Volume 310, number 2, 175–178 FEBS 11598 September 1992

Cloning and sequencing of mouse collagenase cDNA*

Divergence of mouse and rat collagenases from the other mammalian collagenases

Patrick Henriet[a], Guy G. Rousseau[b] and Yves Eeckhout[a]

[a]*Laboratoire de Chimie Physiologique (Connective Tissue Group)* and [b]*Hormone and Metabolic Research Unit, Université de Louvain and International Institute of Cellular and Molecular Pathology, Bruxelles, Belgium*

Received 19 August 1992

Mouse collagenase cDNA was cloned and sequenced. The deduced amino acid sequence was compared to those of the other mammalian collagenases and related matrix metalloproteinases. These comparisons, as well as those of some enzymatic properties, show that the rodent (mouse and rat) interstitial collagenases are very similar but differ more from the other interstitial collagenases than does human neutrophil collagenase. This supports the hypothesis that the order Rodentia is an outgroup to the other eutherian (placental) mammalian orders.

Collagen; Evolution; Homology; Matrix metalloproteinase; Neutrophil; Rodent

1. INTRODUCTION

Vertebrate collagenases belong to a family of matrix metalloproteinases (MMPs) which share several structural and functional characteristics (reviewed by Matrisian [1], Docherty and Murphy [2] and Woessner [3]). On the basis of their amino acid sequences and their substrate specificities, MMPs have been grouped into three major classes: collagenases, gelatinases and stromelysins. Mammalian collagenases have been divided in two groups: interstitial collagenases (EC 3.4.24.7, MMP 1) and neutrophil collagenases (EC 3.4.24.34, MMP 8).

The cDNA sequences of five interstitial collagenases (human, rabbit, porcine, bovine and rat) and of human neutrophil collagenase are known. The importance of murine experimental models in bone, tumor and developmental biology prompted us to clone and sequence the cDNA of mouse collagenase.

2. EXPERIMENTAL

2.1. *Northern blotting analysis*

Calvaria of 5-day-old NMRI (Naval Medical Research Institute) mice were cultured for 4 days in the presence of heparin [4]. Total RNA was isolated by the standard guanidine thiocyanate-caesium chloride method [5] and about 20 μg were electrophoresed on an agarose-formaldehyde gel and blotted onto a nylon membrane (Amersham). A 2.1-kb cDNA probe was prepared by *Xho*I digestion of a pBluescript clone containing a 2.7-kb cDNA of rat collagenase [6]. The membrane was hybridized overnight at 65°C with the ^{32}P-labeled rat probe, washed four times (the most stringent condition used was 0.2× SSC–0.1% SDS at 65°C) and exposed for about 48 h with intensifying screens.

2.2. *Construction of a cDNA library and isolation of a clone*

Poly(A)-rich RNA was purified from total RNA with an oligo(dT)–cellulose column (Pharmacia) and cDNA was synthesized using the Riboclone cDNA Synthesis System *Not*I primer adaptor (Promega Corp.) in the presence of [^{32}P]dATP (Amersham). The products were fractionated on an Agarose A 50 M column and only fractions that did not contain cDNAs smaller than 600 bp were used in the next steps. A lambda gt 11 (*Sfi*I–*Not*I) cDNA library was constructed with this material using Riboclone *Eco*RI Adaptor Ligation System I and Lambda gt 11 Sfi-Not DNA (*Eco*RI–*Not*I arms) Plus Packagene System (Promega Corp.).

Clones for collagenase were isolated from this library by triple screening with the rat cDNA probe using standard procedures [5]. About 1.5×10^5 clones were analyzed in the primary screening. For tertiary screening, membrane duplicates were screened either with the probe described above or with a shorter (390 bp) 5′ fragment of this probe in order to detect the longest inserts. Membranes for tertiary screening were finally washed in 0.1× SSC–0.1% SDS at 65°C.

To identify the longest insert, the 5′ terminal region of four clones positive after tertiary screening was amplified by PCR using Perkin-Elmer Cetus chemicals and procedure. The oligonucleotides used were the lambda gt 11 forward primer (New England Biolabs) and a 36-mer hybridizing a highly conserved sequence of all MMPs, the cysteine switch (5′-ACGAATTCTTTAAACACCCACATCAGGCACTC-CACA-3′). Amplification products were compared by electrophoresis on an agarose gel. The longest insert was subcloned between the *Hind*III and *Not*I sites of pBluescript (Stratagene) after intermediate passage between the *Sfi*I and *Not*I sites of pGEM 13 Zf + (Promega Corp.).

