Gene, 69 (1988) 237–244
Elsevier

GEN 02559

Molecular analysis of the gene encoding α-lytic protease: evidence for a preproenzyme

(Recombinant DNA; extracellular serine protease; *Lysobacter enzymogenes*; cloning in *E. coli*; S1 mapping; sequencing high GC DNA; subtilisin)

J.L. Silen, C.N. McGrath *, K.R. Smith and D.A. Agard

Department of Biochemistry and Biophysics and The Howard Hughes Medical Institute, University of California at San Francisco, San Francisco, CA 94143-0448 (U.S.A.) Tel. (415)476-5143

Received 22 February 1988
Accepted 12 May 1988
Received by publisher 6 June 1988

SUMMARY

A 1.7-kb *Eco*RI fragment containing the structural gene for α-lytic protease has been cloned from *Lysobacter enzymogenes* 495 chromosomal DNA: the first example of a gene cloned from this organism. The protein sequence deduced from the nucleotide sequence encoding this serine protease matches the published amino acid sequence [Olson et al., Nature 228 (1970) 438–442] precisely. Sequence analysis and S1 mapping indicate that, like subtilisin [e.g., Wells et al., Nucleic Acids Res. 11 (1983) 7911–7925] α-lytic protease is synthesized as a pre-pro protein (41 kDa) that is subsequently processed to its mature extracellular form (20 kDa). This first finding of a large N-terminal protease precursor in a Gram-negative bacterial protease strengthens the hypothesis that large precursors may be a general property of extracellular bacterial proteases, and suggests that the N- or C-terminal location of the precursor segment may be significant.

INTRODUCTION

The serine proteases comprise one of the best-characterized group of enzymes, and constitute one of the most convincing examples of convergent evolution. These proteases are characteristically divided into two major evolutionary families on the basis of their three-dimensional structural similarity: the subtilisin family, and the trypsin family. In spite of large variations in amino acid sequence, and a diverse array of sources, enzymes from these two families share the same basic mechanism of action, using an Asp, His, Ser catalytic triad. Among the bacterial representatives of the trypsin family, α-lytic protease, an extracellular enzyme of the Gram-negative soil bacterium *Lysobacter enzymogenes* 495, is of particular interest. It is easily purified from the native source (Whitaker, 1970); its x-ray crystal structure has been determined to high resolution both as a free enzyme

Correspondence to: Dr. D.A. Agard, Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143-0448 (U.S.A.) Tel. (415)476-2521.
* Current address: Department of Biology, Tufts University, Medford, MA 02155 (U.S.A.) Tel. (617)381-3591.

Abbreviations: aa, amino acid(s); bp, base pair(s); Δ, deletion; DEAE, diethylaminoethyl; Ig, immunoglobulin, IPTG, isopropyl-β-D-thiogalactopyranoside; kb, kilobase(s) or 1000 bp; mRNA, messenger RNA; nt, nucleotide(s); RF, replicative form; SSC, 0.15 M NaCl, 0.015 M Na_3 · citrate, pH 7.4; XGal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

0378-1119/88/$03.50 © 1988 Elsevier Science Publishers B.V. (Biomedical Division)

(Brayer et al., 1979; Fujinaga et al., 1985) and as enzyme-inhibitor complex (Bone et al., 1987); and the presence of a single histidine residue, located in the active site, makes it amenable to spectroscopic studies of catalytic mechanism (Robillard and Shulman, 1974; Bachovchin et al., 1981; Bachovchin and Roberts, 1978). As our interests lie in dissecting the relationship between three-dimensional structure and function in this protein, we have cloned and sequenced the gene for α-lytic protease in preparation for expression and site-specific mutagenesis. Examination of the predicted protein sequence upstream from the start of the mature protein suggests that α-lytic protease is expressed as a large precursor (approx. 41 kDa) which is subsequently processed to yield the mature, extracellular form of the enzyme (20 kDa). Northern and S1 nuclease analyses of *Lysobacter* mRNA directly support this hypothesis.

There are several examples of secreted bacterial proteases that are known to be synthesized as large precursors. The subtilisins (Stahl and Ferrari, 1984)* and the neutral proteases of various *Bacillus* species (Yang et al., 1984; Shimada et al., 1985; Tagaki et al., 1985); and the mammalian-like A and B proteases of *Streptomyces griseus* (Henderson et al., 1987) have large N-terminal propeptides. All of these enzymes are serine proteases and are secreted by Gram-positive bacteria. Among the proteases secreted by Gram-negative bacteria only two examples were previously known to be synthesized as precursors: the IgA protease of *Neisseria gonorrhoea* (thought to be a thiol protease, Pohlner et al., 1987) and the extracellular serine protease of *Serratia marcescens* (Yanagida et al., 1986) which are both extremely large proteases (169 kDa and 112 kDa, respectively) with C-terminal extensions of approx. 60 kDa.

