An intracellular protease of the crenarchaeon *Sulfolobus solfataricus*, which has sequence similarity to eukaryotic peptidases of the CD clan

Annamaria GUAGLIARDI^{*1}, Laura CERCHIA^{*} and Mosè ROSSI^{*†}

^{*}Dipartimento di Chimica Biologica, Università “Federico II” di Napoli, Via Mezzocannone 16, 80134 Napoli, Italy, and [†]Istituto di Biochimica delle Proteine, CNR, Via P. Castellino 111, 80131 Napoli, Italy

We purified from crude extracts of the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* a protease that is able to hydrolyse proteins with a pH optimum of 7.5 and a temperature optimum of 70 °C. Assays in the presence of classical protease inhibitors showed that the hydrolytic activity is sensitive to thiol-blocking reagents. Fluorescence assays using synthetic peptides demonstrated that the protease has a preference for cleaving glutamic acid residues. The first 12 residues of the protease match the N-terminus residues of a hypothetical protein in the *S. solfataricus* genome of 95 amino acids in length and calculated molecular mass of 11 072 Da. The whole sequence of the protease

is not related to any known protein, but it bears a segment which is highly similar to one containing the active cysteine residue in eukaryotic peptidases known as legumains. This is the first protease isolated from *S. solfataricus* capable of degrading native proteins effectively. Our results add to the knowledge of the intracellular proteolytic machine in hyperthermophilic microorganisms.

Key words: archaea, hyperthermophile, protease, protein degradation, *Sulfolobus*.

INTRODUCTION

Most of the organisms capable of growth at and above 75 °C (the so-called hyperthermophiles) are prokaryotes which belong to the Archaea domain [1]; hyperthermophilic species can be found in both Crenarchaeota and Euryarchaeota kingdoms. Cytosolic proteases have a key role in the cell physiology of all organisms by modulating the availability of certain regulatory proteins, preventing the accumulation of abnormal proteins, and contributing to the amino acid pool that enables survival under nutrient-limited growth. Hyperthermophiles display a peculiar intracellular proteolytic machine. Genome sequence analyses have shown that the Clp family of proteases which occur in both bacteria and eukaryotes is absent in all hyperthermophilic archaea, and that no homologue of Lon that is ubiquitous across all three domains of life could be identified in the genome of the hyperthermophilic crenarchaeon *Sulfolobus solfataricus*; therefore proteases are believed to be among the genes in hyperthermophilic genomes that are still unassigned in terms of function [2]. Table 1 summarizes the intracellular proteases isolated so far from hyperthermophilic archaea [3–20]: the organisms are all Euryarchaea, except for *Pyrobaculum aerophilum*, *Aeropyrum pernix* and *S. solfataricus*, which belong to the Crenarchaeota kingdom. Some archaeal proteases are novel proteins. The sequence-related cysteine proteases of *Pyrococcus furiosus* [8] and *Pyrococcus horikoshii* [13], as well as the serine prolyl endopeptidase of *P. furiosus* [7], have no sequence similarity to any known protease family; the *Thermococcus litoralis* cysteine pyrrolidone carboxyl peptidase has no sequence or structural similarity to other members of the cysteine protease family, but it shows considerable similarities to other hydrolytic enzymes of widely differing substrate specificity and mechanism,

suggesting that they are the products of divergent evolution from a common ancestor [15].

To increase knowledge of the intracellular protein degradation in hyperthermophilic archaea, we screened crude extracts of *S. solfataricus* for proteases that degrade proteins. A protease was purified to homogeneity, which has a preference for cleaving at glutamic acid residues and is inhibited by cysteine-blocking reagents. The protein primary structure is not related to that of any known protein, but it bears a segment that is very similar to

Table 1 Intracellular proteases described in hyperthermophilic archaea

Organism	Protease	Reference
<i>Thermoplasma acidophilum</i>	20 S proteasome	3
	Tricorn protease and interacting aminopeptidases	4
<i>Methanococcus jannaschii</i>	20 S proteasome	5
<i>Pyrococcus furiosus</i>	Metallo carboxypeptidase	6
	Serine prolyl endopeptidase	7
	Cysteine protease	8
	Cysteine pyrrolidone carboxyl peptidase	9
	Aminopeptidase	10
	Methionine aminopeptidase	11
<i>Pyrococcus horikoshii</i>	Proline dipeptidase	12
	Cysteine protease	13
	Aminopeptidase	14
	Cysteine pyrrolidone carboxyl peptidase	15
<i>Thermococcus litoralis</i>	Serine protease (Aerolysin)	16
<i>Pyrobaculum aerophilum</i>	Serine protease (Pernilase)	17
<i>Aeropyrum pernix</i>	Metallo aminopeptidase	18
<i>Sulfolobus solfataricus</i>	Serine protease	19
	Metallo carboxypeptidase	20

Abbreviations used: Ac, acetyl; AMC, 7-amido-4-methylcoumarin; CBz, benzylloxycarbonyl; E-64, *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; Suc, *N*-succinyl.

