CLONING AND CHARACTERIZATION OF THE GENE ENCODING THE PEPF ENDOPEPTIDASE FROM THE AQUATIC BACTERIUM *CAULOBACTER CRESCENTUS*

Vânia S. Braz; Elza A.S. Lang; Marilis V. Marques*

Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, Brasil

ABSTRACT

The metallopeptidases have a very important role in bacteria, being involved in several processes that rely on protein turnover, such as nutrition, degradation of signal peptides, protein localization and virulence. We have cloned and characterized the gene of the metalloendopeptidase PepF from the aquatic bacterium *Caulobacter crescentus*. The gene upstream of *pepF* (*orf1*) encodes a conserved hypothetical protein found in *Mycobacterium* and *Streptomyces*. *pepF* is co-transcribed with the gene downstream (*orf3*), which encodes a protein that belongs to the ABC1 protein kinase family, suggesting that these two proteins may share a common function in the cell. The *C. crescentus* PepF protein possesses the conserved HEXGH motif present in zinc binding domains of PepF homologs. Disruption of the *pepF* gene by insertion of a vector sequence did not produced any growth defect, but the mutant strain possesses only 30% of the specific activity of endopeptidases present in the wild type strain. Deletions and point mutations in the regulatory region showed that there are two putative promoter regions, and the operon expression is independent of the transcription regulator CtrA. The results indicate that PepF is not essential for either growth or development of this bacterium using peptides as the sole carbon source, suggesting that other peptidases can be sharing this function.

Key words: endopeptidase, M3 family, gene regulation, bacteria

INTRODUCTION

Bacterial peptidases are involved in several important processes in bacterial cells, such as protein localization, hydrolysis of imported peptides, protein turnover and virulence. The metallopeptidases comprise a large group of proteins and are categorized in several families, according to their sequence similarity and function (22). The zinc-dependent metallopeptidases are found in bacteria, protozoa, fungi, plants and animals, and share a common amino acid motif that comprises the zinc-binding site (11). The PepF peptidases are included into family M3, and this class of peptidases is found in all organisms except for viruses and the archaebacteria.

PepF homologs have been found in both Gram negative and Gram positive bacteria, but the role of this protein in the cell was determined only in *Lactococcus lactis*, where two *pepF* genes were found, one encoded in a plasmid and one in the chromosome (16,17). In *L. lactis*, PepF is involved in the ability of the bacterium to utilize milk casein as a carbon source, where it is responsible for degradation of peptides from 7 to 17 amino acids in size (16). The loss of the plasmid copy of *pepF* leads to a 75% reduction in total endopeptidase activity, indicating that this oligopeptidase is responsible for most of this activity in the cell (16).

The aquatic bacterium *Caulobacter crescentus* presents an obligatory differentiation step in each cell cycle, generating by cell division a motile swarmer cell and a sessile stalked cell (3,9). Its unique developmental program has been extensively studied, several key regulatory genes have been identified, and proteolysis plays an important role in regulating *C. crescentus*

development (10,25). Despite being a heterotroph, *Caulobacter* has a preference for growing in very dilute environments, which implies having an efficient way of taking organic matter, and therefore oligopeptidases must have an important role in this bacterium. The recently completed genome sequencing of *C. crescentus* (19) showed 48 genes encoding putative peptidases that belong to different families, but only one was predicted to be from the M3 family, a PepF homolog. In this report, we describe the genetic organization of the locus encoding the only *pepF* homolog found in the genome from *C. crescentus* and characterize its regulatory regions.

MATERIALS AND METHODS

Bacterial strains and media

Caulobacter NA1000 is the wild type, synchronizable strain (7) and the LS2195 strain has a temperature-sensitive allele of the *ctrA* gene (21). *E. coli* S17-1 contains the mobilization genes that allow DNA transfer by conjugation to *C. crescentus* (23). *Caulobacter* strains were grown at 30ºC in PYE complex media or M2 minimal media (6), as indicated.

Cloning and characterization of the *pepF* locus

The first half of the *pepF* gene was identified in a clone previously isolated in our laboratory (14). To obtain the complete region containing *pepF* gene, two partial libraries were constructed in vector pUCBM20 (Boehringer Mannheim) by digestion of the original cosmid clone, one library containing *Hin*dIII/*Pst*I fragments and the other containing *Sac*I/*Bam*HI fragments. The first library was screened with radioactive probes corresponding the beginning of *pepF*. A positive *Hin*dIII/*Pst*I clone of 2.4 kb was obtained and its DNA sequence was determined after subcloning into M13mp18 (Gibco BRL). The nucleotide sequence of this clone showed it contained the complete *pepF* coding region, encompassing 1.6 kb, and part of a third ORF (ORF3). A *Sac*I/*Bam*HI clone of 2.3 kb was isolated, by hybridization with a probe corresponding to the beginning of ORF3. This clone was completely sequenced, showing that ORF3 is 1.2 kb long and no other ORF could be identified immediately downstream of it, using the GeneMark algorithm (1) (Fig. 1).

RT-PCR

In order to determine whether the three genes are organized in an operon, RT-PCR reactions using 1 µg of total RNA and primers flanking each intergenic region were performed with the SuperScript One-Step RT-PCR kit (Gibco BRL). RT-PCR between *orf1* and *pepF* was performed using the pair of primers RT-1/RT-2 (RT-1: 5'-GCGATGGCCTGGAGCTATC-3', RT-2: 5'-ACTAGCAGATAGGCCTCCG-3'), and between *pepF* and *orf3* was performed using the pair RT-3/RT-4 (RT-3: 5'-GCCGCTTTATGAGCAACTGC-3', RT-4: 5'-GGTTGGCACCG

TAAGACAC-3'). As a negative control for each RT-PCR, the reverse transcriptase was ommited from the reactions.

Mapping of regulatory regions

In order to define more precisely the regulatory sequences involved in *pep* expression, different promoter constructions were cloned in front of a reporter *lacZ* gene (plasmid pRK*lacZ*290, (8)), and the levels of expression were determined by measuring b-galactosidase activity (15). Site-directed mutagenesis was performed using primer tac3 (5'-GCATAAGACCCTGTTA AGCCTAAG-3') or tac7 (5'-TCCGAAAGCCGGTGTTTA CCCGCGACCTG-3') as described (12). pVB04 was obtained by cloning the 0.37 kb *Hin*dIII/*Stu*I fragment containing part of the *orf1* gene into vector pRK*lacZ*290. pVB01 was obtained by PCR using primers tac5 (5'-TCGCGCTCGAGCACCGCTTGA-3') and V2 (5'-CAGCACGCCGGGATCCAGCACGCGCGC-3').