Correspondence address: Y. Eeckhout, UCL 75.39, 75 Avenue Hippocrate, B-1200 Bruxelles, Belgium. Fax: (32) (2) 764 7573.

**GenBank/EMBL Database accession number X66473.*

Abbreviations: ME, metalloelastase; MMP, matrix metalloproteinase; NMRI, Naval Medical Research Institute; PCR, polymerase chain reaction; SSC, sodium saline citrate (1× = NaCl 0.15 M, sodium citrate 0.015 M); UTR, untranslated region.

2.3. *DNA sequencing and sequence comparisons*

Progressively deleted clones were generated for each strand with Exonuclease III (Erase A Base System, Promega Corp.). Mouse collagenase cDNA was double-strand sequenced on both strands by the dideoxynucleotide chain termination method [7] with T7 DNA Polymerase or TaqTrack DNA Polymerase Systems (Promega Corp.). The deleted clones were sequenced with T3/T7 primers (Promega Corp.) and fragments of the full-length cDNA with specific primers derived from the rat sequence.

3. RESULTS AND DISCUSSION

When cultured in the presence of heparin, calvaria from 5-day-old mice release about 10 U of collagenase/ml culture medium in 4 days [8], corresponding approximately to 2.5 μg of purified enzyme. As illustrated in Fig. 1, extracts of such cultured calvaria contained a 3.0-kb mRNA which hybridized with a rat collagenase cDNA probe (a gift from Dr. J. Jeffrey, see [6]) under the most stringent conditions used. At lower stringency, a human collagenase cDNA probe (a gift from Dr. H. Nagase) revealed the same band plus a band of 1.9 kb (data not shown), the latter probably corresponding to a stromelysin mRNA [9].

A cDNA library was constructed using RNA extracted from such cultured calvaria and screened with the rat collagenase cDNA probe in order to isolate clones for mouse collagenase. About 1% of the clones were positive at the primary screening. Two subsequent screenings allowed the isolation of four clones with a 2.7-kb insert. The clone with the longest insert was identified by PCR amplification of the 5′ terminal region and this insert was subcloned and sequenced.

The 2,660-bp cDNA (Fig. 2) shares 94% identity with rat collagenase cDNA [6]. Two sequences of 15 and 10 amino acids determined on purified mouse collagenase (Peeters-Joris, personal communication) are identical with the corresponding sequences deduced from our cDNA. These sequences (residues 105–119 and 266–275) are distinct from the highly conserved regions, thus confirming that our cDNA encodes mouse collagenase. Like its rat homologue [6], mouse collagenase cDNA originates from a mRNA longer (2.9–3.0 kb) than the mRNAs encoding the other mammalian collagenases (2–2.5 kb). Therefore, the 2.0-kb mouse embryonic mRNA detected by Brenner et al. [10] with a PCR-derived probe does not code for mouse collagenase but probably corresponds to a related MMP. Our mouse collagenase cDNA contains a 10-bp 5′ untranslated region (UTR), a 1,416-bp open reading frame encoding mouse preprocollagenase, a 1215-bp 3′ UTR containing the polyadenylation signal and a 19-bp poly(A) tail. The initial ATG is preceded by a sequence matching the Kozak consensus [11].

The deduced amino acid sequence of mouse collagenase (Fig. 2) exhibits the domain structure common to MMPs and more particularly to all mammalian collagenases: a signal peptide (residues 1–23), a propeptide

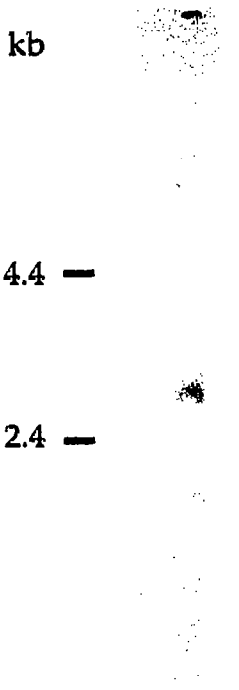

Fig. 1. Northern blot of mouse calvaria total RNA (20 μg) probed with a ^{32}P-labeled rat collagenase cDNA.

