Bacterial aspartic proteinases

Jeffrey Hill *, Lowri H. Phylip

School of Molecular and Medical Biosciences, University of Wales, P.O. Box 911, Cardiff CF1 3US, Wales, UK

Received 2 April 1997; revised version received 28 April 1997

Abstract Regions of genomic DNA encoding putative aspartic proteinase domains were amplified by PCR from the bacterial species, *Escherichia coli* and *Haemophilus influenzae*. Expression of each of these DNA fragments resulted in the accumulation of the corresponding recombinant proteins in insoluble aggregates. Each recombinant protein was solubilised, refolded and shown to be able to cleave synthetic peptides that have been extensively used previously as substrates for aspartic proteinases of vertebrate, fungal and retroviral origin. Each activity was completely blocked by the diagnostic aspartic proteinase inhibitor, acetyl-pepstatin. This is thus the first report demonstrating unequivocally that aspartic proteinases may be present in bacteria.

Key words: Bacterial aspartic proteinase; Expression; Refolding; Molecular evolution; Antibacterial target

1. Introduction

The active site cleft of aspartic proteinases lies between two domains, each of which contributes an aspartic acid residue to the catalytic mechanism [1]. Each aspartic acid residue is found in the characteristic sequence motif ~Hydrophobic–Hydrophobic–Asp–Thr/Ser–Gly~ which, together with a further ~Hydrophobic–Hydrophobic–Gly~ sequence motif forms a structural feature known as a psi-loop [2]. In some enzymes, the two domains are contained within a single polypeptide chain so that each of the two motifs occurs twice within the molecule [3]. In contrast, some aspartic proteinases are homodimers consisting of identical monomers and in these cases, the ~Hydrophobic–Hydrophobic–Asp–Thr/Ser–Gly~ and ~Hydrophobic–Hydrophobic–Gly~ motifs are encoded only once within the corresponding gene [2]. This is the case, for example, with aspartic proteinases encoded within viruses which infect vertebrates, e.g. HIV-1, [4] and plants, e.g. cocao swollen shoot virus [5]. Aspartic proteinases have been found in a wide variety of organisms, ranging from viruses, fungi and parasites to plants, fishes and higher vertebrates including man [6]. Hitherto, however, it has been widely held that bacteria do not produce aspartic proteinases [7].

With the advent of genome sequencing projects, it is now possible to scrutinise bacterial genomes deposited in public domain databases for the occurrence together of the two hallmarks of aspartic proteinases, i.e. the ~Hydrophobic–Hydrophobic–Asp–Thr/Ser–Gly~ and ~Hydrophobic–Hydrophobic–Gly~ motifs. In this report, we show that such information is encoded within the genomes of two bacteria, *Escherichia coli* and *Haemophilus influenzae* and show that the recombinant proteins resulting from expression of each of these DNA regions, are active aspartic proteinases.

2. Materials and methods

Genomic DNA prepared from *H. influenzae* and *E. coli* cells [8] was used as template for PCR. Oligonucleotide primers (incorporating restriction sites for *Bam*HI and *Hin*dIII and stop codons) were designed to amplify each region of DNA encoding the putative aspartic proteinases:

5′-GG ATC CGA TTT ACG ATT TTG TTG TTT AAA C-3′	*H. inf* forward primer
5′-AA AGC TTA AGA ATG GGA ATT ATC AGC TC-3′	*H. inf* reverse primer
5′-CG GAT CCG AGG TCT AAA CCG TGG TCG T-3′	*E. coli* forward primer
5′-GA AGC TTA CAC CGA GAT AAT GGT TTC C-3′	*E. coli* reverse primer

Each PCR product generated was cloned into the pGem-T vector system (Promega, Southampton, UK) and its nucleotide sequence was determined using a Pharmacia ALF automated sequencer. Both sequences were identical to those deposited in the databases. Each DNA fragment was excised with *Bam*HI and *Hin*dIII and cloned into the T7 expression vector pET23c (AMS Biotech., Witney, Oxon, UK) that had been cut with the same enzymes.

