Biochem. J. (1986) 240, 913–916 (Printed in Great Britain)

Comparison of human stromelysin and collagenase by cloning and sequence analysis

Sarah E. WHITHAM,* Gillian MURPHY,† Peter ANGEL,‡ Hans-Jobst RAHMSDORF,‡ Bryan J. SMITH,* Alan LYONS,* Tim J. R. HARRIS,* John J. REYNOLDS†, Peter HERRLICH‡ and Andrew J. P. DOCHERTY*§

*Division of Molecular Biology, Celltech Limited, 244-250 Bath Road, Slough SL1 4DY, U.K., †Cell Physiology Department, Strangeways Research Laboratory, Cambridge CB1 4RN, U.K., and ‡Kernforschungszentrum Karlsruhe, Institut fur Genetik and Toxikologie, P.O. Box 3640, D-7500 Karlsruhe 1, Federal Republic of Germany

A comparison of the cDNA-derived amino acid sequences of human stromelysin and collagenase with the *N*-terminal sequences of purified enzymes reveals that these metalloproteinases are highly conserved and that they are secreted as proenzymes. A putative zinc-binding site was identified by its homology with the zinc-chelating sequence of thermolysin. These sequences permitted the identification of: (1) transin, a protein induced in rat fibroblasts either exposed to growth factors or transformed by oncogenic viruses, as the rat homologue of stromelysin, and (2) XHF1, a protein induced in human fibroblasts after treatment with tumourigenic agents, as collagenase.

INTRODUCTION

Human diploid fibroblasts in culture synthesize a specific set of proteins when treated with tumourigenic agents such as u.v. light, mitomycin c and phorbol esters [1]. These induced proteins are also over-expressed in cells from patients with Bloom's syndrome [1,2], a condition predisposing to a high tumour incidence. One of the proteins, XHF1, was shown to be an M_r 53000 secreted protein [3], but was not further identified. Similarly, rat fibroblasts, either treated with epidermal growth factor (EGF) or transformed by oncogenic viruses, are induced to synthesize a novel mRNA [4]. The nucleotide sequence of the cDNA coded for a polypeptide of M_r 53000, transin [4,5]. Transin was recently shown to have strong homology with the sequence of the metalloproteinase collagenase, cloned from a human skin fibroblast library [6].

There have been numerous independent reports of the induction of the metalloproteinases collagenase and stromelysin in connective tissue cells treated with agents such as interleukin 1 (IL-1) and phorbol 12-myristate 13-acetate (PMA) [7,8]. These two enzymes are important members of the connective tissue metalloproteinase family and are thought to play a role in normal tissue remodelling [9]. Increased metalloproteinase activity has previously been shown to correlate with tumour invasiveness and metastatic potential [10] and it was of interest to establish whether these enzymes could be induced by agents which stimulate cellular transformation. In this paper we use the *N*-terminal amino acid sequences of human and rabbit collagenases and rabbit stromelysin to identify cDNAs isolated from human fibroblasts stimulated with these agents as coding for the metalloproteinases.

METHODS

Enzyme purification and sequencing

Rabbit fibroblast collagenase and stromelysin, activated with APMA, were purified from the culture media of calvariae [11,12]. Human fibroblasts derived from gingival explants (passage 3–6) were cultivated in DMEM supplemented with 0.2% lactalbumin hydrolysate and 5% partially purified pig IL-1. After 48 h conditioned media were collected and human collagenase was purified essentially as described previously [11]. Collagenase had an M_r of 55000 and could be maintained in a latent form in the presence of 0.5 mM-zinc acetate. On activation with APMA, or after storage at 4 °C in the absence of Zn^{2+}, lower M_r species were obtained, the predominant size being 44000 and 28000. Enzymes were reduced and carboxymethylated and subjected to *N*-terminal sequencing by automated Edman degradation on an Applied Biosystems gas-phase sequenator. Activated human collagenase generated two *N*-termini (Figs. 1*d* and 1*e*) which were confirmed by comparison with the predicted cDNA sequence.