(residues 24–104) containing the cysteine switch, a catalytic domain (residues 105–234) containing the zinc binding domain and an hemopexin-like domain (residues 235–472). This sequence shares 97% identity with that of rat collagenase [6]. We compared it with the sequences of other mammalian collagenases and related MMPs available from the GenBank/EMBL Database: (i) interstitial collagenases (MMP 1, EC 3.4.24.7, accession numbers M13509, M17820, X54724, X58256 and M60616), (ii) human neutrophil collagenase (MMP 8, EC 3.4.24.34, accession number J05556), (iii) stromelysins 1 (MMP 3, EC 3.4.24.17, accession numbers X05232, X66402 and PIR International Protein Sequence Database accession numbers A29157 and A00997), (iv) stromelysins 2 (MMP 10, EC 3.4.24.22, X07820 and X05083), (v) homologous domains of gelatinases A (MMP 2, EC 3.4.24.24, accession numbers J03210 and M84324), and (vi) a recently cloned mouse metalloelastase (M82831). The results of this computer-aided comparison are presented in Table I.

These comparisons show that the primary sequences of the same MMPs from different species share 76–98% identity, except the rodent (mouse and rat) collagenases which share no more than 52–55% identity with each of the four other mammalian interstitial collagenases (Table I, closed box). This is not due to differences restricted to a particular region of the protein sequence. This small percentage of identity is equal or even lower than the percentage of identity (52–62%) shared by

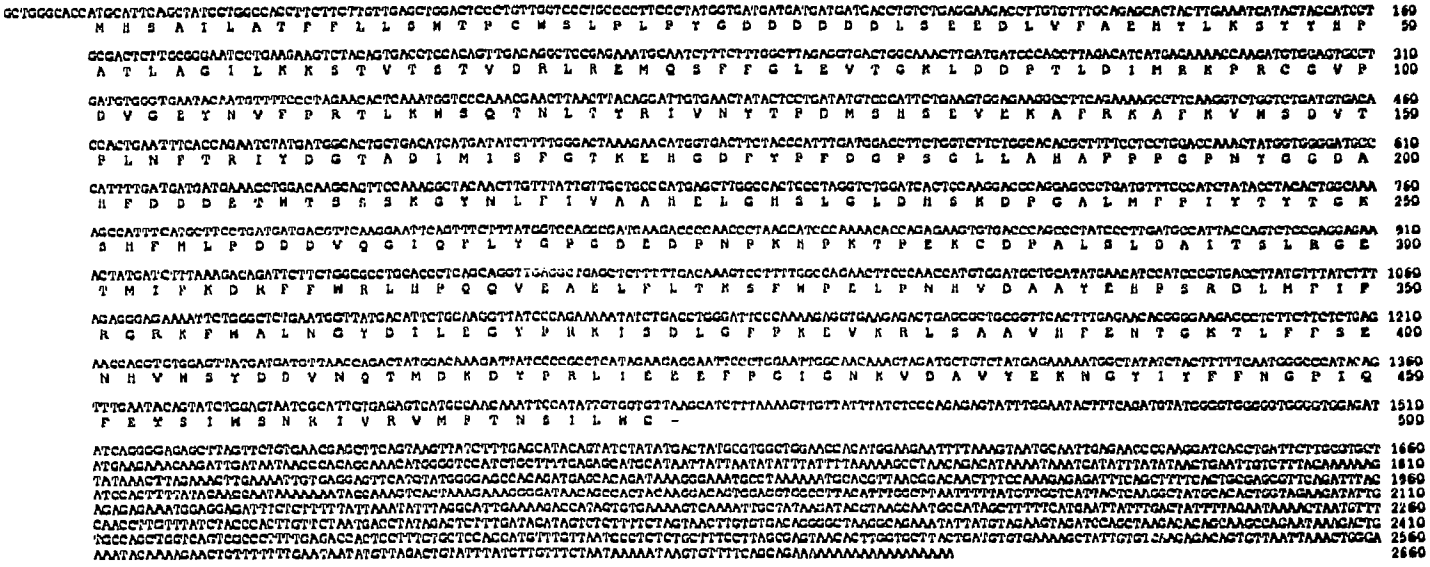

Fig. 2. Nucleotide and deduced amino acid sequence of mouse collagenase. GenBank/EMBL Database accession number X66473.

human neutrophil collagenase (MMP 8) and each of the six interstitial collagenases. This small percentage is also within the range of identity (43–62%) found when different MMPs from the same or from different species are compared, except the stromelysins 1 (MMP 3) and 2 (MMP 10) which share 72–80% identity (Table I, dotted box). The same conclusions were reached when sequences were matched by amino acid similarity instead of identity.