Expression of the putative aspartic proteinases was carried out in *E. coli* strain BL21(DE3)pLysS as described previously [9]. The inductions were initiated by the addition of 1 mM IPTG and the cells were harvested after 2 h. Repeated freeze–thaw cycles followed by sonication were used to ensure complete cell lysis. After centrifugation (7000×*g* for 30 min), the respective cell pellets were washed in buffer A (100 mM Tris-HCl, pH 11.0). The insoluble material from each induction was then dissolved in buffer B (6 M urea, 100 mM Tris-HCl, pH 8.0, 1 mM glycine, 1 mM EDTA).

The two recombinant polypeptides produced were refolded under differing conditions. The *H. influenzae* protein in buffer B was subjected to a rapid dilution in buffer C (10 mM Tris-HCl, pH 8.0) to a final protein concentration of 5 μg/ml. The *E. coli* protein in buffer B was subjected to a rapid dilution in buffer D (100 mM Tris-HCl, pH 8.0, 0.3 mM GSH, 3 mM GSSG) also to a final protein concentration of 5 μg/ml. Both of the resultant solutions were stirred at 20°C for 16 h, re-concentrated using a Filtron Ultrasette 3 kDa cut-off, tangential-flow concentrator (Flowgen Instruments Ltd., Sittingbourne, Kent, UK) and finally in an Amicon stirred cell concentrator (Amicon, Gloucestershire, UK) to final protein concentrations of 1 mg/ml.

The proteolytic activity in each dilute and concentrated protein fraction was assessed using synthetic peptides as substrates:

Ac–Tyr–Arg–Ala–Arg–Val–Phe*NPhe–Val–Arg–Ala–Ala–Lys	(for *E. coli*)
Lys–Pro–Ile–Glu–Phe*NPhe–Arg–Leu	(for *H. influenzae*, where NPhe = *p*-nitrophenylalanine).

The assays were carried out at 37°C in a standard assay buffer (100 mM sodium acetate, pH 4.7, 4 mM EDTA) and the reaction products were analysed by reverse-phase FPLC. The effect of a diagnostic aspartic proteinase inhibitor was investigated by the addition of acetyl pepstatin to a final concentration of 1 μM.

SDS-PAGE was carried out on 15% polyacrylamide gels as de-

*Corresponding author. Fax: (44) 1222-874116.
E-mail: HILLJ@Cardiff.ac.uk

PII S0014-5793(97)00547-4

scribed by Laemmli [10]. Bands corresponding to the expected sizes of the expressed proteins were blotted onto PVDF membranes and loaded into an Applied Biosystems ABI 470 A sequenator for N-terminal sequence analysis by Edman Degradation.

3. Results and discusion

Both of the bacterial sequence entries (Gene Bank accession numbers ECOUW67 and L20805 for *E. coli* and *H. influenzae*, respectively) in the databases encode only one copy of the ~Hydrophobic–Hydrophobic–Asp–Thr/Ser–Gly~ and ~Hydrophobic–Hydrophobic–Gly~ motifs. On this basis, it would be necessary for these putative bacterial aspartic proteinases to exist as homodimers in order to generate catalytic activity, as has been described in detail for the viral aspartic proteinases [11]. In retroviral proteinases, the active site aspartic acid residue is located at approximately residue 25 in the protein sequence (depending on the virus) with the glycine of the ~Hydrophobic–Hydrophobic–Gly~ in the thread of the psi-loop located a further ~60 residues downstream at position ~86 [12]. Inspection of the *E. coli* and *H. influenzae* protein sequences (Fig. 1) shows that approximately 60 residues are interposed also between these two motifs in the bacterial proteins (Fig. 2). It was thus possible to estimate a minimal length for each bacterial (putative) aspartic proteinase domain by comparison with the known cleavage junctions at the N- and C-termini of retroviral aspartic proteinases which consist of 100–130 residues [12]. On this basis, specific primers were designed to enable bacterial DNA fragments to be amplified that were longer than this minimal estimate. For *E. coli*, a 444 bp fragment was amplified and cloned which encoded the 142 amino acids shown in the sequence in Fig. 1 while a 435 bp fragment (encoding 139 residues) was derived in parallel from *H. influenzae*.