¹ To whom correspondence should be addressed (e-mail guaglia@unina.it).

that containing the active cysteine residue in plant legumains, which belong to the CD clan of cysteine peptidase, together with eukaryotic caspases and separins, and clostripains and gingipains from pathogenic bacteria [21].

EXPERIMENTAL

Materials

Lysozyme from hen's-egg white (183 units/mg), the peptides L-Glu-7-amido-4-methylcoumarin (L-Glu-AMC), *N*-succinyl-(Suc-)Ala-Glu-AMC, Suc-Leu-Leu-Val-Tyr-AMC, Suc-Ala-Ala-Phe-AMC, Suc-Ala-Phe-Lys-AMC and benzyloxycarbonyl-(CBz-)Gly-Gly-Arg-AMC, and the inhibitors were from Sigma. The peptides CBz-Ala-Ala-Asn-AMC and acetyl-(Ac-)Asp-Glu-Val-Asp-AMC were from Bachem. *Bacillus stearothermophilus* alcohol dehydrogenase (160 kDa) was purified as described by Guagliardi et al. [22]. The other chemicals were reagent-grade.

Analytical methods

Protein concentration was determined by the assay of Bradford [23] using BSA as the standard. SDS/PAGE analysis [15 % (w/v) acrylamide] was performed as described by Laemmli [24]. N-terminal sequence analysis was performed by automated Edman degradation on an Applied Biosystems Mod.473A Sequencer using the manufacturer's standard programming and chemicals.

Cell growth and purification of the protease

Cells of *S. solfataricus* strain P2 were grown aerobically at 75 °C in DSM 182 culture medium. Cells of the exponential phase (15 g) were added to 30 g of sand (50–150 mesh) and 30 ml of 50 mM Tris/HCl, pH 7.5, containing 0.2 M NaCl, 5 % (v/v) glycerol, and homogenized in a Homni mixer. The sand was removed by centrifugation at 4000 *g* for 20 min at 4 °C, and the supernatant resulting from an ultracentrifugation at 160 000 *g* for 90 min at 4 °C represented the crude extract (800 mg). The crude extract was loaded in separate runs on to a Superose 6 column (2.6 cm × 60 cm; Amersham Biosciences), which was eluted with 10 mM Tris/HCl, pH 7.5, containing 0.1 M NaCl at a flow rate of 1 ml/min; the active fractions were pooled and concentrated with an Amicon ultrafiltration apparatus equipped with a 500 kDa cut-off membrane (35 mg in 10 ml). The sample was dialysed against 100 vols. of 50 mM sulphuric acid for 3 h at 4 °C, and then ultracentrifuged at 140 000 *g* for 45 min at 4 °C. The supernatant was then extensively dialysed against 200 vols. of 10 mM Tris/HCl, pH 7.5 (buffer A). The dialysed sample (5 mg) was loaded on to a Mono S column (0.5 cm × 5 cm; Amersham Biosciences) that was equilibrated in buffer A and connected to an FPLC system; the bound proteins were then eluted by a linear 0–0.4 M NaCl gradient in buffer A at a flow rate of 0.5 ml/min in 40 min. The active fraction (0.1 mg) was dialysed against buffer A, and concentrated by a Savant vacuum centrifuge. Freshly prepared protein was used in all the experiments described in this paper.

Protease-activity assays

During the purification of the protein, the samples were assayed as follows: 20 µg of lysozyme was incubated at 60 °C in 10 mM Tris/HCl, pH 7.5, either alone (the control) or in the presence of

a 5 µg aliquot of the sample (final volume 0.5 ml); after a 30 min incubation, the reactions were stopped in ice and the lytic activity was assayed at room temperature by measuring the absorbance decrease at 350 nm of a 0.1 mg/ml suspension of *Escherichia coli* cells upon the addition of a 100 µl aliquot withdrawn from each mixture. The lytic activity relative to the control mixture was taken as 100 %; the samples containing the protease caused a decrease of lytic activity as a consequence of lysozyme degradation.

