University of Montana

ScholarWorks at University of Montana

Biological Sciences Faculty Publications

Biological Sciences

4-1-1997

A carboxy-terminal processing protease gene is located immediately upstream of the invasion-associated locus from Bartonella bacilliformis

Samuel J. Mitchell
University of Montana, Missoula

Michael F. Minnick
University of Montana, Missoula

Follow this and additional works at: https://scholarworks.umt.edu/biosci_pubs

Part of the Biology Commons

Let us know how access to this document benefits you.

Recommended Citation
Mitchell, Samuel J. and Minnick, Michael F., "A carboxy-terminal processing protease gene is located immediately upstream of the invasion-associated locus from Bartonella bacilliformis" (1997). *Biological Sciences Faculty Publications*. 468.
https://scholarworks.umt.edu/biosci_pubs/468

This Article is brought to you for free and open access by the Biological Sciences at ScholarWorks at University of Montana. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

A carboxy-terminal processing protease gene is located immediately upstream of the invasion-associated locus from *Bartonella bacilliformis*

Samuel J. Mitchell and Michael F. Minnick

Author for correspondence: Michael F. Minnick.Tel: +1 406 243 5972. Fax: +1 406 243 4184.
e-mail: minnick@selway.umt.edu

Division of Biological Sciences, The University of Montana, Missoula, MT 59812-1002, USA

A gene with homology to those encoding an unusual class of C-terminal processing proteases that flanks the invasion-associated locus *ialAB* of *Bartonella bacilliformis* has been identified. The 1302 bp gene, termed *ctpA*, is located immediately upstream of the *ialA* gene and encodes a predicted nascent product of 434 amino acids, producing a mature protein of 411 amino acid residues. The *Bartonella* CtpA appears to undergo autolysis *in vitro*, producing multiple products of 43–46 kDa, and a second group of products of 36–37 kDa. Production of CtpA *in vivo* gives a single product of 41·8 kDa. In addition to a computer-predicted N-terminal secretory signal sequence, the molecular mass difference *in vivo* versus *in vitro* indicates that CtpA is likely to be secreted and post-translationally modified. The full-length CtpA protein shows 30% identity to the CtpA protein of *Synechocystis* sp. 6803 (69% overall sequence similarity). The mature CtpA protein also has significant homology to the tail-specific protease (Tsp) of *Escherichia coli*, with 22% identity and 62% similarity to an internal region of the 660 amino acid Tsp. The CtpA protein does not appear to exhibit haemolysin, collagenase, or caseinase activity. The *ctpA* gene is conserved in all *Bartonella* species examined, as determined by hybridization analyses, but it was not found in *Brucella abortus* or *E. coli*. The *ctpA* gene does not directly affect the erythrocyte-invasion phenotype conferred by *ialAB*, but its homology to other stress-response processing proteases implies an important role in survival of this intracellular pathogen.

Keywords: CtpA, Tsp, *ialAB*, stress-response protease, autolysis

INTRODUCTION

Bartonella bacilliformis is a member of a group of unusual intracellular pathogens of growing medical importance. The bacteria within this group cause a severe haemolytic anaemia (*B. bacilliformis*; Reynafarje & Ramos, 1961), cat-scratch disease and bacillary angiomatosis (*B. henselae*; Regnery *et al.*, 1992; Regnery & Tappero, 1995), and trench fever (*B. quintana*; Vinson & Fuller, 1961). Within the genus *Bartonella*, the ability to invade erythrocytes appears to be limited to *B. bacilliformis* (Reynafarje & Ramos, 1961) and *B. henselae* (Kordick & Breitschwerdt, 1995). We have previously identified a two-gene locus from *B. bacilliformis* (*ialAB*) which confers the ability to invade human erythrocytes upon non-invasive *Escherichia coli* (Mitchell & Minnick, 1995). Because so little is known about the molecular biology of this activity, we undertook the present study in an attempt to identify other genes which might interact with the *ialAB* locus. In addition, the possibility that *ialAB* might lie in the midst of a ‘pathogenicity island’ [Pais; recently reviewed by Falkow (1996)] led us to analyse chromosomal DNA flanking the *ialAB* genes. We have identified a homologue of an interesting class of carboxy-terminal processing proteases (Ctpases) immediately upstream of the *ialAB* locus.

As a whole, carboxy-terminal proteases are not well understood. To date, the best-characterized CtpA is that from the cyanobacterium *Synechocystis* sp. 6803 [SYCCTPA; Shestakov *et al.*, 1994]. This 34 kDa enzyme is responsible for cleavage of 16 C-terminal

Abbreviations: IVTT, *in vitro* transcription/translation; SYCCTPA, CtpA from *Synechocystis* sp. 6803.

The GenBank accession numbers for *ctpA* and the sequence downstream of *ialB* (see Fig. 1, plasmid pSAC2) are L37094 and L46591, respectively.

0002-1153 © 1997 SGM

residues from its target protein, the D1 precursor polypeptide of photosystem II (PSII) (Nixon *et al.*, 1992). The D1 precursor is translated as a polypeptide which binds to the PSII complex and is subsequently processed to its mature form by Ctpase activity (reviewed by Barber & Andersson, 1992). The complex is turned over very rapidly because of light-induced damage to D1 (Barber & Andersson, 1992), thus implying a vital role for CtpA in maintenance of PSII activity in response to environmental stress.

A protein with a conserved core sequence similar to that of the C-terminal proteases is the Tsp (tail-specific protease) of *E. coli* (Silber *et al.*, 1992). Tsp was originally identified as a 76 kDa periplasmic protein involved in processing the penicillin-binding protein 3 (PBP 3) of *E. coli* (Hara *et al.*, 1991). In the absence of Tsp, *E. coli* was extremely sensitive to heat and osmotic stress. More recently, Tsp has been shown to recognize and degrade several proteins with nonpolar C-termini, with strongest preference for alanine at the three C-terminal residues (Keiler & Sauer, 1996). The Tsp of *E. coli* also functions as a universal degradative enzyme for aberrant proteins tagged by a C-terminal hydrophobic sequence (Keiler *et al.*, 1996). Thus, Tsp appears to degrade aberrant proteins arising as a result of environmental stress or anomalous processing. A *tsp* homologue may also play a role in virulence and has been implicated in the survival of *Salmonella typhimurium* within macrophages (Bäumler *et al.*, 1994).

In this study, we describe a *Bartonella* homologue of the carboxy-terminal processing proteases. Although the *ctpA* gene does not appear to directly affect the erythrocyte-invasion phenotype encoded by *ialAB*, the function of C-terminal proteases in response to stress in other organisms implies a significant role for the encoded protein. We also describe several intriguing features of the encoded protein: namely, a putative autocatalytic modification of its carboxy terminus, and a possible alternative translation start site which may give rise to two forms of the protein, perhaps targeted to different cellular locations.

METHODS

Bacterial strains, plasmids and growth conditions. Bacterial strains and plasmids are listed in Table 1. *B. bacilliformis* strains KC583 and KC584 (American Type Culture Collection) were grown on heart infusion agar (Difco) supplemented with 5% (v/v) defibrinated sheep erythrocytes and 5% (v/v) filter-sterilized sheep serum (Quad-Five) (HIBB plates) in a water-saturated atmosphere at 30 °C. *B. quintana*, *B. elizabethae* and *B. vinsonii* were obtained from the American Type Culture Collection in lyophilized form. *B. henselae* was obtained from the Centers for Disease Control & Prevention. *Br. abortus* was obtained as a frozen whole-cell suspension from Burroughs Wellcome.

E. coli strains were grown in Luria–Bertani (LB) medium at 37 °C with constant shaking. When required, ampicillin was added to the growth medium to a final concentration of 100 μg ml^{-1}. Haemolysis was assayed on HIBB plates supplemented with ampicillin as necessary. Collagenase activity was assayed in 5 ml nutrient gelatin tubes [3 g Bacto beef extract (Difco) l^{-1}, 5 g peptone l^{-1}, 120 g gelatin l^{-1}, pH 6·8] with growth at 37 °C and gelling observed at 20 °C.

Nucleic acid isolation and manipulations. Chromosomal DNA was isolated from *B. bacilliformis* using a standard hexadecyl trimethyl ammonium bromide (CTAB) method (Ausubel *et al.*, 1995). Individual DNA fragments were extracted from ethidium-bromide-stained agarose gels using GeneClean II (Bio101). Large-scale purifications of plasmid DNA were done with a Qiagen Midi-kit. Transformation of *E. coli* with plasmids was done by the methods of Chung *et al.* (1989). A bacteriophage λ vector, LambdaGem-11 (Promega), was used to construct a library of partial *Sau*3AI-digested chromosomal DNA fragments from *B. bacilliformis* KC583 by a standard protocol (Ausubel *et al.*, 1995). The library was subsequently probed with the *Bam*HI insert of pIAL1 by standard hybridization protocol (Ausubel *et al.*, 1995) to identify plaques containing the *ialAB* locus plus contiguous DNA. A LambdaGem-11 recombinant identified by this method was subsequently subcloned into the vector pUC19 and named pSAC2 (Fig. 1). After identification of the ORF encoding the putative *ctpA* gene (Fig. 2), subclones pCTPA and pCTPAΔ were constructed by digestion with restriction endonucleases at the appropriate sites (Fig. 1) and ligation into pUC19 by standard methods. Subclone pCTPAΔ was constructed by linker mutagenesis. Following endonuclease excision of an 18 bp *Eco*RV fragment from pCTPA, a 10 bp *Xho*I linker was inserted into the site by blunt-end ligation (Promega; *Xho*I linker: 5′-CCCTCGAGGG-3′). The resulting truncated ORF, *ctpA5′*, contains a mutated sequence beginning at base 520 of the sequence shown in Fig. 2. A subclone containing the two ORFs downstream of the *ialB* gene was constructed and will be described elsewhere (S. J. Mitchell & M. F. Minnick, unpublished).