Overexpression of each of these DNA fragments in *E. coli* resulted in the accumulation of each recombinant protein in the form of insoluble aggregates. Yields were estimated to be 20 and 50 mg/l of culture, respectively, for the recombinant *E. coli* and *H. influenzae* proteins. Each protein was detected by Western blotting using a monoclonal antibody against the 13 N-terminal amino acids (T7 tag) encoded by the expression vector. The protein aggregates were solubilised using a buffer containing 6 M urea (see Section 2). Aliquots were subjected to SDS-PAGE on a 15% gel followed by blotting of the appropriate recombinant protein band onto PVDF membrane for sequencing by Edman degradation. The sequences obtained confirmed that the T7 tag had been retained at the N-terminus of each recombinant protein:

E. coli:	Ala–Ser–Met~T7	tag~**Gly–Leu–Asn–Arg–Gly–Arg–Ser**~
H. influenzae:	Ala–Ser–Met~T7	tag~**Phe–Thr–Ile–Leu–Leu–Phe–Lys**~

Comparison of these sequences with those in Fig. 1 indicates that no proteolytic processing had taken place at the N-terminus of either recombinant protein while it was accumulating (in an insoluble form) in the heterologous cells.

Each of the 6 M urea solutions containing the respective recombinant proteins was used in a variety of experiments that examined a number of different protocols in attempts to refold each protein. Eventually, different protocols were

```
E.coli   ~G L N R G R S M D Q I N - - - - - - D T L V S
H.inf.     ~F T I L L F K Q L S K P T W Q V D T L D V

E.coli     S I K I I A M M L L I I G G G G A F K Q V L V
H.inf.     G Q - - - G L A T L I V K N G - - - K G I L Y

E.coli     D S G V D K Y I A S M M H E T N I S P L L M A
H.inf.     D T G S S W R G G S M A - E L E I L P Y L Q R

E.coli     W S I A A V L - R I I A L G S A T V A A I T A
H.inf.     E G I - - V L E K L I L - - S H D D N D - H A

E.coli     G G I A A - - - - - P - - - L I A T T - G V S
H.inf.     G G A S T I L K A Y P N V E L I T P S R K N Y

E.coli     P E L M V I - - A V G S G S V I - F S H V N D
H.inf.     G E N Y R T F C T A G R D W H W Q G L H F Q -

E.coli     P G F W L F K E Y F N L - T I G E T I K S W S M~
H.inf.     - - - I L S P H - - N V A T R A D N - - S H S~
```

Fig. 1. Amino acid sequences of the putative aspartic proteinase domains from *E. coli* and *H. influenzae*. The ~Asp–Thr/Ser–Gly~ motifs are shown in bold type and the ~Hydrophobic–Hydrophobic–Gly~ motif of the psi-loop is underlined. Gaps (indicated by a hyphen) are inserted to maximise the apparent similarity.

adapted for the two proteins. For the *E. coli* protein, it was necessary to include a combination of reduced and oxidised glutathione in the refolding buffer to prevent the reformation of insoluble aggregates [13]. In both cases, an extensive dilution (detailed in Section 2) step was employed. When aliquots of each of the recombinant proteins was assessed for ability to cleave synthetic peptides, no activity was detected at this dilution stage (final protein concentration = 5 μg/ml; see Section 2). Each dilute solution was concentrated by approximately 2000-fold and then assessed for activity again since it has been demonstrated frequently that the activity of homodimeric aspartic proteinases such as HIV proteinase are enzyme concentration dependent [14,15]. Since the purpose of these investigations was to determine whether the two bacterial proteins were catalytically active as aspartic proteinases, two synthetic chromogenic peptides were utilised in these assays which have been widely used in many previous studies to characterise retroviral [16] and vertebrate/fungal aspartic proteinases [17,18]. Consequently, the activity of the putative homodimeric enzyme from *E. coli* was tested towards the pseudo-symmetrical peptide Ac–Tyr–Arg–Ala–Arg–Val–Phe*NPhe–Val–Arg–Ala–Ala–Lys which has been demonstrated to be among the best substrates for symmetrical homodimeric enzymes, e.g. HIV proteinase [19]. Indeed this peptide was cleaved readily by the recombinant protein from *E. coli* (Fig. 3) and the cleaved products were eluted from the reverse-phase column in identical positions to the well-characterised products resulting from cleavage of this peptide substrate by HIV proteinase [19]. Thus, it is clear that the recombinant *E. coli* protein was attacking this substrate at the intended scissile peptide bond, i.e. between the Phe and *p*-nitrophenylalanine residues.