The finding that α-lytic protease is synthesized as a larger precursor strengthens the evidence that precursors are a general motif of extracellular bacterial proteases. Furthermore, this is the first example of a large N-terminal precursor in a Gram-negative bacterial protease.

* Evidence for propeptides of approx. 77 aa have been found in all subtilisins whose genes have been cloned. These include subtilisin from *Bacillus amyloliquefaciens* (Wells et al., 1983; Vasantha et al., 1984), subtilisin E from *B. subtilis* (Stahl and Ferrari, 1984), and subtilisin Carlsberg from *B. licheniformis* (Jacobs et al., 1985). Subtilisin E has been the best studied example with respect to precursor function (Ikemura et al., 1987), therefore we will use subtilisin E as our prototype for discussions of this enzyme.

MATERIALS AND METHODS

(a) Enzymes

All restriction enzymes were purchased from commercial sources. The Klenow fragment of DNA polymerase I was obtained from Boehringer Mannheim or Amersham. T4 DNA ligase and S1 nuclease were purchased from New England Biolabs. T4 polynucleotide kinase was purchased from Pharmacia. Boehringer Mannheim calf intestinal alkaline phosphatase was purified by Sephadex G-75 column chromatography (Efstratiadis et al., 1977). All enzymes were used in accordance with the manufacturers' instructions.

(b) Strains, plasmids, and media

Lysobacter enzymogenes 495 (ATCC29487) was the generous gift of F.D. Cook. Liquid cultures were grown in modified Whitaker's medium (Hunkapiller et al., 1973). JM105 (Yanisch-Perron et al., 1985) and DG98 (*thi*-1, *hsdR*17, *endA*44, *supE*44, *lacI*^{q}, *lacZΔ*M15, *proC*::Tn*10* [F′ *lacI*^{q},*lacZΔ*-M15, *proC*$^+$]) (D. Gelfand, Cetus Corp., Emeryville, CA) were used to grow M13mp18 and M13mp19 (Norrander et al., 1983) clones. XGal and IPTG were added to YT (Miller, 1972) top agarose when appropriate. Transformation of *E. coli* strains was accomplished by the calcium chloride method (Hershfield et al., 1974).

(c) Oligodeoxynucleotide probe

The 51 base long oligodeoxynucleotide (51-mer) used to identify and isolate α-lytic protease sequences was a generous gift of Dr. John Richards, California Institute of Technology. The oligodeoxynucleotide sequence (5′-TACCAGTGCGGC-ACCATCACGGCCAAGAACGTGACCGCG-AACTACGCGGAG-3′) was deduced from the known sequence of aa 109–126 of α-lytic protease,

with codon bias reflecting the G + C-rich (69 mol %) nature of *Lysobacter* chromosomal DNA (Christensen and Cook, 1978) (see Fig. 2).

Nitrocellulose filters from Southern blots or plaque lifts were hybridized with 10^6 cpm/ml end-labeled probe (10^8 cpm/μg), in standard buffer (Southern, 1975; Benton and Davis, 1977), with 2 μg/ml of poly(rA) (to reduce background) for 12–24 h at 37°C (non-stringent conditions). Washing stringency was established by sequential incubation of chromosomal Southern blots at increasing temperatures in 5 × SSC until only unique bands remained. Once established, 60 min at 60°C was used routinely. Filters were exposed on Kodak XAR 5 film with an intensifying screen at −70°C.

(d) DNA and RNA purification and analysis

Lysobacter chromosomal DNA was purified for Southern blotting and cloning by a simplification of the method of Saito and Miura (1963). For Southern blotting, 5–10 μg of DNA were used per digest, and agarose gels varied in concentration from 0.65 to 1.0%. For *Sau*3A fragment isolation, 50 μg of restricted DNA was resolved on a 6% polyacrylamide gel, and the region between 350 and 550 bp was excised and the DNA electroeluted. *Eco*RI-digested DNA (75 μg) was fractionated on 0.5% agarose, the 2-kb region was concentrated onto and eluted from DEAE paper (Dretzen et al., 1981). M13 RF DNA was prepared by alkaline lysis (Birnboim and Doly, 1979).

RNA was harvested from a single colony of *Lysobacter* grown overnight in 50 ml broth at 27°C. Pellets from 10 ml of cells were processed using the guanidinium isothiocyanate–hot phenol method (Feramisco et al., 1982) and RNA was purified from DNA by precipitation out of guanidinium·HCl (Strohman et al., 1977).