Isolation of stromelysin cDNAs

A single 50-base oligonucleotide probe based on the *N*-terminal amino-acids of activated rabbit stromelysin and collagenase (Fig. 1) was designed to be theoretically capable of hybridizing to cDNAs encoding either enzyme [13,14]. It was assembled by the ligation of two 25-mers synthesized as previously described [15].

Rabbit fibroblasts from explants of normal synovium were grown to confluence in DMEM supplemented with 10% fetal calf serum. After washing they were then maintained in serum-free media for 48 h in the presence of 50–200 ng of PMA/ml. mRNA was isolated [16] and

Abbreviations used: APMA, 4-aminophenylmercuric acetate; DMEM, Dulbecco's modified Eagles's medium; EGF, epidermal growth factor; IL-1, interleukin 1; PMA, phorbol 12-myristate 13-acetate; SSC, 0.15 M-NaCl containing 15 mM-sodium citrate; TIMP, tissue inhibitor of metalloproteinases.

§ To whom reprint requests should be addressed.

used to generate a cDNA library of 75000 recombinant plaques in λgt 10 [17,18] which was screened with the oligonucleotide probe under previously described conditions except that the highest stringency wash was in 1 × SSC [19]. A single clone was isolated which on sequencing [20,21] was found to lack a poly(A) tract but did possess an open reading frame encoding 170 amino acids. Residues 101–127 were identical with the *N*-terminal amino acid sequence of rabbit stromelysin (Figs. 1*b* and 2). It was therefore labelled by the random hexanucleotide method and used to screen for a human stromelysin cDNA [22].

Human gingival fibroblasts were cultivated as described above in the presence of IL-1. After 48 h mRNA was isolated, used to generate a library of 200000 recombinant plaques in λgt 10 and screened with the rabbit cDNA probe, essentially as described above. cDNA inserts from positively hybridizing plaques were subjected to restriction mapping and a complete nucleotide sequence was compiled from related overlapping DNA segments [20,21,23]. This sequence has been recorded at the EMBL data library.

Isolation of a human collagenase cDNA

XHFI is a protein secreted from human fibroblasts after exposure to PMA [1,24], and the isolation of several partial cDNAs corresponding to its mRNA has been previously described [3]. For the present study clone K4 cDNA was nick-translated and used to screen a second library generated in pSP64 [25] by the method of Heidecker & Messing [26] from mRNA isolated from primary human fibroblasts after 6 h treatment with PMA at 20 ng/ml. Restriction mapping demonstrated that eight clones had an identical insert and the complete nucleotide sequence of one clone was determined as described above. Examination of the cDNA revealed an open reading frame with residues encoded by nucleotides 124–180 (Fig. 2) being identical with the *N*-terminal amino acids of latent human collagenase (Fig. 1*a*). This sequence has been recorded at the EMBL data library.

RESULTS AND DISCUSSION

A comparison of the *N*-terminal amino acid sequence of the APMA-activated form of rabit stromelysin with collagenease reveals that they are highly conserved (Figs. 1*b* and 1*c*). This facilitated the design of a probe which was used to isolate a partial rabbit stromelysin cDNA which was sequenced and identified by using the amino acid sequence data. The rabbit cDNA was subsequently used to isolate the corresponding human stromelysin cDNA. It is 1804 nucleotides in length and appears to be complete at the 3′ end because it has a poly(A) tail preceded by the polyadenylation signal AATAAA [27]. It contains an open reading frame (nucleotides 45–1475) which translates into a polypeptide of 477 amino acids of predicted M_r 53916. Evidence that this amino acid sequence represents the primary sequence of human stromelysin comes from the finding that most of the amino acids predicted from the partial rabbit stromelysin cDNA are identical and are found in a similar position (Fig. 2).