The activity of the *H. influenzae* recombinant protein was assessed using a peptide which has been demonstrated to be among the best substrates for archetypal aspartic proteinases, e.g. cathepsin E [17] and cathepsin D [18]. This peptide Lys–Pro–Ile–Glu–Phe*NPhe–Arg–Leu was completely cleaved by the *H. influenzae* protein at the intended scissile peptide bond (Fig. 3b) and the kinetic parameters estimated by the FPLC assay for this reaction at pH 4.7 were $K_m = 105 \pm 14$ μM and

V_m = 100 ± 2 nmol/min/mg protein. In contrast, little or no activity was detected towards the substrate Ac–Tyr–Arg–Ala–Arg–Val–Phe*NPhe–Val–Arg–Ala–Ala–Lys. Similarly, the recombinant *E. coli* protein had little ability to cleave the substrate Lys–Pro–Ile–Glu–Phe*NPhe–Arg–Leu. This suggests that the two bacterial enzymes differ in their substrate specificity, one preferring a retroviral type of substrate while the other prefers an archetypal type of substrate.

The cleavage of the retroviral proteinase substrate by the recombinant *E. coli* protein and the archetypal proteinase substrate by the *H. influenzae* protein were both completely prevented when the diagnostic inhibitor of aspartic proteinases, acetyl pepstatin [20] was included in the assay buffer at a final concentration of 100 nM (Fig. 3, bottom two panels).

This unequivocal demonstration of proteolytic activity intrinsic to each bacterial protein was further confirmed by SDS-PAGE analysis of each concentrated preparation after the the dilution and refolding steps. Less than 5% of the recombinant protein that had been present prior to these steps was recovered in the concentrated material in each case. Thus, it would appear that a large proportion of each recombinant bacterial protein was still misfolded and was digested by the (< 5%) of the protein which had folded properly to generate active proteinase. Autolysis of each correctly folded, active bacterial proteinase may also have taken place, just as has been observed with dimeric aspartic proteinases such as HIV proteinase which are known to undergo autolysis initially at specific sites within the proteinase molecule itself, but eventually leading to complete digestion of the enzyme [21]. Further experiments are now in progress to optimise the refolding conditions yet more, together with production of each recombinant bacterial protein on a larger scale. These should enable the purification of sufficient amounts of each bacterial proteinase to permit a more-detailed characterisation of each enzyme, including crystallisation trials, to be undertaken.

Nevertheless this is the first report identifying a 'typical aspartic proteinase' activity in species of bacteria. Until now, acid proteinase activity in bacteria has either been attributed to proteins that share no primary sequence homology whatsoever to known aspartic proteinases [22] or to proteinases with unknown primary amino acid sequences that are not inhibited by pepstatin [23]. Aspartic proteinases encoded by retroviruses, pararetroviruses [24] and retrotransposons [25] are known to be dimeric whereas eukaryotic aspartic proteinases are predominantly monomeric and are believed to have evolved from a gene duplication with fusion event that places two copies each of the ≐Hydrophobic–Hydrophobic–Asp–Ser/Thr–Gly∼ and ∼Hydrophobic–Hydrophobic–Gly∼ motifs into each gene [3]. Since both of the bacterial

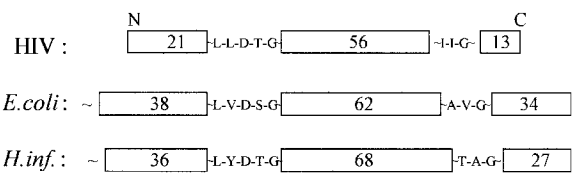

Fig. 2. The relative locations of the two sequence motifs in HIV proteinase and the putative aspartic proteinase domains from *E. coli* and *H. influenzae*. The number of residues adjacent to and in the sequence intervening between the ∼Hydrophobic–Hydrophobic–Asp–Thr/Ser–Gly∼ and ∼Hydrophobic–Hydrophobic–Gly∼ motifs are indicated within each solid block.