Rodent interstitial collagenases, but not the other rodent MMPs (gelatinase A, stromelysins 1 and 2), diverge from the non-rodent homologues studied here not only by their primary sequence but also by the 3′ UTR of their cDNAs, suggesting the existence of two sub-

Table I

Comparison between amino acid sequences of the active form of metalloproteinases

		MMP 1				MMP 1		MMP 8	MMP 3				MMP 10		MMP 2		ME
		human	rabbit	porcine	bovine	mouse	rat	human	human	rabbit	mouse	rat	human	rat	human	mouse	mouse
mouse	ME	46	46	47	47	45	45	45	51	51	49	48	49	49	43	43	100
mouse	MMP 2	44	44	44	46	48	49	44	47	46	46	48	45	47	**98**	100	
human	MMP 2	45	44	45	47	48	49	45	46	46	46	48	45	47	100		
rat	MMP 10	55	52	55	55	50	51	52	72	72	73	73	**77**	100			
human	MMP 10	54	53	55	56	53	54	52	79	80	75	73	100				
rat	MMP 3	56	54	56	56	50	51	51	**76**	**78**	**90**	100					
mouse	MMP 3	57	55	57	58	50	50	52	**79**	**79**	100						
rabbit	MMP 3	56	55	56	57	54	54	54	**85**	100							
human	MMP 3	56	55	56	57	54	54	54	100								
human	MMP 8	60	61	62	62	53	52	100									
rat	MMP 1	**52**	**52**	**54**	**55**	**97**	100										
mouse	MMP 1	**53**	**52**	**54**	**55**	100											
bovine	MMP 1	**88**	**85**	**87**	100												
porcine	MMP 1	**87**	**85**	100													
rabbit	MMP 1	**88**	100														
human	MMP 1	100															

Each value represents the percentage of identity between a sequence indicated at the top of the table and a sequence indicated on the left. Comparison scores between homologous MMPs in different species are in bold. Closed and dotted boxes associate percentages of special interest that are discussed in the text.

groups of interstitial collagenases. Subgroup I includes the rat and mouse interstitial collagenases, subgroup II includes those from human, rabbit, porcine and bovine origin. The 3′ UTR sequences of subgroup I could be aligned by considering 99% of the rat sequence and 96% of the mouse sequence; those of the four members of subgroup II could be aligned by considering 61–88% of their sequences. This comparison showed that the aligned 3′ UTR sequences of the two members of subgroup I share 85% identity and those of the four members of subgroup II, 64–79% identity. Moreover, the rodent 3′ UTRs are longer (1.2 kb) than those of subgroup II (≤ 0.6 kb), accounting for the difference in length of their mRNAs, and no identity was found between any segment of 3′ UTRs of members of the two subgroups.

There are four potential glycosylation sites in the rodent interstitial collagenases and two in those of the other subgroup. The apparent M_r of the proenzyme on SDS-PAGE is higher (66 and 62 kDa) for the rodent than for the other interstitial collagenases (55 and 52 kDa), consistant with a more abundant glycosylation. The interactions of rat [12] and mouse (not shown) collagenases with ion-exchange gels and heparin also differ from those of the other interstitial collagenases.

All vertebrate collagenases, but no other known vertebrate proteinase, cleave the triple helical molecule of collagen types I, II and III at a specific site, generating a three quarter- and a one quarter-length fragment. However, the rodent (rat and mouse) interstitial collagenases, the other mammalian ones and the human neutrophil collagenase differ from each other in their substrate specificity [13–16], for example towards soluble type I. II and III collagens. Rodent collagenases have no marked preference, while the catalytic rates (k_{cat}/K_m) for the degradation of soluble collagens are in the order type III (100%), type I (7%), type II (0.2%) for the interstitial collagenases but in the order type I (100%), type II (20%), type III (5%) for neutrophil collagenase.