(e) Nucleotide sequencing

Both strands of the cloned 1741-bp *Eco*RI fragment were sequenced by the dideoxy chain-termination method (Sanger et al., 1977), using either [α-^{35}S]dATP or [α-^{32}P]dCTP (>1000, >3000 Ci/mmol, Amersham). M13 subclones of convenient restriction fragments were primed by the Pharmacia sequencing primer and, when necessary, oligodeoxynucleotides were used to extend the sequence. A variety of modifications helped to reduce the often severe compression problems due to secondary structure. When the Klenow fragment was used, 7-deazaGTP was substituted for dGTP (Barr et al., 1986), 20 μg of T4 gene *32* protein (O'Farrell, 1987) (a gift of Jack Barry, UCSF) were added per 1 μg template, and sequencing gels included 25% formamide (final concentration). The greatest resolution was achieved with United States Biochemicals (USB) modified T7 polymerase (Sequenase) kit, substituting dITP mixes for dGTP, and adding 5 μg of *E. coli* single-stranded binding protein (USB) per 0.5 μg template (USB sequencing handbook).

RESULTS AND DISCUSSION

(a) Isolation and characterization of a clone containing the α-lytic protease gene

Our strategy for cloning the structural gene for α-lytic protease was to use the known amino acid sequence (Olson et al., 1970) to generate an optimal, unique hybridization probe. This approach is based on the fact that, on average, at least two out of every three bases can be guessed correctly. As discussed by Anderson and Kingston (1983), if the probe is sufficiently long (86 nt), a single copy gene can be identified as a unique band on a mammalian genomic Southern blot and subsequently cloned. Prior knowledge of codon usage or G + C content can be used to improve the accuracy of the probe sequence. We found that with the considerably smaller *Lysobacter* genome, a 51-mer was long enough to identify a single band on a genomic Southern blot. By contrast, a 40-mer probe to the same region hybridized to three bands.

Initially, we were concerned that a foreign bacterial protease cloned into *Escherichia coli* would be lethal. To insure that the cloned gene be inactive, genomic Southern blotting was used to identify a fragment smaller than the known size of the mature protease (600 bp). For this reason, an approx. 500-bp *Sau*3A restriction fragment was initially cloned by ligating size-selected, gel-isolated fragments in *Bam*HI-cut M13mp18 DNA. The ligation mixture was transformed into JM105, plated in top

agarose with IPTG and XGal, and screened with labeled 51-mer. Interestingly, the one positive plaque of the 450 screened, and shown by sequencing to encode an out-of-frame fragment of α-lytic protease, was a blue plaque.

To obtain the full sequence, the 51-mer was again used to screen a size-selected library of 2-kb *Eco*RI fragments cloned into M13mp18. Positive clones from this screening were mapped and were found to have 0.8 kb of upstream DNA and 0.4 kb of DNA beyond the expected end of the protein-coding sequence (Fig. 1).

The entire nucleotide sequence of the *Eco*RI fragment was determined (Fig. 2). When translated, the nucleotide sequence encoding the mature enzyme matched the published amino acid sequence exactly, and was immediately followed by a stop codon (UGA) and a putative RNA termination signal (Rosenberg and Court, 1979). A large open reading frame extended 5′ from the known N terminus of the mature protein. In fact, all three reading frames were open for greater than 300 nt before the enzyme's protein-coding sequence, as stop codons are statistically quite rare in high-G + C content genomes. More intriguing, however, was the translated sequence of amino acids just preceding the N terminus; the number and distribution of polar and charged amino acids did not fit the hydrophobic nature expected of a signal sequence (Watson, 1984).

To determine if α-lytic protease might be produced as part of a larger precursor molecule, like subtilisin E (Stahl and Ferrari, 1984), we examined the upstream sequence. In addition to several possible ATG and many possible GTG start codons, a potential ribosome-binding site and promoter region were also identified (Fig. 2). These putative transcriptional and translational control regions were situated an appropriate distance upstream from one of the ATG codons (nt 248), making it an ideal candidate for a start codon. Initiating at the next available ATG codon (nt 323) would interrupt the very hydrophobic stretch of amino acids comprising the putative signal sequence (aa −182 to −166; Fig. 2), providing further support for the choice of start codon. However, it is important to note that no other gene has been cloned from this microorganism, which by 16S RNA sequence analysis is phylogenetically very distant from *E. coli* (Stackebrandt and Woese, 1981). Furthermore, using affinity-purified α-lytic protease antibodies, no immunologically cross-reacting material to α-lytic protease could be detected in *E. coli* strains carrying the *Eco*RI fragment (not shown). For these reasons, we felt that the unknown nature of *Lysobacter* transcriptional and translational control regions made them unreliable indicators for the start of the α-lytic gene. A direct determination was required.

