Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 343 (2006) 77–84
BBRC
www.elsevier.com/locate/ybbrc

Characterization of a periplasmic S1-like nuclease coded by the *Mesorhizobium loti* symbiosis island

Maxim Pimkin, C. Glenn Miller, Lauryn Blakesley [1], Catherine A. Oleykowski [2], Nagendra S. Kodali [3], Anthony T. Yeung *

Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111, USA

Received 14 February 2006
Available online 28 February 2006

Abstract

DNA sequences encoding hypothetical proteins homologous to S1 nuclease from *Aspergillus oryzae* are found in many organisms including fungi, plants, pathogenic bacteria, and eukaryotic parasites. One of these is the M1 nuclease of *Mesorhizobium loti* which we demonstrate herein to be an enzymatically active, soluble, and stable S1 homolog that lacks the extensive mannosyl-glycosylation found in eukaryotic S1 nuclease homologs. We have expressed the cloned M1 protein in *M. loti* and purified recombinant native M1 to near homogeneity and have also isolated a homogeneous M1 carboxy-terminal hexahistidine tag fusion protein. Mass spectrometry and N-terminal Edman degradation sequencing confirmed the protein identity. The enzymatic properties of the purified M1 nuclease are similar to those of S1. At acidic pH M1 is 25 times more active on single-stranded DNA than on double-stranded DNA and 3 times more active on single-stranded DNA than on single-stranded RNA. At neutral pH the RNase activity of M1 exceeds the DNase activity. M1 nicks supercoiled RF-I plasmid DNA and rapidly cuts the phosphodiester bond across from the nick in the resultant relaxed RF-II plasmid DNA. Therefore, M1 represents an active bacterial S1 homolog in spite of great sequence divergence. The biochemical characterization of M1 nuclease supports our sequence alignment that reveals the minimal 21 amino acid residues that are necessarily conserved for the structure and functions of this enzyme family. The ability of M1 to degrade RNA at neutral pH implies previously unappreciated roles of these nucleases in biological systems.

Keywords: Single-strand specific nuclease; M1; S1; *Mesorhizobium loti*; Symbiosis island; DNase; RNase

Single-strand specific nucleases of the S1 family are eukaryotic DNase/RNase bifunctional metalloenzymes that require Zn^{2+} for activity [1]. The family includes S1 nuclease from *Aspergillus oryzae* [2,3], P1 nuclease from *Penicillium citrinum* [4,5], mung bean nuclease [6–8], and P4 nuclease of *Leishmania* [9]. These enzymes are exo-endonucleases highly specific for single-stranded DNA, have acidic pH optima, and are essentially inactive at alkali pH. They cut double-stranded DNA at relatively large distorted regions, such as the site of an insertion mismatch of three nucleotides or larger, or at AT-rich regions of double-stranded DNA. Typical enzymes of the S1 family do not incise at base-substitution mismatches [10], although there are notable exceptions to this rule [11,12].

The extensive mannosyl glycosylation and multiple disulfide bonds found in these enzymes are common features of eukaryotic extracellular proteins and have greatly complicated our previous attempts to express and purify a recombinant protein for biochemical and structural studies. No S1-like enzymes of bacterial origin have been described. However, analysis of the available complete "small genome" sequences suggests that at least seven bacterial species contain ORFs with distant homology to S1,

* Corresponding author. Fax: +1 215 728 3647.
E-mail address: AT_Yeung@fccc.edu (A.T. Yeung).
[1] Present address: Thomas Jefferson University, Philadelphia, PA, USA.
[2] Present address: Department of Oncology, GlaxoSmithKline, Collegeville, PA, USA.
[3] Present address: Brookdale University Hospital & Med. Ctr., Brooklyn, NY 11212, USA.

0006-291X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2006.02.117

each preceded by a putative signal sequence. The low level of sequence identity between S1 and these hypothetical proteins (25–35%) is not sufficient to provide a reliable functional inference [13] and should be treated as a guide for experimental studies.

S1 homolog genes (*mlr6309*, GenBank Accession No. BAB52626, putative product hereby called the M1 nuclease and *orfE*, GenBank Accession No. AAC24514, putative product hereby called the M2 nuclease) are found on the chromosomes of two strains of *M. loti*, a gram-negative nitrogen-fixing bacterium capable of symbiosis with several *Lotus* species by forming nodules on the roots of the host plant. The symbiotic genes of these *M. loti* strains are clustered in >500 kb contiguous regions on their chromosomes [14,15]. These gene clusters, which have been named "symbiosis islands," encode a wealth of proteins including nitrogen fixation enzymes, symbiosis factors, protein excretion systems, and putative S1-like nucleases; these islands are transmittable to non-symbiotic *Mesorhizobium* species and impart the ability to colonize *Lotus* roots [15]. The presence of putative signal peptide indicates that these proteins may be destined for secretion to the periplasmic space or excretion into the environment.

Single-strand DNase activity cannot be detected in *M. loti* MAFF303099 total cell extracts, periplasmic extracts, or culture media (data not shown), suggesting that the M1 gene is not expressed in laboratory-cultured free-living bacteria. Therefore, we decided to study this gene product by recombinant expression of the ORF. However, to the best of our knowledge, we and others have not been able to express an enzymatically active form of an S1 homolog, including the M1 nuclease, in *Escherichia coli*. Furthermore, it has not been possible to refold the inclusion bodies produced in *E. coli* into active enzymes (data not shown). Because *M. loti* already coded for the M1 nuclease, it appeared a likely bacterial system for the expression of an active protein. Herein, we describe the expression of a recombinant M1 nuclease in *M. loti* and purification of this enzyme to homogeneity. The purified M1 shows a strong preference for single-stranded DNA and RNA, and exhibits different pH optima for its single-strand DNase versus single-strand RNase activities. To the best of our knowledge, this is the first bacterial homolog of the eukaryotic S1 nuclease that has been shown to be enzymatically active.

Materials and methods

Materials and software. Plasmid pUC19 DNA was isolated from DH5α host cells using the QIAGEN Maxi Kit, following the manufacturer's instructions. Calf thymus DNA was obtained from Sigma and purified by repeated cycles of proteinase K digestion and phenol extraction. Chromatography resins and columns were purchased from Pharmacia Biotech. S1 nuclease was purchased from Amersham Biosciences. The *M. loti* strain MAFF303099 was a gift from Dr. C. Ronson. Internet version of ClustalW 1.82 (http://www.ebi.ac.uk/clustalw/) was used for the alignment. The free Cn3D software (http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml) was used for structure visualization.