In conclusion, the cloning of mouse collagenase cDNA and the results of the comparisons made here show that the rodent (mouse and rat) interstitial collagenases are very similar but differ more from the other interstitial collagenases than does neutrophil collagenase. In terms of primary sequence, they are as different from the other mammalian collagenases as are the stromelysins. The divergence of mouse and rat collagenases from the other mammalian ones supports the hypothesis, based on a phylogenetic analysis of DNA sequences [17], that the order Rodentia is an outgroup to the other eutherian (placental) mammalian orders. The other rodent MMPs would have escaped this evolutionary drift through a process of concerted evolution [18] like, for instance, the genes of the growth hormone family [19]. Another possibility would be that present-day rodent collagenase has evolved, as a new enzyme subgroup (then requiring a distinct MMP number), from a gene that separated from the ancestor of the MMP 1 gene [20] prior to the emergence of Rodentia. This raises the intriguing possibility that members of this new subgroup could coexist with MMP 1 in non-rodent species. In any case, our observations raise the issue of reconsidering the distinction between neutrophil (EC 3.4.24.34) and interstitial (EC 3.4.24.7) collagenases.

Acknowledgements: This work was supported by the Fund for Medical Scientific Research (Belgium) and by the Belgian State Prime Minister's Office, Science Policy Programming (inter-university attraction poles, Grant 7bis, and concerted actions, Grant 88/93–122). P.H. is a Research Fellow of the Institut pour l'Encouragement de la Recherche Scientifique dans l'Industrie et l'Agriculture. The generous gifts of a cDNA clone for rat collagenase from Dr. J. Jeffrey and for human collagenase from Dr. H. Nagase are greatly appreciated. The authors are grateful to Drs. K. Crepin, M. Darville, P. Michels, J.N. Octave and F. Opperdoes for their methodological advice and to M.C. Baelden, F. Chenut, Y. Marchand and E. Stevens for their expert assistance.

REFERENCES

[1] Matrisian, L.M. (1990) Trends Genet. 6, 121–125.
[2] Docherty, A.J.P. and Murphy, G. (1990) Annu. Rheum. Dis. 49, 469–479.
[3] Woessner Jr., J.F. (1991) FASEB J. 5, 2145–2154.
[4] Vaes, G. (1972) Biochem. J. 126, 275–289.
[5] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2.d edn., Cold Spring Harbor Laboratory Press.
[6] Quinn, C.O., Scott, D.K., Brinckerhoff, C.E., Matrisian, L.M., Jeffrey, J.J. and Partridge, N.C. (1990) J. Biol. Chem. 265, 22342–22347.
[7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[8] Vaes, G., Eeckhout, Y., Lenaers-Claeys, G., François-Gillet, C. and Druetz, J.E. (1978) Biochem. J. 172, 261–274.
[9] Hammani, K., Henriet, P.M. and Eeckhout, Y. (1992) Gene (in press).
[10] Brenner, C.A., Adler, R.R., Rappolee, D.A., Pedersen, R.A. and Werb, Z. (1989) Genes Dev. 3, 848–859.
[11] Kozak, M. (1991) J. Biol. Chem. 266, 19867–19870.
[12] Roswit, W.T., Halme, J. and Jeffrey, J.J. (1983) Arch. Biochem. Biophys. 225, 285–295.
[13] Welgus, H.G., Kobayashi, D.K. and Jeffrey, J.J. (1983) J. Biol. Chem. 258, 14162–14165.
[14] Welgus, H.G., Fliszar, C.J., Seltzer, J.L., Schmid, T.M. and Jeffrey, J.J. (1990) J. Biol. Chem. 265, 13521–13527.
[15] Hasty, K.A., Jeffrey, J.J., Hibbs, M.S. and Welgus, H.G. (1987) J. Biol. Chem. 262, 10048–10052.
[16] Netzel-Arnett, S., Fields, G., Birkedal-Hansen, H. and Van Wart, H.E. (1991) J. Biol. Chem. 266, 6747–6755.
[17] Li, W.H., Gouy, M., Sharp, P.M., O'hUigin, C. and Yang, Y.W. (1990) Proc. Natl. Acad. Sci. USA 87, 6703–6707.
[18] Lewin, B. (1990) in: Genes IV, pp. 515–516, Oxford University Press, Oxford.
[19] Miller, W.L. and Eberhardt, N.L. (1983) Endocrine Rev. 4, 97–130.
[20] Murphy, G.J.P., Murphy, G. and Reynolds, J.J. (1991) FEBS Lett. 289, 4–7.