Nucleotide sequencing and computer analysis. The *Sac*I insert of pSAC2 was sequenced on both strands by the methods of Sanger *et al.* (1977) using an automated DNA sequencer (Applied Biosystems). In addition to M13 universal forward and reverse primers for pUC19, primers were made on an automated DNA synthesizer (Applied Biosystems). Computer analysis was done using PCGENE 6.8 (Intelligenetics) for ORF identification and analysis; BLAST (Altschul *et al.*, 1990) for database searches; FASTA 2.0 (Pearson, 1990) for dual-sequence alignments; CLUSTAL-W (Thompson *et al.*, 1994) for multiple sequence alignments; and BOXSHADE 3.2 (Hofmann & Baron, 1996) for formatting of aligned sequences.

Nucleic acid hybridization analysis. Total DNA from *E. coli* HB101, and *B. bacilliformis* strains KC583 and KC584, was isolated and digested to completion with *Bam*HI. The plasmid pCTPA was digested with *Sac*I and *Xba*I to release the entire *ctpA*-containing insert. The DNA samples were then separated on an ethidium-bromide-stained 1% (w/v) agarose gel, blotted to nitrocellulose (0·45 μm pore size, Schleicher & Schuell) by the method of Southern (1975) and baked for 1 h at 80 °C. A 1223 bp *Bam*HI–*Eco*RI fragment of pCTPA was purified from an agarose gel and labelled by random-primer extension using [α-^{32}P]dCTP and the Klenow fragment of *E. coli* polymerase I (Feinberg & Vogelstein, 1984). The blot was probed overnight at 50 °C with the ^{32}P-labelled pCTPA fragment and washed at high stringency (approximately 10% DNA mismatch) as previously described (Minnick *et al.*, 1990). The blot was subsequently exposed for 6 h to X-ray film (X-Omat XAR-5, Eastman Kodak) to visualize hybridized DNA fragments.

Table 1. Bacterial strains and plasmids

Strain/plasmid	Description	Source/reference
E. coli		
DH5α	F⁻ ϕ80d*lacZ*ΔM15 *recA1 endA1 gyrA96 thi-1 hsdR17* (r_K^- m_K^+) *supE44 relA1 deoR* Δ(*lacZYA–argF*)*U169*	BRL
HB101	F⁻ *thi-1 hsdS20* (r_B^- m_B^-) *supE44 recA13 ara-14 leuB6 proA2 lacY1 rpsL20* (Str^r) *xyl-5 mtl-1*	Promega/Boyer & Roulland-Dussoix (1969)
Y1090	Δ(*lacU169*) *proA*⁺ Δ(*lon*) *araD139 strA supF* [*trpC22*::Tn*10*(Tet^r)] *hsdR* (r_K^- m_K^+) (pMC9)	Promega/Young & Davis (1983)
Y1090-C	Same as Y1090 except cured of pMC9	This study
Bartonella		
B. bacilliformis KC583	Neotype strain	ATCC/Brenner *et al.* (1991)
B. bacilliformis KC584	Peruvian isolate, 1963	ATCC/Brenner *et al.* (1991)
B. elizabethae	Type strain	ATCC/Daly *et al.* (1993)
B. henselae Houston-1	Type strain	CDC/Regnery *et al.* (1992)
B. quintana Fuller	Type strain	ATCC/Myers *et al.* (1979)
B. vinsonii Baker	Vole agent	ATCC/Weiss & Dasch (1982)
Brucella abortus		
Strain 19	Live attenuated vaccine strain	Burroughs Wellcome
Proteus vulgaris		
LB-53	Laboratory strain	Int. Inst. Biochem. & Biomed. Tech., Chicago, IL, USA
Plasmids		
pUC19	M13mp19 derivative, Amp^r	Yannisch-Perron *et al.* (1985)
pIAL1	pUC19 containing *Bb* KC583 *ialA* and *ialB* on a 1·5 kb *Bam*HI fragment	Mitchell & Minnick (1995)
pSAC2	pUC19 containing *Bb* KC583 *ctpA*, *ialA* and *ialB* on a 4·5 kb *Sac*I fragment	This study
pCTPA	pUC19 containing *Bb* KC583 *ctpA* on a 1·7 kb *Sac*I–(*Hpa*I/*Sma*I fusion) fragment	This study
pCTPAΔ	Same as pCTPA except 10 bp *Xho*I-linker inserted into *Eco*RV site in *ctpA* to create *ctpA5′*	This study

To determine the distribution of the *ctpA*, *ialA* and *ialB* genes within *Bartonella* species, and to investigate their prevalence in other bacteria, a multiple-species hybridization experiment was carried out. Total chromosomal DNA was isolated as described above from five *Bartonella* species (*B. bacilliformis* KC583 and KC584, *B. henselae*, *B. quintana*, *B. elizabethae* and *B. vinsonii*), *Br. abortus* and *E. coli* DH5α. A 3·0 μg sample of each was digested to completion with *Sac*I, as was 50 ng purified pSAC2 as a positive control. The DNA samples were then subjected to electrophoresis on an ethidium-bromide-stained agarose (1%, w/v) gel and vacuum-blotted to a positively charged nylon membrane (ZetaProbe GT; Bio-Rad) under alkaline conditions according to the manufacturer's instructions. Transfer efficiency was verified by post-staining of the agarose gel with ethidium bromide. Probes were constructed as described above, using a 1223 bp *Bam*HI–*Eco*RI fragment of *ctpA*, a 774 bp *Bam*HI–*Swa*I fragment of *ialB*, and a 492 bp *Hpa*I–*Sca*I fragment of *ialA*. Following a 2 h random-primer extension, probes were purified from unincorporated [α-^{32}P]dCTP using a QIAquick Nucleotide Removal Kit (Qiagen) and denatured at 100 °C for 5–10 min prior to addition to fresh hybridization solution. The nylon membrane was pre-hybridized at room temperature for 2 h in 0·25 M Na_2HPO_4, 7% (w/v) SDS, and 5 × Denhardt's solution. Hybridization was carried out overnight at 45 °C with fresh hybridization solution plus the appropriate ^{32}P-labelled probe. The membrane was then washed twice for 15 min in 50 ml 1·0% (w/v) SDS plus sufficient Na_2HPO_4 to give 40%, 30%, 20%, or 10% mismatch according to the DNA:DNA T_m formula of Meinkoth & Wahl (1984) (40% mismatch was obtained with 0·115 M Na_2HPO_4, 45 °C; 30% mismatch with 0·115 M Na_2HPO_4, 55 °C; 20% mismatch with 0·029 M Na_2HPO_4, 55 °C; 10% mismatch with 0·029 M Na_2HPO_4, 65 °C). The membrane was kept damp at all times by wrapping in plastic-wrap, and was exposed to X-Omat XAR-5 film overnight at −70 °C with two intensifying screens. After each 10% mismatch experiment, the probe was stripped from the membrane by washing twice in 500 ml 0·005 M Na_2HPO_4 and 0·5% SDS at 95 °C for 20 min. Removal of the probe was confirmed by a 24 h exposure to X-ray film with two intensifying screens at −70 °C prior to applying the next probe.

Analysis of mRNA by Northern blotting was done by two standard methods (Ausubel *et al.*, 1995): formaldehyde (6·7% w/v)/MOPS gel electrophoresis, or glyoxal/DMSO gel electrophoresis, followed by vacuum-blotting to nylon membranes (ZetaProbe GT; Bio-Rad), and probing of the membranes with an isotopically labelled DNA probe. Briefly, an RNeasy kit (Qiagen) was used to extract total RNA from *B. bacilliformis* strains KC583 and KC584; and from *E. coli* DH5α containing either no plasmid, or pUC19, pSAC2, pCTPA or pCTPAΔ. Ten micrograms of RNA from each sample was denatured and run on 1·2% agarose gels under appropriate conditions for formaldehyde/MOPS or glyoxal/DMSO gels (Ausubel *et al.*, 1995). RNA was transferred by vacuum-blotting to positively charged nylon membranes (ZetaProbe GT) and UV-cross-linked with a total treatment of 150 mJ. A ^{32}P-labelled probe was constructed by random-primer extension on a DNA template constructed from a 1223 bp *Bam*HI–*Eco*RI fragment of pCTPA (Ausubel *et al.*, 1995). The membranes were probed using the manufacturer's recommendations (Bio-Rad) for pre-hybridization, overnight hybridization (45 °C, 50% formamide), and high-stringency washing (final washes of 2 × 20 min at 65 °C in 50 ml 0·1 × SSC, 0·5% SDS). Transfer of RNA to the nylon membranes was verified after probing the membrane by the methylene-blue staining method of Herrin & Schmidt (1988).