Construction and analysis of a *pepF/orf3* mutant

To obtain a null mutant of the *pepF* gene, a 0.65 kb *Sac*I/*Stu*I fragment corresponding to the central region of the gene was cloned into plasmid pNPTS138. The recombinant plasmid was introduced by electroporation into *E. coli* S17-1 and the cells were plated in LB/Km plates. The plasmid was transferred to *Caulobacter* strain NA1000 by conjugation (6). To confirm the correct integration of pNPTS138 in the *pepF* locus, genomic DNA from the transformed clones was purified (5), digested with *Sac*I and transferred to a nylon membrane. The correct integration of the plasmid was confirmed by hybridization with radioactive probes corresponding to either the pNPTS138 plasmid or to the 0.65 kb *Sac*I/*Stu*I fragment. The mutant strain was designated SP2008.

Determination of endopeptidase activity

Endopeptidase activity was assayed in *C. crescentus* strains NA1000 and SP2008 grown to midlog phase. Cells were separated from the culture medium by centrifugation for 15 minutes at 7500 x g, and ressuspended in buffer A (50 mM Tris-HCl pH 7.0/ 20 mM NaCl). Cells were broken by sonication (four pulses of 15 seconds in a Branson Sonifier 250), and cell debris was removed by centrifugation for 5 minutes at 12,000 x g. The enzymatic assay of either cell extracts or culture supernatants was carried out as described (20) in a final volume of 1 ml of buffer A, containing or not 0.5 mM DTT, and the quenched fluorogenic peptide substrate QF7 (orto-aminobenzoyl (Abz)-Gly-Gly-Phe-Leu-Arg-Arg-Val-(2,4,dinitrophenyl) ethylenediamine (EDDnp). Protein concentration was determined by the Bradford assay (BioRad).

RESULTS AND DISCUSSION

Sequencing of the *pepF* gene region showed that it is preceded by a small ORF (ORF1) and that downstream there is

another ORF 1.2 kb long (ORF3) which is separated of *pepF* by only one base pair (Fig. 1). Our results were later confirmed by the complete genome sequencing, where the genes correspond to numbers CC3314 (*orf1*), CC3313 (*pepF*) and CC3312 (*orf3*) in the TIGR database (see http://www.tigr.org).

In order to determine whether the three genes are organized in an operon, RT-PCR reactions using primers flanking each intergenic region were performed. As shown in Figure 2, an amplification product is obtained between *pepF* and ORF3, only when the reverse transcriptase is present, while the negative control without this enzyme did not show any amplified band. This result shows that these two genes are transcribed in a single mRNA species. Confirming this result, no promoter activity was observed when the 1.2 kb *Xho*I/*Stu*I restriction fragment between *pepF* and ORF3 was cloned in front of the reporter gene *lacZ*, indicating that there is no promoter in this
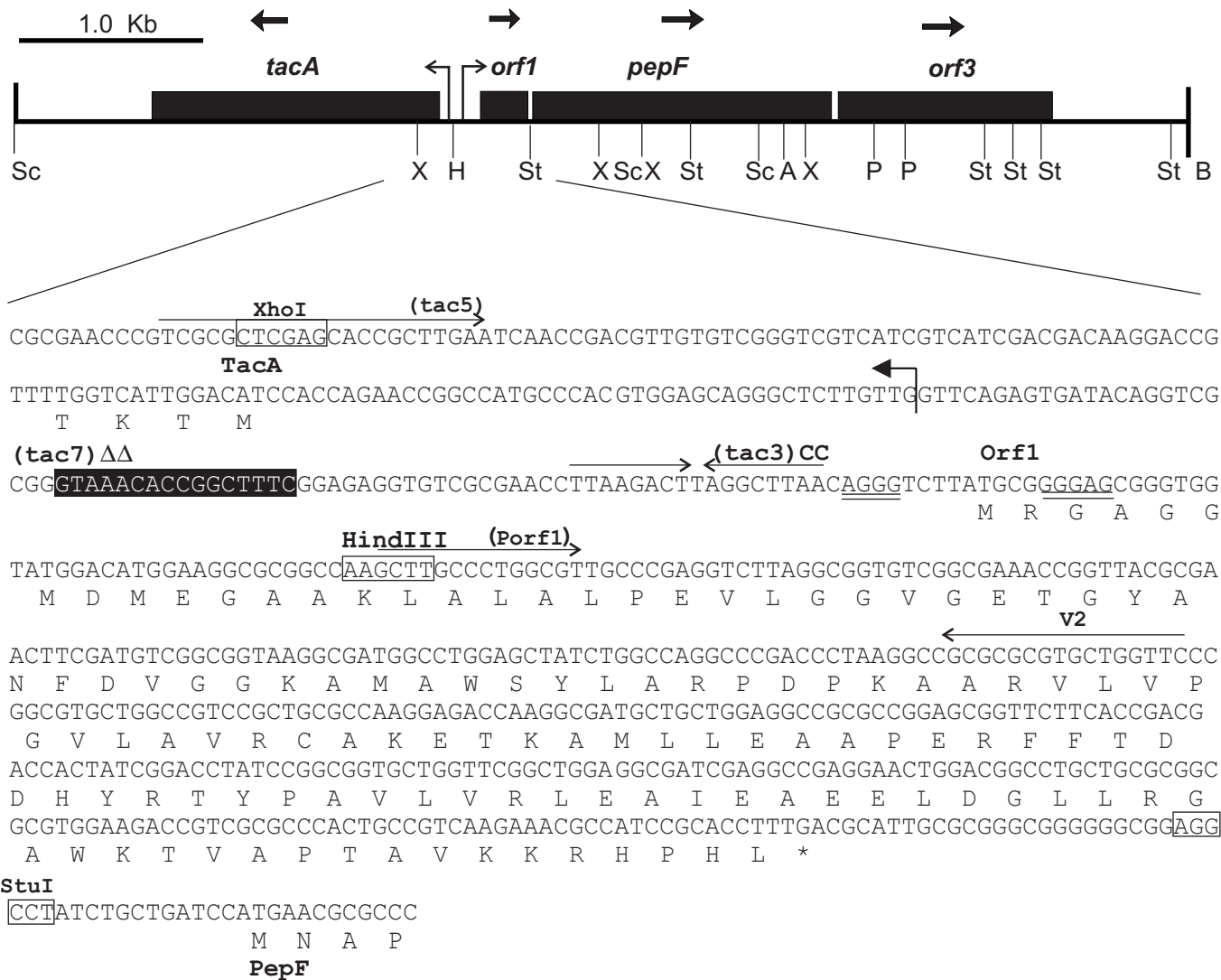
Figure 1. Structural organization of the region of the *C. crescentus* chromosome containing the *pepF* gene, and DNA sequence of the region from *tacA* to *pepF*. The boxes represent the open reading frames, and arrows indicate the direction of transcription. Restriction sites are indicated as follows: Sc, *Sac*I; X, *Xho*I; H, *Hin*dIII; St, *Stu*I; A, *Apa*I; P, *Pst*I, and B, *Bam*HI. Deduced amino acids of the corresponding genes are shown below the respective codons. Restriction sites are boxed. Small arrows mark inverted repeats. Nucleotides altered by site-directed mutagenesis are indicated above the sequence (tac3 and tac7), △ indicates deletion of bases. Putative ribosomal binding sites are double underlined. Bent arrow, *tacA* transcription start site. Single arrows, oligonucleotides used to generate promoter deletions. Consensus sequence for a CtrA-binding site is shaded.