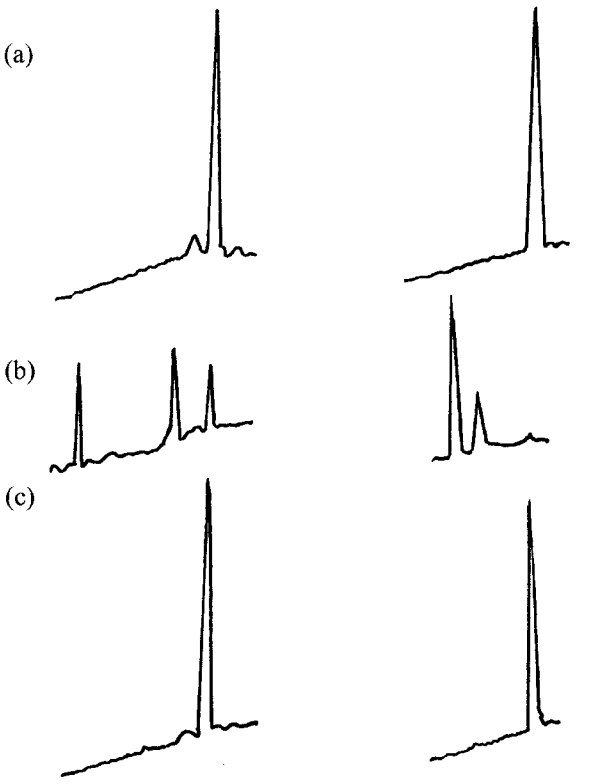

Fig. 3. Cleavage of peptide substrates by recombinant *E. coli* and *H. influenzae* proteins and the effect of pepstatin. The synthetic peptides: Ac–Tyr–Arg–Ala–Arg–Val–Phe*NPhe–Val–Arg–Ala–Ala–Lys (left panels) and Lys–Pro–Ile–Glu–Phe*NPhe–Arg–Leu (right panels) were incubated in 100 mM sodium acetate buffer, pH 4.7, at 37°C (a) alone or with recombinant *E. coli* (left) or *H. influenzae* proteins (right) in the absence (b) or presence (c) of acetyl pepstatin (100 nM).

sequences contain only one copy of each motif, dimeric bacterial aspartic proteinases may represent an important stage in the evolutionary pathway of this family of enzymes. The two DNA fragments that were amplified in this study were themselves only segments of larger open reading frames (1.3 and 2.4 kb, respectively) in the deposited *E. coli* and *H. influenzae* sequences. In these original depositions, it was suggested that these larger genes both encoded membrane-spanning transporter proteins; for gluconate in the case of *E. coli* and DNA in the *H. influenzae* case [26] but no functional evidence in support of this has appeared. Thus, it may be that the aspartic proteinase domain located towards the C-terminal end of the protein encoded by each of these larger genes may be part of of a multifunctional protein and that dimerisation is required not only to generate proteolytic activity but also may be intrinsic to the operation of the respective membrane-spanning transporters. Further work will be necessary to investigate this and to establish whether (and where) the proteinase domain processes itself out from the multifunctional precursor protein, perhaps in a parallel manner to that authenticated so meticulously for HIV polyprotein processing [27]. The significance of such events to bacterial viability also requires investigation, with the incentive of offering potential new antibacterial drug targets. Bacterial infections are becoming an increasingly serious problem with the increased incidence of bacterial meningitis and *E. coli* food poisoning and such problems are exacerbated by the decreasing effectiveness

of conventional antibiotics due to bacterial resistance. The discovery of bacterial members of a widely investigated enzyme family may offer tempting new drug targets.

Acknowledgements: The authors would like to thank Dave Williams (Nevil Hall Hospital, Abergavenny, Gwent, Wales) for supplying cultures of *H. influenzae*, Ben Dunn (University of Florida) for N-terminal sequencing analysis and for the generous gift of peptide substrates, Kohei Oda (Kyoto Institute of Technology) for kindly providing acetyl pepstatin and John Kay, of this School, for his assistance in the preparation of this manuscript.

