Extremophiles (2001) 5:333–341
Digital Object Identifier (DOI) 10.1007/s007920100198
© Springer-Verlag 2001

ORIGINAL PAPER

**Pieter van Solingen · Daan Meijer
Wilhelmus A. H. van der Kleij · Christopher Barnett
Robertus Bolle · Scott D. Power · Brian E. Jones**

Cloning and expression of an endocellulase gene from a novel streptomycete isolated from an East African soda lake

Received: January 10, 2001 / Accepted: April 8, 2001 / Published online: August 11, 2001

Abstract Alkaline cellulase-producing actinomycete strains were isolated from mud samples collected from East African soda lakes. The strains were identified as novel *Streptomyces* spp. by 16S rDNA sequence analysis. A cellulase gene (*cel12A*) from *Streptomyces* sp. strain 11AG8 was cloned by expression screening of a genomic DNA library in *Escherichia coli*. From the nucleotide sequence of a 1.5-kb DNA fragment, an open reading frame of 1,113 nucleotides was identified encoding a protein of 371 amino acids. From computer analysis of the sequence, it was deduced that the Cel12A mature enzyme is a protein of 340 amino acids. The protein contained a catalytic domain, a glycine-rich linker region, and a cellulose-binding domain of 221, 12, and 107 amino acids, respectively. FASTA analysis of the catalytic domain of Cel12A classified the enzyme as a family 12 endoglucanase and the cellulose-binding domain as a family IIa CBD. *Streptomyces rochei* EglS was determined as nearest neighbor with a similarity of 75.2% and 61.0% to the catalytic domain and the cellulose-binding domain, respectively. The *cel12A* gene was subcloned in a *Bacillus* high-expression vector carrying the *Bacillus amyloliquefaciens* amylase regulatory sequences, and the construct was transformed to a *Bacillus subtilis* host strain. Crude enzyme preparations were obtained by ultrafiltration of cultures of the *Bacillus subtilis* recombinant strain containing the 11AG8 *cel12A* gene. The enzyme showed carboxymethylcellulase (CMCase) activities over a broad pH range (5–10) with an optimum activity at pH 8 and 50°C. The enzyme retained more than 95% of its activity after incubation for 30 min under these conditions.

Key words *Streptomyces* · Extremophile · Endocellulase · Cloning · Alkaliphilic · pH profile

Communicated by W.D. Grant

P. van Solingen · D. Meijer · W.A.H. van der Kleij · R. Bolle · B.E. Jones (✉)
Genencor International B.V., Archimedesweg 30, 2333 CN Leiden, The Netherlands
Tel. +31-71568-6123; Fax +31-71568-6130
e-mail: bjones@genencor.com

C. Barnett · S.D. Power
Genencor International Inc., Palo Alto, CA, USA

Introduction

Cellulases are used in several industrial processes, including degrading wood pulp to sugars for ethanol production (Béguin and Aubert 1994), in textile finishing for achieving a worn appearance (for example, the abrasion of denim fabrics), and for modifications of fabric surfaces (Lange 1993) and use in household laundry (Maurer 1997). For use in laundry detergents, the cellulases must fulfill essential requirements such as high activity at pH 7–10, thermostability at 50°C, no calcium or magnesium dependency, and a lack of inhibition by chemical surfactants or chelating or structural agents in the detergent matrix. For textile applications, cellulases must perform high abrasion of denim material while eliciting low tensile strength loss of the fabric.

Representative fungal cellulases like IndiAge (Genencor, Rochester, NY, USA), a *Trichoderma reesei* cellulase, are potent in abrasion but their activities decrease in the pH range 7–10. One of the *Trichoderma* cellulases, EGIII, a family 12 β-1,4-glucanase [EC 3.2.1.4] (Henrissat and Bairoch 1993), is particularly useful in this application. As a single enzyme product, however, it suffers from a reduced performance in the neutral to alkaline pH range. The search for a more alkaline-acting EGIII has led us to screen alkaliphilic actinomycetes from novel and extreme environments. Neutral cellulases from actinomycetes, e.g., *Streptomyces lividans* CelB and *Streptomyces rochei* (A2) EglS, have been described as representatives of family 12 (EGIII) endoglucanases (Tomme et al. 1995a). We thus screened mud samples from East African soda lakes in search of cellulase-producing bacteria. The rationale behind this approach is that the harsh conditions in the textile applications show a certain resemblance to the severe conditions of the natural habitat of these microorganisms. We searched for alkaline cellulases from non-*Bacillus* pro-

karyotes, particularly actinomycetes, which are included in family 12, β-1,4-glucanases. The isolation, cloning, and some relevant characteristics of one such enzyme, Cel12A of *Streptomyces* sp. 11AG8, are described in this article.

Materials and methods

Organisms and growth conditions

Approximately 1 g of alkaline mud sample (in situ conditions: 34°C, pH 10.5, conductivity 12.6 mS cm^{-1}) obtained from Lake Nakuru (0°22′ S and 36°05′ E), an East African soda lake, was suspended in 5 ml of a solution containing 4% w/v NaCl and 1% w/v Na_2CO_3 and incubated for 1 h at 44°C with occasional gentle mixing. Serial dilutions were plated on 25% v/v Soil Extract (Staley et al. 1992) agar medium in 4% w/v NaCl, 1% w/v Na_2CO_3, pH 10, containing 50 mg ml^{-1} rifampicin. The plates were incubated at 20°C for several weeks. The colonies that appeared were examined by microscopy. Bacterial strains that showed compact growth and had a hyphal morphology were isolated as putative actinomycetes. The isolates were replicated to Alkaline Agar Medium, pH 10 (Duckworth et al. 1998) containing 0.3% w/v carboxymethylcellulose (Sigma, St. Louis, MO, USA) and incubated at 30°C for 3 days. The plates were developed with Congo red (Teather and Wood 1982), and colonies surrounded by a clear halo were selected as potential cellulase-producing strains. Several different isolates were obtained, each exhibiting the typical aerial mycelium of *Streptomyces* ("*Streptoverticillium*"). Isolate 11AG8 was selected for further studies.

