & Shotton, D. M. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), 3rd Ed., Vol. 3, pp. 323–375.
33. Shotton, D. M. (1971) in *Proceedings of the International Research Conference on Proteinase Inhibitors*, eds. Fritz, H. & Tscheche, H. (De Gruyter, Berlin), pp. 47–116.
34. Shotton, D. M., White, N. J. & Watson, H. C. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36,** 91–105.
35. Creek, K. E. & Sly, W. S. (1984) in *Lysosomes in Biology and Pathology*, ed. Dingle, J. T. (Elsevier, Amsterdam), pp. 63–82.
36. Ashwell, G. & Morell, A. G. (1977) *Trends. Biochem. Sci.* **2,** 76–78.
37. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227,** 1435–1441.