References

[1] Abdel-Meguid, S.S. Med. Res. Rev., 13 (1993) 731–778.
[2] Pearl, L.H. and Taylor, W.R., Nature 329 (1987) 351–354.
[3] Tang, J., James, M.N.G., Hsu, I.N., Jenkins, J.A. and Blundell, T.L., Nature 271 (1978) 618–621.
[4] Tomasselli, A.G., Howe, W.J., Sawyer, T.K., Wlodawer, A. and Heinrikson, R.L., Chim. Oggi. 9 (1991) 6–27.
[5] Hagen, L.S., Jaquemond, M., Lepingle, A., Lot, H. and Tepfer, M., Virology 196 (1993) 619–628.
[6] Takahashi, K. (ed.) (1995) in: Aspartic proteinases: structure, function, biology and biomedical implications, pp. 1-550, Plenum Press, New York.
[7] Kay, J. (1985) in: Aspartic proteinases and their inhibitors (Kostka, V. ed.), pp. 1-17, Walter de Gruyter, New York.
[8] Hames, B.D. and Higgins, S.J. (eds.) (1985) Nucleic acid hybridisation, pp. 21, IRL Press, Oxford, UK.
[9] Hill, J., Tyas, L., Phylip, L.H., Kay, J., Dunn, B.M. and Berry, C., FEBS Lett. 352 (1994) 155–158.
[10] Laemmli, U.K., Nature 227 (1970) 680–685.
[11] Wlodawer, A., Miller, M., Jakolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. and Kent, S.B.K., Science 245 (1989) 6-6-621.
[12] Gustchina, A. Kervinen, J., Powell, D.J., Zdanov, A., Kay, J., Wlodawer, A., Prot. Sci. 5 (1996) 1453–1465.
[13] Rudolph, R., Lilie, H., FASEB J. 10 (1996) 49–55.
[14] Darke, P.L., Jordan, S.P., Hall, D.L., Zugay, J.A., Shafter, J.A., Kuo, L.C. Biochemistry 33 (1994) 98–105.
[15] Kuzmic, P., Garcìa-Echeverìa, C., Rich, D.H., Biochem. Biophys. Res. Commun. 194 (1993) 301–305.
[16] Dunn, B.M., Gustchina, A., Wlodawer, A., Kay, J., Meth. Enzymol. 241 (1994) 254–278.
[17] Rao-Naik, C., Guruprasad, K., Batley, B., Rapundalo, S., Hill, J., Blundell, T., Kay, J., Dunn, B.M., Prot.: Struct. Funct. Genet. 22 (1995) 168–181.
[18] Beyer, B.M., Dunn, B.M., J. Biol. Chem. 271 (1996) 15590–15596.
[19] Griffiths, J.T., Tomchak, L.A., Mills, J.S., Graves, M.C., Cook, N.D., Dunn, B.M., Kay, J., J. Biol. Chem. 269 (1994) 4787–4793.
[20] Satoi, S., Murao, S., Agric. Biol. Chem. 35 (1971) 1482–1487.
[21] Mildner, A.M., Rothpock, D.J., Leone, J.W., Bannow, C.A., Lull, J.M., Reardon, I.M., Sarcich, J.L., Howe, W.J., Tomich, C.S.C., Smith, C.W., Heinrikson, R.L., Tomasselli, A.G., Biochemistry 33 (1994) 9405–9413.
[22] Lin, X., Tang, J., J. Biol. Chem. 265 (1990) 1490–1495.
[23] Toogood, H.S., Prescott, M., Daniel, M., Biochem. J. 307 (1995) 783–789.
[24] Laco, G.S., Beachy, R.N., Proc. Natl. Acad. Sci. USA 91 (1994) 2654–2658.
[25] Kirchner, J. Sandmeyer, S., J. Virol. 67 (1993) 19–28.
[26] Clifton, S.W., McCarthy, D., Roe, B.A., Gene 146 (1994) 95–100.
[27] Pettit, S.C., Moody, M.D., Wehbie, R.S., Kaplan, A.H., Nantermat, P.V., Klein, C.E., Swanstrom, R., J. Virol. 68 (1994) 8017–8027.