The hydrolytic activity of pure protease was quantified as follows (standard conditions): 50 µg of *B. stearothermophilus* alcohol dehydrogenase was incubated at 70 °C in 10 mM Tris/HCl, pH 7.5, alone (the control) or in the presence of 0.2 µg of protease (final volume of 200 µl); after 15 min incubation, the mixtures were loaded separately on to a Biosep Sec-S4000 column (Phenomenex; dimensions, 0.78 cm × 30 cm; eluent, 0.1 M sodium phosphate, pH 7.0, containing 0.1 M NaCl; flow rate, 0.8 ml/min) connected to an HPLC system equipped with an area integrator (monitoring at 280 nm). The rate of protein degradation (mg of protein degraded/min per mg of protease) was obtained by integrating the area under the peak of the protein substrate in the control mixture and determining the decrease of the area in the run relative to the mixture containing the protease. Under the assay conditions, linear rates of protein hydrolysis were recorded for up to 1 h.

The cleavage of synthetic peptides was assayed as follows: the mixture contained 50 µM freshly prepared AMC-derived peptide and 5 µg of pure protease in 20 mM Tris/HCl, pH 7.5 (final volume of 200 µl); after 20 min incubation at 50 °C, the reaction was stopped by adding 1 ml of 0.1 M sodium acetate, pH 4.3, and free AMC was detected using an excitation wavelength of 365 nm and by measuring fluorescence emission at 450 nm with a PerkinElmer LS 50B spectrofluorimeter. In rate calculations (nmol of AMC released/min per mg of protease), the amount of spontaneous hydrolysis has been corrected for.

RESULTS AND DISCUSSION

Purification of the protein

The *S. solfataricus* crude extract, which was prepared in the absence of any protease inhibitor, was fractionated by gel-filtration chromatography on a Superose 6 column (Figure 1A). In order to recognize proteases active in degrading proteins at high temperatures, we assessed an assay which uses as substrate hen's egg-white lysozyme, a commercial enzyme that exhibits remarkable thermal stability (see the Experimental section). Briefly, lysozyme solutions were incubated for 30 min at 60 °C in the absence (the control mixture) and in the presence of an aliquot from each column fraction, and the lytic activity was then assayed on aliquots withdrawn from each mixture: the fractions that were able to decrease the lytic activity were considered to be 'active' (lysozyme retained full activity in the control mixture). Only the fraction that eluted in the void volume of the column was found active. We confirmed that the decrease in lytic activity resulted from the degradation of lysozyme. At the end of the incubation at 60 °C, the control mixture and the mixture containing lysozyme plus an aliquot of the active sample were loaded on to an analytical gel-filtration column that separates lysozyme from the proteins of the active sample that eluted with the void volume of the column [a Waters 300 SW column (0.8 cm × 30 cm; Millipore) equilibrated in 0.1 M sodium phosphate, pH 7.0, was used, containing 0.1 M NaCl at a flow rate of 0.5 ml/min, and connected to an HPLC system; Figure 1B]. The amount of lysozyme in the control mixture (run a) was higher than the

Figure 1 Identification of the protease which degrades native proteins in crude extracts of *S. solfataricus*

(A) The crude extract (50 mg) was loaded on to a Superose 6 column (2.6 cm $\times$ 60 cm) which was eluted with 10 mM Tris/HCl, pH 7.5, containing 0.1 M NaCl at a flow rate of 1 ml/min, and five fractions were pooled. The fractions (5 μ g/each) were tested for the capability to lower the enzyme activity of a lysozyme solution, as described in the Experimental section; the fraction shaded in the Figure was found to be active. The arrows indicate the elution volumes of the protein markers (sizes in kDa). (B) Elution profiles of the control mixture containing lysozyme alone (run a) and the mixture containing lysozyme plus the aliquot of the active sample from Superose 6 column (run b) after 30 min incubation at 60 $^{\circ}$ C with gel-filtration chromatography (see the Experimental section). (C) SDS/PAGE analysis (5 μ g/lane) of the active sample from Superose 6 column (lane 2): the migration positions of the protein markers (in kDa) are shown (lane 1). After the electrophoresis, the gel was silver-stained.

amount relative to the mixture incubated in the presence of the active sample (run b), which contained trichloroacetic-acid-soluble peptides (results not shown).

The SDS/PAGE analysis of the active sample from the Superose 6 column showed an array of bands (Figure 1C). Electron-microscopy analyses excluded the presence of ordered protein complexes in solution, and ethidium bromide staining of agarose gels showed the presence of large DNA fragments in the sample (results not shown). Incubation of the sample in the presence of DNase I and separation on the Superose 6 column resulted in the disappearance of the large peak, with the

concomitant occurrence of several peaks in the range 30–300 kDa, which were capable of degrading lysozyme and contained DNA (results not shown). This result suggests the existence of an interaction of the species responsible for the protein degradative activity with the nucleic acid or with DNA-bound proteins.

