Volume 289, number 1, 4–7 FEBS 10077 September 1991
© 1991 Federation of European Biochemical Societies 00145793/91/$3.50
ADONIS 00145793910078 8H

Discussion Letter

The origin of matrix metalloproteinases and their familial relationships

George J.P. Murphy[1], Gillian Murphy[2] and John J. Reynolds[2]

[1]*Institute of Plant Science Research, Cambridge Laboratory, Norwich, NR4 7UJ, UK* and [2]*Cell and Molecular Biology Department, Strangeways Research Laboratory, Cambridge, CB1 4RN, UK*

Received 4 June 1991

New computer comparisons of the sequences of mammalian matrix metalloproteinases have established for the first time strong links with bacterial metalloproteinases. We also propose that there are five groups in the family of matrix metalloproteinases, although only three are as yet well-characterized as proteins, and discuss their origin and relationships with other zinc containing proteases.

Matrix metalloproteinase; Sequence alignment; Phylogeny

The family of matrix metalloproteinases (MMPs), which are synthesized primarily by connective tissues, is of great importance in the initial events leading to tissue degradation, both in physiological and pathological situations. Three major groups of MMPs have been well-characterized, each group having more than one distinct gene product, which can be distinguished on immunological and biochemical criteria as well as sequence data [1–3]. The specific collagenases have interstitial collagens almost uniquely as substrates; the second group of MMPs are often referred to as gelatinases because they degrade denatured collagens very efficiently, but are also known as type IV collagenases, degrading native type IV collagen in a relatively specific fashion; members of the third family group, which are now generally referred to as stromelysins, are MMPs which have quite broad proteolytic action but which were originally described as proteoglycanases. Another less clearly characterized metalloproteinase, PUMP, has been described and its sequence indicates that it may be the first member of a fourth group.

Very recently Basset et al. [4] described in detail their interesting findings of a novel putative metalloproteinase gene expressed in stromal cells of human breast carcinomas. Although this putative proteinase was not itself isolated, the mRNA was also found to be expressed in the uterus, placenta and embryonic limb bud, as well as phorbol ester of growth factor stimulated fibroblasts. Based on its sequence and possible function in the progression of malignancy Basset et al. [4] proposed that this gene codes for a new member of the family of MMPs, named 'stromelysin 3'. However, no data on its substrate specificity is available.

Correspondence address: G. Murphy, Strangeways Research Laboratory, Cambridge, CB1 4RN, UK. Fax: (44)(223)411609.

Many mammalian MMPs have now been cloned and sequenced and they show a highly extended degree of similarity with each other (reviewed in [1]). Metalloproteinases with other functions than MMPs, and from many species, have been similarly characterized. Attempts at identification of the basic features of MMP sequences that might be involved in the catalytic mechanism have been based on the detailed studies of the better known bacterial metalloendopeptidases, notably thermolysin, and have preceded any analyses of the MMPs themselves.

In an earlier analysis [5] we suggested that the only feature of MMPs in common with thermolysin was the zinc binding motif HEXXH. With currently published data, within all the known zinc metalloprotease sequences, three groups of enzymes can be delineated according to the presence of absence of identifiable catalytic site motifs. Proteases derived from *Bacillus* sp. and *Pseudomonas* elastase, as well as forms of amino-peptidase N, LTA_4 hydrolase, enkephalinases and angiotensin converting enzymes [9, and references therein], have both the HEXXH motif identified in thermolysin as the two histidine Zn^{2+} binding ligands and the glutamic acid involved in the catalytic mechanism, and a second sequence containing the glutamic acid corresponding to a third Zn^{2+} ligand [9]. A second group including *Serratia* proteinase, bone morphogenetic protein I and the *Erwinia* proteinases B and C contain these motifs as well as a third motif adjacent to the zinc binding site, (HEXXH)XXGXXH, whose function is not known. The third group, which includes the MMPs, *Astacus* proteinase and the snake venom proteinases, contain the zinc box and the adjacent motif of the second group

Published by Elsevier Science Publishers B.V.

Table I

Alignments of metalloproteinases with the new gene product [4].

	Region homologous to PUMP			Full length sequence		
	No. of residues used	No. identical	SD	No. of residues used	No. identical	SD
Stromelysin 3	168			484		
Stromelysin 1	172	86	20.0	378	171	19.9
Stromelysin 2	172	86	22.6	377	166	22.2
Collagenase	170	85	22.6	370	163	22.0
72 kDa Gelatinase	176	86	21.4	382	173	24.6
95 kDa gelatinase	172	73	19.6	431	174	16.5
PUMP	172	73	19.6	172	82	5.6
Erwinia B	222	61	7.1	434	121	3.7
Thermolysin	202	47	3.8	315	93	0.3

The metalloproteinases were aligned with 'stromelysin 3' [4], using either the regions homologous to PUMP or with the full length sequence. SD is the number of standard deviations of the real score above the random scores. The higher the value the greater the similarity to MMP11.

but have no identifiable glutamic acid corresponding to the third Zn^{2+} ligand in thermolysin. Consequently a great deal of speculation currently surrounds the identity of the third Zn^{2+} ligand in these proteinases [6–10] and their possible relationships.

