A gene encoding a thermophilic alkaline serine proteinase from *Thermus* sp. strain Rt41A and its expression in *Escherichia coli*

Gordon K. L. Munro,¹ Ronnie H. McHale,² David J. Saul,²
Rosalind A. Reeves² and Peter L. Bergquist³

Author for correspondence: Peter L. Bergquist. Tel: +61 2 850 8614. Fax: +61 2 850 8799.
e-mail: peter.bergquist@mq.edu.au

¹ School of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

² Centre for Gene Technology, University of Auckland, Private Bag 92019, Auckland, New Zealand

³ School of Biological Sciences, Macquarie University, North Ryde, New South Wales 2109, Australia

The extreme thermophile *Thermus* sp. strain Rt41A produces an extracellular alkaline serine proteinase during growth. This enzyme is stable for more than 24 h at 70 °C and has a pH optimum of 8.0. The proteinase gene was identified using primers designed to amplify a region between two highly conserved amino acid motifs in subtilisin-like proteinases and the PCR product was used to identify a genomic fragment containing the gene. The amino acid sequence deduced from the Rt41A gene contained a region identical to that obtained by amino-terminal sequencing of purified Rt41A proteinase. Comparison of the entire derived peptide sequence with other subtilisin-like serine proteinases revealed significant homologies, especially with aqualysin I from *Thermus aquaticus* YT-1 and with exoprotease A from *Vibrio alginolyticus*. The Rt41A proteinase was expressed in *Escherichia coli* as a fusion protein with glutathione-S-transferase as an aid for purification and to overcome difficulties experienced with other plasmid vectors which produced inactive protein. The enzyme is inactive as synthesized and activation was shown to be temperature-dependent, with shorter incubation times required at higher temperatures; removal of the hydrophobic signal peptide from the start of the gene reduced the time required for activation to less than a third of that required if the signal peptide was present.

Keywords: thermophilic proteinase, *Thermus* Rt41A

INTRODUCTION

Bacterial serine proteinases (EC 3.4.21.-) are a well studied group of endopeptidases characterized by a 'catalytic triad' of amino acid residues at the active site (Wells & Estell, 1988). Most studies have concentrated on the subtilisins, a family of proteinases found initially in *Bacillus* and more recently in other organisms (Cowan *et al.*, 1987a; Frömmel & Sander, 1989; Samal *et al.*, 1990). These extracellular enzymes are produced as 'pre-pro-proteins' which are exported from the cell and are then cleaved by autolysis to release the mature, active enzyme (Vasantha *et al.*, 1984).

Alignment of the peptide sequences of members of the subtilisins and subtilisin-like serine proteinases has

revealed considerable homology between all of the sequences. Generally, the greatest homology is found in the regions involved in substrate catalysis (Asp32, His64 and Ser221; Wells & Estell, 1988) and, to a lesser extent, in substrate specificity. Although the majority of subtilisin-like enzymes described so far have been isolated from mesophilic hosts, there has been an increasing number of subtilisin-like proteinases isolated from thermophilic and hyperthermophilic bacteria (Matsuzawa *et al.*, 1983; Coolbear *et al.*, 1988; Blumentals *et al.*, 1990). One of these enzymes, aqualysin I, has been cloned, sequenced and expressed in *Escherichia coli* (Kwon *et al.*, 1988) and *Thermus thermophilus* HB8 (Touhara *et al.*, 1991).

Thermus sp. strain Rt41A is a thermophilic eubacterium isolated from a geothermal area in New Zealand. The phylogenetic relationship of this strain with other *Thermus* spp. has been examined by others and Rt41A was shown to belong to a distinct cluster of New Zealand isolates and to be significantly different from *Thermus aquaticus* (Saul *et*

Abbreviations: GST, glutathione-S-transferase; PMSF, phenylmethylsulphonyl fluoride.

The GenBank accession number for the sequence reported in this paper is U17342.

al., 1993). *Thermus* sp. strain Rt41A is known to secrete an extracellular alkaline serine proteinase during growth which has a temperature optimum of 75 °C and has greatest activity at pH 8.0 (Janssen *et al.*, 1991). In addition, the enzyme is very thermostable, showing no loss in activity after incubation at 70 °C for 24 h (Peek *et al.*, 1992). The broad specificity of the Rt41A proteinase and its autolytic activity have been demonstrated to be of value in a number of applications in molecular biology, including mRNA and DNA preparation (Borges & Bergquist, 1992; Fung & Fung, 1991; McHale *et al.*, 1991).

In this paper we report the identification and cloning of the gene encoding the alkaline serine proteinase from *Thermus* sp. strain Rt41A and we also report its DNA sequence and the expression of the recombinant enzyme in *E. coli*.

METHODS

Bacterial strains and culture conditions. *E. coli* strains DH5 α [F^- ϕ 80 *dlacZ* Δ M15 Δ (*lacZYA-argF*)]; U169 [*endA1 recA1 hsdR17* ($r_K^- m_K^+$) *deoR thi-1 supE44* λ^- *gyrA96 relA1*]; JM101 [Δ (*lac-proAB*) *supE thi/F'lacI^qZ* Δ M15 *traD36 proAB*]; and C600 [F^- *thr-1 leuB6 thi-1 lacY1 supE44 mcrA*] were used throughout this research. *Thermus* sp. strain Rt41A was grown at 70 °C using modified Medium 162 as previously described (Janssen *et al.*, 1991).