Invasion assays. These were done as previously reported (Mitchell & Minnick, 1995). Briefly, overnight cultures of *E. coli* DH5α containing the appropriate recombinant plasmid (pUC19, pSAC2, pCTPA or pCTPAΔ) were standardized to a concentration of 2·5 × 10^{8} cells ml^{-1}. Recently outdated human erythrocytes (American Red Cross) were washed in PBS and added to a final concentration of 5·0 × 10^{7} cells to yield a final multiplicity of infection of 50 bacteria : 1 erythrocyte. Samples (0·5 ml total volume) were incubated for 1 h at 37 °C, centrifuged at 1320 ***g*** for 20 s to associate bacteria and erythrocytes, and incubated for a further 2 h at 37 °C. Samples were then washed, treated with 200 µg gentamicin ml^{-1} for 2 h at 37 °C, and washed again to remove gentamicin. Intracellular bacteria were then released by lysing the erythrocytes with 100 µl deionized water. Samples were then plated onto LB/ampicillin plates for overnight growth at 37 °C and colonies were counted the following morning. All experiments were done in triplicate.

Protein expression *in vitro* and *in vivo*. Expression of protein products *in vitro* was done using two commercially available *E. coli*-derived *in vitro* transcription/translation (IVTT) kits (Amersham and Promega). Briefly, 2 µg purified plasmid DNA was incubated with 10·0 µCi (37 MBq) [^{35}S]Met/Cys (EXPRESS, NEN Dupont) and the IVTT kit for 2 h at 37 °C, followed by immediate addition of 80 µl Laemmli sample buffer (LSB) (Laemmli, 1970), heating to 100 °C for 10 min, and storage at −20 °C. Aliquots of 15 µl were heated to 100 °C for 10 min, quickly centrifuged to collect the supernatant, and subjected to 0·1% SDS-PAGE (12·5%, w/v, acrylamide) analysis. The gel was vacuum-dried and subjected to autoradiography overnight (X-Omat XAR-5).

Because of the autolytic processing observed *in vitro* (see Fig. 5), two separate protease-inhibitor treatments were applied during subsequent IVTT incubations. In the first experiment, a mixture of inhibitors was made with a final concentration of 6 µg aprotinin ml^{-1} (Sigma), 0·5 µg leupeptin ml^{-1} (Sigma), 0·7 µg pepstatin ml^{-1} (Sigma). An IVTT reaction was conducted as described above and divided into two equal aliquots. The protease-inhibitor mix was applied to one aliquot at 0 min and to the other aliquot at 10 min, and the reaction allowed to proceed for a total time of 2 h. The samples were then treated as described above and analysed by SDS-PAGE and autoradiography. The second protease inhibitor mixture was prepared using a Complete Protease Inhibitor Cocktail Tablet (Boehringer Mannheim, no. 1697498). One tablet was dissolved in 2·0 ml deionized water and 1 µl of this solution was used per inhibition reaction, to give a final concentration of 1·0 µg aprotinin ml^{-1}, 0·5 µg leupeptin ml^{-1}, 2·0 µmol Pefablock ml^{-1} (a water-soluble PMSF analogue), and 1·0 µmol EDTA ml^{-1}. The experiment was done in an identical fashion to that described above, with one half of the IVTT sample treated at 0 min and the second at 10 min, followed by a 2 h incubation at 37 °C. The samples were then treated as described above and analysed by SDS-PAGE and autoradiography.

Expression of the *ctpA*-encoded protein *in vivo* was accomplished by the Maxicell procedure of Sancar *et al.* (1979). *E. coli* HB101 containing recombinant plasmids were irradiated with 195 µW cm^{-2} UV light for 50 s at a 15 cm exposure distance. Maxicells containing the ^{35}S-labelled proteins were then pelleted, resuspended in 75 µl LSB, and heated to 100 °C for 10 min. Samples were centrifuged at 16000 ***g*** for 1 min, and 20 µl of each supernatant was then subjected to 0·1% SDS-PAGE (12·5%, w/v, acrylamide) and visualized by overnight autoradiography of the dried gel.

Haemolysis, collagenase and caseinase assays. Haemolysis assays were done by streaking *E. coli* HB101 containing either no plasmid, or pUC19, pSAC2, pCTPA or pCTPAΔ, onto HIBB solid medium and growing overnight at 37 °C. Single plates were scored visually for haemolytic clearing.

Collagenase assays were done by stabbing the above *E. coli* HB101 strains into 5 ml nutrient gelatin tubes in duplicate. A positive control was done with *Proteus vulgaris*. Samples were grown at 37 °C for 7 d and assayed daily by chilling the tubes to 20 °C for 30 min to test for liquefaction or gelling.

Caseinase assays were done using the Lon^{-} *E. coli* strain Y1090 (Promega) after curing the strain of pMC9 (Amp^{r}, Promega). Curing was done by growing the strain at 37 °C in LB for 30 h with periodic replenishment of LB to maintain exponential growth. At the end of this period, colonies were identified which had become Amp^{s} and curing was confirmed by loss of the plasmid. The strain Y1090-C was then transformed with pUC19, pSAC2, pCTPA or pCTPAΔ as described above, and transformants confirmed by plasmid isolation. *E. coli* Y1090-C transformants containing the above plasmids were then plated in duplicate onto skim-milk plates and grown at 37 °C overnight to assay for casein proteolysis.

RESULTS

A 4·3 kb *Sac*I fragment of the *B. bacilliformis* chromosome contains ORFs flanking *ialAB*

In order to determine if the invasion-associated locus (*ialAB*) of *B. bacilliformis* is flanked by other virulence factors in a 'pathogenicity island', we isolated regions upstream and downstream of the original 1469 bp *Bam*HI fragment containing *ialAB*. By probing a bacteriophage λ library of *Sau*3AI-derived *B. bacilliformis* KC583 fragments with a ^{32}P-labelled 1469 bp *Bam*HI fragment containing the *ialAB* locus, approximately 10 kb of the *Bartonella* chromosome was isolated. A 4·3 kb *Sac*I fragment of the clone was found to contain

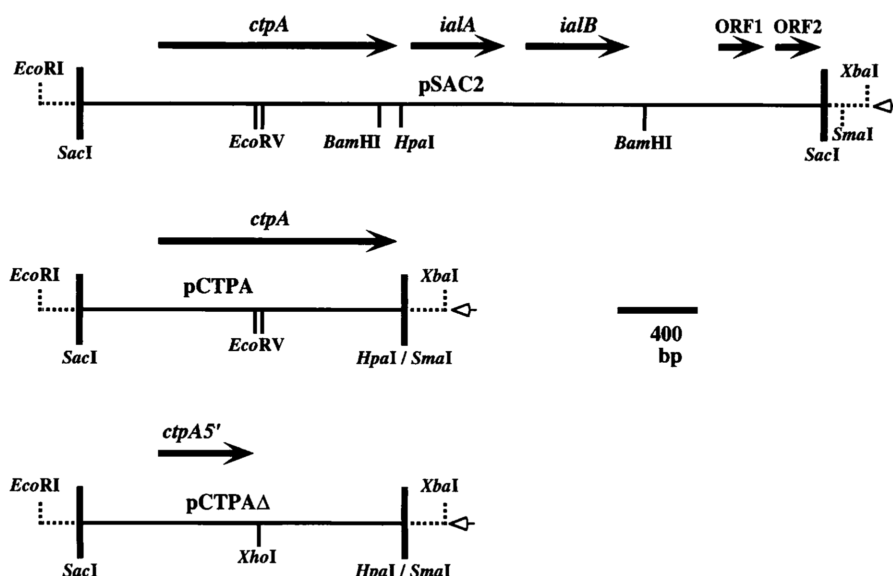

Fig. 1. Partial restriction endonuclease map of pSAC2, pCTPA and pCTPAΔ inserts. Positions of the *ctpA*, *ialA*, and *ialB* genes are indicated by the boldface arrows, as are two novel ORFs (ORF1 and ORF2). The outlined arrowheads indicate the direction of the *lacZ'* promoter in the cloning vector, pUC19. Subclone pCTPA is indicated below the respective restriction sites used in its construction from pSAC2. Dotted lines indicate endonuclease sites within the pUC19 polylinker used to construct subclones. pCTPAΔ was constructed by *Xho*I-linker mutagenesis of pCTPA. The resulting ORF, *ctpA5'*, contains a mutated sequence beginning at base 520 of the sequence shown in Fig. 2, resulting in early translation termination.

the *ialAB* locus by DNA hybridization and was subsequently removed from the LambdaGem-11 arms and ligated into the *Sac*I site of pUC19 to create the pSAC2 recombinant (Fig. 1). Sequencing from primers near the *Bam*HI sites surrounding *ialAB* revealed that approximately 1·8 kb of the total insert was located upstream of *ialA* and 1·0 kb was located downstream of *ialB*. Digestion of the pSAC2 clone with *Hpa*I and *Sma*I followed by ligation of the blunt ends allowed us to clone the DNA fragment upstream of *ialA* in a recombinant plasmid named pCTPA (Fig. 1). Two minor ORFs of unknown significance were found between the *Bam*HI and *Sac*I sites of pSAC2 (Fig. 1). We have continued to sequence downstream of the 3' *Sac*I site of pSAC2 and have recently identified an inorganic pyrophosphatase gene which we will describe elsewhere (S. J. Mitchell & M. F. Minnick, unpublished).