(b) Northern-blot and S1 analysis

The hypothesis that α-lytic protease is biologically synthesized as a larger precursor was substantiated by both Northern-blot (not shown) and S1 nuclease analysis (Fig. 3). These experiments indicated that the mRNA was approx. 1.4 kb, and that the message started at nt 118, within 7 nt of the putative −10 sequence, and more than 700 nt before the start of the mature protein sequence. This evidence, in addition to the position of the putative ribosome-binding site and signal peptide sequence, makes the ATG at nt 248 the most likely start codon.

(c) Codon usage

This is the first report of a gene cloned from the Canadian soil bacterium *L. enzymogenes* 495. The

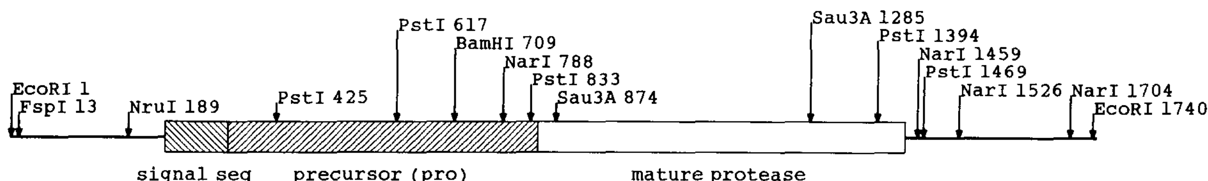

Fig. 1. Restriction map of the cloned 1.7-kb *Eco*RI fragment. The fragment contains the coding sequence for the mature protease (open box) and upstream regions corresponding to the proposed signal sequence and precursor (pro) regions (hatched boxes). All sites shown were confirmed by nucleotide sequence determination (see MATERIALS AND METHODS, section **e**, and Fig. 2). Only two *Sau3*A sites in the mature coding region (RESULTS, section **a**), and the *Fsp*I site (base 13) used for S1 mapping (Fig. 3, legend), are shown.

```
                                                                                             -35                    -10
GAATTCGTACAATGCGCAGGCGAATGCGCCGTCTTGTGACACGGCTCGAATTGCGCGTGTGATCCGGCCTCGCGCAAGCGGCTTGCGCTGGCGGCGCATCTTTTTATCTGGCATA

*-->
GCCCCGCGAGGCCGCGCCCAACGACGGCGTTGCGCCTCGCCGCTTACGGCCGCCGTTCGCGCAACGGCGTTCGCGAAGCCGACCGGCTTCACCGACGATCCCATGAAAGATTCCTCT

                -199                                -190                                    -180
 rbs            Met Tyr Val Ser Asn His Arg Ser Arg Arg Val Ala Arg Val Ser Val Ser Cys Leu Val Ala Ala Leu Ala Ala Met Ser
AGGAGTTTTCGCATG TAC GTA TCG AAC CAC CGT TCG CGC CGC GTC GCG CGC GTG TCC GTT TCC TGC CTC GTC GCC GCG CTG GCG GCG ATG TCC

    -170                                    -160                                    -150
Cys Gly Ala Ala Leu Ala Ala Asp Gln Val Asp Pro Gln Leu Lys Phe Ala Met Gln Arg Asp Leu Gly Ile Phe Pro Thr Gln Leu Pro
TGC GGC GCT GCG CTG GCG GCC GAT CAG GTC GAT CCT CAG CTA AAA TTC GCA ATG CAG CGC GAT CTG GGC ATC TTC CCG ACC CAG CTG CCG

    -140                                    -130                                    -120
Gln Tyr Leu Gln Thr Glu Lys Leu Ala Arg Thr Gln Ala Ala Ala Ile Glu Arg Glu Phe Gly Ala Gln Phe Ala Gly Ser Trp Ile Glu
CAG TAC CTG CAG ACC GAA AAA CTC GCG CGC ACC CAG GCT GCG GCG ATC GAG CGC GAG TTC GGT GCT CAG TTC GCC GGC AGC TGG ATC GAG

    -110                                    -100                                    -90
Arg Asn Glu Asp Gly Ser Phe Lys Leu Val Ala Ala Thr Ser Gly Ala Arg Lys Ser Ser Thr Leu Gly Gly Val Glu Val Arg Asn Val
CGC AAC GAA GAC GGC AGC TTC AAG CTC GTC GCC GCG ACC TCG GGC GCG CGC AAG TCG AGC ACG CTG GGC GGG GTG GAA GTG CGC AAC GTG

    -80                                     -70                                     -60
Arg Tyr Ser Leu Lys Gln Leu Gln Ser Ala Met Glu Gln Leu Asp Ala Gly Ala Asn Ala Arg Val Lys Gly Val Ser Lys Pro Leu Asp