Figure 2. Analysis of the transcripts of the *pepF* operon. RT-PCR reactions were performed using 1 µg of total RNA as template with the respective primers, and analyzed on a non-denaturing 6% polyacrylamide gel. Lanes 1-3, analysis of transcript between *orf1* and *pepF*: Lane 1, PCR reaction with primers RT-1/RT-2 without reverse transcriptase (negative control); lane 2, RT-PCR reaction using primers RT-1/RT-2 (a specific band of 314 bp was produced); lane 3, PCR reaction with Taq DNA polymerase, primers RT-1/RT-2 and plasmid DNA template (positive control). Lanes 4-6, analysis of tr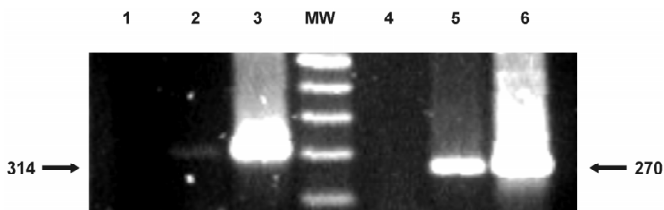anscript between *pepF* and *orf3*: Lane 4, PCR reaction with primers RT-3/RT-4 without reverse transcriptase (negative control); lane 5, RT-PCR reaction using primers RT-3/RT-4 (a specific band of 270 bp was produced); lane 6, PCR reaction with Taq DNA polymerase, primers RT-3/RT-4 and plasmid DNA template (positive control).

region (not shown). RT-PCR between ORF1 and *pepF* (Fig. 2, lane 2) produced a very faint band, although specific, and this result could not be improved after several attempts. The intergenic space between *orf1* and *pepF* is small (41 bp); no putative regulatory sequences could be identified in this region; and there is no intrinsic transcription terminator between *orf1* and *pepF*. These results altogether indicate that *orf1* is co-transcribed with *pepF* and *orf3*.

The deduced amino acid sequences of the three ORFs were compared to the available databases. ORF1 is similar to hypothetical proteins from *Mycobacterium tuberculosis* (35% identity) and *Streptomyces coelicolor* (39% of similarity). PepF presents strongest similarity to PepF from *Helicobacter pilori* (33% identity) and from *Lactococcus lactis* (29% identity) (Fig. 3). A domain analysis of PepF showed that it presents a common domain to several other metallopeptidases (PRODOM domain PD005933), sharing with these proteins the zinc-binding region signature (HEXGH) (22), and it also contains a putative transmembrane domain from amino acids 71 to 91.

ORF3 was annotated as a conserved hypothetical protein in the genome analysis (19), but a further analysis showed that it belongs to the large ABC1 protein kinase family, which includes representatives both in eukaryotes and prokaryotes (13). ORF3 presents strongest similarity to a hypothetical protein from *S. coelicolor* (29% identity), to a hypothetical protein from *C. elegans* (28% identity), to hypothetical protein Rv3197 from *M.

tuberculosis* (32% identity), to a putative ABC transporter from *Methanobacterium thermoautotrophicum* (29% identity) and to several eukaryotic ABC1 homologs. The ABC1 family includes several eukaryotic proteins that are essential for the efficient functioning of the mitochondrial *bc1* complex, as well as the *E. coli ubiB* gene, that is required for ubiquinone biosynthesis (2). The *pepF* gene disruption in strain SP2008 (see below) could lead to a polar effect in the *orf3* gene, since no promoter activity was detected between these two genes. *Caulobacter* is an obligatory aerobic prokaryote, therefore any mutations that impair respiration should be deleterious to the cell. Strain SP2008 shows no growth defect that could indicate a loss of respiratory function, which suggests that the role of ORF3 is probably not related to respiratory chain function.

A *pepF* null mutant was constructed by disrupting the gene with a plasmid encoding a kanamycin resistance marker, the insertion of the plasmid in the *pepF* locus was confirmed by Southern blotting (not shown). The *pepF*::pNPTS138 strain (SP2008) showed no growth defects, with a doubling time of 3 h, equivalent to the wild type strain NA1000, and normal swimming behaviour was observed by light microscopy. To assess whether a *pepF/orf3* mutation would have an effect on survival under stress conditions, a nutritional stress was induced by allowing the cells to enter stationary phase, and cell viability was determined. SP2008 showed no reduction in viability during the stationary phase when compared to strain NA1000, indicating that PepF and ORF3 are not essential for cell viability during prolonged growth. Strain SP2008 is also able to grow in peptone/yeast extract medium, indicating that PepF activity is not essential for nutrition under these conditions.