PCR. Two primers (RM-5 and RM-6) were designed to amplify the region between the highly conserved active site histidine (His64) and the conserved active site serine (Ser221) (numbering according to the sequence of subtilisin BPN'; Wells *et al.*, 1983; Vasantha *et al.*, 1984; also see Fig. 6). The primer sequences were: RM-5, 5'-CA(TC) GG(AGCT) ACC AA(CT) GTG GC(GCT) GG-3'; RM-6, 5'-(AGC)GG GGT (AGC)GC CAT GGA GGT (GCT)CC-3'. The primers were designed taking into account the codon usage pattern of *Thermus*. PCR was performed using 10 pmol of each primer plus genomic DNA from *Thermus* sp. strain Rt41A. The reaction conditions were: (94 °C/1 min; 55 °C/1 min; 72 °C/2 min) $\times$ 40. Buffer conditions were: 1 U *Taq* polymerase (Ampli Taq , Perkin-Elmer Cetus), 250 μ M dNTPs, 1.5 mM $MgCl_2$, 100 mM Tris/Cl (pH 8.8), 500 mM KCl and 0.01 % gelatin in a reaction volume of 50 μ l. PCR products were polished and phosphorylated and ligated into the *Sma*I site of pUC19. *E. coli* strain DH5 α was transformed by this DNA to ampicillin resistance and the transformants were screened for inserts.

Cloning of the gene. Genomic DNA from *Thermus* sp. strain Rt41A was digested to completion with various restriction endonucleases and ligated to appropriately cut pGEM5Zf(+) or pUC19. Transformant colonies from this DNA were screened by colony hybridization using the 470 bp PCR product labelled with [32 P]dCTP by random priming as the probe.

DNA sequencing and sequence analysis. Pseudo-random and random shear libraries were created in M13mp10 from the genomic fragments in the recombinant plasmids pNZ1266 and pNZ1271 (see Results). Isolates from these libraries were sequenced. Manual sequencing was performed using Sequenase (USB) and *Taq* polymerase. Automated sequencing was done on an Applied Biosystems 373A. Both strands of the DNA were sequenced and computer analysis was performed on a Silicon Graphics 4D/30 using the GCG package of Devereux *et al.* (1984).

Construction of expression plasmids. Oligonucleotide primers were designed for use in a PCR reaction to amplify the gene encoding the Rt41A proteinase. Berg80, 5'-GGA GGG AAT TCT GAA GCG AGG CGG TCT TTG G-3', bound at the start of the ORF encoding the proteinase; Berg89, 5'-CGC CTG AAT TCT CAA CCC GCC GGC GGC GAG-3', bound 57 bp down from the start of the ORF at the predicted start of the 'pro-region' of the proteinase. Berg89 was designed to produce an enzyme lacking the leader sequence. Both primers have an *Eco*RI site (underlined) to simplify insertion of the PCR product into the expression vector pGEX-KG (Guan & Dixon, 1991). Berg81, 5'-GCA AGA AGC TTT AGC CGT TTA CTG TCA GA-3', was used as reverse primer in both PCR reactions and it has a *Hind*III site. PCR was performed as follows: 94 °C/4 min (hot start) followed by (94 °C/1 min; 50 °C/1 min; 72 °C/4 min) $\times$ 30. The final cycle was followed by a step of 72 °C for 8 min. PCR products were gel purified, digested with *Eco*RI and *Hind*III and ligated into the same sites in the vector. Recombinant plasmids were used to transform *E. coli* strain DH5 α to ampicillin resistance. The two expression plasmids produced were designated pNZ1997 (GST:Rt41A proteinase fusion) and pNZ1989 (GST:Rt41A(Δ 1-57) proteinase fusion).

Expression of Rt41A proteinase in *E. coli*. Induction of expression was performed according to the protocol of Guan & Dixon (1991). Cells were harvested by centrifugation, resuspended in 50 mM HCT buffer [50 mM HEPES, pH 8.1 adjusted at 70 °C, 5 mM $CaCl_2$, 0.01 % (v/v) Triton X-100] and lysed in a French pressure cell. The resulting lysate was separated into soluble (cell extract) and insoluble (cell pellet) fractions by centrifugation at 20000 *g* for 30 min.

Purification of recombinant Rt41A proteinase. The GST:Rt41A proteinase fusion was purified by incubation at 70 °C for 12 h to precipitate *E. coli* proteins and activate the proteinase. Alternatively, the GST:Rt41A fusion protein was affinity purified on glutathione agarose beads and then incubated at high temperature (70 °C or above) to activate the proteinase.