Fig. 2 also demonstrates that there is 75% overall homology between rabbit stromelysin, human stromelysin and another amino acid sequence predicted from a rat cDNA, referred to as transin [4,6]. This cDNA corresponds to a mRNA which is enriched in normal fibroblasts after infection with either polyoma virus or Rous sarcoma virus or transfection with either the middle T oncogene or the cellular oncogene H-*ras*. The same mRNA was also reported to be specifically induced after exposure of fibroblasts to EGF [4]. The predicted M_r of the rat protein encoded by pTRI is approx. 53000, in close agreement with the human enzyme and the *in*

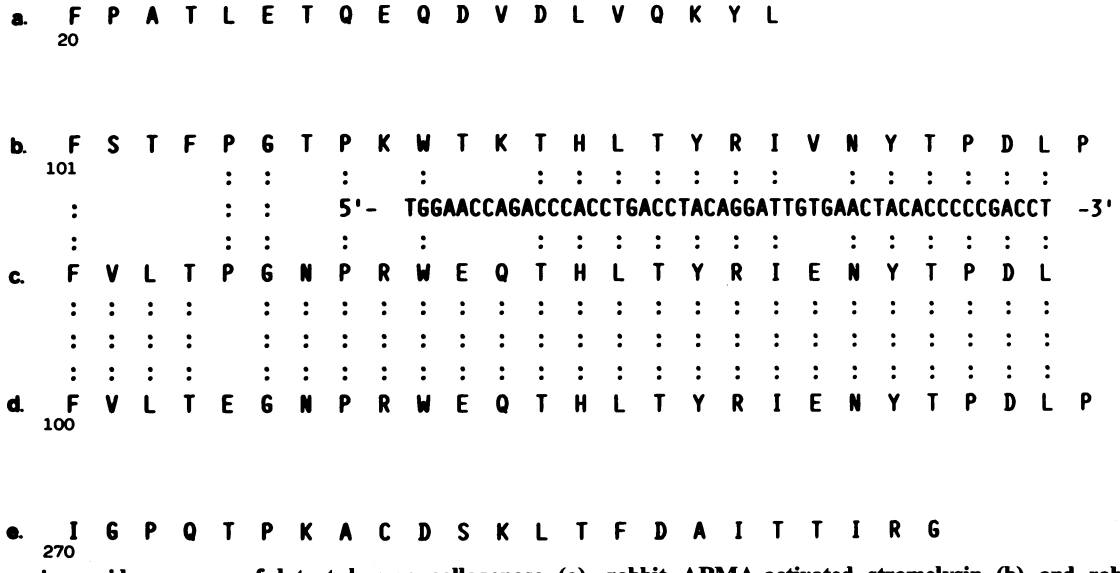

Fig. 1. *N*-Terminal amino acid sequence of latent human collagenase (a), rabbit APMA-activated stromelysin (b) and rabbit APMA-activated collagenase (c)

The rabbit sequences are compared with the main (54% yield) *N*-terminal sequence obtained with the APMA-activated human collagenase (d). The less abundant (46% yield) human sequence present in this preparation is also shown (e). The oligonucleotide probe based on the rabbit sequences is shown between (b) and (c). The numbers refer to the position of these amino acids in the cDNA-predicted sequences shown in Fig. 2.