Growth experiments with *Streptomyces* sp. 11AG8 were carried out in Brain Heart Infusion (BHI) agar (Oxoid, Basingstoke, England). Shake flask fermentations were carried out in Alkaline Broth (Duckworth et al. 1998) at 30°C and 250 rpm shaking for 72 h in an orbital shaker (New Brunswick Scientific, Edison, NJ, USA). *Bacillus subtilis* host and recombinant strains were grown in Tryptone Soya Broth (TSB), 3% w/v (Oxoid).

DNA procedures

Cells from liquid cultures were harvested by centrifugation. Chromosomal DNA was isolated as described by Saito and Miura (1963). The 16S rDNA was fully sequenced on both DNA strands using primers derived from the conserved regions for prokaryote 16S ribosomal RNA (Lane et al. 1985). The reactions were performed on a LiCor automated DNA sequencer using dye-primer chemistry and a ABI 310 Genetic Analyzer using dye-terminator chemistry, and the resulting contigs were assembled and corrected by hand.

Plasmid DNA was isolated using the QIAprep Spin Plasmid Kit protocol (Qiagen, Chatsworth, CA, USA). The plasmid vectors used in this study are given in Table 1. The fragments of recombinant DNA in the cloning vectors were sequenced using vector-derived primers and chromosome walking techniques (Shyamala and Ames 1989). Nucleotide sequence analysis was carried out by BaseClear (Leiden, The Netherlands). Nucleotide and amino acid sequences were analyzed by the computer programs Clone 4.1 (Scientific and Educational Software, Durham, NC, USA), DNASTAR (DNASTAR, Brighton, MA, USA), and GCG Sequence Analysis Software Package (Genetics Computer Group, Wisconsin, MA, USA). Comparison of the sequences was carried out with FASTA (Pearson and Lipman 1988). The sequences of the 16S rDNA and the *cel12A* gene of *Streptomyces* strain 11AG8 have been submitted to the GenBank database under accession numbers AF233375 and AF233376, respectively.

DNA molecular techniques were carried out according to standard protocols (Sambrook et al. 1989). *Escherichia coli* transformation was carried out by electroporation using Epicurian Coli XL1Blue competent cells and an Electroporator 1000 (Stratagene, La Jolla, CA, USA). Transformation of *Bacillus subtilis* DS12367, an amylase- and sporulation-negative mutant of DB104 (Kawamura and Doi 1984), with plasmid DNA was carried out as described by Vehmaanperä (1988).

Cloning of cellulase genes in *E. coli*

Chromosomal DNA of the alkaliphilic *Streptomyces* strain was partially digested with *Sau*3A. Fragments were size-fractionated by agarose gel electrophoresis and the 2- to 6-kb fraction was isolated from the gel using the QIAquick Gel Extraction Kit protocol (Qiagen). The *E. coli–Bacillus* shuttle vector pPNeoOri, a pUC19-derived plasmid (Yanish-Perron et al. 1985) carrying the origin of replication and the neomycin marker of pUB110 (Gryczan et al. 1978) for replication and selection in *Bacillus* was digested with *Sal*I. The protruding 5′-ends of both fragments and vector were partially filled using the Klenow fragment of *E. coli* polymerase I. The partially filled-in fragments and vector were ligated and transformed to *E. coli*. The cells were mixed with 4 ml top layer (0.3% w/v carboxymethylcellulose [CMC], 0.5% w/v agarose) and poured on selective Luria-Bertani (LB) plates, containing 100 μg ml^{-1} ampicillin. The plates were incubated at 37°C for 2 days. The plates were developed with Congo red (Teather and Wood 1982), and the clear halo-forming colonies were picked and rescued by streaking on to fresh selective plates. Single colonies were grown overnight at 37°C in LB medium, containing 100 μg ml^{-1} ampicillin, for plasmid DNA isolation and tested for carboxymethylcellulase (CMCase) activity by agar diffusion assay. The inserted DNA was characterized by restriction enzyme analysis and the nucleotide sequence was determined.

Subcloning in *Bacillus subtilis*

*Pst*I and *Sma*I cloning sites were linked to the mature 11AG8 *cel12A* gene by polymerase chain reaction (PCR) using *Streptomyces* sp. 11AG8 as template DNA and the

Table 1. Plasmids used in this study

Plasmid	Properties	Reference
pPNeoOri	AmpR, NeoR, pUB replicon	Gryczan et al. (1978)
pPNeoOri	pUC replicon	Yanish-Perron et al. (1985)
pHPLT	NeoR, pUB replicon	Gryczan et al. (1978)
pHPLT	*Bacillus licheniformis* promoter, signal sequence, terminator	van Ee (1991)
pHP11AG8	*Streptomyces* sp. 11AG8 cel12A in pHPLT	This study

following primers: 5′-CTTCATTCTGCAGCATCGGCGA ACCAGCAGATCTGCGACCGCTACGGC-3′ (forward) and 5′-TAAGACAGTGCCCGGGGGTGCCCCATTGT CACTCAC-3′ (reverse). PCR was carried out with *Pwo* DNA polymerase (Boehringer, Mannheim, Germany) according to the procedure recommended by the supplier. The 1.6-kb PCR fragment was digested with *Pst*I and *Sma*I and ligated into *Pst*I–*Hpa*I-digested *Bacillus* expression vector pHPLT (Table 1). The ligation mixture was transformed to *Bacillus subtilis* DS12367 using the method of Vehmaanperä (1988).