The 5' 1·8 kb *Sac*I–*Hpa*I fragment of pSAC2 contains a homologue of C-terminal processing protease genes, including *ctpA* and *tsp*

The 1761 bp *Sac*I–*Hpa*I insert of pCTPA contains a 1302 bp ORF which we identified as the gene for a putative C-terminal protease by computer database searches and subsequently named *ctpA*. The sequence shown in Fig. 2 shows the close proximity of the *ctpA* gene to the *ialA* gene (separated by only 71 nt) and oriented in the same direction as *ialA*. The region upstream of the putative start codon of *ctpA* (nucleotide 1, Fig. 2) contains hexamers with similarity to the *E. coli* consensus promoters (McLure, 1985). The putative −35 (TTTACA) and −10 (TGAATT) hexamers are followed by a putative ribosome-binding site (AGGA) with perfect identity to the *E. coli* consensus sequence (Gold *et al.*, 1981) located 6 nt upstream of the start codon (Stormo *et al.*, 1982). A possible rho-independent transcription terminator (Rosenberg & Court, 1979) is found 21 nt downstream of the stop codon and is composed of a 14-base inverted repeat separated by 3 nt. The hairpin predicted to form by this repeat has a thermostable free energy of $-16{\cdot}5$ kcal mol^{-1} ($-69{\cdot}0$ kJ mol^{-1}) (PCGENE) and is immediately followed by four thymidines, further supporting the notion that this is the transcription-termination site.

Computer analysis of the ORF indicated that the predicted protein has a length of 434 amino acid residues, with a molecular mass of approximately 47 kDa. A probable secretory signal cleavage site (Von Heijne, 1986) was found between residues 23 and 24, which would result in a secreted protein (411 residues) with a predicted molecular mass of 45 kDa. Computer analysis with PCGENE also indicated that there is an alternative start site at nucleotide 310, which would possibly lead to a non-secreted protein of 331 residues (predicted molecular mass 38 kDa).

Further analysis of the predicted full-length CtpA protein (434 residues) was done by searching various databases with BLAST (Altschul *et al.*, 1990). Several high-probability alignments were found to a group of carboxy-terminal processing proteases, including the 46 kDa CtpA protease of *Synechocystis* 6803 (SYCCTPA; Shestakov *et al.*, 1994) and the 73 kDa Tsp

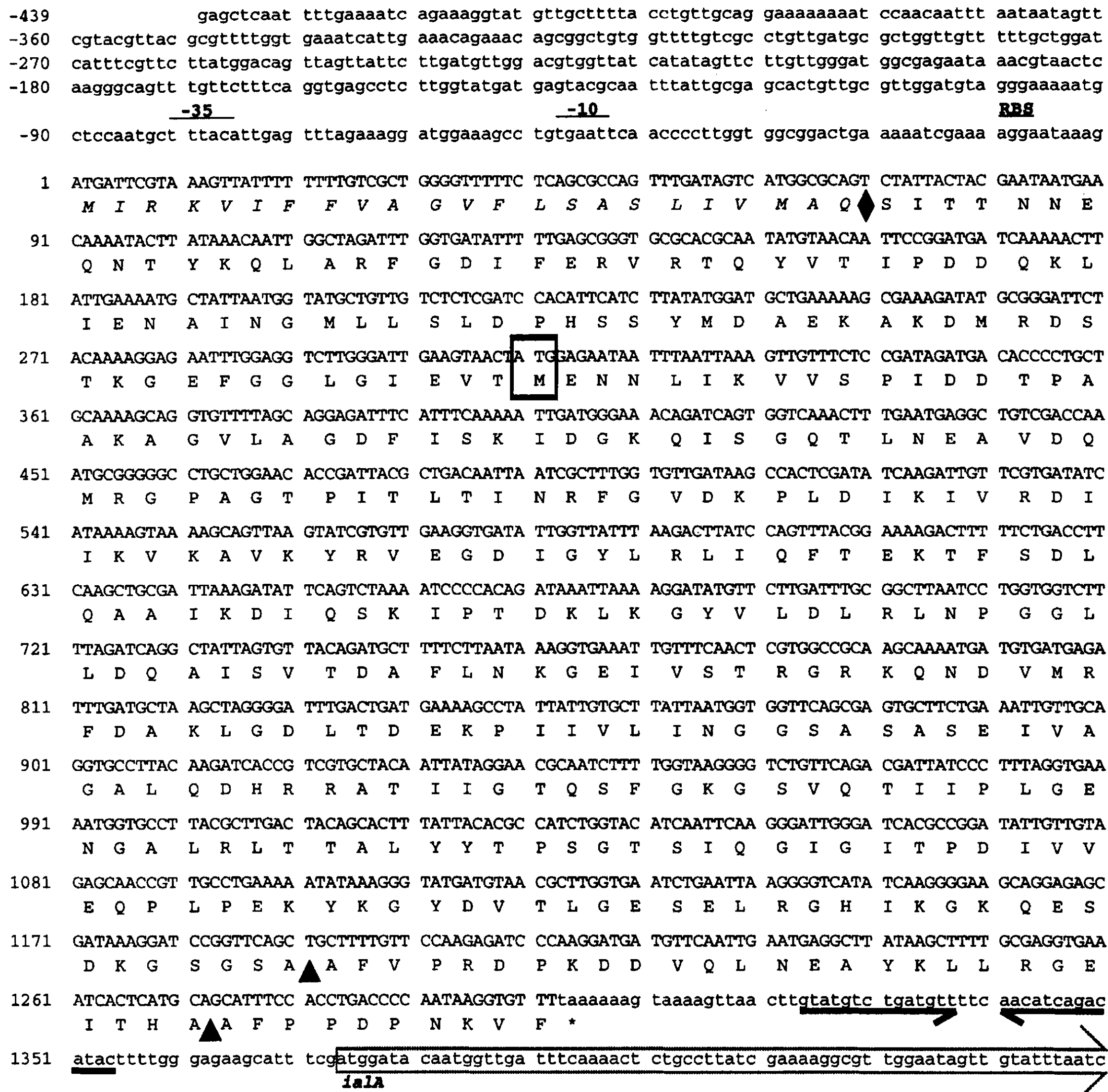

Fig. 2. Nucleotide sequence of the 1761 bp *SacI–HpaI* insert of pCTPA plus 3′ flanking regions containing the 5′ end of *ialA*. Nucleotides within the *ctpA* gene are given in uppercase letters. Putative −35 and −10 promoter hexamers and the potential ribosome-binding site (RBS) are shown above the sequence and an alternative translational start site is indicated by the box. The stop codon (taa) following the ORF is marked with an asterisk, and single-sided arrows underneath the sequence indicate a putative rho-independent transcription termination site. The 5′ end of the *ialA* gene is marked by an open-ended double arrow. The deduced amino acid sequence is shown below the second base of its respective triplet codon. Residues constituting the predicted secretory signal sequence are shown in italics and the predicted cleavage site is denoted by a diamond. Possible C-terminal processing sites are indicated by upward-pointing triangles in the protein sequence.

of *E. coli* (Silber *et al.*, 1992). Alignments of all three predicted protein sequences are shown in Fig. 3, and indicate multiple areas of strong homology. The individual alignment of the *B. bacilliformis* CtpA with SYCCTPA had an identity of 30% (FASTA) and overall similarity of 69% (PCGENE). Alignment of the *B. bacilliformis* CtpA with the *E. coli* Tsp revealed 19% identity (FASTA) and 46% similarity (PCGENE). When the N- and C-terminal overhanging portions of Tsp were excluded from the analysis, alignment of the *B. bacilliformis* CtpA with Tsp increased to 22% identity (FASTA) and 62% overall similarity (PCGENE). The BLAST analysis also revealed moderate homology of the *B. bacilliformis* CtpA to several interphotoreceptor retinol-binding proteins (IRBPs).