Determination of proteinase activity. Agar plates containing 2 % (w/v) casein (technical grade), 1.5 % (w/v) milk agar, (100 μ g ampicillin ml^{-1}) and IPTG (100 μ M) were used to screen for proteolytic activity of recombinants. Recombinants were grown on the plates at 37 °C for 12 h and then transferred to 70 °C for 2-3 h. Proteolytic activity was characterized by a ring of precipitated casein around the colony. In liquid samples, activity was determined using a colorimetric assay with 0.2 % azocasein in 50 mM HCT buffer (Cowan *et al.*, 1987b).

Amino terminal sequencing. Recombinant Rt41A proteinase was purified from lysates of *E. coli* containing the expression plasmid pNZ1997 and the amino terminal sequence of the protein determined using an Applied Biosystems 470A Gas Phase Analyser.

RESULTS

Cloning of the Rt41A proteinase gene

A PCR of genomic DNA from *Thermus* sp. strain Rt41A amplified two products of about 200 bp and 470 bp. The expected product size was 470 bp, based on the predicted binding sites of the two primers to the aqualysin I gene sequence. Dot blot analysis revealed that the two products hybridized to each other under conditions of high stringency. Southern blots of Rt41A genomic DNA digested with *Bam*HI and *Apa*I showed that the 470 bp PCR product hybridized to single bands of 1.6 kb and

[illegible]

Fig. 1. The *aprA* sequence of *Thermus* sp. Rt41A. The flanking sequence and the putative ribosome-binding site are also included. The deduced peptide sequence of Rt41A proteinase is shown in standard, single letter notation. The putative signal sequence has been underlined and the mature portion of the peptide is boxed. Also shown is a large inverted repeat upstream of the *aprA* coding region.

2.0 kb, respectively. Fragments of appropriate size were recovered from digested genomic DNA by separation on agarose gels and elution onto DEAE NA45 membranes (Schleicher & Schuell). This DNA was inserted into pGEM5Zf(+) (*Apa*I fragments) or pUC19 (*Bam*HI fragments), DH5 α was transformed with the ligation mixtures and plated onto LB containing ampicillin. Transformant colonies were screened as described in Methods. Plasmids carrying the correct *Apa*I and *Bam*HI fragments were designated pNZ1271 and pNZ1266, respectively. Sequencing of the two different inserts showed that they overlapped by 1251 bp. The DNA sequences of the two inserts were combined to produce a contiguous sequence of 1850 bp (Fig. 1).

Sequence analysis

A 1227 bp ORF encoding a peptide of 408 amino acids (42.9 kDa calculated molecular mass) was identified on the forward strand of the sequence. Comparisons with other proteinase sequences showed it to be 74 % similar (65 % identity) to aqualysin I. The gene was designated *aprA*. The ORF has a G + C content of 69.2 mol % which is very similar to other genes from *Thermus* and the GC bias in the third position of the codon is very high (93.2 %). An *E. coli*-like ribosome-binding site was also identified 5 bp upstream of the potential initiation codon and an inverted repeat sequence was found upstream of *aprA* between positions 282 and 307. This sequence consists of 22 perfectly matched nucleotides. The first 20 amino acids of AprA may function as a leader sequence and a possible signal peptidase processing site was

Fig. 2. A schematic representation of the products of the two expression plasmids pNZ1989 and pNZ1997. L, linker region of pGEX-KG; S, signal peptide of the Rt41A proteinase.

identified between Ser19 and Ser20 using the '-3,-1' rule (von Heijne, 1983). The start of the mature proteinase was identified by homology with the amino terminal sequence obtained from Rt41A proteinase purified from the native organism (Peek *et al.*, 1992). The deduced mature peptide sequence is 276 amino acids long and contains Asp39, His72 and Ser224 which are tentatively identified as the 'catalytic triad' by analogy with other serine proteinases. It is proposed that the structure of the peptide produced by *aprA* is: a 19 amino acid signal peptide followed by a 'pro-region' of 113 amino acids and a catalytic domain of 276 amino acids (Fig. 1).

Expression of the proteinase in *E. coli*

aprA was amplified in PCR reactions and its product was inserted into a variety of *E. coli* expression vectors. Active protein was obtained only when using the plasmid vector pGEX-KG (Guan & Dixon, 1991), which produced the protein as a carboxy-terminal fusion with glutathione-S transferase (GST). Two different fusions were produced, pNZ1997 and pNZ1989 (Fig. 2). These plasmids differed in that the signal peptide of *aprA* was deleted in pNZ1989. Cell extracts prepared from induced cultures containing pNZ1997 displayed no detectable proteolytic activity when assayed with azocasein using standard conditions, although a small amount of activity could be seen in the casein agar plate assay. However, when cell extracts were incubated at 85 °C, there was a progressive increase in the level of proteolytic activity. With pNZ1997, extracts needed to be incubated at 85 °C for at least 1 h before active proteinase could be detected. Activation was temperature-dependent, with the fusion protein requiring a 4 h incubation at 70 °C or a 5 min incubation at 100 °C to produce mature active proteinase; however, at the higher temperature activity was rapidly lost on prolonged incubation (Fig. 3). The protein produced from pNZ1989 (lacking the Rt41A signal peptide) had a different response to temperature. This enzyme required only a 20 min incubation at 85 °C before maximum proteolytic activity was detected (data not shown).