We were unable to fractionate the proteins by conventional chromatographic techniques even upon exposures of the sample to high-salt concentrations, chaotropic agents or detergents, this demonstrating the strong DNA–protein and protein–protein interactions. We succeeded in purifying the protease as follows.

Figure 2 Fractionation of the acid-soluble sample by cation-exchange chromatography

(A) The active sample after an acid wash (5 mg) was extensively dialysed against 10 mM Tris/HCl, pH 7.5, and loaded on to a Mono S column (0.5 cm $\times$ 5 cm) equilibrated in 10 mM Tris/HCl, pH 7.5, and connected to an FPLC system; the proteins bound to the resin were eluted with a linear 0–0.4 M NaCl gradient (shown by the diagonal line in the Figure), and six fractions were pooled. The fractions (5 μ g/each) were tested for the capability to lower the enzyme activity of a lysozyme solution, as described in the Experimental section; the shaded fraction was found to be active. (B) SDS/PAGE analysis (5 μ g/lane) of the active sample from the column (lane 1); the migration positions of the protein markers (in kDa) are shown (lane 2). After the electrophoresis, the gel was silver-stained.

The active sample from the Superose 6 column was dialysed for 3 h against 50 mM sulphuric acid, and the acid-insoluble proteins and DNA were removed by centrifugation; the DNA-free supernatant was found to be active in the assay of lytic activity decrease, and it contained non-interacting proteins, which were fractionated by using Mono S cation-exchange chromatography. The flow-through of the column was found to be inactive, and the bound proteins were eluted by a 0–0.4 M NaCl linear gradient (Figure 2A); the peak which eluted at 0.34 M NaCl was the only one capable of decreasing the lytic activity. The sample showed a single band of approx. 10 kDa on SDS/PAGE analysis (Figure 2B) in the absence or presence of reducing agents. The yield from a typical purification procedure was 0.1 mg of pure protease starting from 0.8 g of crude extract; a 3 h dialysis against 0.2 M sulphuric acid did increase the protein content in the acid-soluble fraction, but it caused irreversible inactivation of the protease activity.

Gel-filtration chromatography on a Biosep Sec-S2000 column (Phenomenex; 0.78 cm $\times$ 30 cm) indicated that the protein exists as a monomer; no higher-molecular-mass forms in solution were observed, thus excluding a propensity of the protease for self-aggregation (results not shown).

Functional studies

The assay on the basis of lytic activity decrease allowed us to identify a protease that degrades proteins in crude extracts of the archaeon. However, this assay is inappropriate for quantifying the protease activity. In order to obtain protein-degradation rates, we measured the amount of protein substrate hydrolysed/min on the basis of chromatographic separations of the reaction mixtures (see the Experimental section). We chose the thermostable alcohol dehydrogenase from *B. stearothermophilus* as a substrate to characterize the protease activity.

The activity was determined at different pH values; higher hydrolysis rates were calculated at pH 7.5 (Figure 3A). When the activity was assayed at various temperatures (the thermophilicity of the enzyme), the protease exhibited its maximum hydrolysis rate at 70 °C (Figure 3B), which is close to the physiological temperature for the optimal growth of *S. solfataricus*.

Experiments aimed at investigating the heat resistance of the protease activity (the thermal stability of the enzyme) showed

Figure 3 Activity of the *S. solfataricus* protease as a function of pH (A) and temperature (B), and effect of temperature on protease stability (C)

The assays were performed under standard conditions (see the Experimental section), except that either (A) the pH (10 mM sodium acetate buffer was used for the experiment at pH 5.5; 10 mM Tris/HCl was used for the experiments in the pH range 7–9) or (B) the assay temperature was altered. (C) The activity was assayed following pre-incubation of the protease (0.02 mg/ml) in 10 mM Tris/HCl, pH 7.5, at 60 °C (○), 70 °C (●) or 80 °C (■).

that prior heating of the protease solution to 80 °C for 30 min or to 70 °C for 90 min stimulated the hydrolysis rates (Figure 3C). The activity stimulation upon pre-heating is a feature of some

Table 2 Effects of different inhibitors on the activity of the *S. solfataricus* protease

Standard assays were performed as described in the Experimental section in the presence of the compound indicated. The residual activity shown in the Table as '100' (in the absence of any added compound) represents 0.4 mg of protein degraded/min per mg of protease.