Construction of the expression plasmids pNOF-M1-Kan and pM1-H. A PCR product of the *mlr6309* gene of *M. loti* strain MAFF303099, GenBank Accession No. AP003008 [14], was cloned in pNOF100-G plasmid between *Nco*I and *Pst*I restriction endonuclease sites to place protein expression under the control of *tac* promoter (GL Biotech, Germany). A modification of Hetero-Stagger PCR cloning technique was used [16]. Two pairs of primers (TY973 and TY974, sequences: 5′ AAATCTTTGGTGC TGGGGAT and 5′ CGAGCGCGCGTAGG, respectively; TY975 and TY976, sequences 5′ CATGAAATCTTTGGTGCTGGGGAT and 5′-TGCACGAGCGCGCGTA, respectively) were used in two separate PCRs to amplify the *mlr6309* gene (including putative signal peptide-coding sequence) from *M. loti* genomic DNA. The PCR products were column-purified using the QIAquick PCR Purification Kit, mixed in equimolar amounts, and subjected to thermal denaturation followed by slow cooling to allow formation of heteroduplexes that possess *Nco*I and *Pst*I sticky ends. The resulting product was cloned in the pNOF100-G plasmid that had been digested with *Nco*I and *Pst*I restriction enzymes, yielding the pNOF-M1 plasmid. Because *M. loti* cells transformed with pNOF-M1 could not be easily selected, we replaced the *aacC1* gene for gentamicin phosphotransferase of pNOF-M1 with the gene for kanamycin phosphotransferase from the pACYC177 plasmid [17] to yield the plasmid pNOF-M1-Kan which was able to transform *M. loti* to kanamycin resistance. The recombinant protein produced was designated M1 and has a calculated molecular weight of 28,228.8.

To construct a His-tagged derivative of M1, a sequence encoding a hexahistidine tag and a 5-amino acid linker peptide was attached to the 3′-end of the M1 coding gene by a modified QuikChange™ site-directed mutagenesis procedure (Stratagene). The point of insertion was immediately after the codon for the last amino acid of M1. A pair of complementary synthetic oligonucleotides (TY1163 and TY1164, sequences 5′ CCACAGCTCCCGCATGCTCCCTGCCTGAATTCGAAGAAGAAC ACCACCACCACCACCACTAAAGCCGTTCTGGCGCCTACGCGC GCTC and 5′ GAGCGCGCGTAGGCGCCAGAACGGCTTTAGT GGTGGTGGTGGTGGTGTTCTTCTTCGAATTCAGGCAGGGAG CATGCGGGAGCTGTGG, respectively) was purified by PAGE before use. The 3′ region of TY1163 consists of the first 26 nucleotides of the destination pNOF-M1-Kan plasmid downstream from the point of insertion. The 5′ region of TY1163 consists of the first 26 nucleotides of pNOF-M1-Kan upstream of the insertion point. The middle region is formed by the sense-strand of five amino acid linker followed by a hexahistidine tag and a stop codon (EFEEEHHHHHH∗). The whole-plasmid PCR mutagenesis reaction was conducted in accordance with manufacturer's recommendations. Fifty nanogram of recipient plasmid and 100 ng of each mutagenic oligonucleotide were used in a 50 μl reaction volume. The thermal cycling protocol consisted of 30 s initial denaturation step at 95 °C, followed by 18 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, and elongation at 68 °C for 16 min (2 min/kb plasmid template DNA). Transformation of 30 μl chemically competent XL-1-blue *E. coli* cells followed by selection of Kan^R clones yielded 14–25 colonies per plate, about 70% of which contained recombinant DNA. The new expression vector was 8056 bp in size and was designated pM1-H. The M1 nuclease produced was designated M1His6 with a theoretical molecular weight of 29051.6.

Expression of the M1 nuclease in M. loti. Preparation of electro-competent *M. loti* cells and electroporation were performed essentially as described [18]. *M. loti* containing pNOF-M1-Kan or pM1-H was grown in 1 L batches of TY broth (5 g bactotryptone, 3 g yeast extract, and 0.875 g $CaCl_2{\cdot}2H_2O$ per 1 L of deionized water) supplemented with 20 mg/L kanamycin sulfate at 27 °C in a shaking water bath. When OD_{600} reached 0.4–0.6 (typically after 5 days of cultivation), the cells were harvested and resuspended in 1 L of TY broth, 20 mg/L kanamycin sulfate, 10 μM $ZnCl_2$, and 0.5 mM IPTG. The cells were harvested at 24 h at a culture density of 1.0–1.2. All subsequent procedures were done at 4 °C or on ice.

Protein extraction from periplasmic space. Mesorhizobium loti cells were washed with 20% sucrose, 50 mM Tris–HCl, pH 7.5, pelleted by centrifugation, and resuspended in 100 mL of 20% sucrose, 50 mM Tris–HCl, pH 7.5, containing 10,000 U/mL T4 Ready-Lyse Lysozyme (Epicentre).

The suspension was incubated on ice for 1 h. After an equal volume of deionized water was added, the suspension was shaken briefly and then the spheroplasts were pelleted by centrifugation at 14,000*g* for 30 min and discarded. The supernatant was filtered through a 0.22 μm filter (Millipore). When necessary, the crude protein extract was concentrated by ultra filtration through a cellulose YM-10 membrane (Millipore) before loading onto a Mono Q column.

Recombinant M1 purification. Purification was performed on a microbore HPLC SMART system (Pharmacia Biotech). The filtered crude periplasm extract was loaded on a Mono Q column (1.6 × 55 mm) that had been equilibrated with 50 mM Tris–HCl, pH 8.2. A salt gradient (0–0.5 M NaCl in 4 mL of 50 mM Tris–HCl, pH 8.2) was passed through the column. Fractions containing M1 were determined by the RF I nicking assay and SDS–PAGE with Coomassie staining, pooled, and concentrated using a Centricon® YM-3 device. Superdex 200 size-exclusion chromatography was performed in 50 mM Tris–HCl, pH 8.2, 1 M NaCl, and 10 μM $ZnCl_2$.