Since the *pepF* gene encodes a peptide highly similar to metalloendopeptidases of the M3 family, endopeptidase activity of strain SP2008 was assayed towards a peptide that is a substrate for the thimet mammalian oligopeptidases (20). In order to have an indication whether PepF localization is cytoplasmatic or extracellular, endopeptidase activity was determined in the culture supernatants as well as in total cell extracts. Some mammalian metalloendopeptidases that belong to the M3 family are the so-called thimet peptidases, whose activity is dependent on thiol (22). Activity was assayed using a quenched peptide substrate that becomes fluorescent when cleaved by an endopeptidase, in the presence and absence of 0.5 mM DTT. As observed in Table 1, the supernatants of both NA1000 and SP2008 cultures showed a similar specific activity, indicating that a similar set of enzymes are secreted in both cases. The absence of PepF activity in the cell culture medium does not support the hypothesis of the peptidase being secreted, and so it is probably localized in the cytoplasm or associated with the cell membrane. However, endopeptidase activity measured in cell extracts using this substrate, showed a 30% reduction on specific activity of strain SP2008 when compared to strain NA1000, suggesting that disruption of *pepF* caused a loss of a significant cellular
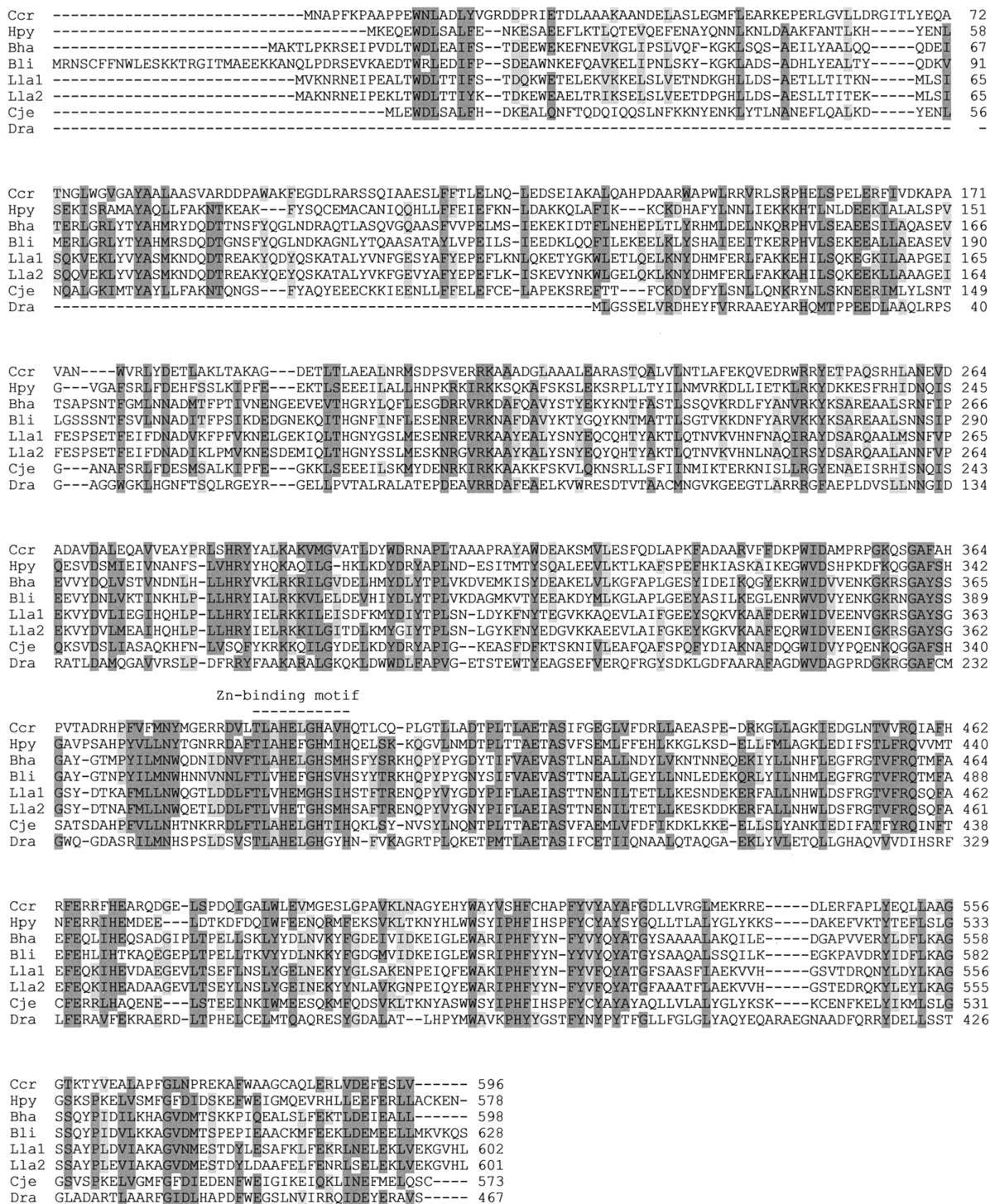
```
Ccr  --------------------------MNAPFKPAAPPEWNLADLYVGRDDPRIETDLAAAKAANDELASLEGMFLEARKEPERLGVLLDRGITLYEQA  72
Hpy  --------------------------MKEQEWDLSALFE--NKESAEEFLKTLQTEVQEFENAYQNNLKNLDAAKFANTLKH-----YENL  58
Bha  ---------------------MAKTLPKRSEIPVDLTWDLEAIFS--TDEEWEKEFNEVKGLIPSLVQF-KGKLSQS-AEILYAALQQ-----QDEI  67
Bli  MRNSCFFNWLESKKTRGITMAEEKKANQLPDRSEVKAEDTWRLEDIFP--SDEAWNKEFQAVKELIPNLSKY-KGKLADS-ADHLYEALTY-----QDKV  91
Lla1 --------------------------MVKNRNEIPEALTWDLTTIFS--TDQKWETELEKVKKELSLVETNDKGHLLDS-AETLLTITKN-----MLSI  65
Lla2 --------------------------MAKNRNEIPEKLTWDLTTIYK--TDKEWEAELTRIKSELSLVEETDPGHLLDS-AESLLTITEK-----MLSI  65
Cje  ----------------------MLEWDLSALFH--DKEALQNFTQDQIQQSLNFKKNYENKLYTLNANEFLQALKD-----YENL  56
Dra  --------------------------------------------------------------------------------------------  -
```

```
Ccr  TNGLWGVGAYAALAASVARDDPAWAKFEGDLRARSSQIAAESLFFTLELNQ-LEDSEIAKALQAHPDAARWAPWLRRVRLSRPHELSPELERFIVDKAPA  171
Hpy  SEKISRAMAYAQLLFAKNTKEAK---FYSQCEMACANIQQHLLFFEIEFKN-LDAKKQLAFIK---KCKDHAFYLNNLIEKKKHTLNLDEEKIALALSPV  151
Bha  TERLGRLYTYAHMRYDQDTTNSFYQGLNDRAQTLASQVGQAASFVVPELMS-IEKEKIDTFLNEHEPLTLYRHMLDELNKQRPHVLSEAEESILAQASEV  166
Bli  MERLGRLYTYAHMRSDQDTGNSFYQGLNDKAGNLYTQAASATAYLVPEILS-IEEDKLQQFILEKEELKLYSHAIEEITKERPHVLSEKEEALLAEASEV  190
Lla1 SQKVEKLYVYASMKNDQDTREAKYQDYQSKATALYVNFGESYAFYEPEFLKNLQKETYGKWLETLQELKNYDHMFERLFAKKEHILSQKEGKILAAPGEI  165
Lla2 SQQVEKLYVYASMKNDQDTREAKYQEYQSKATALYVKFGEVYAFYEPEFLK-ISKEVYNKWLGELQKLKNYDHMFERLFAKKAHILSQKEEKLLAAAGEI  164
Cje  NQALGKIMTYAYLLFAKNTQNGS---FYAQYEEECKKIEENLLFFELEFCE-LAPEKSREFTT---FCKDYDFYLSNLLQNKRYNLSKNEERIMLYLSNT  149
Dra  -----------------------------------------------------------MLGSSELVRDHEYFVRRAAEYARHQMTPPEEDLAAQLRPS   40
```

```