CGC TAC AGC CTC AAG CAG CTG CAG TCG GCG ATG GAG CAG CTC GAC GCC GGT GCC AAC GCG CGG GTG AAG GGC GTC AGC AAG CCG CTC GAC

    -50                                     -40                                     -30
Gly Val Gln Ser Trp Tyr Val Asp Pro Arg Ser Asn Ala Val Val Val Lys Val Asp Asp Gly Ala Thr Asp Ala Gly Val Asp Phe Val
GGC GTG CAG AGC TGG TAC GTG GAT CCG CGC AGC AAC GCG GTG GTG GTC AAA GTC GAC GAC GGC GCG ACG GAT GCC GGC GTC GAC TTC GTC

    -20                                     -10                                     1
Ala Leu Ser Gly Ala Asp Ser Ala Gln Val Arg Ile Glu Ser Ser Pro Gly Lys Leu Gln Thr Thr Ala Asn Ile Val Gly Gly Ile Glu
GCC CTC AGC GGC GCC GAC AGT GCG CAG GTG CGG ATC GAA AGC TCG CCG GGC AAG CTG CAG ACC ACG GCC AAC ATC GTC GGC GGC ATC GAA

    10                                      20                                      30
Tyr Ser Ile Asn Asn Ala Ser Leu Cys Ser Val Gly Phe Ser Val Thr Arg Gly Ala Thr Lys Gly Phe Val Thr Ala Gly His Cys Gly
TAC TCG ATC AAC AAC GCC TCG CTG TGC TCG GTC GGC TTC TCG GTC ACC CGC GGC GCG ACC AAG GGC TTC GTC ACC GCC GGC CAC TGC GGC

    40                                      50                                      60
Thr Val Asn Ala Thr Ala Arg Ile Gly Gly Ala Val Val Gly Thr Phe Ala Ala Arg Val Phe Pro Gly Asn Asp Arg Ala Trp Val Ser
ACG GTC AAT GCG ACC GCG CGC ATC GGC GGC GCG GTG GTC GGC ACC TTC GCC GCG CGC GTG TTC CCC GGC AAC GAC CGC GCC TGG GTC AGC

    70                                      80                                      90
Leu Thr Ser Ala Gln Thr Leu Leu Pro Arg Val Ala Asn Gly Ser Ser Phe Val Thr Val Arg Gly Ser Thr Glu Ala Ala Val Gly Ala
CTG ACC AGC GCG CAG ACT CTG CTG CCG CGC GTG GCC AAC GGC AGC AGC TTC GTC ACC GTG CGC GGC AGC ACC GAG GCC GCG GTC GGC GCG

    100                                     110                                     120
Ala Val Cys Arg Ser Gly Arg Thr Thr Gly Tyr Gln Cys Gly Thr Ile Thr Ala Lys Asn Val Thr Ala Asn Tyr Ala Glu Gly Ala Val
GCG GTG TGC CGC TCG GGC CGC ACC ACC GGT TAC CAG TGC GGC ACC ATC ACC GCC AAG AAC GTC ACC GCC AAC TAC GCC GAA GGC GCG GTG

    130                                     140                                     150
Arg Gly Leu Thr Gln Gly Asn Ala Cys Met Gly Arg Gly Asp Ser Gly Gly Ser Trp Ile Thr Ser Ala Gly Gln Ala Gln Gly Val Met
CGC GGC CTG ACC CAG GGC AAC GCC TGC ATG GGC CGC GGC GAT TCG GGC GGT TCG TGG ATC ACC AGC GCC GGC CAG GCG CAG GGC GTG ATG

    160                                     170                                     180
Ser Gly Gly Asn Val Gln Ser Asn Gly Asn Asn Cys Gly Ile Pro Ala Ser Gln Arg Ser Ser Leu Phe Glu Arg Leu Gln Pro Ile Leu
TCG GGC GGC AAC GTG CAG TCC AAC GGC AAC AAC TGC GGC ATC CCG GCC TCG CAG CGC AGC AGC CTG TTC GAG CGT CTG CAG CCG ATC CTG

    190                             198
Ser Gln Tyr Gly Leu Ser Leu Val Thr Gly OP                                                            TERMINATOR
AGC CAG TAC GGC CTG AGC CTG GTC ACC GGC TGA TCGCCGACAAGATCAGCGGCGCCGCGCCTGCAGCGCCGTCGATGCGAACGCCCCGCCTTGCGCGGGGCGTTCG

TTTTTCAGCGCGGCGCCGTCGCGGCGCGCGCTACCACTGCGGCCCGTACAGCAGCGGCTCGGCTTCCCACGCGGCGCGGCTCGCCGGCGCGAGTTCGCGCGCGTCGCGGGTACGTTGCG

GGTCGAGCTTGAGTGTCACCATCTCGCGCTCGTCGACTTCCTGCGGACGGTAGCCGTTGCGGTGGAGCAGGGCGCCGCTGAGGTTTTTCGCGCTCAGCCGGCCGTCGAATTC
```