Compound added (concn)	Residual activity
None	100
PMSF (3 mM)	84
EDTA (5 mM)	100
Pepstatin A (10 μ M)	100
DTNB (5 mM)	51
NEM (5 mM)	5
E-64 (10 μ M)	100

Table 3 Cleavage specificity of the *S. solfataricus* protease

The assays were performed as described in the Experimental section. The rate of hydrolysis is shown as nmol of AMC released/min per mg of protease.

Peptide	Hydrolysis rate	Activity (%)
L-Glu-AMC	4.2	100
Suc-Ala-Glu-AMC	3.4	81
Ac-Asp-Glu-Val-Asp-AMC	0.6	14
Suc-Leu-Leu-Val-Tyr-AMC	0.5	12
Suc-Ala-Ala-Phe-AMC	0.5	12
Suc-Ala-Phe-Lys-AMC	0.46	11
CBz-Gly-Gly-Arg-AMC	0.4	9.5
CBz-Ala-Ala-Asn-AMC	0.3	7.1

enzymes from thermophilic sources, and is believed to reflect the activation of the low-activity status adopted by these catalysts at moderate temperatures.

We performed protein degradation assays in the presence of classical inhibitors of proteases (Table 2). The serine-protease-specific reagent, PMSF, slightly lowered the activity, whereas EDTA and the inhibitor of aspartic proteases, pepstatin A, had no effect. The presence in the reaction mixture of 5,5'-dithiobis-(2-nitrobenzoic acid) ('DTNB'), a reversible thiol-blocking reagent, reduced the proteolytic activity by approx. 50 %, whereas the irreversible thiol-blocking reagent, *N*-ethylmaleimide ('NEM') almost completely suppressed the activity. This behaviour suggested the involvement of cysteine residue(s) in the hydrolytic reaction. However, the enzyme proved resistant to inhibition by *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)-butane (E-64), an irreversible inhibitor of the papain family of cysteine proteases. The presence of ATP or other nucleotides in the reaction mixtures had no effect on the velocity of hydrolysis (results not shown).

The cleavage specificity of the protease was investigated by fluorescence assays with AMC-derived peptides, which were performed at 50 °C because of the instability of the peptides at higher temperatures. The assays demonstrated that the protease had a preference for cleaving after glutamic acid residues; it was capable of cleaving after tyrosine and phenylalanine residues (chymotrypsin-like activity), after lysine and arginine residues (trypsin-like activity), and after asparagine and aspartic acid residues, albeit to a much lower extent (Table 3). We obtained evidence to show that the protease degrades relatively large proteins (50 and 200 kDa) to yield small (< 500 Da) peptides,

this reflecting the cleavage of a protein molecule at different residues that may be accessible to proteolytic attack.

Amino acid sequence

A pure protease solution underwent N-terminal micro-sequencing, and 12 residues were identified: AKKRSKLEIIQA. This sequence matches residues 2–13 (the initiation methionine residue is missing) of a hypothetical protein in the *S. solfataricus* genome [25] of 95 amino acids in length and with a calculated molecular mass of 11072 Da and pI of 10.52 (Figure 4A). The basic character of the protein matches its solubility in acid, and its strong binding to the cation-exchange column, which was developed at pH 7.5.

The whole sequence does not reveal similarity to any other protein in data banks; the residues from 8 to 29 share 66 % identity with the residues from 10 to 30 of a hypothetical 109-amino-acid protein present in the *Sulfolobus tokodaii* genome [26]. The residues from positions 15 to 23, LEACKSGSP (including the only cysteine residue), are almost identical with the sequence segments that contain the catalytic cysteine residue in eukaryotic proteases that are known as legumains (Figure 4B). Legumains are a family of cysteine endopeptidases, members of which have been found in plants, blood flukes, mammalian lysosomes and the nematode *Caenorhabditis elegans*. Legumains, eukaryotic caspases and separins, and clostripains and gingipains from pathogenic bacteria, are members of the CD clan of cysteine peptidases, which have specific protein substrates in the cell [21]. It is noteworthy that the *S. solfataricus* protease shares with the members of this clan an insensitivity to inhibition mediated by E-64, which inactivates other families of cysteine peptidases.