Ccr  VAN----WVRLYDETLAKLTAKAG---DETLTLAEALNRMSDPSVERRKAAADGLAAALEARASTQALVLNTLAFEKQVEDRWRRYETPAQSRHLANEVD  264
Hpy  G---VGAFSRLFDEHFSSLKIPFE---EKTLSEEEILALLHNPKRKTRKKSQKAFSKSLEKSRPLLTYILNMVRKDLLIETKLRKYDKKESFRHIDNQIS  245
Bha  TSAPSNTFGMLNNADMTFPTIVNENGEEVEVTHGRYLQFLESGDRYAPLND-ESITMTYSQALEEVLKTLKAFSPEFHKIASKAIKEGWVDSHPKDFKQGGAFSH  266
Bli  LGSSSNTFSVLNNADITFPSIKDEDGNEKQITHGNFINFLESENREVRKNAFDAVYKTYGQYKNTMATTLSGTVKKDNFYARVKKYKSAREAALSNNSIP  290
Lla1 FESPSETFEIFDNADVKFPFVKNELGEKIQLTHGNYGSLMESENREVRKAAYEALYSNYEQCQHTYAKTLQTNVKVHNFNAQIRAYDSARQAALMSNFVP  265
Lla2 FESPSETFEIFDNADIKLPMVKNESDEMIQLTHGNYSSLMESKNRGVRKAAYKALYSNYEQYQHTYAKTLQTNVKVHNLNAQIRSYDSARQAALANNFVP  264
Cje  G---ANAFSRLFDESMSALKIPFE---GKKLSEEEILSKMYDENRKIRKKAAKKFSKVLQKNSRLLSFIINMIKTERKNISLLRGYENAEISRHISNQIS  243
Dra  G---AGGWGKLHGNFTSQLRGEYR---GELLPVTALRALATEPDEAVRRDAFEAELKVWRESDTVTAACMNGVKGEEGTLAERRRGFAEPLDVSLLNNGID  134
```

```
                            Zn-binding motif
                            ----------
Ccr  PVTADRHPFVFMNYMGERRDVLTLAHELGHAVHQTLCQ-PLGTLLADTPLTLAETASIFGEGLVFDRLLAEASPE-DRKGLLAGKIEDGLNTVVRQIAFH  462
Hpy  GAVPSAHPYVLLNYTGNRRDAFTIAHEFGHMIHQELSK-KQGVLNMDTPLTTAETASVFSEMLFFEHLKKGLKSD-ELLFMLAGKLEDIFSTLFRQVVMT  440
Bha  GAY-GTMPYILMNWQDNIDNVFTLAHELGHSMHSFYSRKHQPYPYGDYTIFVAEVASTLNEALLNDYLVKNTNNEQEKIYLLNHFLEGFRGTVFRQTMFA  464
Bli  GAY-GTNPYILMNWHMNNLFTLVHEFGHSVHSYYTRKHQPYPYGDIVIDKEIGLEWSRIPHFYYN-YVVQYATGYSSAAQALSSQILK-----EGKPAVDRYIDFLKAG  488
Lla1 GSY-DTKAFMLLNWQGTLDDLFTLVHEMGHSIHSTFTRENQPYVVYGDYPIFLAEIASTTNENILTETLLKESNDEKERFALLNHWLDSFRGTVFRQSQFA  462
Lla2 GSY-DTNAFMLLNWQETLDDLFTLVHETGHSMHSAFTRENQPYVVGNYPIFLAEIASTTNENILTETLLKESKDDKERFALLNHWLDSFRGTVFRQSQFA  461
Cje  SATSDAHPFVLLNHTNKRRDLFTIAHELGHTIHQKLSY-NVSYLNQNTPLTTAETASVFAEMLVFDFIKDKLKKE-ELLSLYANKIEDIFATFYRQINFT  438
Dra  GWQ-GDASRILMNHSPSLDSVSTLAHELGHGYHN-FVKAGRTPLQKETPMTLAETASIFCETIIQNAALQTAQGA-EKLYVLETQLLGHAQVVVDIHSRF  329
```

```
Ccr  RFEERRFHEARQDGE-LSPDQIGALWLEVMGESLGPAVKLNAGYEHYWAYVSHFCHAPFYVYAYAFGDLLVRGLMEKRRE-----DLRERFAPLYEQLIAAG  556
Hpy  NFEERRIHEMDEE---LDTKDFDQIWFEENQRMFEKSVKLTKNYHLWWSYIPHFIHSPFYCYAYSYGQLLTLALYGLYKKS----DAKEFVKTYTEFLSLG  533
Bha  EEFEQLIHEQSADGIPLTPELLSKLYDLNKKYFGDEIVIDKEIGLEWARIPHFYYN-FYVYQYATGYSAAAALAKQILE-----DGAPVVERYLDFLKAG  558
Bli  EFEHLIHTKAQEGEPLTPELLTKVYYDLNKKYFGDGMVIDKEIGLEWSRIPHFYYN-YYVYQYATGYSAAQALSSQILK-----EGKPAVDRYIDFLKAG  582
Lla1 EFEQKIHEVDAEGEVLTSEFLNSLYGELNEKYYGLSAKENPEIQFEWAKIPHFYYN-FYVFQYATGFSAASFIAEKVVH-----GSVTDRQNYLDYLKAG  556
Lla2 EFEQKIHEADAAGEVLTSEYLNSLYGEINEKYYNLAVKGNPEIQYEWARIPHFYYN-FYVFQYATGFAAATFLAEKVVH-----GSTEDRQKYLEYLKAG  555
Cje  CFERRLHAQENE---LSTEEINKIWMEESQKMFQDSVKLTKNYASWWSYIPHFIHSPFYCYAYAYAQLLVLALYGLYKSK----KCENFKELYIKMLSLG  531
Dra  LFERAVFEKRAERD-LTPHELCELMTQAQRESYGDALAT--LHPYMWAVKPHYYGSTFYNYPYTFGLLFGLGLYAQYEQARAEGNAADFQRRYDELLSST  426
```

```
Ccr  GTKTYVEALAPFGLNPREKAFWAAGCAQLERLVDEFESLV------  596
Hpy  GSKSPKELVSMFGFDIDSKEFWEIGMQEVRHLLEEFERLLACKEN-  578
Bha  SSQYPIDILKHAGVDMTSKKPIQEALSLFEKTLDEIEALL------  598
Bli  SSQYPIDVLKKAGVDMTSPEPIEAACKMFEEKLDEMEELLMKVKQS  628
Lla1 SSAYPLDVIAKAGVNMESTDYLESAFKLFEKRLNELEKLVEKGVHL  602
Lla2 SSAYPLEVIAKAGVDMESTDYLDAAFELFENRLSELEKLVEKGVHL  601
Cje  GSVSPKELVGMFGFDIEDENFWEIGIKEIQKLINEFMELQSC----  573
Dra  GLADARTLAARFGIDLHAPDFWEGSLNVIRRQIDEYERAVS-----  467
```