Recombinant M1His6 purification. Hexahistidine-tagged M1 was purified from crude periplasm extract by metal affinity chromatography using Talon™ resin (Clontech). The crude extract was loaded on a column that had been equilibrated with 50 mM Tris–HCl, pH 7.6, 300 mM NaCl, and 10 μM $ZnCl_2$. After sample loading, the column was washed with the same buffer and then with 50 mM Tris–HCl, pH 7.6, 300 mM NaCl, 10 μM $ZnCl_2$, and 5 mM imidazole. M1His6 was eluted with 50 mM Tris–HCl, pH 7.6, 300 mM NaCl, 10 μM $ZnCl_2$, and 150 mM imidazole. The nuclease was further purified on Mono Q anion-exchange column as described in the native M1 purification section. Protein concentration was measured by Bio-Rad Bradford Protein Assay [19]. Computer-based densitometry analysis was used to assess the purity of M1 preparations following SDS–PAGE with GelCode staining. The enzymatic properties of M1His6 were found to be identical to those of M1 (data not shown), therefore only the data for the unmodified M1 nuclease are presented in this manuscript.

Confirmation of the purified recombinant M1 nuclease protein sequence. The purified recombinant M1 nuclease was resolved on SDS–PAGE and the main protein band was excised and subjected to tryptic digestion. The N-terminal peptide sequence was determined by Edman degradation. An independent de novo sequencing of the N-terminal peptide was performed in the LC/MS/MS mode and the MALDI-TOF mode on an Applied Biosystems Q-Star® XL mass spectrometer. Peptide coverage for the rest of the protein was performed by peptide mass mapping on both a Bruker Reflex IV mass spectrometer and the Q-Star® XL mass spectrometer. The Mascot search engine (http://www.matrixscience.com) was used to determine the peptide match in Swiss-Prot database.

In-gel refolding of M1 nuclease and activity staining. An SDS polyacrylamide gel was prepared as described [20] except 0.15 mg/mL heat-denatured calf thymus DNA was added in the separating gel and no reducing agents were used in either the gel or the buffers [12]. After sufficient separation of the protein samples had been achieved, the gel was washed three times for 20 min with 25% isopropanol, 10 mM Tris–HCl, pH 7.4, and then three times with 10 mM Tris–HCl, pH 7.4. For protein refolding the gel was transferred into refolding buffer (40 mM Tris–HCl, pH 7.4, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 2 μM $ZnCl_2$) and incubated for 24 h at room temperature followed by 1 h incubation in the same buffer at 37 °C. After washing in 10 mM Tris–HCl, pH 7.4, for 10 min, the gel was stained in 0.2% toluidine blue, 10 mM Tris–HCl, pH 7.4, for 30 min, and destained in 10 mM Tris–HCl, pH 7.4, for 2 h. Transparent DNA-free bands indicative of enzyme activity were visualized through Toluidine Blue staining of surrounding background undigested DNA. This assay displays the ability of a protein to refold into an active nuclease. It confirms the nuclease activity on single-stranded high molecular weight DNA.

RF-I nicking assay. 0.5 μg of pUC19 plasmid RF-I (Promega, USA) was incubated with the designated amounts of M1 for 30 min at 37 °C in 20 μl of Buffer A (30 mM sodium acetate, pH 5.0, 10 mM NaCl, and 10 μM $ZnCl_2$). To stop the reaction, 6.25 μl of stop solution (50 mM Tris–HCl, pH 6.8, 3% SDS, 4.5% β-mercaptoethanol, 30% glycerol, and 0.001% Bromophenol Blue) was added. Ten microliter of the final mixture was loaded onto a 1% agarose gel in 1 × TAE buffer. After electrophoresis at 6 V/cm for 1 h, the gel was stained with ethidium bromide and the DNA was visualized on a UV transilluminator.

Single-strand DNA solubilization assay. The nuclease DNA solubilization assay was similar to that previously described [21]. Unless otherwise specified, 50 mg of heat-denatured purified calf thymus DNA (Invitrogen #15633-019; results obtained using Calbiochem # 2618 purified by proteinase K treatment and repeated phenol extraction were similar) was incubated with 3.5 U of M1 nuclease for 10 min in buffer A. One hundred milliliter of cold 20 mM $LaCl_3$ in 0.2 N HCl was added to stop the reaction. After centrifugation at 21,000*g* for 40 min, the absorbance of the supernatant was measured at 260 nm to determine the amount of DNA that had become acid-soluble. One Unit of M1 activity is defined, according to the traditional S1 nuclease Unit (http://www1.amershambiosciences.com), as the amount of enzyme that produces 1 μg of acid-soluble material in "S1 Unit Definition Buffer" (30 mM sodium acetate, pH 4.6, 100 mM NaCl, and 1 mM $ZnCl_2$) in 1 min at 37 °C when purified sheared single-stranded calf thymus DNA is used as substrate.

Single-strand RNA solubilization assay. Fifty microgram of purified Torula Yeast RNA (Ambion) was incubated with 1 U of M1 in 100 μl of Buffer A at 37 °C. At the designated times, 13 μl of cold 3 M sodium acetate, pH 5.2, and 282 μl of ethanol were added. The mixture was incubated at −20 °C overnight. After centrifugation at 21,000*g* for 45 min to precipitate the RNA, the absorbance of the supernatant was measured at 260 nm to quantify the amount of soluble nucleotides.

Results

ClustalW alignment of M1 protein sequence with homologous sequences

Table 1 shows a ClustalW alignment of the bacterial S1 homologs with the well-characterized fungal enzymes S1 and P1, the plant homologs CEL I of celery [12] and BEN1 of barley, and the P4 nuclease of the eukaryotic pathogen *Leishmania amazonensis*. Fig. 1 illustrates the active site of a crystal structure of P1 [21,22] with conserved active site amino acid residues shown in ball and stick representation. It is apparent that sequence conservation is restricted to a few amino acids in the enzyme active site.