Fig. 3. Activation of recombinant Rt41A proteinase at different temperatures: ○, 100 °C; ●, 85 °C; □, 75 °C. Samples were assayed after increasing periods of time. The points are mean values from duplicate experiments and are expressed as a percentage of the maximum value measured for the sample. Error bars show the deviations between the duplicate values.

Fig. 4. SDS-PAGE analysis of recombinant Rt41A proteinase produced by *E. coli*. Lanes: 1, molecular mass markers; 2, unpurified total cell extract; 3, eluate from glutathione agarose beads; 4, eluate treated at 70 °C.

Characterization of recombinant Rt41A proteinase

The recombinant Rt41A enzyme produced by pNZ1997 (full length) was purified and partially characterized. SDS-PAGE analysis of affinity-purified, heat-activated samples showed that the sample contained a single protein of 32 kDa (Fig. 4). The amino terminal sequence of this protein was found to match that of Rt41A proteinase purified from the native organism and also matched the deduced peptide sequence of *aprA*. Thermostability of the recombinant proteinase was determined at 70 °C and 100 °C. After 24 h at 70 °C, no loss in activity was detected, whereas at 100 °C, the proteinase had a half-life of 7.7 min. In 10 min assays, Peek *et al.* (1992) showed that

Fig. 5. (a) Effect of EDTA or EGTA on proteolytic activity. (b) Effect of PMSF on proteolytic activity. Samples were pre-treated with the chelating agents or with PMSF as indicated. Error bars show the deviations between duplicate experiments.

the native proteinase had greatest activity at 90 °C. The same result was obtained by us with the recombinant enzyme (data not shown).

The chelators EDTA and EGTA were tested for their ability to inhibit the activity of the proteinase. Little loss in proteolytic activity was observed after 30 min incubations of the recombinant proteinase with 5 mM EDTA or EGTA on ice (Fig. 5a). However, incubation at 70 °C for 10 min with 10 mM EDTA reduced the level of remaining activity to less than 10% of the original value. Incubation of the recombinant proteinase with phenyl-

methylsulphonyl fluoride (PMSF) (1 mM) at 70 °C or on ice resulted in complete inhibition (Fig. 5b).

DISCUSSION

Primers RM5 and RM6 were designed to bind to regions of the gene which are highly conserved between a variety of genera, and redundancies were included to cope with differing codon usage. Evidence of the success of these primers is that they have also been used to isolate the gene encoding an alkaline serine proteinase from a thermophilic *Bacillus* sp. AK1 (MacIver *et al.*, 1994). The G + C content of the gene isolated from the *Bacillus* is 40 mol % whereas the Rt41A proteinase gene has a G + C content of 69 mol %. These results imply that the primers are broadly applicable for screening members of other genera for the presence of serine proteinases.

The high G + C content is characteristic of sequences obtained from the genus *Thermus*. It has been proposed that the high G + C content and non-random codon usage leads to poor translational efficiency in *E. coli* (Sørensen *et al.*, 1989). The preference for G or C in the third position of the codon is exceptionally high, although not as high as that found in some other genes from *Thermus* (see Kagawa *et al.*, 1984).

The Rt41A proteinase gene contains four cysteine residues which have been shown to form two disulphide bonds (Peek *et al.*, 1992). Of the known bacterial and fungal subtilisin-like serine proteinases, thermitase (Meloun *et al.*, 1985), proteinases K (Jany *et al.*, 1986), R and T (Samal *et al.*, 1991) and exoprotease A (Deane *et al.*, 1989) are the only other enzymes containing cysteine residues which are thought to form or have been shown to form disulphide bonds. With the exception of exoprotease A, these proteinases are thermostable, exhibiting temperature optima ranging from 55 to 75 °C. In addition, all of these proteinases show greater stability in SDS than the non-cysteine-containing Subtilisin BPN' (Deane *et al.*, 1987). Rt41A proteinase retains nearly 90 % activity after a 1 h incubation in 5 % SDS at 4 °C and 73 % after one hour at 70 °C in 1 % SDS (Peek *et al.*, 1992). Disulphide bonds are thought to contribute to thermal stability; however, they are not the sole determinant (see, for example, Takagi *et al.*, 1990; Pantoliano *et al.*, 1987; Wells & Powers, 1986), and exoprotease A has two disulphide bonds in the same position as aqualysin I and Rt41A proteinase, but this enzyme is not thermostable (Deane *et al.*, 1987).

Alignment of the deduced peptide sequence for the Rt41A proteinase with other serine proteases (Fig. 6) shows that considerable homology exists, particularly around the residues of the active site 'catalytic triad' (Asp, His and Ser) and Asp155 (the oxyanion pocket). Homology around the catalytic residues is greatest for the His72 and Ser224 residues and is less highly conserved for the Asp38 residue. Aqualysin I and Rt41A share a 71 % homology, with greatest homology at the amino-terminus of the mature enzyme, i.e. close to the active site. Homology in the pre-pro-region is much lower.