Figure 3. Amino acid sequence alignments of the product of the *pepF* gene. Alignment was done using the ClustalW program (22), and edited using the GeneDoc program (23). Amino acids are shaded according to degree of conservation. PepF from *Caulobacter* (Ccr) is compared to putative metallopeptidases from *H. pylori* (Hpy; B71934), *B. halodurans* (Bha; BAB06575), *B. licheniformis* (Bli; T44581), *L. lactis* chromosomal gene product (Lla1; P94876) and *L. lactis* plasmid gene product (Lla2; P54124), *C. jejuni* (Cje; F81313), *D. radiodurans* (Dra; B75322). The conserved zinc-binding motif is indicated above the sequence. Amino acid numbers are indicated at the right of each sequence.

Table 1. Assays for endopeptidase activity of the wild type strain NA1000 and the mutant strain SP2008.

Samples[a]	Specific activity[b] (nmol.min^{-1}.μg^{-1})	Relative activity[c]
NA1000 extract (+DTT)	548 ± 14	1.0
NA1000 extract (-DTT)	949 ± 47	1.73
SP2008 extract (+DTT)	368 ± 13	0.67
SP2008 extract (-DTT)	443 ± 17	0.80
NA1000 Sn (+DTT)	192 ± 31	1.0
NA1000 Sn (-DTT)	361 ± 13	1.88
SP2008 Sn (+DTT)	262 ± 22	1.36
SP2008 Sn (-DTT)	372 ± 57	1.93

[a] Activity was measured for cell extracts and for the cultures supernatants (Sn); [b] The numbers represent the average followed by the standard deviation; [c] Activities are shown relative to the specific activity of NA1000 cell extract in the upper part of the Table and to the specific activity of NA1000 supernatant in the lower part

endopeptidase activity in the mutant. This result also indicates that there are other peptidases in the cell that are able to cleave this substrate, which agrees with the presence of so many genes encoding peptidases in the genome (19). Forty-eight gene products were annotated as putative peptidases, out of which nine are from the M20/25/40 families, seven from the M23/27 families, four from the M16 family and one representative of each of the M1, M3, M13, M22 and M24 families.

An increase in activity was observed when the assays were carried out in the absence of DTT in all cases, indicating that even if there is a thiol-dependent endopeptidase in *Caulobacter*, its activation is not very pronounced, and some of the enzymes are in fact inhibited by thiol. Our results show that *C. crescentus* PepF hydrolyzes the peptide of 7 residues that is a substrate for the mammalian thimet oligopeptidase (4). The *L. lactis* PepF is also capable of hydrolyzing peptides very similar to this, containing 7 to 17 amino acids, with a cleavage site preference after a Phe residue (16). The substrate specificity of *C. crescentus* PepF has not been determined, but since it is similar to the *L. lactis* enzyme, some of its catalytic properties should be conserved.