Determination of the cellulase activity

Cellulase transformants were grown in shake culture for 72 h. *E. coli* transformants were cultured in yeast extract (10 g l^{-1})–peptone (20 g l^{-1}) medium (sYEP) supplemented with isopropyl-β-D-thiogalactoside (0.1 mM), ampicillin (10 μg ml^{-1}), and lactose (2 g l^{-1}). *Bacillus* strains were cultivated in tryptone soya broth. Cells were separated by centrifugation, and enzyme assays were performed either directly on the supernatant or after ultrafiltration in a Filtron Minisette (Pall, Northborough, MA, USA) using a 10-kDa cutoff membrane.

A very sensitive, semiquantitative assay for cellulase production was performed by an agar diffusion method. A solution (20 ml) of agar (1.5% w/v) containing CMC (0.5% w/v, low viscosity; Sigma) in glycine-NaOH buffer (0.2 M, pH 10) was poured into 9-cm Petri dishes and allowed to solidify. Wells (5 mm diameter) were punched in the gel and filled with 50 μl supernatant or ultrafiltrate and incubated overnight. Halos revealing cellulase activity were detected with Congo red (Teather and Wood 1982).

Assays for CMCase activity were performed in 0.1 M phosphate-citrate buffer over the pH range 5–10. Enzyme preparations in 0.1 M buffer (0.1 ml) were mixed with 0.5 ml 1% w/v CMC solution and 0.5 ml 0.2 M buffer. After incubation for 30 min at 40°C, the mixture was heated in a boiling water bath for 15 min. The reducing sugar formed was quantified as glucose by the parahydroxybenzoic acid hydrazide procedure (Lever 1972). One CMCase unit was defined as the amount of enzyme required to liberate 1 μmol reducing sugar, measured as glucose, per minute. To determine the thermostability of the enzyme the CMCase assay was carried out after incubation of the crude enzyme preparations for 30 min at temperatures between 30° and 70°C with 10°C intervals. Protein was quantified by the method of Smith et al. (1985) using the BCA Protein Assay Kit (Pierce, Rockford, IL, USA) with bovine serum albumin as standard.

Gel electrophoresis

Proteins were separated by denaturating polyacrylamide gel electrophoresis (SDS-PAGE) using Mini-Protean II ready gels (Bio-Rad, Hercules, CA, USA) following the procedure described by the supplier. Cellulase activity in native gels was detected by an overlay technique using 0.3% w/v CMC, 0.5% w/v agarose, and the Congo red staining technique (Béguin 1983).

Phylogenetic analysis

GenBank was searched for related sequences using the algorithm FASTA (Pearson and Lipman 1988). 16S rRNA gene sequences were aligned using the MEGALIGN program (DNASTAR, Brighton, MA, USA) by the CLUSTAL method. Aligned sequences were edited using EDITSEQ (DNASTAR) and checked manually to give fully overlapping sequences of 1,462 bp. Evolutionary distances between pairs of DNA sequences were determined using the algorithm of Jukes and Cantor (1969). Phylogenetic trees were constructed from the evolutionary distances by the neighbor-joining method (Saitou and Nei 1987) with bootstrap analysis of 1,000 trees using the TREECON program package (Van de Peer and De Wachter 1994).

The following GenBank DNA sequences were used during the phylogenetic analysis: *Streptomyces vellosus* NRRL8037 (X99942), *Streptomyces thermoviolaceus* DSM 40443^T (Z68096), *Streptomyces thermodiastaticus* DSM 44061^T (Z68101), *Streptomyces thermocarboxydovorans* DSM 44296^T (U94489), *Streptomyces setonii* ATCC25497 (D63872), *Streptomyces sampsonii* ATCC25495 (D63871), *Streptomyces* ("*Streptoverticillium*") *mashuensis* DSM 40221 (X79323), *Streptomyces lividans* TK21 (Y00484), *Streptomyces galbus* DSM 40089 (X79852), *Streptomyces coelicolor* DSM 40233^T (Z76678), *Streptomyces caelestris* NRRL2418 (X80824), *Streptomyces bluensis* ISP5564 (X79324), and *Amycolatopsis orientalis* DSM 43187 (X81573).

Results

Phenotypic characterization of *Streptomyces* sp. strain 11AG8

When grown on soil extract medium at pH 10, the initial small, round, glistening transparent colony developed a white to gray-white aerial mycelium after a few days incubation. On alkaline agar medium, the strain formed a dry, leathery, cream-colored, opaque colony producing an aerial mycelium on maturity. Under the microscope, the strain exhibited extensively branched pseudomycelium that fragmented into irregular rods, isolated spores, and spores in chains. Growth was observed on BHI medium between pH 7.4 and pH 10.5. The strain did not grow at pH 6.7 or below and at pH 11.0 and above. The temperature range for growth on both BHI medium (pH 7.4) and alkaline agar medium (pH 10) appeared to be between 20° and 40°C with an optimum at about 30°C.