In all peptidases of the CD clan, a block of four predominantly hydrophobic residues precedes the segments containing the catalytic cysteine [21]; in *S. solfataricus* protease, the segment LEACKSGSP is preceded by the hydrophobic residues KLEIIQA (from residue positions 7 to 13), which are similar to the residues that immediately precede the catalytic cysteine in human caspases (Figure 4C). Moreover, the cysteine residue present in the *S. solfataricus* protease and the catalytic cysteine residue in human caspases are followed by the basic residues of lysine and arginine respectively. The *S. solfataricus* protease shares the post-glutamyl-peptide-hydrolysing activity with caspases; in this respect, legumains cleave after asparagine residues, clostripain, gingipain R and separin cleave after arginine or lysine residues [21], and human legumain cleaves after aspartate residues [27].

One important difference between the archaeal protease and the cysteine peptidases of the CD clan must be emphasized. The activities of almost all CD clan enzymes depend on catalytic dyads of cysteine and histidine: strikingly, the *S. solfataricus* protease lacks histidine residues.

The involvement *in vivo* of the *S. solfataricus* protease described here is under investigation. It is noteworthy that, on occasion, the protease co-elutes from Mono S column together with a 14 kDa protein and a 7 kDa protein, which may be among its natural substrates (A. Guagliardi, unpublished work). Micro-sequence analyses identified the 14 kDa protein as a transcriptional regulator (SSO0048), and the 7 kDa protein as Sso7d, which is known to affect DNA topology [28,29] and to have a role as a chaperone in rescuing inactive protein assemblies [30]. Specific proteins which control DNA-related events (including transcriptional regulators) are among the physiological substrates of the proteases Lon and Clp in bacteria and eukarya.

A

MAKKRSKLEIIQAILEACKSGSPKTRIMYGANLSYALTGRYIKMLMDLEIKQEGKQYML

TKKGEELEDIRRFNDRKNMDQLKEKINSVLSIK

B

<i>S.solfataricus</i>	S	K	L	E	I	I	Q	A	I	L	E	A	C	K	S	G	S	P
<i>S.japonicum</i>	S	K			L	V	I	Y		V	E	A	C	E	S	G	S	M
Mouse	Q	K			M	V	F	Y	I		E	A	C	E	S	G	S	M
Bovine	Q	K			M	V	F	Y	I		E	A	C	E	S	G	S	M
Rat	Q	K			M	V	F	Y	I		E	A	C	E	S	G	S	M
Human		K			M	V	F	Y	I		E	A	C	E	S	G	S	M
<i>C.elegans</i>	Q				L	T	F	Y		L	E	A	C	E	S	G	S	M
jack bean					E	M	V	I	Y	I		E	A	C	E	S	G	S
soy bean					E	M	V	I	Y		V	E	A	C	E	S	G	S
castor bean		K				M	V	I	Y		V	E	A	C	E	S	G	S
<i>V.narbonensis</i>	S					L	V	F	Y		L	E	A	C	E	S	G	S
<i>P.vulgaris</i>						L	V	F	Y		L	E	A	C	E	S	G	S
<i>P.vulgaris</i>					E	M	V	I	Y		V	E	A	C	E	S	G	S
<i>Z.mays</i>	S					L	V	F	Y		L	E	A	C	E	S	G	S
<i>N.tabacum</i>	S	K				L	V	F	Y		M	E	A	C	E	S	G	S
<i>N.tabacum</i>					E	M	V	L	Y	I		E	A	C	E	S	G	S
<i>A.thaliana</i>	S					L	V	F	Y		L	E	A	C	E	S	G	S
<i>V.sativa</i>	S					L	V	F	Y		L	E	A	C	E	S	G	S

C

<i>S.solfataricus</i>	K	L	E	I	I	Q	A	I	L	E	A	C	K	S	G	S
caspase 3	K	L	F	I	I	Q	A					C	R		G	T
caspase 7	K	L	F	F	I	Q	A					C	R		G	T
caspase 6	K	I	F	I	I	Q	A					C	R		G	N
caspase 1	K	V	I	I	I	Q	A					C	R		G	