Purification of M1 and M1His6 nucleases and analysis by activity gel

An SDS–PAGE of the time-course of a protein induction experiment is presented in Fig. 2C to show the accumulation of a ~29 kDa native M1 protein in the *M. loti* periplasm extract after the induction. A Coomassie stained PAGE gel, and a single-strand nuclease activity gel that illustrates the activity of the M1 and M1His6 nucleases during their expression and purification, are shown in Figs. 2A and B. The M1 band from both crude extracts and purified samples was excised from the Coomassie stained gel and analyzed by MALDI-TOF mass spectrometry after tryptic digestion to confirm protein identity. The purified native M1 sample was more than 90% pure. GroEL chaperonin 4 of *M. loti* was identified by mass spectrometry as the major contaminant. The enzyme purification table for M1 and M1His6 is shown in Table 2. In the Superdex S200 step, the M1 protein formed two distinct peaks, a minor one corresponding to monomer or dimer and the major peak corresponding to aggregates (estimated MW more than 500,000 Da). The major peak fractions (aggregates) were pooled and used in

Table 1
A ClustalW alignment of the M1 sequence with homologous sequences

```
P1      WGALGHATVAYVAQHYVSPEAASWAQGILGS-------SSSSYLASIASWADEYRLTSAG------ 53
M1      WGPEGHSIVAEIAQRRLSSTALMEVKRILGG---------EVAMASVASWADDVRYAIH------- 50
M2      WGQEGHAAVAEIAQHRLTSSASDVVQRLLRAHLGL-TGQQVVSMASIASWADDYRADGH------- 58
X1      WGPQGHRLVARVAETELTPQARAQVSQLLAG-------EADPSLAGVATWADELRANDPD------ 53
X2      WGPQGHRLVARIAETELSTQARTQVAQLLAG-------EPDPTLHGVATWADELREHDPD------ 53
Ch      WGQEGHRITGYIAQQLLSSKAKAEVKKLIP----------NADFAQLALYMDQHKQELKQ------ 50
BEN1    WGKEGHYMTCKIADGFLTSEASAAVKDLLPS-------WANGELAEVCAWAD--RQRFRY------ 51
CEL I   WSKEGHVMTCQIAQDLLEPEAAHAVKMLLPD-------YANGNLSSLCVWPDQIRHWYKY------ 53
S1      WGNLGHETVAYIAQSFVASSTESFCQNILGD-------DSTSYLANVATWADTYKYTDAG------ 53
P4      WGCVGHMLLAEIARRQLDLENEEKIELMAAVFSGSGPFPMSPSMVQAACWADDVKLWRQY------ 60
        *.  **     :*   :          :               :   . : *  :

P1      --KWSASLHFIDAEDNP------PTNCNVDYERDCGSSG-----CSISAIANYTQRVSDSSL---- 102
M1      --PESYNWHFVDIPLA---------DSKYDPVSQCAANVQGD--CAIAEIDRAEHEITCAT----- 98
M2      --KDTSNWHFVDIPLASLPGGS-SATTDYDAIRDCADDATYGS-CLLKALPAQEAILSDATK---- 116
X1      LGKRSGPWHYVNLGEHDCGYVP---------PRDCPD--GN---CVIAALEQQTAVLADRNQ---- 101
X2      LGKRSGPWHYVNLGEHDCAYSP---------PRDCPD--GN---CVIAALDQQAALLADRTQ---- 101
Ch      TLPGSDQWHYN—DEPVCSGVT----------EDECPD--GN---CAANQIDRYRKVLADRGA---- 96
BEN1    --RWSSPLHFADTPG----------DCNFSYARDCHDTKGNKDVCVVGAINNYTAALEDPS----- 100
CEL I   --RWTSSLHFIDTPDQ---------ACSFDYQRDCHDPHGGKDMCVAGAIQNFTSQLGHFRHGTSD 108
S1      --EFSKPYHFIDAQDNP------PQSCGVDYDRDCGSAG-----CSISAIQNYTNILLESPN---- 102
P4      ---AMSTWHFYAMPYNPGNINITDPVNTVNAVTVCLDMV-----TS-----------LKNSKA---- 104
              *:                          *          :

P1      SSENHAEALRFLVHFIGDMTQPLHD-EAYAV-------GGNKINVTFDGYH---------------- 145
M1      DPLQRRDSLRYLIHIVGDLHQPFHTVADNT--------GENALAVTVKFGGLIKSPPKTPAD---- 152
M2      DDESRWKALAFVIHLTGDLAQPLHCVQRVDGSQKDQ--GGNTLTVTFNVTRPAPDNSTFR------ 174
X1      PLAARRQALKFVVHFVGDIHQPMHAGYAHDK-------GGNDFQLQVAGKG---------------- 145
X2      PLDVRRQALKFVVHFVGDIHQPMHAGYAHDK-------GGNDFQLQIDGKG---------------- 145
Ch      AKADRAQALTFLIHMVGDIHQPLHAADNLDR-------GGNDFKVQLPGSSKI-------------- 142
BEN1    SPYDPTESLMFLAHFVGDVHQPLHCGHVDDL-------GGNTIKLRWYRRK---------------- 144
CEL I   RRYNMTEALLFLSHFMGDIHQPMHVGFTSDM-------GGNSIDLRWFRHK---------------- 152
S1      GSE-ALNALKFVVHIIGDTHQPLHD-ENLEA-------GGNGIDVTYDGET---------------- 144
P4      PLYLLNFAWVNLVHIFGDLHQPLHTISRYTTAYPHGDQGGNAISVRVGGKK---------------- 155
              :   : *: **  **:*               *.* . :

P1      --DNLHSDWDTYMPQ------KLIGGHALSDAESWAKTLVQNIESGNYTAQAIGWIKGDN------ 197
M1      ---NLHAVWDSTIIK------QTTYAWGSYVDRLETDWLLKHPEASETLD---------------- 193
M2      DFTTFHSVWDTDLIT------FKYYDWG--LAAAEAEKLLPTLAADLLADDT-------------- 218
X1      --SNLHALWDSGMLN------DRHLSDDAYLKVLLALP-ATAPTSPALPPP--------------- 187
X2      --SNLHALWDSGMLN------DRHLSDDAYLQRLLALP-AATAVSAALPPP--------------- 187
Ch      --SNLHSVWDTALVQ------QELNGADEKSWAAADLQRYQRNVSGWQGGG--------------- 185
BEN1    --SNLHHVWDSDVIT------QAMKDFFNRDQDTMIEAIQRNITD-DWSSEEKQWEACGS------ 195
CEL I   --SNLHHVWDREIIL------TAAADYHGKDMHSLLQDIQRNFTEGSWLQDVESWKECDD------ 204
S1      --TNLHHIWDTNMPE------EAAGGYSLSVAKTYADLLTERIKTGTYSSKKDSWTDGID------ 196
P4      --VKLHALWDNICSATPPRYQRPLSHTDLFALSATADGLVETYTFSEALETLVD------------ 207
            :*  **

P1      ISEPITTATRWASDANALVCTVVMPHGAAALQ------TGDLYPTYYDSVIDTIELQIAK------ 251
M1      -------PVAWALEAHTLAQEMAAGITNG----------ANLDNDYYAKALPVVDEQLGR------ 236
M2      -------PEKWLAECHRQAEAAYQALPAGTPLKSDIGHPVILDQAYFEKFHPVVTQQLAL------ 271
X1      -------GAAWAQASCRIAITPGVYPDT-----------HVLPPTYIATYRPIAETQLRV------ 229
X2      -------AAAWAQASCKIAVTPRVYPSA-----------HVLPATYIATYRPIAETQLRI------ 229
Ch      -------VMDWVHESNQYARADVYGPLAGFSCGASPSTPVYLDNTYLRAGGLLVDQQLAK------ 238
BEN1    RTK-ITCAEKYAKESALLACD-AYEGVEQGD---------TLGDDYYFRALPVVEKRIAQ------ 244
CEL I   IS---TCANKYAKESIKLACNWGYKDVESGE---------TLSDKYFNTRMPIVMKRIAQ------ 252
S1      IKDPVSTSMIWAADANTYVCSTVLDDGLAYIN------STDLSGEYYDKSQPVFEELIAK------ 250
P4      -------VMAIHEESYMFAVNTSYPGVTPGG---------TLSRAYLDKCKRVAEARLTL------ 251
                     .   .                      *   *           :

P1      GGYRLANWINEIHGSEIAK---------------- 270
M1      AGLRLAAVLNRWLATAPACSLP------------- 258
M2      GGLHLAAELNEALKGGK------------------- 288
X1      AGDRLAALLNAALRSP------------------- 245
X2      AGDRLAAILNAALASP------------------- 245
Ch      AGARIAAVINQALN--------------------- 252
BEN1    GGVRLAVILNQIFSGKNSRLQSI------------ 267
CEL I   GGIRLSMILNRVLGSSADHSLA------------ 274
S1      AGYRLAAWLDLIASQPS------------------ 267
P4      GGYRLGYLLNQLLSGITVDKAALEAHRAARPKWSA 286
        .* ::.  ::
```