DNA hybridization analyses indicate that the *ctpA*, *ialA* and *ialB* genes are highly conserved within *Bartonella* species, but not in a related intracellular pathogen or in *E. coli*

In order to verify that the *ctpA*-containing DNA fragment was of *Bartonella* origin, DNA hybridization analysis was done using *Bam*HI-digested DNA from *B.*

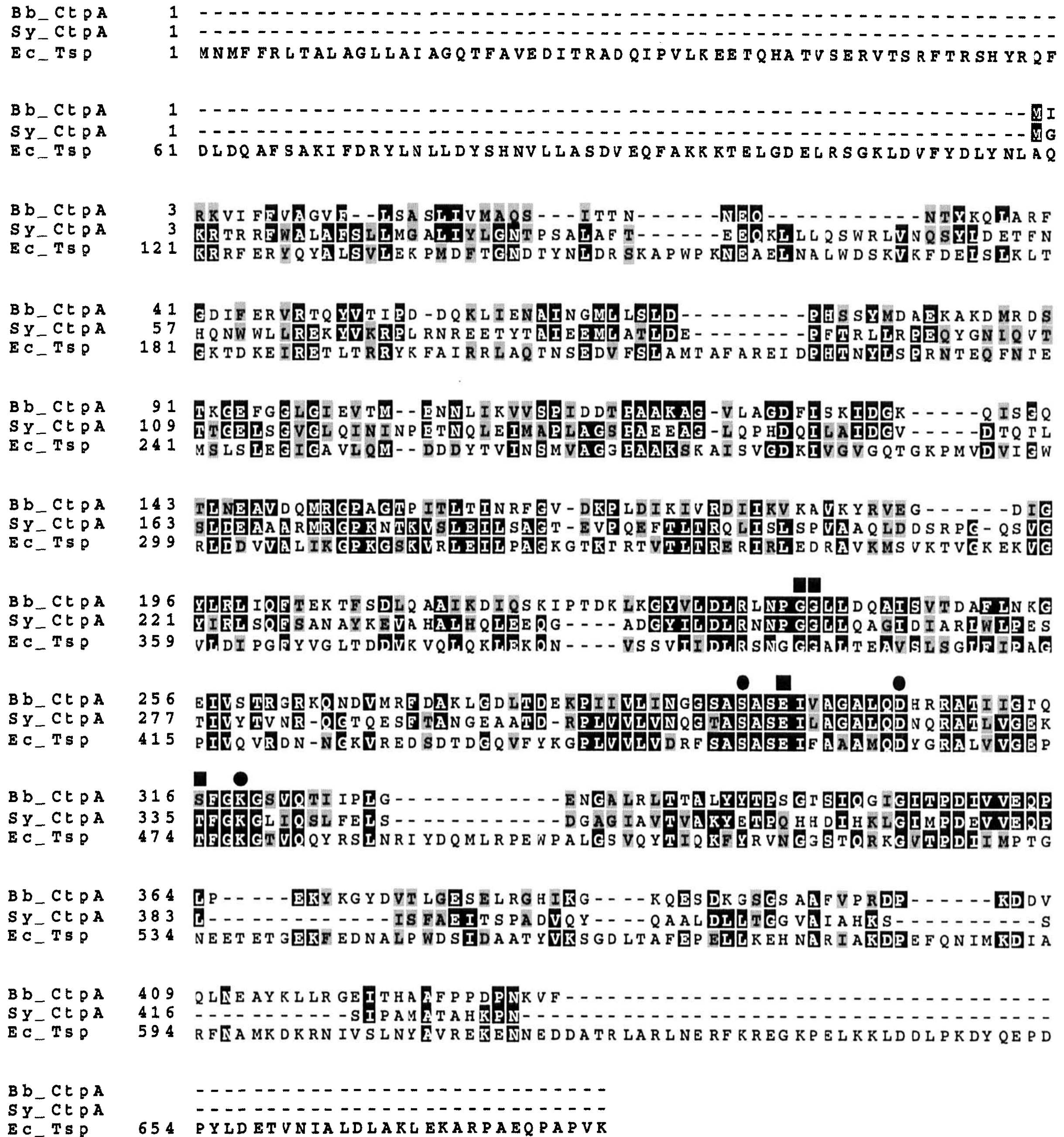

Fig. 3. Multiple alignment of the *B. bacilliformis* CtpA (Bb_CtpA) with the *Synechocystis* sp. 6803 CtpA (Sy_CtpA) and the *E. coli* Tsp (Ec_Tsp) proteins (all are shown in their immature form). Identical amino acid residues are noted in black, conserved residues in grey, and introduced gaps by hyphens. The three catalytic residues and four structurally essential residues of Tsp (Keiler & Sauer, 1995) are indicated above the sequence by black circles and squares, respectively. The GenBank accession numbers for Bb_CtpA, Sy_CtpA, and Ec_Tsp are L37094, A53964, and A41798, respectively.

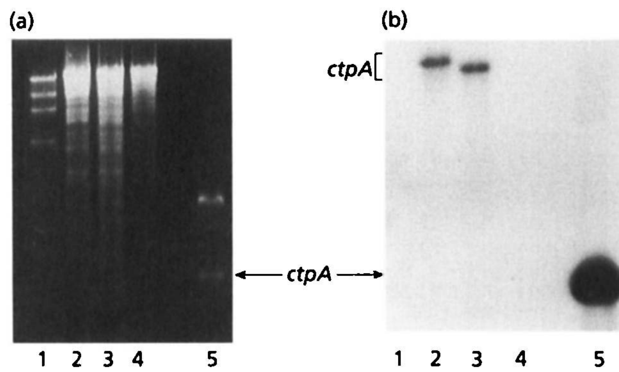

Fig. 4. Detection of the *ctpA* gene in the *B. bacilliformis* chromosome by DNA hybridization. (a) Ethidium-bromide-stained agarose gel (1 %, w/v) containing: 1, *λ*/*Hin*dIII DNA size standards; 2–4, *Bam*HI-digested chromosomal DNA of *B. bacilliformis* KC583 (2), *B. bacilliformis* KC584 (3) and *E. coli* HB101 (4); and 5, *Sac*I-*Xba*I digested pCTPA. (b) The corresponding autoradiograph following DNA hybridization with a ^{32}P-labelled *Bam*HI–*Eco*RI subfragment of the pCTPA DNA insert. Lanes are the same as in (a). The double arrow points to the pCTPA fragment producing the hybridization signal. *ctpA*-containing chromosomal fragments of the two *B. bacilliformis* strains are bracketed.

bacilliformis KC583 and KC584 and from *E. coli* HB101. As shown in Fig. 4(b), probing the blotted DNA with an internal *ctpA* fragment clearly demonstrated that the DNA fragment had a *Bartonella* origin (lanes 2 and 3) and was not found in *E. coli* (lane 4) at approximately 10 % mismatch. This is further supported by the overall 39·5 mol % G + C content of the *ctpA* gene (Fig. 2), a value which is very close to the 39 mol % G + C content of *B. bacilliformis* KC583 (Brenner *et al.*, 1991). A slight difference in sizes of the hybridization signals between the two *B. bacilliformis* strains (lanes 2 and 3) is presumably due to restriction-fragment length polymorphism.

Because little is known about the genus *Bartonella* and we were interested in determining the abundance of the three closely linked genes (*ctpA*, *ialA* and *ialB*) within this group, we then conducted a pan-specific hybridization analysis. DNAs from several species of *Bartonella* – *B. bacilliformis* KC583 and KC584, *B. henselae*, *B. quintana*, *B. elizabethae* and *B. vinsonii* – from *Brucella abortus* (an intracellular pathogen closely related to *Bartonella*), and from *E. coli* DH5α, were included to determine the distribution of these genes. A positive control lane containing the *Sac*I insert of pSAC2 was also included. As summarized in Table 2, probing with fragments of the *ctpA*, *ialB*, and *ialA* genes (in that order) at a calculated DNA:DNA mismatch $\leqslant$ 40 % revealed that the three genes are found in all of the *Bartonella* species examined, but are absent in *Br. abortus* and *E. coli*. In addition, increasing levels of stringency indicated that the *ctpA* gene is more highly conserved within the *Bartonella* species than are the *ialA* or *ialB* genes.

The *ctpA* gene does not affect the erythrocyte-invasion phenotype conferred by *ialAB*

To determine if the *ctpA* gene had a direct effect upon erythrocyte invasion, invasion assays were done using the modified gentamicin-protection method we described previously (Mitchell & Minnick, 1995). Experiments with pCTPA (containing only *ctpA*) indicated that this gene does not confer an invasive phenotype upon *E. coli* (data not shown). An additional experiment with pSAC2 (containing *ctpA*, *ialA*, *ialB*, ORF1 and ORF2) revealed that there was no statistically

Table 2. Distribution of the *ctpA*, *ialA* and *ialB* genes in various *Bartonella* species and other bacteria

Target DNA	Probe and percentage mismatch:											
	ctpA				*ialA*				*ialB*			
	40	30	20	10	40	30	20	10	40	30	20	10
B. bacilliformis KC583	+	+	+	+	+	+	+	ND*	ND	+	+	+†
B. bacilliformis KC584	+	+	+	+	+	+	+	ND*	ND	+	+	+†
B. quintana	+	+	+	±	+	±	−	ND*	ND	+	−	ND
B. henselae	+	+	±	−	+	+	−	ND*	ND	+	−	ND
B. vinsonii	+	+	±	−	+	±	−	ND*	ND	±	−	ND
B. elizabethae	±	−	−	−	±	−	−	ND*	ND	−	−	ND
Br. abortus	−	−	−	−	−	−	−	ND*	ND	−	−	ND
E. coli	−	−	−	−	−	−	−	ND*	ND	−	−	−†
pSAC2	+	+	+	+	+	+	+	ND*	ND	+	+	ND

ND, Not done; ND*, not done due to membrane degradation.

† Done in a previous study (Mitchell & Minnick, 1995).