The promoter region of the *pep* operon shows several notable features (Fig. 1). Upstream of the divergent *tacA* transcription start site there is a binding site for the global transcription regulator CtrA (GAAA-N7-TTTAC; (21)) centered 25 bp upstream of the *tacA* start site (14). The region upstream of *orf1* contains one inverted repeat that could have a regulatory role, centered at position -18 relative to *orf1* most upstream ATG. Two alternative start codons for *orf1* are located 15 bp downstream, and at least the first of these has a good ribosome binding site consensus sequence.
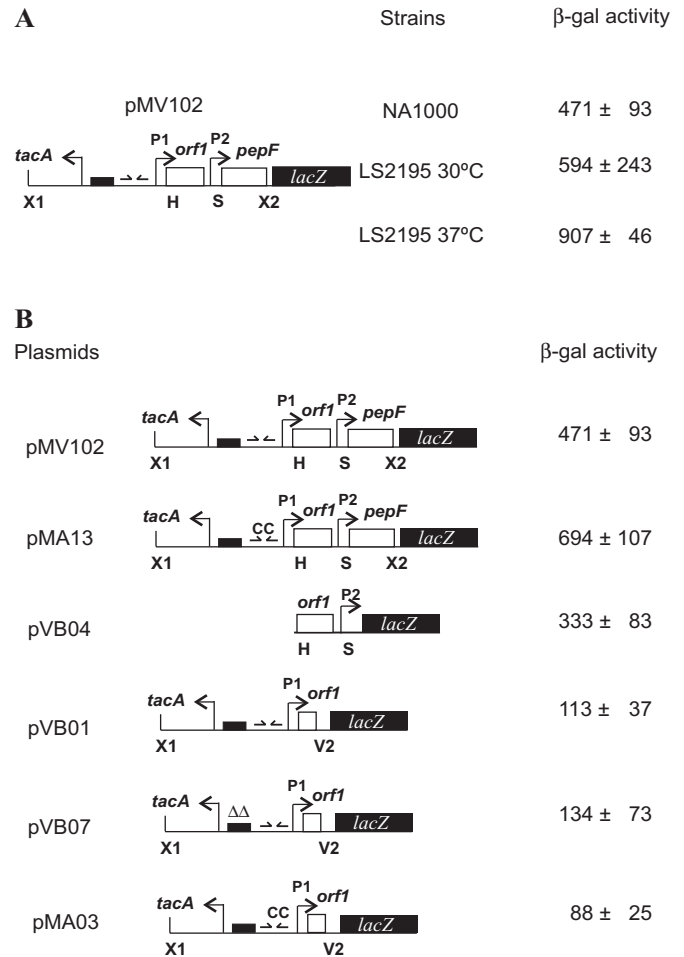

Figure 4. Determination of the transcriptional regulatory regions in the *pepF* locus. (**A**) Schematic representation of plasmid pMV102, which contains DNA from the beginning of the *tacA* gene (*Xho*I site) to the middle of the *pepF* gene cloned into the reporter vector pRK*lacZ*290. This plasmid was introduced into *C. crescentus* strains NA1000 and LS2195, and transcription levels were determined by b-galactosidase activity assays. (**B**) Different constructions of the *pep* regulatory region were cloned into pRK*lacZ*290 and these plasmids were introduced into *C. crescentus* NA1000 by conjugation. The transcriptional activity of each construction was determined by b-galactosidase activity assays. Numbers are the means of at least three independent measurements, followed by the respective standard deviation. Restriction sites are indicated as follows: X$_1$ and X$_2$, *Xho*I; H, *Hin*dIII; S, *Stu*I. V2, oligonucleotide used to generate the cloned fragment. Shaded box indicates the CtrA-binding site. Inverted arrows indicate the inverted repeat. White boxes indicate the open reading frames. △ indicates deletion of two nucleotides in the pVB07 mutant. CC indicates substitution of two nucleotides in the pMA13 mutant.

The expression of *pep* in two *C. crescentus* strains was determined through b-galactosidase activity assays in cells containing the pMV102 construction (Fig. 4A). This plasmid was introduced into strain NA1000 (wild type) and into strain LS2195, in which the CtrA protein is inactive at the temperature of 37ºC. In order to investigate whether *pepF* transcription was somehow affected by CtrA, its expression was determined by b-galactosidase activity assay both at 30ºC and at 37ºC. Figure 4A shows that the levels of b-galactosidase activity were even a little higher at 37ºC, indicating that *pepF* expression is not affected by the absence of CtrA.

The RT-PCR results have indicated that two distinct promoters could be used to transcribe the *pepF* gene, one before *orf1* (P1) and another between *orf1* and *pepF* (P2). A two-base alteration in the inverted repeat localized at position –18 (from *orf1* first ATG) did not alter the levels of expression, indicating that this sequence does not have a regulatory role. In order to analyze each promoter separately, deletions were made both upstream and downstream the *orf1* gene (Fig. 4B). A deletion of the region upstream the *Hin*dIII site (pVB04) caused a little pronounced reduction in the b-galactosidase levels, suggesting that the promoter P2 between *orf1* and *pepF* is responsible for most the transcription activity. This can be also confirmed by deletion of downstream sequences, indicating that the P1 promoter is not very strong; the construction that removes the region downstream the beginning of *orf1* confers less than half the activity of b-galactosidase (compare pVB01 and pMV102). A mutation that inactivates the CtrA binding site has no effect in P1 activity (pVB07), but the two-base alteration in the inverted repeat causes a reduction in expression (pMA03). The results with pVB04 indicate that the region comprised between the *Hin*dIII site and the *Stu*I site contains promoter activity, which is probably responsible for transcription of the *pepF/orf3* genes, explaining the increased amount of mRNA corresponding to *pepF* and *orf3* observed in the RT-PCR.

The structure of the *pep* operon in *C. crescentus* is very interesting, since the *pepF* gene is co-transcribed with a gene encoding a putative protein kinase, which could indicate a coordinate role for the two peptides in the cell. The absence of cell cycle regulation indicates that this endopeptidase is probably involved in basal metabolic functions, as described for *L. lactis* (16,17). However, we could see no gene induction by the presence of peptides in the culture medium, and the *pepF* mutant strain is able to grow using peptides as the sole carbon source (not shown), suggesting that there could be other peptidases in the cell which are responsible for nutrition.