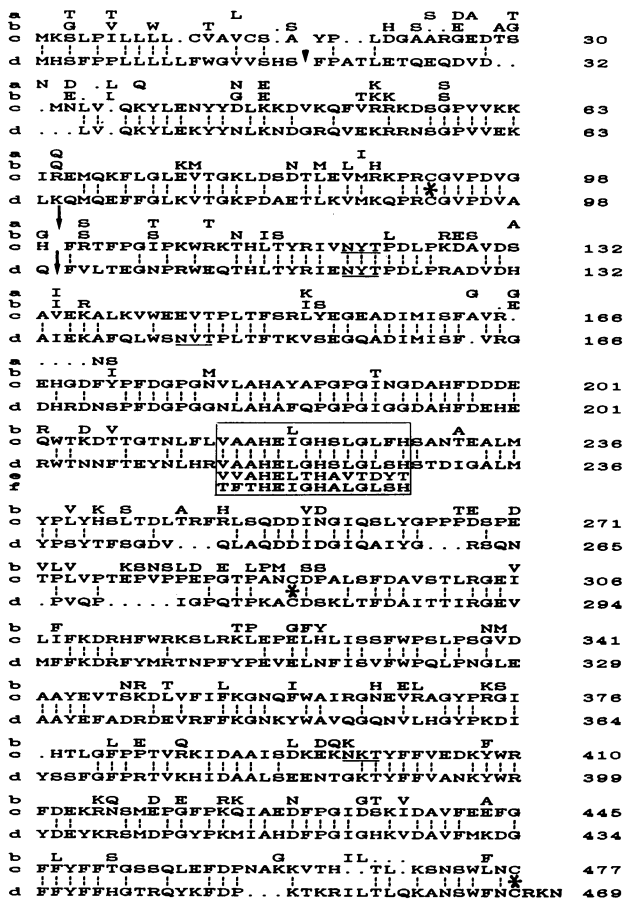

Fig. 2. Comparison of the predicted amino-acid sequences of human stromelysin (c) and human collagenase (d)

Those amino acids in rat stromelysin (b) and the *N*-terminal half of rabbit stromelysin (a) which differ from human stromelysin (c) are also shown. The triangle denotes the signal peptidase cleavage site generating the *N*-terminus of human procollagenase. Arrows indicate the cleavage which occurs on activation of human collagenase and rabbit stromelysin in the presence of APMA. The amino acid sequences thought to be involved in co-ordinating the zinc atom at the active site and corresponding sequences in thermolysin (e) [29] and *Serratia* metalloproteinase (f) [30] are boxed. Potential glycosylation sites (underlined) and the conserved cysteine residues in the proenzymes (asterisk) are also shown. The collagenase sequence is identical with that reported by Goldberg *et al.* [6] except for positions 115 (T), 200 (H), 208 (T) and 410 (G): in three positions (115, 208, 410) our designations in collagenase are identical with those in stromelysin.

vitro translation product of rabbit stromelysin mRNA [28, and results not shown]. We conclude, therefore, that the protein encoded by pTRI is rat stromelysin.

The latent form of human collagenase was purified from the culture supernatants of IL-1-stimulated human gingival fibroblasts, and subjected to *N*-terminal amino acid sequencing (Fig. 1*a*). This sequence information allowed us to identify a cDNA, which corresponds to a mRNA whose abundance is increased in primary human fibroblasts after exposure to either IL-1, PMA, u.v. light or mitomycin c [3,24] as coding for human collagenase. Examination of the cDNA reveals an open reading frame encoding a protein of 469 amino acids (calculated M_r 53828), with residues 20–38 being identical with the *N*-terminal amino acids of latent human collagenase (Fig. 1*a* and 2).

Alignment of the amino acid sequences of human stromelysin and human collagenase reveals that they are highly homologous (55%; Fig. 2). Conserved features included the two cysteine residues found in the active form of these enzymes (see below) and one in the propeptide as well as one of the two potential glycosylation sites found in collagenase. Hydrophobicity plots suggest that these proteins are all water-soluble, except for a hydrophobic *N*-terminal sequence. Confirmation that this represents a signal sequence, which is cleaved off after the serine at position 19 in collagenase, comes from the finding that the *N*-terminal amino acid of latent collagenase can be found immediately downstream from this residue (Fig. 1*a*). In human, rabbit and rat stromelysin there are analogous signal peptidase cleavage sites but these have not been confirmed by *N*-terminal amino acid sequencing of the latent enzymes.