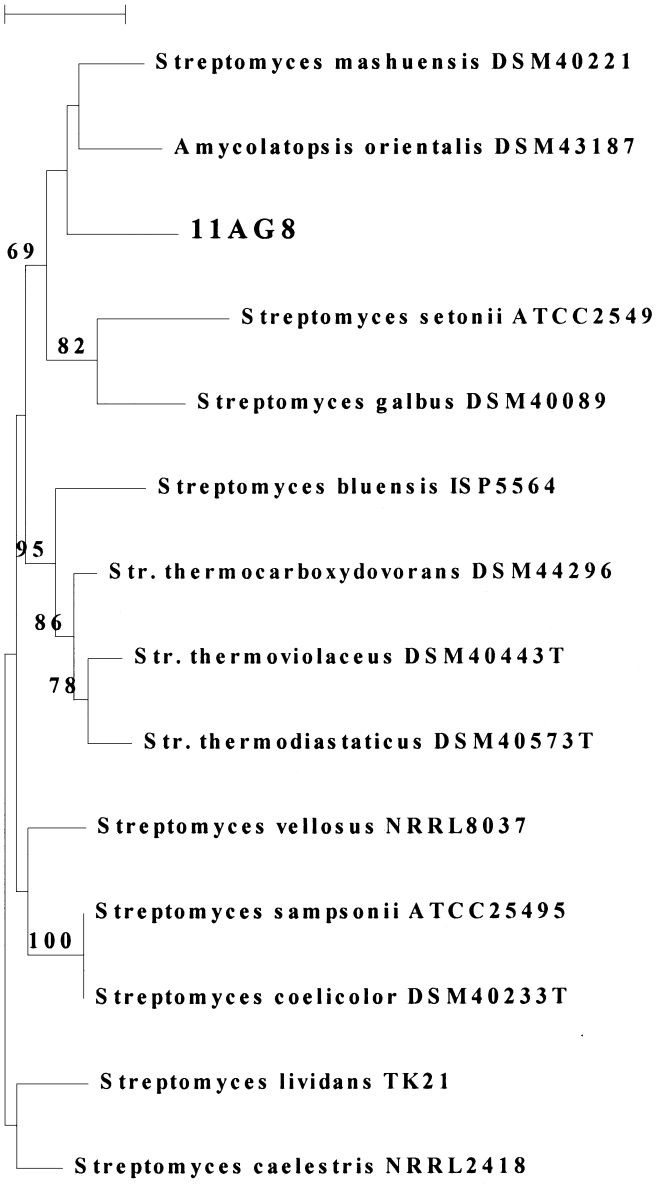

Fig. 1. Unrooted phylogenetic tree showing the relationship between strain 11AG8 within the genus *Streptomyces* and related taxa. The *numbers* at the nodes are bootstrap values (>65%) based on 1,000 resamplings of the data. *Bar* indicates 2 inferred substitutions per 100 nucleotides

Phylogenetic analysis of *Streptomyces* sp. strain 11AG8

FASTA analysis of the 1,468-bp 16S rDNA sequence of strain 11AG8 revealed homology to *Streptomyces* spp. The nearest neighbor of the strain was *Streptomyces thermoviolaceus* with an identity of 97.2% in a 1,460-bp overlap. The percentage of identity of the 16S rDNA, the morphological appearance, and the growth characteristics classified strain 11AG8 as an apparently novel species of *Streptomyces*. The relationship between *Streptomyces* sp. strain 11AG8 and related taxa is illustrated in Fig. 1.

Cloning of a cellulase gene by expression library screening

A library of genomic DNA of strain 11AG8 was cloned in an *E. coli–Bacillus* shuttle vector. One cellulolytic clone was isolated from 17,000 transformants. Plasmid DNA isolated from this clone carried a 3.5-kb insert. The nucleotide sequence analysis of the flanking regions of the insert was determined using vector-derived primers. The flanking region adjacent to the *lacZ* promoter of the vector showed similarity to the 5′-end of *Streptomyces* cellulases. Sequence analysis of the gene was completed using internal primers. The nucleotide sequence of the gene is given in Fig. 2.

The nucleotide sequence contains an open reading frame of 1,113 bp starting with an ATG and ending in the TGA that encodes a protein of 371 amino acids. The protein has a putative signal peptide that would allow signal peptidase processing after Ala-31 to give a mature protein of 35.8 kDa. A catalytic domain of 221 amino acids, a glycine-rich linker of 12 amino acids, and a cellulose-binding domain (CBD) of 107 amino acids can be detected. The catalytic domain of Cel12A of strain 11AG8 is similar to that of several *Streptomyces* spp. and fungi included in family 12 of the glycosyl hydrolases (Henrissat and Bairoch 1993). Likewise, by homology, the CBD can be assigned to family IIa (Tomme et al. 1995b) together with *Streptomyces lividans* CelB and *Streptomyces rochei* EglS. A review of the FASTA results is given in Table 2. The amino acid alignment of the catalytic domain and CBD of Cel12A and that of the various family 12 endoglucanases is shown in Fig. 3A,B.

Expression of the Cel12A gene in *E. coli* and *Bacillus subtilis*

Liquid cultures in sYEP medium of the *E. coli* recombinant with the 3.5-kb insert carrying the *cel12A* gene under