Despite the obvious sequence similarities between the mature aqualysin I and Rt41A proteinases, there are substantial DNA differences upstream and downstream of the functional proteinase sequences. There are also a number of functional differences that are evident from studies of the recombinant enzymes. Both have identical molecular masses for the mature enzymes (28 400 Da), but different temperature optima (90 °C for Rt41A proteinase and 95 °C for aqualysin). Aqualysin I has a pH optimum of 10.4 as compared with pH 8.0 for Rt41A proteinase, and they have different cleavage patterns for the oxidized insulin B chain (Peek *et al.*, 1992). Furthermore, the precursor of aqualysin I consists of four discrete regions: an amino-terminal signal peptide, an amino-terminal pro-sequence, a proteinase region and a carboxy-terminal pro-sequence. Lee *et al.* (1991) showed that deletions in the amino-terminal sequence produced by restriction enzyme digestion and re-ligation of the gene, eliminated proteolytic activity and decreased the stability of the recombinant enzyme in *E. coli*. These experiments suggested that the amino-terminal pro-sequence is necessary for the production of the active enzyme. In other experiments, it was shown that the carboxy-terminal pro-sequence was not essential for the production of active aqualysin I in *E. coli*. In contrast, no carboxy-terminal pro-region has been identified for Rt41A proteinase.

A preliminary copy of the deduced peptide sequence of the Rt41A proteinase (which contained errors) was aligned with prokaryotic and eukaryotic subtilisin-like serine proteinases by Siezen *et al.* (1991). We have used the correct sequence of Rt41A proteinase in the alignment with other subtilisin-like proteinases (Fig. 6). This alignment further highlights the conserved regions and shows that these regions appear to be conserved throughout the Archaea, Bacteria and Eukarya.

There has been increasing interest in the production of thermophilic enzymes including thermostable proteinases in the last 10–15 years because of their potential in industrial applications (Cowan *et al.*, 1985; Bergquist *et al.*, 1987; Coolbear *et al.*, 1992). Alkaline serine proteases have been described from several *Thermus* isolates including *T. ruber* (Skjenstad *et al.*, 1992; Souter *et al.*, 1992) and *Thermus* strains TOK3 (Saravani *et al.*, 1989), Ok6 (Jones *et al.*, 1988), ITI243 (Ævarsson *et al.*, 1991), IS-15 (Bjarnason *et al.*, 1992), Rt6 (Cowan *et al.*, 1987b) and Rt41A (Peek *et al.*, 1992). Only two of the serine proteases from *Thermus* have been expressed in *E. coli*: aqualysin I (Touhara *et al.*, 1991) and the Rt41A proteinase (this work). Expression of the Rt41A proteinase in *E. coli* was successful when it was produced as a protein fusion to GST. Normally, thermophilic enzymes that are fused to GST are inactive at their optimum temperature because at elevated temperatures the mesophilic GST portion of the fusion denatures and precipitates the protein. With fusions produced by the vector pGEX-KG, it is normal to remove the GST portion by cleavage with thrombin. In the case of the Rt41A protease, the GST is removed along with the pre-pro region during activation. As a result, a one-step purification process using glutathione agarose can be employed to produce exceptionally pure enzyme. A

```

1
Rt41A .....HUSPATWGLDRDQRTPLDGRITYTATRCUHAUUDTGILLSHQETGRIGKSVDAITPGGSA..QDCNGHG1
Aql .....ATGSPAPWGLDRDQRTPLDGRITYTATRCUHAUUDTGILLSHQETGRIGKSVDAITPGGSA..GGNG..QDCNGHG1
proA .....ANOTNAWGLDRDQRTPLDNNISANFQSTETAVUUDTGUNNAHVEEGGSUSGVDFUNDADA..SDONGHG1
proK .....AATGNAPWGLDRDQRTSPGTSTVYVESAGQSCUUYUUDTGLEASHREFEGRQMUKTYVY...SS..ADONGHG1
bpn .....AQSUUYVSQKAPAL...HSQGYTGSNUKJADTSGDSSHADL...KUGAGASMPSETNPFQDNHSG1
Carl .....AQTVYVSLPLKADKV...QAQGFKSNUKJADTSGDSSHADL...NUUGGASFGAGEAYN..TONGHG1
Therm VTPNDPVFSSRQVPPQKIQAPQA...WDIA.ESSAKIAIUDTGUQSNHFDLASKUUGGLDFUNDSTP..QNGNGHG1
81
Rt41A HUAGTUGS.....TIVGURKQUTHPURULQCNSSCSNSEDIAQLDHT....QHUKPAIINMSLGG..DASTALDTG
Aql HUAGTUGS.....UTVGRKQUNLYAURULQCNSSCSNSEDIAQLDHT....RHHARPAIINMSLGG..DUSTALDNR
proA HUAGTUGS.....SLYGURKQUNLYGURULSCSSCSNSEDIAQLDHT....AHASGPSAIINMSLGG..DQVULQSA
proK HUAGTUGS.....RTYGURKQTQIFGUKULQNNSSGOVETIAGMFAASDKNNRCPKGVASLSLGG..GVSSUNSA
bpn HUAGTUAALNNSIGULGURPSASVYKULGADSSGOVETIAGMFAASDKNNRCPKGVASLSLGG..GVSSUNSA
Carl HUAGTUAALDNTTIGULGURPSASVYKULGADSSGOVETIAGMFAASDKNNRCPKGVASLSLGG..GVSSUNSA
Therm HUAGTUAALNNSIGULGURPSASVYKULGADSSGOVETIAGMFAASDKNNRCPKGVASLSLGG..GVSSUNSA
161
Rt41A UNNAINASUUVVAAAGNDADACFVS...PARUTAAITUGATITSDYASFSNVCRLD..FAPGOSTISAVTSSTATNT
Aql UNNAINASUUVVAAAGNDADACFVS...PARUTAAITUGATITSDYASFSNVCRLD..FAPGOSTISAVTSSTATNT
proA UNNAINASUUVVAAAGNDADACFVS...PARUTAAITUGATITSDYASFSNVCRLD..FAPGOSTISAVTSSTATNT
proK UNNAINASUUVVAAAGNDADACFVS...PARUTAAITUGATITSDYASFSNVCRLD..FAPGOSTISAVTSSTATNT
bpn UNNAINASUUVVAAAGNDADACFVS...PARUTAAITUGATITSDYASFSNVCRLD..FAPGOSTISAVTSSTATNT
Carl UNNAINASUUVVAAAGNDADACFVS...PARUTAAITUGATITSDYASFSNVCRLD..FAPGOSTISAVTSSTATNT
Therm UNNAINASUUVVAAAGNDADACFVS...PARUTAAITUGATITSDYASFSNVCRLD..FAPGOSTISAVTSSTATNT
241
Rt41A ISGTSMATPHUAGVAGLAAVLQVETATPSQASALLVYETPNVUKNAGRYSPNLLVITPF...
Aql ISGTSMATPHUAGVAGLAAVLQVETATPSQASALLVYETPNVUKNAGRYSPNLLVITPF...
proA ISGTSMATPHUAGVAGLAAVLQVETATPSQASALLVYETPNVUKNAGRYSPNLLVITPF...
proK ISGTSMATPHUAGVAGLAAVLQVETATPSQASALLVYETPNVUKNAGRYSPNLLVITPF...
bpn ISGTSMATPHUAGVAGLAAVLQVETATPSQASALLVYETPNVUKNAGRYSPNLLVITPF...
Carl ISGTSMATPHUAGVAGLAAVLQVETATPSQASALLVYETPNVUKNAGRYSPNLLVITPF...
Therm ISGTSMATPHUAGVAGLAAVLQVETATPSQASALLVYETPNVUKNAGRYSPNLLVITPF...