Figure 4 Primary structure of the *S. solfataricus* protease

(A) The amino acid sequence of the *S. solfataricus* protease. (B) Alignment of the residues from 6 to 23 of *S. solfataricus* protease with the sequence segments containing the catalytic cysteine residue of legumains. The accession numbers of the sequences are: Q97VX8, *S. solfataricus* protease; P42665, *Schistosoma japonicum* (blood fluke) legumain; O89017, mouse (*Mus musculus*) legumain; Q95M12, bovine (*Bos taurus*) legumain; Q9JLN3, rat (*Rattus norvegicus*) legumain; Q99538, human (*Homo sapiens*) legumain; Q9U589, *C. elegans* legumain; P49046, jack bean (*Canavalia ensiformis*) legumain; P49045, soy bean (*Canavalia ensiformis*) legumain; P49042, castor bean (*Canavalia ensiformis*) legumain; Q9XGB9, *Vicia narbonensis* legumain; O24325, *Phaseolus vulgaris* legumain; O24326, *Phaseolus vulgaris* legumain; Q9FER6, *Zea mays* legumain; Q9XG75, *Nicotiana tabacum* legumain; Q9XG76, *Nicotiana tabacum* legumain; P49047, *Arabidopsis thaliana* legumain; P49044, *Vicia sativa* legumain. (C) Alignment of the residues from 7 to 23 of *S. solfataricus* protease with the sequence segments containing the catalytic cysteine residue of human caspases. The accession numbers of the sequences are: P42574, caspase 3; P55210, caspase 7; P55212, caspase 6; and P29466, caspase 1. The amino acids were aligned manually: residues were considered to be similar if they were hydrophobic and non-polar (A, V, I, L, P, M, F and W), polar and non-charged (G, S, T, C, Y, N and Q), basic (H, R and K) or acidic (E and D) (amino acid residues are shown using the single-letter code). Residues identical with the archaeal protease are indicated in bold type.

In addition to the role in controlling cellular events by the selective removal of key proteins and enzymes, Lon and Clp are responsible for the non-specific elimination of abnormal proteins generated by environmental injuries (e.g. those resulting from heat shock and exposures to toxic compounds), and their cellular levels are increased by various stresses [31]. The fact that the *S. solfataricus* protease is not overexpressed in heat-shocked cells of the archaeon [32] does not rule out *a priori* its involvement in a pathway of removal of damaged proteins.

This work was supported by the Program 'Inactivation of the causative agents of transmissible spongiform encephalopathies by thermophilic and hyperthermophilic

proteases' of the European Union, and by Ministero dell'Università e della Ricerca Scientifica.

REFERENCES

- 1 Wheelers, M. L., Kandler, O. and Woese, C. R. (1992) On the nature of global classification. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2930–2934
- 2 Ward, D. E., Shockley, K. R., Chang, L. S., Levy, R. D., Michel, J. K., Connors, S. B. and Kelly, R. M. (2002) Proteolysis in hyperthermophilic microorganisms. *Archaea* **1**, 63–74
- 3 Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. and Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533–539