Fig. 2. Nucleotide sequence and deduced amino acid sequence of chromosomal DNA fragment of *Streptomyces* sp. 11AG8 carrying the *cel12A* gene. Typical prokaryotic transcription signals, –35, –10, and the Shine–Delgarno sequence are shown in *bold*. An inverted repeat, typical for actinomycete regulatory regions (Lao et al. 1991), is shown in *bold* and *underlined with a dashed line*. The stop codon is given by an *asterisk*; the *downward pointing arrow* indicates the cleavage site for the signal peptidase

```
CGTCTCGGCCCTCTCGGATGGGGCTGGCTGGAGCGCAGGACGCGCGGCCGGGCGGTCTCG 60

GACCAGGCAGTGGAGGAGATCCGGGCCCTGATCCACTACCCGGAGCTGCGTCCCGCTCAG 120

GAGGCGGACCCGGCCCAGCAGGGCATGCCGCTGGGCCCGGCCGCCGTGGCGGCCGGTGTC 180

CTGTGGGGTGCGGCGGTGCTGTTCCTCCTGTGAGGCGGCCCCCGAGCGGGACGCGGGCCG 240

GGGCACGGTGAAGCACGGATGCCGCCGCCCGTCTTGTCATGGGCCGGGTATCGCCCTACA 300
                                    -35                  -10
GTCCGGCACGACAGTGCATGGGAGCGCTCCCAGAGTCCGCACCGCTCCTGCTGAAGGGAC 360
                 <-----||----->             Shine-Dalgarno
ACCATGAGATCCCATCCCCGCTCCGCGACGATGACCGTCCTCGTCGTCCTGGCCTCGCTC 420
   M  R  S  H  P  R  S  A  T  M  T  V  L  V  V  L  A  S  L
                                       ⇓
GGCGCGCTGCTCACCGCAGCGGCTCCCGCCCAGGCGAACCAGCAGATCTGCGACCGCTAC 480
 G  A  L  L  T  A  A  A  P  A  Q  A  N  Q  Q  I  C  D  R  Y

GGCACCACCACGATCCAGGACCGGTACGTGGTGCAGAACAACCGCTGGGGCACCAGCGCC 540
 G  T  T  T  I  Q  D  R  Y  V  V  Q  N  N  R  W  G  T  S  A

ACCCAGTGCATCAATGTGACCGGCAACGGTTTCGAGATCACCCAGGCCGACGGTTCGGTG 600
 T  Q  C  I  N  V  T  G  N  G  F  E  I  T  Q  A  D  G  S  V

CCGACCAACGGCGCCCCGAAGTCCTATCCCTCGGTCTACGACGGCTGCCACTACGGCAAC 660
 P  T  N  G  A  P  K  S  Y  P  S  V  Y  D  G  C  H  Y  G  N

TGCGCGCCCCGCACGACGCTGCCCATGCGGATCAGCTCGATCGGCAGCGCGCCCAGCAGT 720
 C  A  P  R  T  T  L  P  M  R  I  S  S  I  G  S  A  P  S  S

GTCTCCTACCGCTACACCGGCAACGGCGTCTACAACGCCGCGTACGACATCTGGCTGGAC 780
 V  S  Y  R  Y  T  G  N  G  V  Y  N  A  A  Y  D  I  W  L  D

CCGACACCCCGCACCAACGGGGTGAACCGGACCGAGATCATGATCTGGTTCAACCGGGTC 840
 P  T  P  R  T  N  G  V  N  R  T  E  I  M  I  W  F  N  R  V

GGCCCGGTCCAGCCCATCGGTTCGCCGGTCGGCACGGCCCACGTCGGCGGCCGCAGCTGG 900
 G  P  V  Q  P  I  G  S  P  V  G  T  A  H  V  G  G  R  S  W

GAGGTGTGGACCGGCAGCAACGGTTCGAACGACGTGATCTCCTTCCTGGCGCCCTCCGCG 960
 E  V  W  T  G  S  N  G  S  N  D  V  I  S  F  L  A  P  S  A

ATCAGCAGCTGGAGCTTCGACGTCAAGGACTTCGTCGACCAGGCCGTCAGCCACGGCCTG 1020
 I  S  S  W  S  F  D  V  K  D  F  V  D  Q  A  V  S  H  G  L

GCCACCCCGGACTGGTACCTCACCAGCATCCAGGCGGGCTTCGAGCCGTGGGAGGGCGGC 1080
 A  T  P  D  W  Y  L  T  S  I  Q  A  G  F  E  P  W  E  G  G

ACCGGTCTGGCCGTGAACTCGTTCTCCTCCGCGGTGAACGCCGGGGGCGGGAACGGCGGC 1140
 T  G  L  A  V  N  S  F  S  S  A  V  N  A  G  G  G  N  G  G

ACTCCGGGGACACCGGCGGCCTGCCAGGTCTCCTACAGCACCCACACCTGGCCCGGCGGC 1200
 T  P  G  T  P  A  A  C  Q  V  S  Y  S  T  H  T  W  P  G  G

TTCACCGTCGACACCACCATCACCAATACCGGCTCCACACCCGTCGACGGCTGGGAACTG 1260
 F  T  V  D  T  T  I  T  N  T  G  S  T  P  V  D  G  W  E  L

GACTTCACCCTCCCCGCCGGTCACACGGTCACCAGCGTGTGGAACGCGCTGATCAGCCCC 1320
 D  F  T  L  P  A  G  H  T  V  T  S  V  W  N  A  L  I  S  P

GCCTCGGGCGCGGTCACGGCACGCAGCACCGGCTCCAACGGCCGGATCGCGGCCAACGGC 1380
 A  S  G  A  V  T  A  R  S  T  G  S  N  G  R  I  A  A  N  G

GGGACCCAGTCCTTCGGTTTCCAGGGCACCTCCAGCGGAGCGGGGTTCACCGCACCGGCC 1440
 G  T  Q  S  F  G  F  Q  G  T  S  S  G  A  G  F  T  A  P  A

GGGGCCCGGCTCAACGGCACCTCCTGCACAGTGAGATGACAATGGGGCACCGCCGGGCAC 1500
 G  A  R  L  N  G  T  S  C  T  V  R  *

TGTCTTA                                                      1507
```