Using computer analyses we have now extended the comparison between the MMPs and other metallo-proteases in an attempt to gain further information on the origin and relationships of these enzymes. Thermolysin was included in the comparisons, together with the sequence of the metalloproteinase B derived from the plant bacterium *Erwinia chrysanthemi* [11]. The catalytic domain of the MMPs can be defined by that of the punctuated (previously putative) metalloproteinase, PUMP, which lacks the C-terminal domain present in all the other known MMPs (reviewed in [1,3]). The regions corresponding to the sequence of PUMP in both stromelysin-1 and interstitial collagenase have been shown to retain catalytic activity [1]. Since we wished to compare only the catalytic domains, the fibronectin-like domains of the gelatinases were not included in the sequence comparison, since these appear to have arisen by recombination of a common ancestor with fibronectin [12]. Comparison of the sequence of the catalytic domain of the proposed new 'stromelysin 3' [4] with those of the other members of the family (Table I) indicates that it might better be described as the first member of a new sub-group. The similarity between pairs of sequences was tested using the program ALIGN [13], employing the PAM250 matrix and with a gap penalty of three. To test the statistical significance of the alignments the sequences were randomized and realigned 100 times. The matrix bias parameter was varied until the number of standard deviations of the real score over the random score (SD) reached a maximum. Stromelysin 1 (MMP3) and stromelysin 2 (MMP10) are very similar, in that when compared with the methods used to produce Table I the pairwise SD value is 45 for the catalytic domain and 62.8 for the full-length sequence. In contrast the catalytic region of 'stromelysin 3' shows significant similarity to all the other known matrix metalloproteinase sub-groups, with the highest homology, as indicated by the SD value, being to collagenase and stromelysin 2. This was also true when the full-length sequence was compared, with the highest homology being to the 72 kDa gelatinase rather than to the stromelysins. The homology to thermolysin and the *Erwinia* protease was lower, though still statistically significant. Since the precise biochemistry and function of the new enzyme are unknown we suggest that it be systematically named matrix metalloproteinase 11 (MMP11), although future work may suggest a more descriptive title.

The conclusion that 'stromelysin 3' (MMP11) represents the first member of a new subgroup was reinforced when the sequences of the catalytic domains were subsequently used to derive a phylogenetic tree. The tree was constructed using TREEALIGN [14], with a gap penalty function of 1 + (length of insertion or deletion) ×3