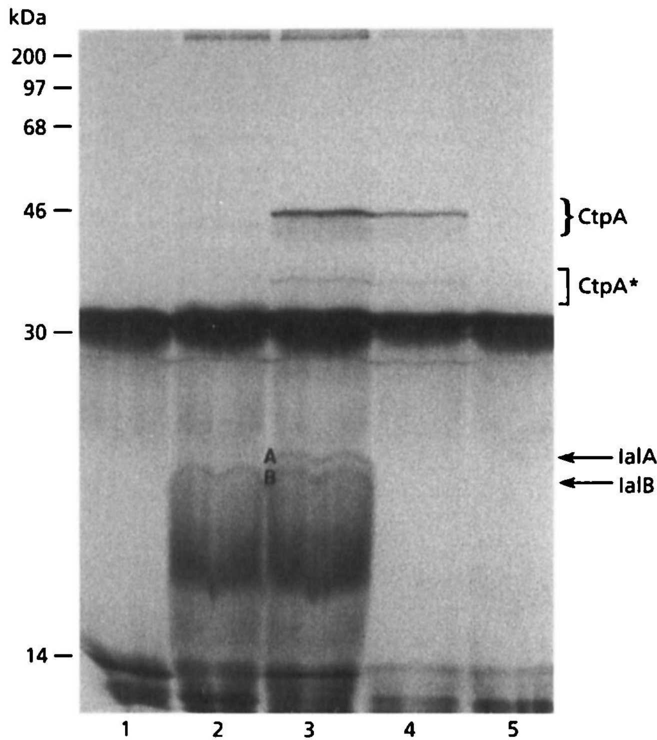

Fig. 5. Expression of the *ctpA* gene *in vitro*. A prokaryote-directed IVTT kit (Promega) was used to produce [^{35}S]Met/Cys-labelled protein products and analysed by SDS-PAGE (12·5%, w/v, acrylamide). Molecular mass values determined from ^{14}C-labelled protein standards are indicated on the left. Lanes and products: 1, pUC19; 2, pIAL1 (encoding IalA and IalB); 3, pSAC2 (encoding CtpA, IalA, and IalB); 4, pCTPA (encoding CtpA); 5, pCTPAΔ (encoding an N-terminal CtpA fragment of 183 residues). The IalA (20 kDa) and IalB (18 kDa) protein products in lanes 2 and 3 are marked A and B, respectively. A group of products encoded by the full-length *ctpA* ORF is indicated by 'CtpA' at 43–46 kDa, and a smaller group of products presumably encoded from the alternative start site by 'CtpA*' at 36–37 kDa.

significant increase in the level of erythrocyte invasion conferred by *ctpA*, as compared to pIAL1 (containing only the *ialA* and *ialB* genes) (data not shown).

The protein product of *ctpA* does not affect IalA or IalB, but may autolytically process itself *in vitro* where it cannot be easily inhibited

Because *ctpA* is so closely linked to *ialA* in the genome, we conducted several experiments to determine if *ctpA* might affect *ialA* and/or *ialB* gene products in a more subtle fashion than was detectable by the invasion assay. As it is possible that the *ctpA* gene product might function incorrectly in the *E. coli* environment, we analysed both the transcription and translational products. Two separate analyses of RNA transcription products from *B. bacilliformis* KC583 and KC584, as well as from *E. coli* DH5α containing pSAC2, pCTPA, pIAL1 or pUC19, were done. In both cases, probing with a single-stranded ^{32}P-labelled DNA probe constructed from an internal *ctpA* template revealed that the transcript exists in both *B. bacilliformis* strains and in *E. coli* containing the *ctpA* gene cloned within pSAC2 or pCTPA (data not shown). No signal was seen on lanes containing control RNA from *E. coli* DH5α, *E. coli* DH5α(pUC19) or *E. coli* DH5α(pIAL1) in either experiment. However, because of anomalous migration of the transcripts, presumably caused by association with the 23S and 16S rRNA, we were unable to size the transcripts accurately. Transfer of total RNA in all lanes was verified by post-staining the membrane with methylene blue after probe hybridization and washing.

In order to analyse the expected protein products from *ctpA*, a prokaryote-directed IVTT kit was used to express the *ctpA* and/or *ialA* and *ialB* genes. As shown in Fig. 5, several protein products were produced by the *ctpA* gene (lanes 3 and 4) that were not found in either a control (pUC19, lane 1), or clones containing other *Bartonella* DNA (pIAL1, lane 2). One group of bands in the 43–46 kDa range appears to correspond to the full-length 1302 bp ORF, whereas a less intense group of bands at approximately 36–37 kDa may represent the protein encoded by the 993 bp ORF beginning with the alternative start site (Fig. 2). Because SYCCTPA has been reported to possibly undergo post-translational processing in addition to removal of its leader peptide (Anbudurai *et al.*, 1994), we were interested to see if the *Bartonella* CtpA might be modifying itself or other CtpA molecules. Therefore, we mutated the *ctpA* gene in pCTPA by enzymically removing an 18 bp *Eco*RV fragment and inserting a 10 bp *Xho*I linker fragment in its place (Fig. 1). This insertion resulted in a nonsense mutation in the *ctpA* gene, thus creating *ctpA5′*. The truncated protein product encoded by the 549 bp *ctpA5′* contains only 173 N-terminal amino acid residues, followed by a 10-residue missense sequence (NH_2-PSRYHKCKCC-COOH) and an early stop codon, and would have a predicted molecular mass of 20 kDa. As shown in Fig. 5, lane 5, it is apparent that the multiple protein products seen in both the 36–37 kDa and 43–46 kDa ranges all disappear when only *ctpA5′* is expressed. Subsequent analysis of the *ctpA5′* product on a longer SDS-PAGE system revealed that it was a single polypeptide migrating at approximately 16–18 kDa (data not shown). The analysis was repeated with a second prokaryote-directed IVTT kit from a different manufacturer with identical results (data not shown). Neither ORF1 nor ORF2 appeared to produce visible protein products (predicted molecular masses of 11 and 12 kDa, respectively) with either IVTT system (Fig. 5, lane 3).

Because SYCCTPA has been reported to possibly undergo removal of not only its N-terminal secretory leader peptide, but also additional residues during maturation (Anbudurai *et al.*, 1994), the *in vitro* results in Fig. 5 (lanes 3 and 4) may indicate a similar C-terminal processing event for CtpA. It is known that SYCCTPA cleaves a known target (protein D1) between two alanine residues located nine residues from the C-terminus. In addition, the *E. coli* Tsp shows strong

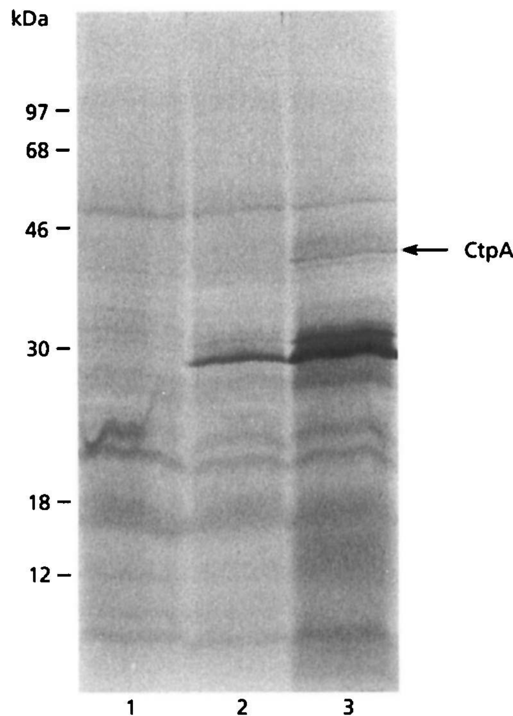

Fig. 6. Expression of the *ctpA* gene product *in vivo* using *E. coli* HB101 maxicells analysed on SDS-PAGE (12·5%, w/v, acrylamide). Molecular mass values determined from ^{14}C-labelled protein standards are indicated on the left. Lanes and products: 1, *E. coli* HB101; 2, *E. coli* HB101 containing pUC19; 3, *E. coli* HB101 containing pCTPA. The arrow indicates the mature 42 kDa CtpA protein.

preference for cleaving target proteins containing alanine near the C-terminus (Keiler & Sauer, 1996). As shown in Fig. 2, the *B. bacilliformis* CtpA contains several alanine pairs near its C-terminus, which could serve as targets for either autolytic processing or proteolysis of other CtpA molecules *in vitro*.

Because the *ctpA* gene is predicted to contain a secretory signal leader peptide (Fig. 2), we performed a maxicell analysis to better define the mature form of CtpA. As shown in Fig. 6, lane 3, the mature CtpA is approximately 41–42 kDa in mass, slightly less than the predicted molecular mass of the mature form of the secreted protein (45 kDa). The CtpA protein appears to be insert-specific, as shown by the lack of a corresponding band in *E. coli* HB101 alone or *E. coli* HB101 containing pUC19 (Fig. 6, lanes 1 and 2, respectively).