```

Fig. 6. Comparison of the deduced mature peptide sequence of the Rt41A proteinase with subtilisins and subtilisin-like serine proteinases. Aql, aqualysin (Kwon *et al.*, 1988); proA, exoprotease A (Deane *et al.*, 1989); proK, proteinase K (Jany *et al.*, 1986); bpn, subtilisin BPN' (Wells & Estell, 1988); Carl, subtilisin Carlsberg (Jacobs *et al.*, 1985); Therm, thermitase (Meloun *et al.*, 1985).

heat-treatment step similar to that used for aqualysin I was required for the activation of the proteinase from the protein fusion.

Assays of the activation of the proteinase showed that there was a temperature-dependent 'lag' period where no proteinase was detectable. Rt41A proteinase is highly homologous with exoprotease A from *Vibrio alginolyticus* which was shown also to undergo a lag of up to 8 h between synthesis and the appearance of fully active protease during expression of the recombinant gene (Deane *et al.*, 1987, 1989). Hence the functional features of these two proteinases are remarkably similar despite their disparate origins and large differences in temperature stability.

Removal of the putative signal peptide from Rt41A proteinase virtually eliminated the lag period, suggesting that the signal peptide somehow interferes with the processing of the pro-enzyme to the mature enzyme. Activation of subtilases has been shown to be an intramolecular process with the pro-region responsible for the correct folding of the molecule (Kobayashi & Inouye, 1992; Ohta & Inouye, 1990). It is possible that the signal peptide (which is normally removed before processing in the native organism, Fujishige *et al.*, 1992; Terada *et al.*, 1990) prevents the processing of the pro-Rt41A to the mature enzyme, perhaps by causing the expressed protein to fold incorrectly, thereby inhibiting access of the proteinase to the pro-enzyme processing site.

ACKNOWLEDGEMENTS

This research was funded in part by Pacific Enzymes and the University of Auckland Research Grants Committee.