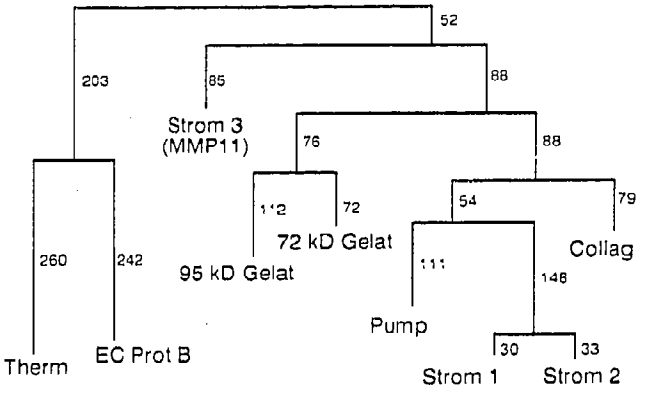

Fig. 1. Phylogenetic tree for metalloproteinases. The most parsimonious tree is shown here, with the branch lengths being indicated on each branch. The tree is rooted at the midpoint of the longest branch and is calculated assuming that all branches evolve at the same rate.

```
Strom 3 (MMP11)  FVLSGG--RWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADI
Strom 1          FRTFPGIPKWRKTHLTYRIVNYTPDLPKDAVDSAVEKALKVWEEVTPLTFSRLYEGEADI
Strom 2          FSSFPGMPKWRKTHLTYRIVNYTPDLPRDAVDSAIEKALKVWEEVTPLTFSRLYEGEADI
Collag           FVLTEGNPRWEQTHLTYRIENYTPDLPRADVDHAIEKAFQLWSNVTPLTFTKVSEGQADI
72kD Gelat       YNFFPRKPKWDKNQITYRIIGYTPDLDPETVDDAFARAFQVWSDVTPLRFSRIHDGEADI
95kD Gelat       FQTFEGDLKWHHHNITYWIQNYSEDLPRAVIDDAFARAFALWSAVTPLTFTRVYSRDADI
PUMP 1           YSLFPNSPKWTSKVVTYRIVSYTRDLPHITVDRLVSKALNMWGKEIPLHFRKVVWGTADI
                 +       +*    +****  ++ +*    +    + ** +**  +** *  +    ***

Strom 3 (MMP11)  MIDFARYWDGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAH
Strom 1          MISFAVREHGDFYPFDGPGNVLAHAYAPGPGINGDAHFDDDEQWT-KDTTGTNLFLVAAH
Strom 2          MISFAVKEHGDFVSFDGPGHSLAHAYPPGPGLYGDIHFDDDEKWT-EDASGTNLFLVAAH
Collag           MISFVRGDHRDNSPFDGPGGNLAHAFQPGPGIGGDAHFDEHERWT-NNFTEYNLHRVAAH
72kD Gelat       MINFGRWEHGDGYPFDGKDGLLAHAFAPGTGVGGDSHFDDDELWTLGEGQGYSLFLVAAH
95kD Gelat       VIQFGVAEHGDGYPFDGKDGLLAHAFPPGPGIQGDAHFDDDELWSLGKGVGYSLFLVAAH
PUMP 1           MIGFARGAHGDSYPFDGPGNTLAHAFAPGTGLGGDAHFDEDERWTDGSSLGINFLYAATH
                 +* *    + *  +*** +  *****+ * +   ** *** +* **  +  + ++  +***

Strom 3 (MMP11)  EFGHVLGLQHTTAAKALMSAFYTFRYP----LSLSPDDCRGVQHLYGQP--WPTVT
Strom 1          EIGHSLGLFHSANTEALMYPLYH-SLTDLTRFRLSQDDINGIQSLYGPPPDSPET-
Strom 2          ELGHSLGLFHSANTEALMYPLYN-SFTELAQFRLSQDDVNGIQSLYGPPPASTEEP
Collag           ELGHSLGLSHSTDIGALMYPSY--TFSG--DVQLAQDDIDGIQAIYGRS-QNPVQP
72kD Gelat       EFGHAMGLEHSQDPGALMAPIYT--YTK--NFRLSQDDIKGIQELYGASPD-IDLG
95kD Gelat       EFGHALGLDHSSVPEALMYPMYR--FTE--GPPLHKDDVNGIRHLYGPRPE-PEPR
PUMP 1           ELGHSLGMGHSSDPNAVMYPTYG-N-GDPQNFKLSQDDIKGIQKLYGKRSNSRKK-
                 *+** +*+ **    *+* + *          *  **  *++ +** +
```

Fig. 2. Multiple alignments of metalloproteinases. The regions of the matrix metalloproteinases homologous to PUMP-1 were aligned. Residues identical in all sequences are marked with an *, while those in which conservative substitutions have occurred are indicated by +.

```
      ~              ~#           ~     ##~#   ~          ~ ~#
A   1 FV----LSGG------RWE-------KTDL-----TYRILR-F---------PWQLV-QE  27
B   1 ITGTSTV--GVGR--------GVLGDQKNINTTYSTYYYLQ------DNTRGDGIFTYDA  44
C   1 YTTDKAVSEGLTRPHTTWNGDNVFGKAANL-----TYSFLNTFSSTPNGHTGPVKFT-PV  54
       +    +  *                    ++      **  *+               ++

       ~    ~ #~ ~#~  ~## #  ~     ###~# #    # #  ~###   ~  ####~
A  28 QVRQTMAEALKVWSDVTPLTFTEV-HEGRADIMIDFARYWHGDDLPFDGP--GGILAHAF  84
B  45 KYRTTLPGSL--WADADN-QF-------------FASY---DAPAVDAHYYAGV---TY  81
C  55 QMQQAKL-SLQSWADAANLTFTEVSPNQKANI--TFANYTRNADGSLNTD------TQAY 105
      + + +   +*  *+*++   *                ***+   ++ + ++         ++

        # ~               ## ###   ~#  #~      ~   ~ ~~  ~#      ~
A  85 --FPKTHR----------EGDVHFD--YDET-WT-----IGDDQGTDLLQVA------A 117
B  82 DYYKNVHNRLSYDGNNAAIRSSVHYSQGYNNAFWNGSEMVYGDGDGQTFIPLSGGIDVVA 141
C 106 AAYPGTHP----------VSGSAWFN--YNQS--S-----IRNPDTDEYGRHS-----FT 141
        +  +*             +++ ++  *+++  +       + ++ +  + +       +

      ##~## ~#~ #~    #                    ~# ~ :
A 118 HEFGHVLGLQHTTAAKA----------------------LMSAFY---------TFR-- 143
B 142 HELTHAV-----TDYTAG---LIYQNESGAINEAISDIFGTLVEFYANKN---------- 184
C 142 HEIGHALGLSHPAEYNAGEGDISYKNSAAYAED--SRQFSIMS--Y-WEVENTGGDFKGH 196
      *++++++     ++ +*                        +   *

         #  ##  #~~ ~## ~
A 144 YPLSLSPDDCRGVQHLYGQPWPTVT- 168
B 185 -------PD-WEIGEDVYTPGISGDSL 202
C 197 YSAGPLMDDIAAIQKLYGANMTTRTG 222
            *     +++*++  ++ +
```