Amino acid numbering is given with respect to the primary structure of mature P1. GenBank Accession Nos. are given in parentheses. P1, P1 nuclease of *Penicillium citrinum* (P24289); M1, gene product of *mlr6309* of *M. loti* strain MAFF303099 (BAB52626); M2, gene product of *orfE* of *M. loti* stain ICMP3153 (AAC24514); X1, gene product of XCC3198 of *Xanthomonas campestris* pv. campestris strain ATCC 33913 (AAM42468); X2, gene product of XAC3356 of *Xanthomonas axonopodis* pv. citri strain 306 (AAM38199); Ch, gene product of CV0060 of *Chromobacterium violaceum* strain ATCC 12472 (AAQ57739); BEN1, BEN1 nuclease of barley (BAA82696); CEL I, CEL I nuclease of celery (AAF42954); S1, S1 nuclease of *Aspergillus oryzae* (AAB20216); P4, P4 nuclease of *Leishmania amazonensis* (AAO65599). Symbols: ∗, identity; :, strong similarity; ., weak similarity.

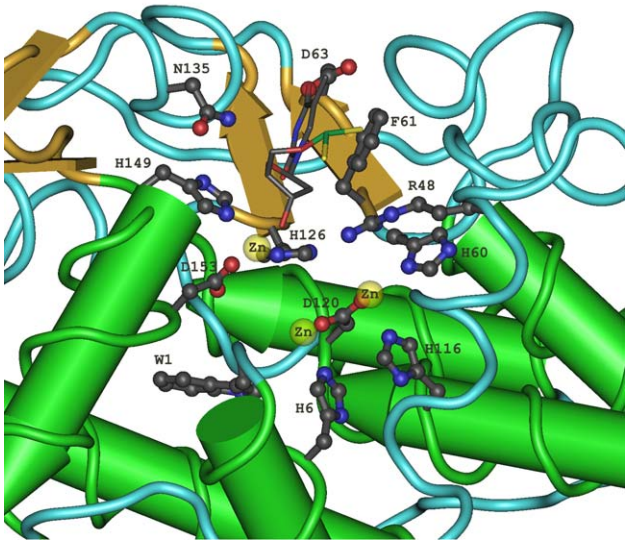

Fig. 1. Crystal structure of the active site of P1 nuclease in complex with a substrate analog [20], Protein Data Bank code 1AK0. The substrate analog is shown as CPK-colored tubes, only the nucleotide located 5′ of the incised phosphodiester bond is shown for convenience. Amino acid numbering is given with respect to the primary structure of mature P1.

Table 2
Purification of M1 nuclease from the periplasm of *M. loti*

	Volume (mL)	Total protein (μg)	Total activity (U[a])	Specific activity (U/μg)
Native M1				
Crude	400.00	ND	ND	ND
Crude, YM-10	15.00	4080.0	29,000	7.1
MonoQ	1.24	167.4	9,424	56.3
Superdex 200	0.35	110.0	2,730	24.8
M1His6				
Crude[b]	50.00	1100.0	14,850	13.5
Metal affinity	2.30	195.5	3,799	19.4
MonoQ	0.80	60.9	468	7.7

We observed a decrease in the specific activity of M1 at the final stages of purification. Reversible protein aggregation in the final preparations of concentrated M1 is a likely cause of this decline in the specific activity. ND, not determined.

[a] Enzyme activity was measured in conventional S1 Units.

[b] Additional step of extraction of the spheroplasts with hypotonic buffer (50 mM Tris–HCl, pH 7.5) improved protein yields. Consequently, in the M1His6 protein purification scheme "crude extract" refers to pooled extracts after the additional hypotonic wash.

biochemical assays as purified M1. When the latter was diluted and applied back to the Superdex S200 column, the estimated size of M1 aggregates decreased and a larger fraction of the protein was found in the monomer or dimer region, suggesting that the aggregation is reversible and M1 may be a monomer or dimer under assay dilutions (data not shown). Indeed, the enzyme shows no qualitative changes in enzymatic activity upon dilution or aggregation and therefore does not appear to be inactivated by aggregation during enzyme purification.

Identification of recombinant protein

The N-terminal peptide was determined by Edman degradation sequencing to be: (S/G)GPEGH(S/A)I(A/V)(A/V).