REFERENCES

- Ævarsson, A., Holst, O. & Kristjánsson, J. K. (1991). Growth behaviour and protease production by an Icelandic *Thermus* sp. isolate in batch and continuous culture. *Curr Microbiol* **23**, 1–6.
- Bergquist, P. L., Love, D. R., Croft, J. E., Streiff, M. B., Daniel, R. M. & Morgan, W. H. (1987). Genetics and potential biotechnological applications of thermophilic and extremely thermophilic micro-organisms. *Biotechnol Genet Eng Rev* **5**, 199–244.
- Bjarnason, J. B., Kristjánsson, H., Olafsdóttir, S., Gísladóttir, G., Fox, J. W., Kristjánsson, J. K. & Alfreðsson, G. (1992). Purification and characterisation of a thermophilic proteinase produced by *Thermus* strain IS-15. In *Thermophiles: Science and Technology*. IceTec: Reykjavík, Iceland.
- Blumentals, I. I., Robinson, A. S. & Kelly, R. M. (1990). Characterisation of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *Appl Environ Microbiol* **56**, 1992–1998.
- Borges, K. M. & Bergquist, P. L. (1992). A rapid method for preparation of bacterial chromosomal DNA in agarose plugs using *Thermus* Rt41A proteinase. *Biotechniques* **12**, 222–223.
- Coolbear, T., Daniel, R. M., Cowan, D. A. & Morgan, H. W. (1988). Proteases from extreme thermophiles. *Enzyme Eng* **9**, 279–281.
- Coolbear, T., Daniel, R. M. & Morgan, H. W. (1992). The enzymes from extreme thermophiles: bacterial sources, thermostabilities and industrial relevance. In *Advances in Biochemical Engineering/Biotechnology*, vol. 45, pp. 57–98. Edited by A. Fiechter. Berlin: Springer-Verlag.
- Cowan, D. A., Daniel, R. M. & Morgan, H. W. (1985). Thermophilic proteases: properties and potential applications. *Trends Biotechnol* **3**, 68–72.
- Cowan, D. A., Smolenski, K. A., Daniel, R. M. & Morgan, H. W. (1987a). An extremely thermostable extracellular proteinase from a strain of the archaeobacterium *Desulfurococcus* growing at 88 °C. *Biochem J* **247**, 121–133.
- Cowan, D. A., Daniel, R. M. & Morgan, H. W. (1987b). A