Fig. 2. The complete sequence of the 1746-bp *Eco*RI fragment containing the α-lytic protease gene. Putative promoter −35 and −10, ribosome-binding site, and terminator sequences are shown in bold type, underlined. The transcriptional signals are poor matches to *E. coli* consensus sequences (see RESULTS, section **b**). The transcription start point as determined by S1 mapping (see Fig. 3, and RESULTS, section **b**) is indicated by an asterisk.

average G + C content for the entire 1745-bp cloned fragment is 69.0%, with little difference between the coding (69.8%) and non-coding (68.4%) regions. In the non-coding regions the base doublets GC + CG account for 35.6% of all base doublets; this value is decreased to 32.2% within the coding portion, presumably due to the constraints of codon choices. As expected from a gene with such high % G + C, the codon usage is strongly biased (codon bias index = 82.8%; Bennetzen and Hall, 1982) towards

Fig. 3. S1 nuclease mapping of α-lytic protease message. The strand-specific probe (Burke, 1984) as well as the dideoxy nucleotide sequence ladder were synthesized from M13 clones using the oligodeoxynucleotide 5′-CTCCTAGAGGAATC-TTTC-3′, which hybridizes to the coding strand in the region of the putative ribosome-binding site (base 248). The probe was delineated at the 3′ end by an *Fsp*I digestion (base 13, see Fig. 1), and separated from its complementary strand on an alkaline agarose gel. 5×10^5 cpm of probe was hybridized with 50 μg total RNA at 62°C in 80% formamide (Berk and Sharp, 1977). The sample was digested with 1000 units of S1 nuclease (lane a), and the products were resolved on a 6% denaturing polyacrylamide gel. The RNA transcription start point is indicated by the asterisk.

those codons that utilize G or C in the third position (Table I). Interestingly, the charged amino acids Glu, Asp, and Lys are exceptions, showing, respectively, only 50%, 57%, and 73% bias towards G or C in the third position.

(d) Conclusion

The discovery of a precursor for α-lytic protease reinforces the speculation that 'pro' sequences are a common feature of prokaryotic serine proteases, just as zymogen forms are common in their mammalian counterparts. However, the large size of these precursors suggests either a more complicated mechanism of temporary inactivation than in the mammalian proteases or a more complex role for the precursor. Indeed, in the two cases where the role of the propeptide has been studied, it was either essential for the production of active protein and was probably involved in the correct folding of the mature domain (subtilisin E; Ikemura et al., 1987); or it was necessary for the selective secretion of the protease (IgA protease; Pohlner et al., 1987).

Preliminary experiments using the cloned gene of α-lytic protease expressed in *E. coli* indicate that, as for subtilisin E (Ikemura et al., 1987), the precursor domain is necessary for proper folding of the mature domain (manuscript in preparation). The apparent similarities in the requirement of an N-terminal precursor domain for proper folding of such evolutionarily unrelated, but functionally similar, proteins from two different classes of bacteria suggests a common role for these precursor sequences. Furthermore, the discovery of an N-terminal precursor on a Gram-negative protease indicates that the difference between proteases having N- and C-terminal precursors is not simply related to differences in export between the two families of bacteria. Future experiments with α-lytic protease will address these issues.

ACKNOWLEDGEMENTS

We thank Drs. Roger Bone and Marc Navre for helpful comments on the manuscript, John Vasquez for helpful discussions on codon usage, and Ms. Mei Lie Wong for her expert assistance with the figures. This work has been supported by grants from the Searle Scholars program, the NSF Presidential Young Investigator Program, and the Howard Hughes Medical Institute.