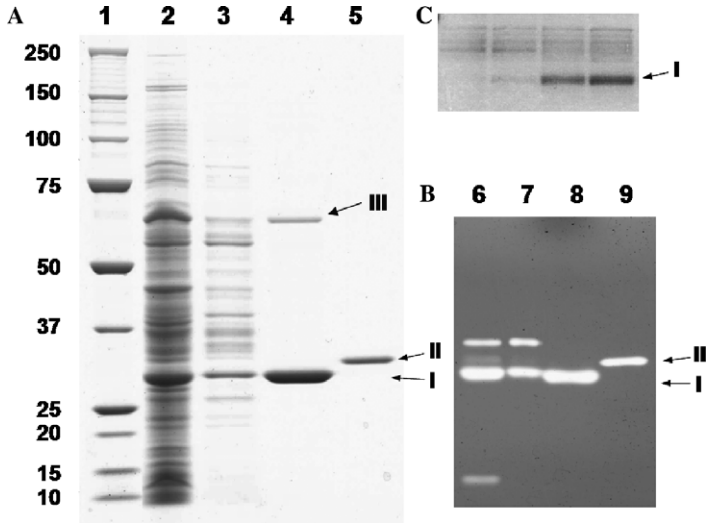

Fig. 2. Purification of M1 and M1His6 nucleases. (A) A SDS–PAGE gel stained with GelCode reagent (Pierce). (B) Detection of single-strand DNase activities in M1 samples after in-gel refolding. It can be seen that, unlike the total cell extract, the periplasm extract contains one major contaminating single-strand specific DNase which is removed during subsequent purification. Lane 1, molecular mass standards shown in kDa on the side. Lanes 2 and 6, total cellular protein of *M. loti* over-expressing M1 nuclease; lanes 3 and 7, crude periplasm extract of *M. loti* over-expressing M1 nuclease. Lanes 4 and 8, about 1 μg of M1 enzyme (band I) after purification on Mono Q column and Superdex 200 gel filtration. Band I was used for the identification of M1 nuclease by mass spectrometry and N-terminal Edman degradation sequencing. These samples of M1 are more than 90% pure and contain *M. loti* GroEL chaperonin 4 as the major contaminant (band III); lanes 5 and 9 show the mobility change after hexahistidine tag was fused to the COOH terminus of M1 nuclease to produce M1His6 (band II). The sample is homogeneous M1His6 after metal affinity purification followed by Mono Q. (C) SDS–PAGE followed by staining with GelCode (Pierce) illustrating time-course of induction of M1 in *M. loti* periplasm. The lanes are, from the left, periplasm extracts of *M. loti* pNOF-M1-Kan culture before induction with IPTG, 1, 16, and 24 h after induction, respectively.

De novo sequencing of the N-terminal peptide was performed both in the LC/MS/MS mode and the MALDI-TOF mode on the Q-Star® mass spectrometer and yielded: WGPEGHSIVAEIAQR. The amino acid coverage of the entire M1 nuclease by tryptic peptide mass mapping was 68%. Both amino and carboxy terminal peptides were identified.

NaCl and $MgCl_2$ concentration optima of M1 nuclease

The activity of M1 nuclease digestion of single-stranded DNA is maximal at NaCl concentrations below 50 mM. The enzyme is about 20% inhibited at 100 mM NaCl. This property is common among S1 homologs [23,2] which have evolved to work optimally at the conditions of low ionic strength where short double-stranded oligonucleotides generated by previous stages of high molecular weight DNA degradation would separate into single-strands. M1 nuclease is neither inhibited nor activated at Mg^{2+} concentrations less than 10 mM and is only 25% inhibited at 50 mM $MgCl_2$ (data not shown).

Single-strand DNA versus RNA and double-strand DNA specificities of M1 nuclease

M1 nuclease is about 25 times more active on single-stranded DNA than on double-stranded DNA as measured in the DNA solubilization assay (Fig. 3A). Furthermore, at pH 5, it is more active on single-stranded DNA than on single-stranded RNA (Fig. 3B).

M1 nuclease nicking of RF-I and cutting RF-II across the DNA nick

The M1 nuclease is similar to S1 and mung bean nuclease in its ability to nick supercoiled plasmid RF-I DNA, presumably at destabilized AT-rich regions [7]. It can be seen in the experiment shown in Fig. 4 that the RF-II (nicked covalently closed circle) from the first DNA nick is rapidly converted to RF-III (linear form) before other RF-I molecules are nicked. The successive linearization

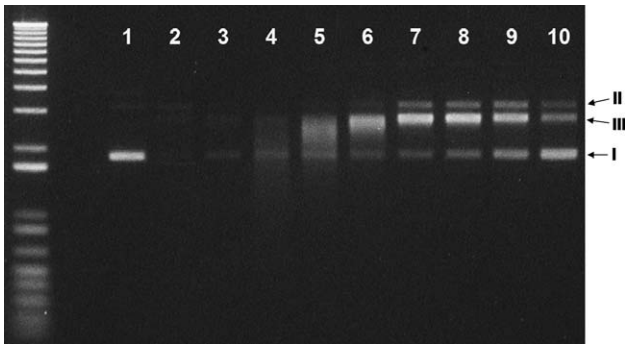

Fig. 4. Nuclease properties of M1, cutting supercoiled DNA and nicked-circular DNA across the nick. Lane 1, no enzyme; lanes 2–10, 160, 16, 3.2, 1.6, 0.8, 0.32, 0.16, 0.08, and 0.032 ng of M1 nuclease, respectively. (I) RF-I supercoiled plasmid DNA; (II) RF-II nicked-circular plasmid DNA; (III) RF-III linearized plasmid DNA. Molecular weight marker: 1 kb Plus DNA Ladder (Invitrogen).

suggests that the second nick is very close to the first nick on the opposite strand. The M1 nuclease plasmid nicking activity was completely inhibited by the presence of 0.24 mM of the Zn^{2+} chelator *o*-phenanthroline (data not shown), suggesting that Zn^{2+} is required for activity.

M1 is an RNase at neutral pH

S1 nuclease and mung bean nuclease are traditionally perceived as single-strand specific DNases that are active at acidic pH. However, their RNase activities are largely unexplored. We found that M1 is mainly an RNase in a physiological pH range (Fig. 3C). For example, it is about 10 times more active as an RNase than DNase at pH 7.5. The optimum of RNase activity of M1 appears at pH 6–6.5 as opposed to the single-strand DNase activity optimum of pH 5. However, the largest RNase/DNase activity ratio is observed at pH values above 7.5. For instance, at pH 8.2 the amount of M1 solubilized RNA was 14 times greater than that of DNA. Surprisingly, the same pH dependence was observed for the activities of S1 nuclease. For instance, at pH 6.5, S1 was at least 3.5 times more active as RNase than DNase (data not shown).