ACKNOWLEDGEMENTS

RESUMO

Clonagem e caracterização do gene que codifica a endopeptidase PepF da bactéria aquática *Caulobacter crescentus*

As metalopeptidases têm papel importante em bactérias, estando envolvidas em vários processos que dependem de reciclagem de proteínas, como nutrição, degradação de peptídios-sinal, localização de proteínas e virulência. Nós clonamos e caracterizamos o gene da metaloendopeptidase PepF da bactéria aquática *Caulobacter crescentus*. O gene a montante de *pepF* (*orf1*) codifica uma proteína hipotética conservada encontrada em *Mycobacterium* e *Streptomyces*. *pepF* é co-transcrito com o gene a jusante (*orf3*), que codifica uma proteína pertencente à família ABC1 de proteínas quinases, sugerindo que estas duas proteínas podem compartilhar uma função comum na célula. A proteína PepF de *C. crescentus* possui o motivo HEXGH presente nos domínios de ligação de zinco de homólogos de PepF. A interrupção do gene *pepF* por inserção de sequência de DNA de um vetor não gerou nenhum defeito de crescimento, mas a linhagem mutante possui apenas 30% da atividade especifica de endopeptidases presentes na linhagem selvagem. Deleções e mutações pontuais na região regulatória mostraram que há duas possíveis regiões promotoras, e a expressão do operon é independente do regulador de transcrição CtrA. Os resultados indicam que PepF não é essencial para o crescimento desta bactéria utilizando peptídios como fonte de carbono, sugerindo que outras peptidases possam estar realizando esta função.

Palavras-chave: endopeptidase, família M3, regulação gênica, bactéria

REFERENCES

1. Borodovsky, M.; McIninch, J. Recognition of genes in DNA sequence with ambiguities. *Biosystems*, 30: 161-171, 1993.
2. Brasseur, G.; Tron, G.; Dujardin, G.; Slonimski, P.P.; Brivet-Chevillote, P. The nuclear ABC1 gene is essential for the correct conformation and functioning of the cytochrome bc1 complex and the neighbouring complexes II and IV in the mitochondrial respiratory chain. *Eur. J. Biochem.*, 246: 103-111, 1997.
3. Brun, Y.V.; Marczynski, G.; Shapiro, L. The expression of asymmetry during *Caulobacter* cell differentiation. *Annu. Rev. Biochem.*, 63: 419-450, 1994.
4. Camargo A.C.; Gomes M.D.; Reichl A.P.; Ferro E.S.; Jacchieri S.; Hirata I.Y.; Juliano L. Structural features that make oligopeptides susceptible substrates for hydrolysis by recombinant thimet oligopeptidase. *Biochem. J.*, 324: 517-522, 1997.

5. Chen, W.P.; Kuo, T.T. A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. *Nucleic Acids Res.*, 21: 2260, 1993.
6. Ely, B. Genetics of *Caulobacter crescentus. Methods Enzymol.*, 204: 372-384, 1991.
7. Evinger, M.; Agabian, N. Envelope-associated nucleoid from *Caulobacter crescentus* stalked and swarmer cells. *J. Bacteriol.*, 132: 294-301, 1977.
8. Gober, J.W.; Shapiro, L. A developmentally regulated *Caulobacter* flagellar promoter is activated by 3' enhancer and IHF binding elements. *Mol. Biol. Cell*, 3: 913-926, 1992.
9. Gober, J.W.; Marques, M.V. Regulation of cellular differentiation in *Caulobacter crescentus. Microbiol. Rev.*, 59: 31-47, 1995.
10. Jenal, U.; Fuchs, T. An essential protease involved in bacterial cell-cycle control. *EMBO J.*, 17: 5658-69, 1998.
11. Jongeneel, C.V.; Bouvier, J.; Bairoch, A. A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Lett.*, 242: 211-214, 1989.
12. Kunkel, T.A.; Bebenek, K.; McClary, J. Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.*, 204: 125-139, 1991.
13. Leonard, C.J.; Aravind, L.; Koonin, E.V. Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily. *Genome Res.*, 8: 1038-1047, 1998.
14. Marques, M.V.; Gomes, S.L.; Gober, J.W. A gene coding for a putative sigma 54 activator is developmentally regulated in *Caulobacter crescentus. J. Bacteriol.*, 179: 5502-5510, 1997.
15. Miller, J.H. *Experiments in molecular genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York pp. 352-355, 1972.
16. Monnet, V.; Nardi, M.; Chopin, A.; Chopin, M.-C.; Gripon, J.-C. Biochemical and genetic characterization of PepF, an oligopeptidase from *Lactococcus lactis. J. Biol. Chem.*, 269: 32070-32076, 1994.
17. Nardi, M.; Renault, P.; Monnet, V. Duplication of the *pepF* gene and shuffling of DNA fragments on the lactose plasmid of *Lactococcus lactis. J. Bacteriol.*, 179: 4164-4171, 1997.
18. Nicholas K.B.; Nicholas Jr., H.B.; Deerfield II, D.W. GeneDoc: Analysis and Visualization of Genetic Variation, *EMBNEW. NEWS*, 4: 14, 1997.
19. Nierman, W.C.; Feldblyum, T.V.; Laub, M.T.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.R.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.B.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Venter, J.C.; Shapiro, L.; Fraser, C.M. Complete genome sequence of *Caulobacter crescentus. Proc. Natl. Acad. Sci. USA*, 98: 4136-41, 2001.
20. Portaro, F.C.V.; Gomes, M.D.; Cabrera, A.; Fernandes, B.L.; Silva, C.L.; Ferro, E.S.; Juliano, L.; Camargo, A.C.M. Thimet oligopeptidase and the stability of MHC class I epitopes in macrophage cytosol. *Biochem Biophys Res Commun.*, 255: 596-601, 1999.
21. Quon, K.C.; Marczynski, G.; Shapiro, L. Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell*, 84: 83-93, 1996.
22. Rawlings, N.D.; Barrett, A.J. Evolutionary families of metallopeptidases. *Meth. Enzymol.*, 248: 183-228, 1995.
23. Simon, R.; Priefer, U.; Puhler, A. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. *Biotechnology*, 1: 784-790, 1983.
24. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22: 4673-4680, 1994.
25. Wright, R.; Stephens, C.; Zweiger, G.; Shapiro, L.; Alley, M.R. *Caulobacter* Lon protease has a critical role in cell-cycle control of DNA methylation. *Genes Dev.*, 10: 1532-42, 1996.