Fig. 3. Comparison of stromelysin 3 (MMP11) with bacterial metalloproteinases. The sequences aligned were: A, Stromelysin 3 (MMP11); B, thermolysin and C, *Erwinia* protease B, aligned as in Fig. 2. Residues identical in all three sequences are marked with an * below the sequences, while those in which conservative substitutions have occurred are indicated by +. Residues in MMP11 common to all MMPs are indicated above the sequence with a #, while conservative substitutions are indicated by ~. The ungapped residue numbers are indicated on the left and right margins. Residues conserved across al sequences are in bold.

Fig. 1 indicates that in the most parsimonious tree, the new MMP [4] did not group with any of the other MMP sequences, forming instead a separate branch between the thermolysin and *Erwinia* sequences and those of the other MMPs. It thus provides a link between the MMPs previously known [1] and the well-known bacterial metalloproteinase, thermolysin. Previously, algorithms used to detect evolutionary relationships between zid proteases failed to define the presence of such a link [15].

We have also compared in more detail the predicted amino acid sequence of a number of human MMPs including the new gene product [4] in a multiple alignment, using regions corresponding to PUMP. Multiple alignments were carried out using CLUSTAL [16]. A pairwise gap penalty of 1 was used, with multiple alignment gap penalties of 7 (fixed) and 9 (varying). The alignments (Fig. 2) clearly demonstrate that the similarities extend over far greater regions of the molecules than just the zinc binding motif previously identified. Since the phylogenetic tree suggests that MMP11 is the MMP nearest to thermolysin in an evolutionary sense, it was aligned with thermolysin and the *Erwinia* protease (Fig. 3). Although only twelve residues are conserved throughout the sequences, a distinct clustering of conservatively replaced residues can be observed. These comparisons establish for the first time an extended relationship between the mammalian MMPs and a bacterial metalloproteinase, although it does not involve the other catalytic site motifs identified in thermolysin [17]. Thermolysin has other regions of similarity with the *E. chrysanthemi* enzymes but not all of these have similarity with the mammalian MMPs.

Further detailed computer analysis should be rewarding now that the human MMPs can clearly be related back to enzymes of primitive origin. Such work could not only give new insights into the molecular mechanisms of action of MMPs and their functions but also provide a considerable challenge to the computer expert.

REFERENCES

[1] Docherty, A.J.P. and Murphy, G. (1990) Ann. Rheum. Dis. 49, 469–479.
[2] Murphy, G., Reynolds, J.J. and Hembry, R.M. (1989) Int. J. Cancer 44, 757–760.
[3] Matrisian, L.M. (1990) Trends Genet. 6, 121–125.
[4] Basset, P., Bellocq, J.P., Wolf, C., Stoll, I., Hutin, P., Limacher, J.M., Podhajcer, O.L., Chenard, M.P., Rio, M.C. and Chambon, P. (1990) Nature 348, 699–704.
[5] Whitham, S.E., Murphy, G., Angel, P., Rahmsdorf, H.-J., Smith, B.J., Lyons, A., Harris, T.J.R., Reynolds, J.J., Herrlich, P. and Docherty, A.J.P. (1986) Biochem. J. 240, 913–916.
[6] Vallee, B.L. and Auld, D.S. (1990) Biochemistry 29, 5647–5659.
[7] Springman, E.B., Angleton, E.L., Birkedal-Hansen, H. and Van Wart, H.E. (1990) Proc. Natl. Acad. Sci. USA 87, 364–368.
[8] 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.
[9] Shannon, J.D., Baramova, E.N., Bjarnason, J.B. and Fox, J.W. (1989) J. Biol. Chem. 264, 11575–11583.
[10] Takeya, H., Onikura, A., Nikai, T., Sugihara, H. and Iwanaga, S. (1990) J. Biochem. 108, 711–719.
[11] Delepelaire, P. and Wandersman, C. (1989) J. Biol. Chem. 264, 9083–9089.
[12] Collier, I., Bruns, G.A.P., Goldberg, G.I. and Gerhard, D.S. (1991) Genomics 9, 429–434.
[13] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.
[14] Hein, J. (1990) Methods Enzymol. 183, 626–645.
[15] Jongeneel, C.V., Bouvier, J. and Bairoch, A. (1989) FEBS Lett. 242, 211–214.
[16] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237–244.
[17] Weaver, L.H., Kester, W.R. and Matthews, B.W. (1977) J. Mol. Biol. 114, 119–132.