TABLE I

Codon usage for the α-lytic protease gene of *Lysobacter enzymogenes*[a]

No.	Codon	aa	No.	Codon	aa	No.	Codon	aa	No.	Codon	aa
0/	UUU/	phe	0/	UCU/	ser	0/	UAU/	tyr	0/	UGU/	cys
12/	UUC/	phe	4/	UCC/	ser	8/	UAC/	tyr	8/	UGC/	cys
0/	UUA/	leu	0/	UCA/	ser	0/	UAA/	OC	1/	UGA/	OP
0/	UUG/	leu	15/	UCG/	ser	0/	UAG/	AM	4/	UGG/	trp
0/	CUU/	leu	1/	CCU/	pro	0/	CAU/	his	2/	CGU/	arg
7/	CUC/	leu	1/	CCC/	pro	2/	CAC/	his	22/	CGC/	arg
1/	CUA/	leu	0/	CCA/	pro	0/	CAA/	gln	0/	CGA/	arg
18/	CUG/	leu	8/	CCG/	pro	23/	CAG/	gln	2/	CGG/	arg
0/	AUU/	ile	1/	ACU/	thr	1/	AAU/	asn	1/	AGU/	ser
12/	AUC/	ile	21/	ACC/	thr	17/	AAC/	asn	19/	AGC/	ser
0/	AUA/	ile	0/	ACA/	thr	3/	AAA/	lys	0/	AGA/	arg
6/	AUG/	met	4/	ACG/	thr	8/	AAG/	lys	0/	AGG/	arg
1/	GUU/	val	3/	GCU/	ala	6/	GAU/	asp	4/	GGU/	gly
20/	GUC/	val	22/	GCC/	ala	8/	GAC/	asp	42/	GGC/	gly
1/	GUA/	val	1/	GCA/	ala	6/	GAA/	glu	0/	GGA/	gly
18/	GUG/	val	27/	GCG/	ala	6/	GAG/	glu	1/	GGG/	gly

[a] Columns "No." specify how many times a given codon is used in the enzyme-coding gene.

REFERENCES

Anderson, S. and Kingston, I.B.: Isolation of a genomic clone for bovine pancreatic trypsin trypsin inhibitor by using a unique-sequence synthetic DNA probe. Proc. Natl. Acad. Sci. USA 80 (1983) 6838–6842.

Bachovchin, W.W. and Roberts, J.D.: Nitrogen-15 nuclear magnetic resonance spectroscopy. The state of histidine in the catalytic triad of α-lytic protease. Implications for the charge-relay mechanism of peptide-bond cleavage by serine proteases. J. Am. Chem. Soc. 100 (1978) 8041–8097.

Bachovchin, W.W., Kaiser, R., Richards, J.H. and Roberts, J.D.: Catalytic mechanism of serine proteases: reexamination of the pH dependence of the histidyl 1J13C2-H coupling constant in the catalytic triad of a-lytic protease. Proc. Natl. Acad. Sci. USA 78 (1981) 7323–7326.

Barr, P.J., Thayer, R.M., Laybourn, P., Najarian, R.C., Seela, F. and Tolan, D.R.: 7-Deaza-2′-deoxyguanosine-5′-triphosphate: enhanced resolution in M13 dideoxy sequencing. Biotechniques 4 (1986) 428–432.

Bennetzen, J.L. and Hall, B.D.: Codon selectivity in yeast. J. Biol. Chem. 257 (1982) 3026–3031.

Benton, W.D. and Davis, R.W.: Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196 (1977) 180–182.

Berk, A.J. and Sharp, P.A.: Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12 (1977) 721–732.

Birnboim, H.C. and Doly, J.: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7 (1979) 1513–1523.

Bone, R., Shenvi, A.B., Kettner, C.A. and Agard, D.A.: Serine protease mechanism: structure of an inhibitory complex of α-lytic protease and a tightly bound peptide boronic acid. Biochemistry 26 (1987) 7609–7614.

Brayer, G.D., Delbaere, L.T.J. and James, M.N.G.: Molecular structure of the α-lytic protease from *Myxobacter 495* at 2.8 Å resolution. J. Mol. Biol. 131 (1979) 743–775.

Burke, J.F.: High-sensitivity S1 mapping with single-stranded [^{32}P]DNA probes synthesized from bacteriophage M13mp templates. Gene 30 (1984) 63–68.

Christensen, P. and Cook, F.D.: *Lysobacter*, a new genus of nonfruiting, gliding bacteria with a high base ratio. Int. J. Syst. Bacteriol. 28 (1978) 367–393.

Dretzen, G., Bellard, P., Sassone-Corsi, P. and Chambon, P.: A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. 112 (1981) 295–298.

Efstratiadis, A., Vournakis, J.N., Doris-Keller, H., Chaconas, G., Dougall, D.K. and Kafatos, F.C.: End labeling of enzymatically decapped mRNA. Nucleic Acids Res. 4 (1977) 4165–4174.

Feramisco, J.R., Smart, J.E., Burridge, K., Helfman, D.M. and Thomas, G.P.: Co-existence of vinculin and a vinculin-like protein of higher molecular weight in smooth muscle. J. Biol. Chem. 257 (1982) 11024–11031.

Fujinaga, M., Delbaere, L.T.J., Brayer, G.D. and James, M.N.G.: Refined structure of α-lytic protease at 1.7 Å resolution.