Table 2. FASTA analysis of *Streptomyces* sp.11AG8 Cel12A catalytic domain and cellulose-binding domain (CBD)

Species	Catalytic domain		CBD	
	Identity (%)	Overlap (AA)	Identity (%)	Overlap (AA)
Streptomyces rochei EglS	75.2	218	61.0	105
Streptomyces lividans CelB	72.4	217	59.0	105
Aspergillus aculeatus Egl	34.4	227	–	–
Aspergillus kawachii Egl	30.2	232	–	–
Aspergillus oryzae Egl	29.6	235	–	–
Humicola insolens EGIII	33.2	232	–	–
Trichoderma reseei EGIII	31.5	224	–	–
Cellulomonas fimi CbhA	–	–	46.2	106
Thermonospora fusca Egl	–	–	46.2	106
Clostridium cellulovorans Egl D	–	–	43.5	108
Cellulomonas fimi Egl A	–	–	47.1	104

control of its own promoter did not express measurable amounts of CMCase, although cellulase production could be detected by the sensitive agar diffusion assay. *Bacillus subtilis* transformants of the shuttle vector carrying *cel12A* were also unsuccessful at cellulase production. Therefore, the gene sequence encoding the mature protein of Cel12A was subcloned in the *Bacillus* expression vector pHPLT. This vector contains the promoter, the signal sequence, and the terminator sequence of *Bacillus amyloliquefaciens* thermostable amylase, together with the origin of replication and the neomycin marker of pUB110 (van Ee 1991). The construct was transformed to *Bacillus subtilis* DS12367.

The *Bacillus* recombinant strain harboring the plasmid pHP11AG8 was grown overnight in TSB; cells were removed by centrifugation and the supernatant was concentrated fivefold by ultrafiltration. The crude enzyme preparation was analyzed by SDS-PAGE (Fig. 4). In addition to a 36-kDa band representing the putative full-length enzyme, an extra band of 24 kDa was also detected. Both these bands exhibited CMCase activity on the zymograms of the native gel (data not shown). The N-terminal amino acid sequences of both bands were determined and appeared to be identical, confirming the deduced N-terminal amino acid sequence as given in Fig. 2. It is possible that the C-terminal end of the protein is partially clipped during fermentation without disturbing CMCase activity.

Properties of the Cel12A

Carboxymethyl cellulase activity of the crude recombinant *Bacillus* enzyme preparation revealed an optimal activity at pH 8 at 40°C, and the preparation retained more than 80% of its activity at pH 9 (Fig. 5A). The thermostability of the enzyme was measured as residual activity after preincubation between 30° and 70°C for 30 min. The crude enzyme was stable up to 30 min at 55°C (pH 5–8) but was completely inactivated by incubation for 30 min at 70°C (Fig. 5B).

Discussion

Streptomyces sp. strain 11AG8 is apparently a novel alkaliphilic actinomycete isolated from a mud sample from Lake Nakuru, an East African soda lake. The isolation of actinomycetes from these samples was made possible by suppression of the dominant population, in particular endospore-forming alkaliphilic bacilli, by addition of rifampicin to the selection medium. The actinomycete colonies were recognized by their compact growth on agar medium, detected by their hyphal morphology under the microscope, and further identified by their typical growth and metabolic characteristics. FASTA nucleotide sequence comparison of the 16S rDNA and of the *cel12A* gene confirmed that strain 11AG8 belongs to the actinomycetes. It seems likely that strain 11AG8 is a novel species of the genus *Streptomyces*.

The gene coding for Cel12A of strain 11AG8 was isolated from an *E. coli* library because the recombinant strain carrying this gene produced sufficient CMCase activity to be detected by the very sensitive Congo red reagent in the CMC-agar diffusion assay. In liquid culture, however, measurable activity (i.e., >0.01 units CMCase ml^{-1}) was observed in neither *E. coli* nor *Bacillus* recombinants if the *cel12A* gene was expressed under its own promoter. This finding is in agreement with the observations of Walter and Schrempf (1995) for the *cel1* gene of *Streptomyces reticuli*. In contrast, however, the *cel12A* of strain 11AG8 could be expressed (8–10 units CMCase ml^{-1}) under the control of *Bacillus* regulatory sequences including the *Bacillus amyloliquefaciens aprE* signal sequence. It is perhaps possible that the minor expression (<0.01 units CMCase ml^{-1}) in *E. coli* observed in the very sensitive but only semiquantitative agar diffusion assay could have been caused by "readthrough" from the *lacZ* promoter of the shuttle vector, but no experiments were carried out to confirm this point. However, enhanced CMCase activity (0.01 units CMCase ml^{-1}) could be demonstrated on addition of IPTG (isopropylthiogalactoside) to the medium, a result consistent with this hypothesis (Miller 1978).