All attempts at sequencing the APMA-activated form of human collagenase gave rise to mixed *N*-termini which could only be confirmed by comparison with the amino acid sequence predicted from the cDNA. A major sequence present in these mixtures (Fig. 1*d*) was, however, found to share significant homology with the *N*-terminal sequences of both activated rabbit collagenase and stromelysin (Figs. 1*b* and 1*c*). As shown in Fig. 2, this sequence is found immediately downstream from a conserved sequence of eight residues (90–97 in human stromelysin). We confirm therefore that the enzymes are all secreted as proenzymes as previously reported [7] and conclude that the amino acids found between the putative signal sequence and the phenylalanine at the *N*-termini of the APMA-activated form of these enzymes (position 100 in human stromelysin and collagenase) are cleaved off during activation. These data are essentially in agreement with the observations of Goldberg *et al.* [6] on the sequence of human collagenase, although they were unable to sequence the latent form of the enzyme and the trypsin-activated form started at position 101. A second *N*-terminus in activated human collagenase preparations is found at residues 270–292 in the complete sequence (Figs. 1*e* and 2). The occurrence of lower M_r forms with *N*-termini which can be identified within the predicted amino acid sequence suggests that after APMA treatment, or storage of the enzyme at 4 °C, some degree of self-cleavage occurs. Whether self-cleavage of the proenzyme is an important activation mechanism *in vivo* in normal tissue remains to be confirmed.

Another conserved region (residues 215–226 in human stromelysin) has significant homology with part of the zinc-chelating region of the bacterial metalloproteinase thermolysin [29] and a similar sequence is found in *Serratia* metalloproteinase [30]. It therefore seems likely that the conserved histidine residues (218 and 222) participate in binding the zinc atom known to be at the active site of all metalloproteinase studied in any detail.

Metalloproteinases have previously been shown to be induced in connective tissue cells by either IL-1 or tumour promoters such as PMA [7,8]. Our new data show that the stimulated transcription of mRNAs for metalloproteinases is a common feature in fibroblasts either transformed with oncogenes or exposed to EGF.

Plasminogen activator [31] and a lysosomal protein [32] are also known to be induced after transformation. The production of metalloproteinases by cells exposed to such a wide variety of stimuli is indicative of their importance in the development of the invasive potential of transformed cells [33]. In normal tissues the production and activity of metalloproteinases are tightly regulated [7,34] but in connective tissue disorders the regulatory mechanisms appear to be altered. The levels of the naturally occurring tissue inhibitor of metalloproteinases, TIMP [35,19], which can block the destruction of type I collagen [36] and the invasion of amnion membrane by tumour cells in model systems [37], may be a significant controlling factor. The structural information on the family of metalloproteinases presented in this paper, together with an understanding of their interaction with TIMP, should aid in the design of drugs which can block metalloproteinase activity and therefore be of therapeutic value.

The authors would like to thank the Fonds der Chemischen Industrie and the Medical Research Council for financial support, George Murphy, Helmut Ponta and Marina Schorpp for helpful discussions, George Murphy for computer analyses, and Mary Harrison for expert technical assistance.