Analysis of hydrogen bonding and solvent structure. J. Mol. Biol. 183 (1985) 479–502.

Henderson, G., Krygsman, P., Liu, C.J., Davey, C.C. and Malek, L.T.: Characterization and structure of genes for proteases A and B from *Streptomyces griseus*. J. Bacteriol. 169 (1987) 3778–3784.

Hershfield, V., Boyer, H.W., Yanofsky, C., Lovett, M.A. and Helinski, D.R.: Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. Proc. Natl. Acad. Sci. USA 71 (1974) 3455–3459.

Hunkapiller, M.W., Smallcombe, S.H., Whitaker, D.R. and Richards, J.H.: Carbon nuclear magnetic resonance studies of the histidine residue in α-lytic protease. Implications for the catalytic mechanism of serine proteases. Biochemistry 12 (1973) 4732–4743.

Ikemura, H., Takagi, H. and Inouye, M.: Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*. J. Biol. Chem. 262 (1987) 7859–7864.

Jacobs, M., Eliasson, M., Uhlen, M. and Flock, J.-I.: Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. Nucleic Acids Res. 13 (1985) 8914–8926.

Miller, J.: Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1972.

Norrander, J., Kempe, T. and Messing, J.: Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26 (1983) 101–106.

O'Farrell, P.: Dideoxy DNA sequencing using T4 gene *32* protein. BMBiochemica 4 (1987) 13.

Olson, M.O.J., Nagabhushan, N., Dzwiniel, M., Smillie, L.B. and Whitaker, D.R.: Primary structure of α-lytic protease: a bacterial homologue of the pancreatic serine proteases. Nature 228 (1970) 438–442.

Pohlner, J., Halter, R., Beyreuther, K. and Meyer, T.F.: Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature 325 (1987) 458–462.

Robillard, G. and Shulman, R.G.: High resolution nuclear magnetic resonance studies of the active site of chymotrypsin. I. The hydrogen bonded protons of the 'charge relay' system. J. Mol. Biol. 86 (1974) 519–540.

Rosenberg, M. and Court, D.: Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13 (1979) 319–353.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Saito, H. and Miura, K.I.: Preparation of transforming deoxyribonucleic acid by phenol treatment. Biochim. Biophys. Acta 72 (1963) 619–629.

Sequenase: step by step protocols for DNA sequencing with Sequenase. USB Sequencing Handbook, p. 13.

Shimada, H., Honjo, M., Mita, I., Nakayama, A., Akaoka, A., Manabe, K. and Furutani, Y.: The nucleotide sequence and some properties of the neutral protease gene of *Bacillus amyloliquefaciens*. J. Biotechnol. 2 (1985) 75–85.

Southern, E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98 (1975) 503–517.

Stackebrandt, E. and Woese, C.R.: The evolution of prokaryotes. In Carlile, M.J., Collins, J.F. and Moseley, B.E.B. (Eds.), Molecular and Cellular Aspects of Microbial Evolution. Symp. Soc. Gen. Microbiol., Vol. 32, Cambridge University Press, Cambridge, 1981, pp. 1–31.

Stahl, M.L. and Ferrari, E.: Replacement of the *Bacillus subtilis* subtilisin structural gene with an in-vitro-derived deletion mutation. J. Bacteriol. 158 (1984) 411–418.

Strohman, R.C., Moos, P.S., Micou-Eastwood, J., Spector, D., Przybyla, A. and Paterson, B.: Messenger RNA for myosin polypeptides: isolation from single myogenic cell cultures. Cell 10 (1977) 265–273.

Takagi, M., Imanaka, T. and Aiba, S.: Nucleotide sequence and promoter region for the neutral protease gene from *Bacillus stearothermophilus*. J. Bacteriol. 163 (1985) 824–831.

Vasantha, N., Thompson, L.D., Rhodes, C., Banner, C., Nagle, J. and Filpula, D.: Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bacteriol. 159 (1984) 811–819.

Watson, M.E.E.: Compilation of published signal sequences. Nucleic Acids Res. 12 (1984) 5145–5164.

Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A. and Chen, E.Y.: Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. Nucleic Acids Res. 11 (1983) 7911–7925.

Whitaker, D.R.: The α-lytic protease of a Myxobacterium. Methods Enzymol. 19 (1970) 599–613.

Yang, M.Y., Ferrari, E. and Henner, D.J.: Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. J. Bacteriol. 160 (1984) 15–21.

Yanagida, N., Uozumi, T. and Beppu, T.: Specific excretion of *Serratia marcescens* protease through the outer membrane of *Escherichia coli*. J. Bacteriol. 166 (1986) 937–944.

Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33 (1985) 103–119.

Communicated by F. Barany.