- 4 Tamura, N., Lottspeich, F., Baumeister, W. and Tamura, T. (1998) The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell* **95**, 637–648
- 5 Wilson, H. L., Ou, M. S., Aldrich, H. C. and Maupin-Furlow, J. (2000) Biochemical and physical properties of the *Methanococcus jannaschii* 20S proteasome and PAN, a homolog of the ATPase (Rpt) subunits of the eucaryal 26S proteasome. *J. Bacteriol.* **182**, 1680–1692
- 6 Cheng, T. C., Ramakrishnan, V. and Chan, S. I. (1999) Purification and characterization of a cobalt-activated carboxypeptidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Protein Sci.* **8**, 2474–2486
- 7 Harris, M. N., Madura, J. D., Ming, L. J. and Harwood, V. J. (2001) Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile *Pyrococcus furiosus*. *J. Biol. Chem.* **276**, 19310–19317
- 8 Chang, L. S., Hicks, P. M. and Kelly, R. M. (2001) Protease I from *Pyrococcus furiosus*. *Methods Enzymol.* **330**, 403–413
- 9 Tsunasawa, S., Nakura, S., Tanigawa, T. and Kato, I. (1998) Pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Pyrococcus furiosus*: cloning and overexpression in *Escherichia coli* of the gene, and its application to protein sequence analysis. *J. Biochem. (Tokyo)* **124**, 778–783
- 10 Tsunasawa, S. (1998) Purification and application of a novel N-terminal deblocking aminopeptidase (DAP) from *Pyrococcus furiosus*. *J. Protein Chem.* **17**, 521–522
- 11 Tahirrov, T. H., Oki, H., Tsukihara, T., Ogasahara, K., Yutani, K., Libeu, C. P., Izu, Y., Tsunasawa, S. and Kato, I. (1998) High-resolution crystals of methionine aminopeptidase from *Pyrococcus furiosus* obtained by water-mediated transformation. *J. Struct. Biol.* **121**, 68–72
- 12 Ghosh, M., Grunden, A. M., Dunn, D. M., Weiss, R. and Adams, M. W. (1998) Characterization of native and recombinant forms of an unusual cobalt-dependent proline dipeptidase (prolidase) from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **180**, 4781–4789
- 13 Du, X., Choi, I. G., Kim, R., Wang, W., Jancarik, J., Yokota, H. and Kim, S. H. (2000) Crystal structure of an intracellular protease from *Pyrococcus horikoshii* at 2-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14079–14084
- 14 Ando, S., Ishikawa, K., Ishida, H., Kawarabayasi, Y., Kikuchi, H. and Kosugi, Y. (1999) Thermostable aminopeptidase from *Pyrococcus horikoshii*. *FEBS Lett.* **447**, 25–28
- 15 Singleton, M., Isupov, M. and Littlechild, J. (1999) X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Thermococcus litoralis*. *Structure Fold. Des.* **7**, 237–244
- 16 Volkl, P., Markiewicz, P., Stetter, K. O. and Miller, J. H. (1994) The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. *Protein Sci.* **3**, 1329–1340
- 17 Chavez Crocker, P., Sako, Y. and Uchida, A. (1999) Purification and characterization of an intracellular heat-stable proteinase (pernilase) from the marine hyperthermophilic archaeon *Aeropyrum pernix* K1. *Extremophiles* **3**, 3–9
- 18 Hanner, M., Redl, B. and Stoffler, G. (1990) Isolation and characterization of an intracellular aminopeptidase from the extreme thermophilic archaeobacterium *Sulfolobus solfataricus*. *Biochim. Biophys. Acta* **1033**, 148–153
- 19 Burlini, N., Magnani, P., Villa, A., Macchi, F., Tortora, P. and Gueritore, A. (1992) A heat-stable serine proteinase from the extreme thermophilic archaeobacterium *Sulfolobus solfataricus*. *Biochim. Biophys. Acta* **1122**, 283–292
- 20 Colombo, S., Toietta, G., Zecca, L., Vanoni, M. and Tortora, P. (1995) Molecular cloning, nucleotide sequence, and expression of a carboxypeptidase-encoding gene from the archaeobacterium *Sulfolobus solfataricus*. *J. Bacteriol.* **177**, 5561–5566
- 21 Barrett, A. J. and Rawlings, N. D. (2001) Evolutionary lines of cysteine peptidases. *Biol. Chem.* **382**, 727–733
- 22 Guagliardi, A., Martino, M., Iaccarino, M., De Rosa, M., Rossi, M. and Bartolucci, S. (1996) Purification and characterization of the alcohol dehydrogenase from a novel strain of *Bacillus stearothermophilus* growing at 70 °C. *Int. J. Biochem. Cell Biol.* **28**, 239–246
- 23 Bradford, M. M. (1976) A rapid and sensitive method for quantitation of microgram quantities of proteins utilizing dye binding of proteins. *Anal. Biochem.* **72**, 248–254
- 24 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680–685
- 25 She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A. et al. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7835–7840
- 26 Kawarabayasi, Y., Hino, Y., Horikawa, H., Jinno, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A. et al. (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res.* **8**, 123–140
- 27 Halfon, S., Patel, S., Vega, F., Zurawski, S. and Zurawski, G. (1998) Autocatalytic activation of human legumain at aspartic acid residues. *FEBS Lett.* **438**, 114–118
- 28 Guagliardi, A., Napoli, A., Rossi, M. and Ciaramella, M. (1997) Annealing of complementary DNA strands above the melting point of the duplex promoted by an archaeal protein. *J. Mol. Biol.* **267**, 841–848
- 29 Lopez-Garcia, P., Knapp, S., Ladenstein, R. and Forterre, P. (1998) *In vitro* DNA binding of the archaeal protein Sso7d induces negative supercoiling at temperatures typical for thermophilic growth. *Nucleic Acids Res.* **26**, 2322–2328
- 30 Guagliardi, A., Cerchia, L., Moracci, M. and Rossi, M. (2000) The chromosomal protein Sso7d of the crenarchaeon *Sulfolobus solfataricus* rescues aggregated proteins in an ATP hydrolysis-dependent manner. *J. Biol. Chem.* **275**, 31813–31818
- 31 Gottesman, S., Wickner, S. and Maurizi, M. R. (1997) Protein quality control: triage by chaperones and proteases. *Genes Dev.* **11**, 815–823
- 32 Cerchia, L., Rossi, M. and Guagliardi, A. (2000) An archaeal chaperonin-based reactor for renaturation of denatured proteins. *Extremophiles* **4**, 1–7

Received 1 July 2002/1 August 2002; accepted 6 August 2002

Published as BJ Immediate Publication 6 August 2002, DOI 10.1042/BJ20021017