The CtpA protein does not appear to be an extracellular haemolysin, collagenase or caseinase

Computer analysis revealed strong homology between the *Bartonella* CtpA, the *Synechocystis* 6803 CtpA, and the *E. coli* Tsp. Therefore, several assays were done to determine if the *Bartonella* CtpA might have similar functions to these proteases, or unique features reflecting the haemotrophic lifestyle of *B. bacilliformis*. A haemolysis assay revealed that the products encoded by *ctpA* (and *ialAB*) did not possess haemolytic activity when expressed in *E. coli*. In addition, a collagenase assay indicated that the gene products did not act to break down gelatin in the growth medium. A proteolysis assay was done by introducing the recombinant plasmids into the Lon$^-$ *E. coli* strain Y1090-C. Overnight growth at 37 °C on skim-milk plates did not reveal any clearing around colonies, indicating that CtpA is not secreted extracellularly.

DISCUSSION

We have identified a putative carboxy-terminal processing protease gene, *ctpA*, immediately upstream of a two-gene locus previously implicated in erythrocyte invasion by *B. bacilliformis* (Mitchell & Minnick, 1995). Our original goal was to explore the sequences flanking *ialAB* to determine if the locus was part of a larger pathogenicity island. By exploring the distribution of *ctpA*, *ialA* and *ialB*, we found that these genes are conserved throughout the genus *Bartonella*, but do not appear to be highly conserved across genera. However, computer analysis of the predicted CtpA protein product indicates that it belongs to a novel class of C-terminal proteases, enzymes which all appear to be involved in stress response. Therefore, despite the lack of a direct connection to the erythrocyte-invasion virulence mechanism, this gene may be an essential component for the bacteria to survive the stressful lifestyle imposed upon an intracellular pathogen.

We have determined that *ctpA* does not directly impact the *ialAB* locus. Invasion, haemolysin, collagenase and caseinase assays all demonstrated that CtpA is probably not active extracellularly. Despite the close physical proximity of the genes, CtpA does not appear to modify either IalA or IalB directly, as demonstrated by the constant molecular mass of these proteins in the presence or absence of CtpA (Figs 5 and 6). This is less surprising in light of the recent description of the *Synechocystis* sp. 6803 genome (Churin *et al.*, 1995), in which mapping of *ctpA* revealed that it was over 300 kbp away from its nearest known target-encoding gene (*psbA*-2). However, one caveat is that *ialA* expression appears to be strikingly enhanced in the presence of *ctpA*, as shown in Fig. 5, lane 3 (*ialA* with *ctpA* upstream), versus Fig. 5, lane 2 (*ialA* without *ctpA*). Although there is a potential rho-independent terminator between the genes (Fig. 2), it is possible that this hairpin may function as a transcriptional pause site (Landick *et al.*, 1996). Alternatively, there may simply be a basal level of transcription read-through in the IVTT system. If there is transcriptional pausing between the genes, it may imply the existence of an operon.

Based upon computer analysis of the predicted CtpA protein, it seems clear that this enzyme probably acts by binding substrates with a hydrophobic C-terminus and enzymically cleaving them. As seen in Fig. 3, the *Bartonella* CtpA contains several highly conserved motifs found in a group of enzymes which all bind

hydrophobic ligands (IRBPs, Tsp, and the photosystem II Ctpases). Recent work by Keiler & Sauer (1995) identified the active-site residues of mature *E. coli* Tsp protease as Ser-430, Asp-441 and Lys-455. All three amino acids lie within a region of IRBP homology (Silber *et al.*, 1992). As seen in Fig. 3, each active-site residue of Tsp has an identical counterpart in the *Bartonella* CtpA sequence. In addition, three of the four residues identified as playing a structural role essential for Tsp activity are identical in *Bartonella* CtpA, while the fourth residue is conservatively substituted. These data suggest that the *Bartonella* CtpA contains a similar active site, and that the enzyme may function by the serine-lysine dyad mechanism proposed for Tsp (Keiler & Sauer, 1995).

In addition to the sequence-predicted binding of hydrophobic target proteins, the *Bartonella* CtpA shares other unusual characteristics with Tsp and SYCCTPA. It has been shown that the common protease inhibitors PMSF, iodoacetamide, EDTA and pepstatin have little effect upon Tsp (Miller, 1996) or SYCCTPA (Bowyer *et al.*, 1992), as we observed in attempting to inhibit the autolytic process of CtpA *in vitro*. In combination with the lack of homology between the C-terminal proteases and other known proteases, this insensitivity to inhibitors has led at least one group to postulate that these enzymes possess an unusual catalytic site (Taguchi *et al.*, 1995). It is also worth noting that both SYCCTPA and Tsp are believed to be stress-response proteases. It is clear that the extremely rapid degradation of D1 by light energy must be balanced by an equally rapid proteolytic maturation of nascent D1 by SYCCTPA (Barber & Andersson, 1992). Similarly, *E. coli* mutants lacking a functional Tsp are sensitive to both heat shock and osmotic stress (Hara *et al.*, 1991). A Tsp homologue from *Salmonella typhimurium* has been implicated in intracellular survival of the bacteria within the acidic environment of macrophages (Bäumler *et al.*, 1994). Very recently, Tsp has also been shown to be responsible for degrading aberrant proteins synthesized from damaged mRNA in *E. coli* (Keiler *et al.*, 1996). The proteins are modified by addition of a short hydrophobic C-terminal marker (AANDENYALAA) from the structural RNA molecule *ssrA*, after which they are specifically recognized by Tsp and degraded. In all of these cases, it is apparent that this class of C-terminal proteases acts to remove or replace 'stressed' protein(s). Although the target of the *Bartonella* CtpA has not yet been identified, the pathogen's passage from insect vector to human circulation and subsequently to host cells must subject it to a variety of stressors.

Because the target of the other C-terminal proteases was identified prior to the actual protease, generalized techniques for finding Ctpase targets have not been developed. A further complication is that no method of directly manipulating *Bartonella* genomes exists, so all work must necessarily be done in *E. coli*. Therefore, at this time we can only speculate upon possible targets. However, the autolytic *in vitro* behaviour bears further investigation. Anbudurai *et al.* (1994) proposed that SYCCTPA may be further modified after loss of the secretory-signal leader peptide, based upon the difference between the predicted molecular mass of the mature protein (43 kDa) and the observed mass (39 kDa). We obtained similar results with CtpA, wherein the predicted molecular mass of the mature protein (45 kDa) was considerably larger than that observed *in vivo* by SDS-PAGE (42 kDa; Fig. 6, lane 3). Although this may be due to aberrant migration in the SDS-PAGE system, the multiple *in vitro* protein products (Fig. 5, lanes 3 and 4) support the notion that the C-terminus may be modified. It is apparent that these products arise only from an ORF encoding the full C-terminal end, as an early truncation mutant, *ctpA5'*, did not produce any of these proteins. In addition to the full-length, presumably periplasmically directed, CtpA the alternative start site may create a subpopulation of cytoplasmically bound proteases. In support of this hypothesis, Tsp activity has been noted both periplasmically and cytoplasmically, although there is some disagreement as to whether the cytoplasmic activity is artifactual (Miller, 1996). Although the products of both the full-length and alternative start site were clearly observed *in vitro* (Fig. 5, lanes 3 and 4), only a single product was found *in vivo* (Fig. 6, lane 3). It is possible that the alternative start site is insignificant within the cellular milieu. It is not clear whether CtpA is truly autocatalytic during folding and maturation, or whether it is the target of its fraternal neighbours *in vitro*, or both.

Sequence analysis also supports a possible autocatalytic activity of CtpA (see Fig. 2, putative C-terminal cleavage sites). The site of D1 cleavage by SYCCTPA has been the subject of intense investigation, beginning with identification of a scissile bond between two alanine residues 9 amino acids from the C-terminus of spinach (Takahashi *et al.*, 1988). In the green alga *Scenedesmus obliqus*, the mature form of CtpA appears to cleave at a conserved motif consisting of a small nonpolar residue, a larger nonpolar residue, a negatively charged residue, a nonpolar residue, and a nearby proline, respectively (Bowyer *et al.*, 1992). We note a slightly different, but conserved, domain on the carboxy side of both putative scissile bonds in Fig. 2: A-AFVPRD, and A-AFPPD. In both cases, cleavage may occur between the two alanine residues, and the C-terminal phenylalanine, proline and aspartic acid may be responsible for correct site geometry. The CtpA from *Synechocystis* sp. 6803 cleaves between an Ala-Ser bond 16 residues from the carboxy terminus (Nixon *et al.*, 1992), and Ala-Ala mutants were cleaved equally well. Nixon *et al.* (1992) demonstrated that SYCCTPA can also cleave the D1 precursor protein from other organisms (including higher plants), indicating that there is probably a conserved structural motif in the target that is necessary for scission, rather than an exact primary sequence. The substrate specificity of Tsp appears to be based upon two critical factors: a free carboxyl group at the target protein's C-terminus, and one to three small hydrophobic residues ($A > S > V$ preferred) at the C-terminus (Keiler *et al.*,

1995). CtpA contains non-polar C-terminal residues, but the bulky phenylalanine group at the CtpA terminus may indicate a recognition scheme different from that of Tsp. CtpA and SYCCTPA may function in a similar, but not identical, fashion to Tsp, as the CtpA enzymes recognize residues immediately surrounding the scissile bond, whereas Tsp appears to recognize the C-terminal residues.