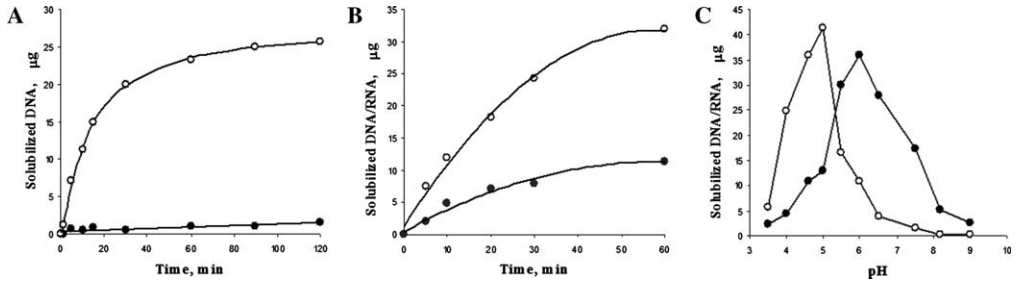

Fig. 3. M1 nuclease specificities. (A) Single-strand versus double-strand activity: 3.9 U of M1 nuclease was incubated in buffer A, at pH 5.0, with 25 µg of heat-denatured purified calf thymus DNA (hollow circles) or 25 µg of purified phage λ DNA fragmented with *Hin*dIII restriction enzyme (solid circles). (B) DNase versus RNase activity. One Unit of M1 nuclease was incubated in buffer A with 50 µg of heat-denatured purified calf thymus DNA (hollow circles) or 50 µg of purified Torula yeast RNA (solid circles). (C) M1 is an RNase at neutral pH. Fifty microgram of Torula Yeast RNA (solid circles) or heat-denatured calf thymus DNA (hollow circles) was incubated with 15 U of M1 nuclease in the presence of 10 mM NaCl and 10 µM $ZnCl_2$ at varying pH.

Discussion

M1 nuclease

We have shown that the gene product of the *mlr6309* gene of *M. loti* is an active nuclease with properties similar to fungal nuclease S1 and mung bean nuclease. The recombinant enzyme was purified from the periplasmic space of *M. loti* where upon overproduction it accounted for more than 10% of the total protein.

It should be noted that the NH_2 terminal tryptophan is highly conserved within the family. Trp1 is submerged deep in the hydrophobic core in the P1 crystal structure [21,22], suggesting that the NH_2 terminus authenticity is a critical factor for folding and stability of these proteins (Fig. 1). Processing of a signal peptide appears to be an effective way of providing a bacterial protein with a bulky amino terminal residue such as tryptophan. We infer from the processing of the leader sequence in the active recombinant M1 that the native endogenous enzyme should also be driven into *M. loti* periplasm and have an amino terminal tryptophan residue. The stability and solubility of M1 in the absence of the extensive mannosyl-glycosylation found in S1 nuclease homologs of eukaryotic origin is a useful property.

The newly established *M. loti* expression system is potentially useful for the study of other symbiosis genes that may not be easily expressed in *E. coli* and for the expression of other S1 nuclease homologs in bacteria.

Implications of the sequence conservation

Although the overall sequence conservation is low (primary structures of mature P1 and M1 proteins share only 24% sequence identity), all nine amino acids identified in the crystal structure of the mature P1 protein (Trp1, His6, Asp45, His60, His116, Asp120, His126, His149, and Asp153) [21,22] as essential for the binding of the three Zn^{2+} atoms and for catalysis, as well as one disulfide bridge between Cys80 and Cys85 (except in P4 nuclease), were conserved in M1 (Fig. 1). Three residues (Phe61, Asp63, and Arg48) which were involved in DNA binding in P1 are conserved in M1 but are less conserved among other homologs. In addition to these conserved functional residues, Ala13 and four glycine residues are conserved. The P1 crystal structure indicates that they allow the polypeptide chain to change directions within the limited space of the P1 protein core. The hydrophobic side chains of conserved Pro124 and Trp152 are anchored in the hydrophobic regions of the protein so that the conserved functional amino acids next to them are correctly positioned. Most other residues of M1 have diverged from the P1 sequence. The demonstration that the bacterial homolog M1 is indeed a functional S1-like nuclease supports the validity of this alignment. The roles of three conserved residues Asn135, Leu233, and Tyr237 were not previously explained. The latter two residues appear to orient an external polypeptide loop in the P1 structure. Asn135 may assist His149 in its bonding with Zn^{2+}. Taken together, the highly structure- and function-specific amino acid conservations among distant members of S1 superfamily strongly suggest that the S1-like ORFs in various organisms, including bacteria, encode for catalytically active enzymes with overall structures similar to P1.

The RNase activity at neutral pH

Because of their importance in molecular biological applications, studies of these nucleases have mainly revolved around their DNase properties, their activity with RNA being comparatively neglected. However, the ability of S1 homologs to degrade RNA at physiological pH ranges further suggests potential roles of these nucleases in RNA metabolism. A previous suggestion that the tertiary structure of P1 nuclease is unstable at neutral pH due to internal uncompensated carboxylate residue pairs [21] is not applicable to M1 or S1 which seem to be stable and active at this pH. Moreover, Glu128 and Asp66 that are proposed to cause charge repulsion to destabilize the S1 and P1 structures unless they are protonated [21,22] are not conserved in M1. Thus, a more complex explanation is needed for the different pH optima of M1 and S1 DNase versus RNase activities. Future structural studies may be able to identify the factors influencing the pH optima, such as substrate binding affinities, metal ion usage, substrate structures, or binding or catalytic residue efficiencies. It is noteworthy that P1 binds DNA by stacking the base immediately 5′ of the phosphodiester bond to be hydrolyzed in a hydrophobic pocket without any direct interaction of the enzyme with the deoxyribose moiety [21]. Conserved amino acid residues not implicated in the DNase activity, such as Asn135, may be involved in the interaction with the ribose moiety. This hypothesis could be tested with site-directed mutagenesis and substrate specificity studies.