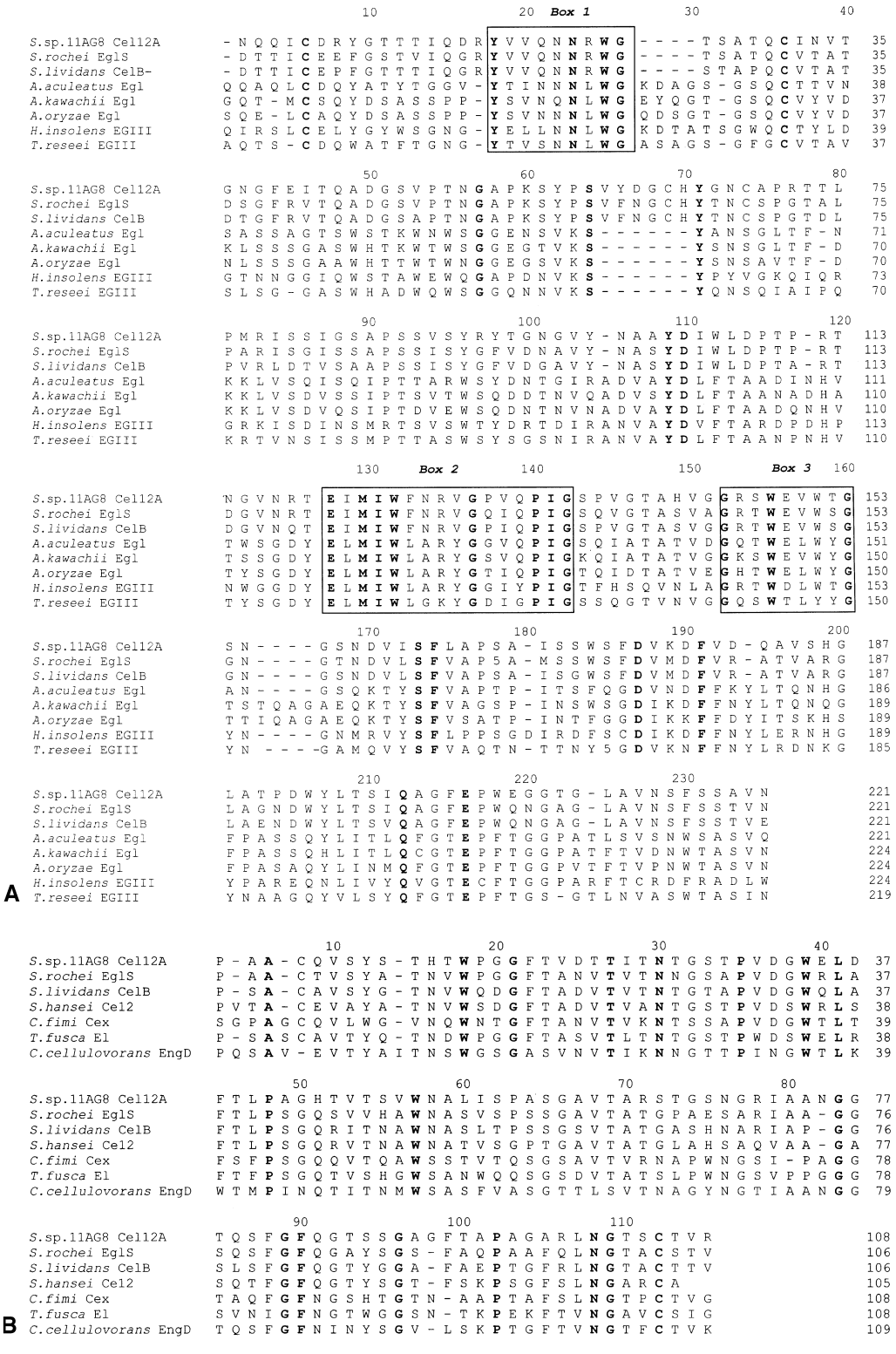

Fig. 3. **A** Sequence alignment of catalytic domains of various *Streptomyces* and fungal family 12 endoglucanases showing three major regions (*boxes*) of similarity. **B** Sequence alignment of various family IIa cellulose-binding domains shows conserved amino acids in electrophoresis of concentrated culture supernatant of *Bacillus* recombinant strain containing the plasmid pHP11AG8

The *cel12A* gene was recognized in the *E. coli* clone from the deduced sequence of the 1,113-bp open reading frame. A Shine–Dalgarno sequence and typical prokaryotic –10 and –35 regions can be recognized 5′ upstream of the ATG translation initiation codon. Further, the upstream sequence revealed a conserved inverted repeat sequence: TGGGAGC GCTCCCA. This sequence is also found upstream of the start codon of other actinomycete cellulase genes and is described as a regulator sequence on which an activator protein can bind (Lao et al. 1991). The deduced Cel12A protein has 371 amino acids, with a calculated molecular mass of 38.5 kDa. The putative signal sequence has a length of 31 amino acids with a predicted signal peptidase cleavage site after Ala-Gln-Ala (von Heijne

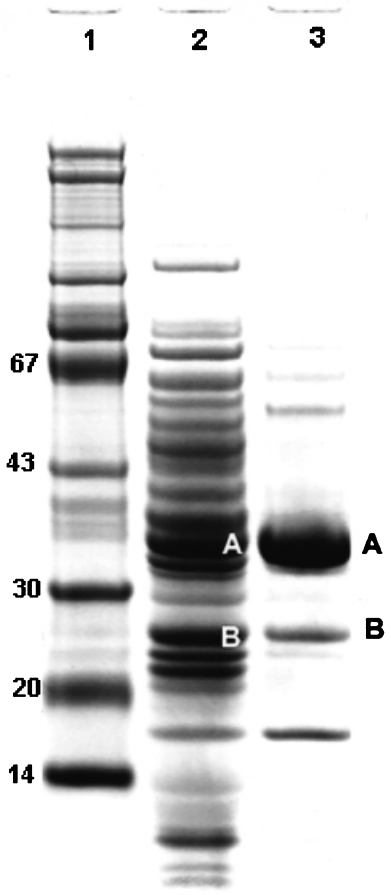

Fig. 4. SDS-PAGE electrophoresis of concentrated culture supernatants of *Bacillus* recombinant strain carrying the plasmid pHP11AG8. *Lane 1*, molecular weight markers (kDa); *lane 2*, culture supernatant (ultrafiltrate); *lane 3*, partially purified material from a 1.2 M ammonium sulfate precipitate. *A*, 36-kDa band representing the full-length mature cellulase Cel12A from *Streptomyces* sp. 11AG8; *B*, 24-kDa band representing a truncated form of the cellulase lacking the cellulose-binding domains (CBD)