Due to its varied lifestyle, *B. bacilliformis* is subjected to many environmental stresses. The *ctpA* gene we have identified in this study is probably involved in maintenance of a highly stressed or rapidly degraded protein. The exact target(s) of the enzyme remains to be elucidated, as does the possible interaction of *ctpA* with *ialA* transcription. Although it does not appear to be directly involved in the erythrocyte-invasion virulence locus *ialAB* of *B. bacilliformis*, *ctpA* may well prove to have a vital function for bacterial survival within the host environment. Studies on the possible role of the protease in intracellular survival are presently under way.

ACKNOWLEDGEMENTS

We thank Joan Strange (The University of Montana Murdoch Molecular Biology Facility) for excellent technical assistance with nucleic acid sequencing, Kay Crull (American Red Cross) for providing us with outdated human erythrocytes, and Dr Russell Regnery (CDC-P) for the kind gift of *B. henselae*. This work was supported by Public Health Service grants AI34050 and RR10169 from the National Institutes of Health (NIAID).

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.

Anbudurai, P. R., Tsafrir, S. M., Ohad, I., Shestakov, S. V. & Pakrasi, H. B. (1994). The *ctpA* gene encodes the C-terminal processing protease for the D1 protein of the photosystem II reaction center complex. *Proc Natl Acad Sci USA* 91, 8082–8086.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Strahl, K. (1995). *Current Protocols in Molecular Biology*. New York: John Wiley.

Barber, J. & Andersson, B. (1992). Too much of a good thing: light can be bad for photosynthesis. *Trends Biochem Sci* 17, 61–66.

Bäumler, A. J., Kuster, J. G., Stojiljkovic, I. & Heffron, F. (1994). *Salmonella typhimurium* loci involved in survival within macrophages. *Infect Immun* 62, 1623–1630.

Bowyer, J. R., Packer, J. C. L, McCormack, B. A., Whitelegge, J. P., Robinson, C. & Taylor, M. A. (1992). Carboxyl-terminal processing of the D1 protein and photoactivation of water-splitting in Photosystem II. *J Biol Chem* 267, 5424–5433.

Boyer, H. W. & Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J Mol Biol* 41, 459–472.

Brenner, D. J., O'Connor, S. P., Hollis, D. G., Weaver, R. E. & Steigerwalt, A. G. (1991). Molecular characterization and proposal of a neotype strain for *Bartonella bacilliformis*. *J Clin Microbiol* 29, 1299–1302.

Chung, C. T., Niemela, S. L. & Miller, R. H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci USA* 86, 2172–2175.

Churin, Y. N., Shalak, I. N., Börner, T. & Shestakov, S. V. (1995). Physical and genetic map of the chromosome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *J Bacteriol* 177, 3337–3343.

Daly, J. S., Worthington, M. G., Brenner, D. J., Moss, C. W., Hollis, D. G., Weyant, R. S., Steigerwalt, A. G., Weaver, R. E., Daneshvar, M. I. & O'Connor, S. P. (1993). *Rochalimaea elizabethae* sp. nov. isolated from a patient with endocarditis. *J Clin Microbiol* 31, 872–881.

Falkow, S. (1996). The evolution of pathogenicity in *Escherichia, Shigella*, and *Salmonella*. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn., pp. 2723–2729. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.

Feinberg, A. P. & Vogelstein, B. (1984). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 137, 266–267.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. & Stormo, G. (1981). Translation initiation in prokaryotes. *Annu Rev Microbiol* 35, 365–407.

Hara, H., Yamamoto, Y., Higashitani, A., Suzuki, H. & Nishimura, Y. (1991). Cloning, mapping, and characterization of the *Escherichia coli prc* gene, which is involved in the C-terminal processing of penicillin-binding protein 3. *J Bacteriol* 173, 4799–4813.

Herrin, D. L. & Schmidt, G. W. (1988). Rapid, reversible staining of Northern blots prior to hybridization. *Biotechniques* 6, 196–200.

Hofmann, K. & Baron, M. D. (1996). BOXSHADE: 3.2. Lausanne, Switzerland: http://ulrec3.unil.ch/software/BOX_form.html.

Keiler, K. C. & Sauer, R. T. (1995). Identification of active site residues of the Tsp protease. *J Biol Chem* 270, 28864–28868.

Keiler, K. C. & Sauer, R. T. (1996). Sequence determinants of C-terminal substrate recognition by the Tsp protease. *J Biol Chem* 271, 2589–2593.

Keiler, K. C., Silber, K. R., Downard, K. M., Papayannopoulos, I. A., Biemann, A. & Sauer, R. T. (1995). C-terminal specific protein degradation: activity and substrate specificity of the Tsp protease. *Protein Sci* 4, 1507–1515.

Keiler, K. C., Waller, P. R. H. & Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271, 990–993.

Kordick, D. L. & Breitschwerdt, E. B. (1995). Intraerythrocytic presence of *Bartonella henselae*. *J Clin Microbiol* 33, 1655–1656.

Laemmli, U. K. (1970). Cleavage of stuctural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.

Landick, R., Turnbough, C. L., Jr & Yanofsky, C. (1996). Transcription attenuation. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn., pp. 1263–1286. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.

McLure, W. R. (1985). Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem* 54, 171–204.

Meinkoth, J. & Wahl, G. (1984). Hybridization of nucleic acids immobilized on solid supports. *Anal Biochem* 138, 267–284.

Miller, C. G. (1996). Protein degradation and proteolytic modification. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn, pp. 938–954. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.

Minnick, M. F., Heinzen, R. A., Frazier, M. E. & Mallavia, L. P. (1990). Characterization and expression of the *cbbE′* gene of *Coxiella burnetii*. *J Gen Microbiol* **136**, 1099–1107.

Mitchell, S. J. & Minnick, M. F. (1995). Characterization of a two-gene locus from *Bartonella bacilliformis* associated with the ability to invade human erythrocytes. *Infect Immun* **63**, 1552–1562.

Myers, W. F., Wisseman, C. L., Jr, Fiset, P., Oaks, E. V. & Smith, J. F. (1979). Taxonomic relationship of vole agent to *Rochalimaea quintana*. *Infect Immun* **26**, 976–983.

Nixon, P. J., Trost, J. T. & Diner, B. A. (1992). Role of the carboxy terminus of polypeptide D1 in the assembly of a functional water-oxidizing manganese cluster in photosystem II of the cyanobacterium *Synechocystis* sp. PCC 6803: assembly requires a free carboxyl group at C-terminal position 344. *Biochemistry* **31**, 10859–10871.

Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* **183**, 63–98.

Regnery, R. & Tappero, J. (1995). Unraveling mysteries associated with cat-scratch disease, bacillary angiomatosis, and related syndromes. *Emerg Infect Dis* **1**, 16–20.

Regnery, R. L., Anderson, B. E., Clarridge, J. E., III, Rodriguez-Barradas, M. C., Jones, D. C. & Carr, J. H. (1992). Characterization of a novel *Rochalimaea* species, *R. henselae* sp. nov., isolated from blood of a febrile, human immunodeficiency virus-positive patient. *J Clin Microbiol*, **30**, 265–274.

Reynafarje, C. & Ramos, J. (1961). The hemolytic anemia of human Bartonellosis. *Blood*, **17**, 562–578.

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

Sancar, A., Hack, A. M. & Rupp, W. D. (1979). Simple method for identification of plasmid-coded proteins. *J Bacteriol* **137**, 692.

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

Shestakov, S. V., Anbudurai, P. R., Stanbekovas, G. E., Gadzhiev, A., Lind, L. K. & Pakrasi, H. B. (1994). Molecular cloning and characterization of the *ctpA* gene encoding a carboxyl-terminal processing protease. *J Biol Chem* **269**, 19354–19359.

Silber, K. R., Keiler, K. C. & Sauer, R. T. (1992). Tsp: a tail-specific protease that selectively degrades proteins with nonpolar C termini. *Proc Natl Acad Sci USA* **89**, 295–299.

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

Stormo, D. G., Schneider, T. D., Gold, L. & Ehrenfeucht, A. (1982). Use of the ‘perceptron’ algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res* **10**, 2997–3011.

Taguchi, F., Tamamoto, Y. & Satoh, K. (1995). Recognition of the structure around the site of cleavage by the carboxyl-terminal processing protease for the D1 precursor protein of the photosystem II reaction center. *J Biol Chem* **270**, 10711–10716.

Takahashi, M., Shiraishi, T. & Asada, K. (1988). COOH-terminal residues of D1 and the 44 kDa CPa-2 at spinach photosystem II core complex. *FEBS Lett* **240**, 6–8.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Vinson, J. W. & Fuller, H. S. (1961). Studies on trench fever. I. Propagation of rickettsia-like microorganisms from a patient's blood. *Pathol Microbiol*, 152–166S.

Von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res* **14**, 4683–4690.

Weiss, E. & Dasch, G. A. (1982). Differential characteristics of strains of *Rochalimaea*: *Rochalimaea vinsonii* sp. nov., the Canadian Vole Agent. *Int J Syst Bacteriol* **32**, 305–314.

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

Young, R. A. & Davis, R. W. (1983). Yeast RNA polymerase II genes: isolation with antibody probes. *Science* **222**, 778–782.

Received 22 July 1996; revised 2 November 1996; accepted 12 November 1996.