REFERENCES

1. Mallick, U., Rahmsdorf, H. J., Yamamoto, N., Ponta, H., Wegner, R.D. & Herrlich, P. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 7886–7890
2. Schorpp, M., Mallick, U., Rahmsdorf, H. J. & Herrlich, P. (1984) Cell **37**, 861–868
3. Angel, P., Poting, A., Mallick, U., Rahmsdorf, H. J., Schorpp, M., & Herrlich, P. (1986) Mol. Cell Biol. **6**, 1760–1766
4. Matrisian, L. M., Glaichenhaus, N., Gesnel, M.-C. & Breathnach, R. (1985) EMBO J. **4**, 1435–1440
5. Matrisian, L. M., Leroy, P., Ruhlmann, C., Gesnel, M.-C. & Breathnach, R. (1986) Mol. Cell. Biol. **6**, 1679–1686
6. Goldberg, G. I., Wilhelm, S. M., Kronberger, A., Bauer, E. A., Grant, G. A. & Eisen, A. Z. (1986) J. Biol. Chem. **261**, 6600–6605
7. Harris, E. D., Welgus, H. G. & Krane, S. M. (1984) Collagen Relat. Res. **4**, 493–512
8. Murphy, G., Hembry, R. M. & Reynolds, J. J. (1986) Collagen Relat. Res. **6**, in the press
9. Murphy, G., Cawston, T. E., Galloway, W. A., Barnes, M. J., Bunning, R. A. D., Mercer, E., Reynolds, J. J. & Burgeson, R. E. (1981) Biochem. J. **199**, 807–811
10. Liotta, L. A., Tryggvason, K., Garbisa, S., Hart, I., Foltz, C. M. & Shafie, S. (1980) Nature (London) **284**, 67–68
11. Cawston, T. E. & Murphy, G. (1981) Methods Enzymol. **80**, 711–722
12. Galloway, W. A., Murphy, G., Sandy, J. D., Gavrilovic, J., Cawston, T. E., & Reynolds, J. J. (1983) Biochem. J. **209**, 741–752
13. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) Nucleic Acids Res. **9**, r43–r47
14. Lathe, R. (1985) J. Mol. Biol. **183**, 1–12
15. Patel, T. P., Millican, T. A., Bose, C. C., Titmas, R. C., Mock, G. A. & Eaton, M. A. W. (1982) Nucleic Acids Res. **10**, 5605–1619
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York
17. Gubler, U. & Hoffman, B. J. (1983) Gene **25**, 263–269
18. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in DNA Cloning Techniques, A Practical Approach (Glover, D. M., ed.), vol. 1, pp. 49–78, IRC Press, Oxford
19. Docherty, A. J. P., Lyons, A., Smith, B. J., Wright, E. M., Stephens, P. E., Harris, T. J. R., Murphy, G. & Reynolds, J. J. (1985) Nature (London) **318**, 66–69
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
21. Messing, J. & Vieira, J. (1982) Gene **19**, 269–276
22. Feinberg, A. P. & Vozelstein, B. (1983) Anal. Biochem. **132**, 6–13
23. Maxam, A. M. & Gilbert, W. (1980) Methods Enzymol. **65**, 499–560
24. Herrlich, P., Angel, P., Rahmsdorf, H. J., Mallick, U., Poting, A., Hieber, L., Lucke-Huhle, C. & Schorpp, M. (1986) Adv. Enzyme Regul. **25**, 485–504
25. Melton, D. A., Krieg, P. A., Rebagliati, N. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) Nucleic Acids Res. **12**, 7035–7056
26. Heidecker, G. & Messing, J. (1983) Nucleic Acids Res. **11**, 4891–4906
27. Proudfoot, N. & Brownlee, G. G. (1981) Nature (London) **252**, 359–362
28. Frisch, S. M., Chin, J. R. & Werb, Z. (1983) J. Cell Biol. **97**, Suppl. **2:5**, 430a (abstr.)
29. Matthews, B. W., Weaver, L. H. & Kester, W. R. (1974) J. Biol. Chem. **249**, 8030–8044
30. Lee, I. S., Wakabayashi, S., Miyata, K., Tomada, K., Yoneda, M., Kangawa, K., Minamino, N., Matsuo, H. & Matsubura, H. (1984) J. Biochem. (Tokyo) **96**, 1409–1418
31. Unkeless, J. C., Tobia, A. Ossowski, L., Quigley, J. P., Rifkin, D. B. & Reich, E. (1973) J. Exp. Med. **17**, 85–112
32. Gal, S., Willingham, M. C. & Gottesman, M. M. (1985) J. Cell Biol. **100**, 535–544
33. Liotta, L. A. (1986) Cancer Res. **46**, 1–7
34. Reynolds, J. J. (1985) Br. J. Dermatol. **112**, 715–723
35. Reynolds, J. J., Bunning, R. A. D., Cawston, T. E. & Murphy, G. (1981) in Cellular Interactions (Dingle, J. T. & Gordon, J. L., eds.), pp. 205–213. Elsevier/North-Holland, Amsterdam
36. Gavrilovic, J., Reynolds, J. J. & Murphy, G. (1985) Cell Biol. Int. Rep. **9**, 1097–1107
37. Thorgeirsson, U. P., Liotta, L. A., Kalebic, T., Margulies, I. M., Thomas, K., Rios-Candelore, M. & Russo, R. G. (1982) J. Natl. Cancer Inst. **69**, 1049–1054

Received 13 August 1986/3 October 1986; accepted 10 October 1986