Biological roles

It is striking that all organisms, other than plants and fungi, that are known to encode S1 nuclease homologs are either pathogens or symbionts. Among them are several species of the widely disseminated human eukaryotic parasites *Leishmania* and *Plasmodium*. Examples of organisms encoding S1 homologs in the bacteria kingdom include a citric plant pathogen *Xanthomonas axonopodis*, human pathogens *Leptospira interrogans*, which is the cause of leptospirosis, and *Chromobacterium violaceum*, which causes severe purulent infections in mammals, including humans. It is noteworthy that P4 nuclease of *Leishmania* is apparently essential for the survival of amastigotes, the *Leishmania* form within the mammalian host [9]. S1 homologs are encoded by symbiotic bacteria *Mesorhizobium* and *Nostoc*. It has been noticed that symbionts and pathogens share many genetic homologies for their interaction with the eukaryotic hosts. In fact, rhizobia are similar in many

ways to certain bacterial intracellular pathogens that reside within a vacuolar compartment (i.e., brucella) ([24,25] http://jura.ebi.ac.uk:8765/ext-genequiz/genomes/mel0103/). Thus, it appears likely that M1 as well as other microbial S1 homologs will be assigned a function within the microbe-host interaction framework.

Acknowledgments

This work was supported, in part, by Grants NIH CA71426 and US. Army DMAD17-97-1-7286 to A.T.Y, by institutional grants from the National Institute of Health to Fox Chase Cancer Center (CA06927, RR05539), and an appropriation from the Commonwealth of Pennsylvania. We thank Drs. C. Ronson and J. Sullivan for the gift of *Mesorhizobium loti* clones. We thank Drs. G.D. Markham, E. Jaffe, and A. Bellacosa for critical reading of this manuscript, Dr. S. Seeholzer of the Fannie E. Rippel Biochemistry and Biotechnology facility and Dr. C. Krishnan of Applied Biosystems for assistance in the mass spectrometry protein identification. We thank Dr. J. Fernandez of the Protein/DNA Technology Center of the Rockefeller University for N-terminal micro-sequencing.

References

[1] M.J. Frazer, R.L. Low, Fungal and mitochondrial nucleases, in: S.M. Linn, R.S. Lloyd, R.J. Roberts (Eds.), Nucleases, Cold Spring Harbor Laboratory Press, 1993, pp. 171–207.
[2] V.M. Vogt, Purification and properties of S1 nuclease from *Aspergillus*, Methods Enzymol. 65 (1980) 248–255.
[3] R.C. Wiegand, G.N. Godson, C.M. Radding, Specificity of the S1 nuclease from *Aspergillus oryzae*, J. Biol. Chem. 250 (1975) 8848–8855.
[4] M. Fujimoto, A. Kuninaka, H. Yoshino, Substrate specificity of nuclease P1, Agric. Biol. Chem. 38 (1974) 1555–1561.
[5] K. Maekawa, S. Tsunasawa, G. Dibo, F. Sakiyama, Primary structure of nuclease P1 from *Penicillium citrinum*, Eur. J. Biochem. 200 (1991) 651–661.
[6] D. Kowalski, W.D. Kroeker, M. Laskowski Sr., Mung bean nuclease I. Physical, chemical, and catalytic properties, Biochemistry 15 (1976) 4457–4462.
[7] W.D. Kroeker, D. Kowalski, M. Laskowski Sr., Mung bean nuclease I. Terminally directed hydrolysis of native DNA, Biochemistry 15 (1976) 4463–4467.
[8] L.G. Sheflin, D. Kowalski, Mung bean nuclease cleavage of a dA + dT-rich sequence or an inverted repeat sequence in supercoiled PM2 DNA depends on ionic environment, Nucleic Acids Res. 12 (1984) 7087–7104.
[9] S. Kar, L. Soong, M. Colmenares, K. Goldsmith-Pestana, D. McMahon-Pratt, The immunologically protective P-4 antigen of *Leishmania* amastigotes. A developmentally regulated single strand-specific nuclease associated with the endoplasmic reticulum, J. Biol. Chem. 275 (2000) 37789–37797.
[10] J.R. Silber, L.A. Loeb, S1 nuclease does not cleave DNA at single-base mis-matches, Biochim. Biophys. Acta 656 (1981) 256–264.
[11] C.A. Oleykowski, C.R. Bronson Mullins, D.W. Chang, A.T. Yeung, Incision at nucleotide insertions/deletions and base pair mismatches by the SP nuclease of spinach, Biochemistry 38 (1999) 2200–2205.
[12] B. Yang, X. Wen, N.S. Kodali, C.A. Oleykowski, C.G. Miller, J. Kulinski, D. Besack, J.A. Yeung, D. Kowalski, A.T. Yeung, Purification, cloning, and characterization of the CEL I nuclease, Biochemistry 39 (2000) 3533–3541.
[13] B. Rost, Enzyme function less conserved than anticipated, J. Mol. Biol. 318 (2002) 595–608.
[14] T. Kaneko, Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, S. Tabata, Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*, DNA Res. 7 (2000) 331–338.
[15] J.T. Sullivan, C.W. Ronson, Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene, Proc. Natl. Acad. Sci. USA 95 (1998) 5145–5149.
[16] Z. Liu, Hetero-stagger cloning: efficient and rapid cloning of PCR products, Nucleic Acids Res. 24 (1996) 2458–2459.
[17] M. Bagdasarian, R. Lurz, B. Ruckert, F.C. Franklin, M.M. Bagdasarian, J. Frey, K.N. Timmis, Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*, Gene 16 (1981) 237–247.
[18] B. Garg, R.C. Dogra, P.K. Sharma, High-efficiency transformation of *Rhizobium leguminosarum* by electroporation, Appl. Environ. Microbiol. 65 (1999) 2802–2804.
[19] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding, Anal. Biochem. 72 (1976) 248–254.
[20] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[21] A. Volbeda, A. Lahm, F. Sakiyama, D. Suck, Crystal structure of *Penicillium citrinum* P1 nuclease at 2.8 A resolution, EMBO J. 10 (1991) 1607–1618.
[22] C. Romier, R. Dominguez, A. Lahm, O. Dahl, D. Suck, Recognition of single-strand DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs, Proteins 32 (1998) 414–424.
[23] S.C. Sung, M. Laskowski Sr., A nuclease from mung bean sprouts, J. Biol. Chem. 237 (1962) 506–511.
[24] B. Finlay, S. Falkow, Common themes in microbial pathogenesis revisited, Microbiol. Mol. Biol. Rev. 61 (1997) 136–169.
[25] R.A. Ugalde, Intracellular lifestyle of *Brucella* spp. Common genes with other animal pathogens, plant pathogens, and endosymbionts, Microbes Infect. 1 (1999) 1211–1219.