- comparison of extracellular serine proteases from four strains of *Thermus aquaticus*. *FEMS Microbiol Lett* **43**, 155–159.
- Deane, S. M., Robb F. T. & Woods, D. R. (1987). Production and activation of an SDS-resistant alkaline serine exoprotease of *Vibrio alginolyticus*. *J Gen Microbiol* **133**, 391–398.
- Deane, S. M., Robb, F. T., Robb, S. M. & Woods, D. R. (1989). Nucleotide sequence of the *Vibrio alginolyticus* calcium-dependent, detergent-resistant alkaline serine exoprotease A. *Gene* **76**, 281–288.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.
- Frömmel, C. & Sander, C. (1989). Thermitase, a thermostable subtilisin: comparison of predicted and experimental structures and the molecular cause of thermostability. *Proteins Struct Funct Genet* **5**, 22–37.
- Fujishige, A., Smith, K. R., Silen J. L. & Agard, D. A. (1992). Correct folding of alpha-lytic protease is required for its extracellular secretion from *Escherichia coli*. *J Cell Biol* **118**, 33–42.
- Fung, M. & Fung, K. Y. (1991). PCR amplification of mRNA directly from a crude cell lysate prepared by thermophilic protease digestion. *Nucleic Acids Res* **19**, 4300.
- Guan, K. L. & Dixon, J. E. (1991). Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione-S-transferase. *Anal Biochem* **192**, 262–267.
- von Heijne, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur. J Biochem* **133**, 17–21.
- Jacobs, M., Eliasson, M., Uhlen, M. & Flock, J. (1985). Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. *Nucleic Acids Res* **13**, 8913–8926.
- Janssen, P. H., Morgan, H. W. & Daniel, R. M. (1991). Effects of medium composition on extracellular proteinase stability and yield in batch cultures of a *Thermus* sp. *Appl Microbiol Biotechnol* **34**, 789–793.
- Jany, K., Lederer, G. & Mayer, B. (1986). Amino acid sequence of proteinase K from the mold *Tritirachium album* Limber. Proteinase K – a subtilisin-related enzyme with disulphide bonds. *FEBS Lett* **199**, 139–144.
- Jones, C. W., Morgan, H. W. & Daniel, R. M. (1988). Aspects of protease production by *Thermus* strain Ok6 and other New Zealand isolates. *J Gen Microbiol* **134**, 191–198.
- Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T. & Oshima, T. (1984). High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *J Biol Chem* **259**, 2956–2960.
- Kobayashi, T. & Inouye, M. (1992). Functional analysis of the intramolecular chaperone – mutational hot spots in the subtilisin pro-peptide and a 2nd-site suppressor mutation within the subtilisin molecule. *J Mol Biol* **226**, 931–933.
- Kwon, S., Terada, I., Matsuzawa, H. & Ohta, T. (1988). Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1 and characteristics of the deduced primary structure of the enzyme. *Eur J Biochem* **173**, 491–497.
- Lee, Y.-C., Miyata Y., Terada I., Ohta T. & Matsuzawa, H. (1991). Involvement of NH-terminal pro-sequence in the production of active aqualysin I (a thermophilic serine protease) in *Escherichia coli*. *Agric Biol Chem* **55**, 3027–3032.
- McHale, R. H., Stapleton, P. M. & Bergquist, P. L. (1991). A rapid method for the preparation of samples for PCR. *Biotechniques* **10**, 21–22.
- MacIver, B., McHale, R. H., Saul, D. J. & Bergquist, P. L. (1994). Cloning and sequencing of a serine proteinase gene from a thermophilic *Bacillus* species and its expression in *Escherichia coli*. *Appl Environ Microbiol* **60**, 3981–3988.
- Matsuzawa, H., Hamaoki, M. & Ohta, T. (1983). Production of thermophilic extracellular proteases (aqualysins I and II) by *Thermus aquaticus* YT-1, an extreme thermophile. *Agric Biol Chem* **47**, 25–28.
- Meloun, B., Baudys, M., Kostka, V., Hausdorf, G., Frömmel, C. & Höhne, W. E. (1985). Complete primary structure of thermitase from *Thermoactinomyces vulgaris* and its structural features related to the subtilisin-type proteinases. *FEBS Lett* **183**, 195–200.
- Ohta, Y. & Inouye, M. (1990). Pro-subtilisin E: purification and characterization of its autoprocessing to active subtilisin E *in vitro*. *Mol Microbiol* **4**, 295–304.
- Pantoliano, M. W., Ladner, R. C., Bryan, P. N., Rollence, M. L., Wood, J. F. & Poulos, T. L. (1987). Protein engineering of subtilisin BPN', enhanced stabilization through the introduction of two cysteines to form a disulfide bond. *Biochemistry* **26**, 2077–2082.
- Peek, K., Daniel, R. M., Monk, C., Parker, L. & Coolbear, T. (1992). Purification and characterisation of a thermostable proteinase isolated from a *Thermus* sp. strain Rt41A. *Eur J Biochem* **207**, 1035–1044.
- Samal, B. B., Karan, B., Boone, T. C., Osslund, T. D., Chen, K. K. & Stabinsky, Y. (1990). Isolation and characterization of the gene encoding a novel, thermostable serine proteinase from the mould *Tritirachium album* Limber. *Mol Microbiol* **4**, 1789–1792.
- Samal, B. B., Karan, B., Parker, C. & Stabinsky, Y. (1991). Isolation and thermal studies of two novel serine proteinases from the fungus *Tritirachium album* Limber. *Enzyme Microb Technol* **13**, 66–70.
- Saravani, G., Cowan, D. A., Daniel, R. M. & Morgan, H. W. (1989). Caldolase, a chelator-insensitive extracellular serine proteinase from a *Thermus* sp. *Biochem J* **262**, 409–416.
- Saul D. J., Rodrigo, A. G., Reeves, R. A., Williams, L. G., Borges, K. M., Morgan, H. W. & Bergquist, P. L. (1993). Phylogeny of twenty *Thermus* isolates constructed from 16S rRNA gene sequence data. *Int J Syst Bacteriol* **43**, 754–760.
- Siezen, R. J., de Vos, W. M., Leunissen, J. A. M. & Dijkstra, B. W. (1991). Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng* **4**, 719–737.
- Skjenstad, T., Hreggvidsson, G. O., Gudjónsson, T. & Kristjánsson, J. K. (1992). Isolation and some properties of a protease from *Thermus ruber*. In *Thermophiles: Science and Technology*. IceTec: Reykjavík, Iceland.
- Sørensen, M. A., Kurland, C. G. & Pedersen, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol* **207**, 365–377.
- Souter, N. H., Sharp, R. J. & Marks, T. S. (1992). Characterisation of an alkaline serine protease from *Thermus ruber*. In *Thermophiles: Science and Technology*. IceTec: Reykjavík, Iceland.
- Takagi, H., Takahashi T., Momose H., Inouye M., Maeda Y., Matsuzawa, H. & Ohta, T. (1990). Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease. *J Biol Chem* **265**, 6874–6878.
- Terada, I., Kwon, S., Miyata, Y., Matsuzawa, H. & Ohta, T. (1990). Unique precursor structure of an extracellular protease, aqualysin I, with NH₂- and COOH-terminal pro-sequences and its processing in *Escherichia coli*. *J Biol Chem* **265**, 6576–6581.
- Touhara, N., Taguchi, H., Koyama, Y., Ohta, T. & Matsuzawa, H. (1991). Production and extracellular secretion of aqualysin-I (a

thermophilic subtilisin-type protease) in a host-vector system for *Thermus thermophilus*. *Appl Environ Microbiol* **57**, 3385–3387.

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

Wells, J. A. & Estell, D. A. (1988). Subtilisin – an enzyme designed to be engineered. *Trends Biochem Sci* **13**, 291–297.

Wells, J. A. & Powers, D. B. (1986). In vivo formation and stability

of engineered disulfide bonds in subtilisin. *J Biol Chem* **261**, 6564–6570.

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

Received 25 November 1994; revised 24 February 1995; accepted 17 March 1995.