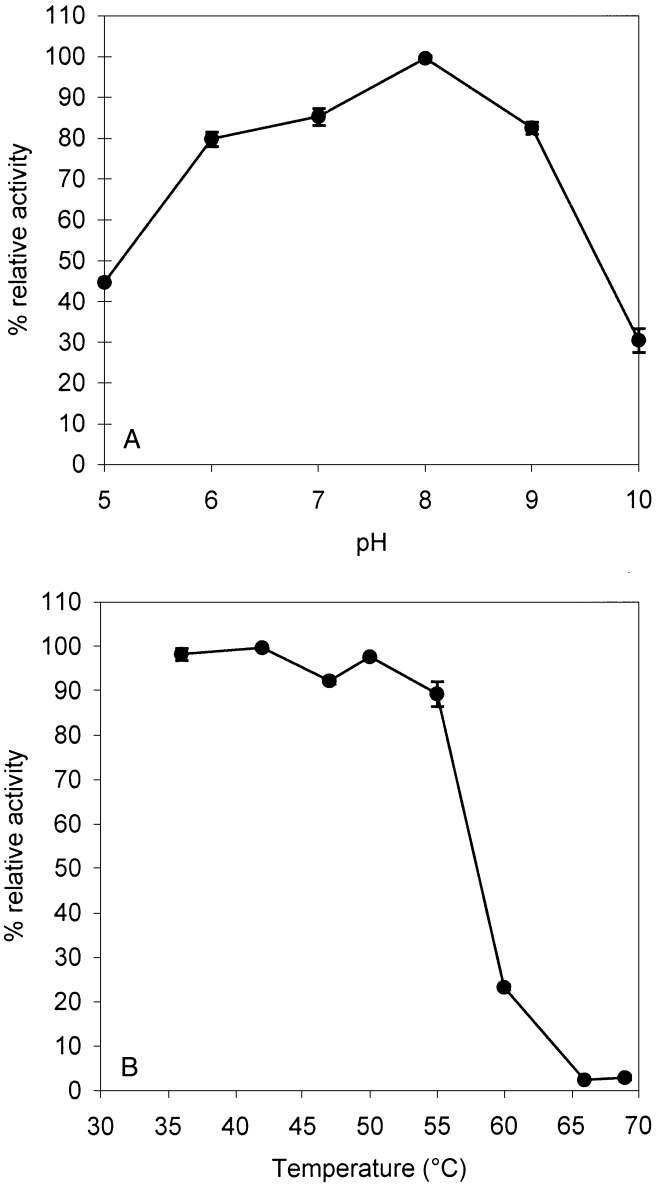

Fig. 5. **A** Relative enzyme activity (as a percentage of maximum activity) as a function of pH for *Streptomyces* sp. 11AG8 Cel12A at 40°C. Values are averages of duplicate assays; *error bars* indicate SD. **B** Stability of *Streptomyces* sp. 11AG8 Cel12A at pH 8 expressed as the percent remaining activity (relative to the maximum activity) after 30 min incubation at each temperature point. Values are averages of duplicate assays; *error bars* indicate SD

1986). The deduced mature Cel12A enzyme is a protein of 35.8 kDa, a value close to the size of the upper CMCase active band in the PAGE gel of the crude enzyme preparation (see Fig. 4).

Cel12A was compared with other proteins in the SWISS PROT database by using FASTA. A significant level of similarity was observed for the two separate domains (see Table 2). The first domain (residues 32 to 250) of Cel12A showed similarity with the catalytic domains of several streptomycete and fungal endoglucanases. A second domain (residues 272 to 340) showed similarity to CBDs of endo- and exoglucanases of a wide variety of prokaryote microorganisms. The cellulases that give the best scores in FASTA analysis of the catalytic domain belong to family 12 of glycosylhydrolases as classified by Henrissat and Bairoch (1993).

Alignment of the amino acid sequence of the catalytic domain of seven family 12 cellulases (Tomme et al. 1995a) and that of Cel12A show three major regions of similarity (see Fig. 3A), placing Cel12A from *Streptomyces* sp 11AG8 clearly in this glycosylhydrolase family. The presence of various discrete bands of CMCase active material in the zymograms of native PAGE gels of crude enzyme preparations of Cel12A indicate the presence of proteolytic activity in the culture broth of the *Bacillus* recombinant strain. The N-terminal sequences of both the full-length and the major (24-kDa) truncated Cel12A were identical. The position of cleavage was assessed by mass spectroscopy determination of protein molecular weight. The purified core 24-kDa species consisted of two molecules, one with a MW of 24,468.8 (comprising 230 amino acid residues, 2 disulfide bridges) and one with a MW of 24,724.1 (comprising 233 amino acid

residues, 2 disulfide bridges). Both molecules terminate with a proline residue that implies a proline-specific protease may be responsible for cleavage. The cleavage sites are located at the downstream end of the linker sequence, adjacent to the CBD, leaving the CMCase activity intact. Proteolytic cleavage of the CBD in *Streptomyces* cellulases is not uncommon and has been described earlier by Schlochtermeier et al. (1992).

By comparison to the homologous EGIII cellulases of *Trichoderma reesei* and *Humicola insolens*, which do not contain a cellulose-binding domain, it would appear that the activity of these enzymes is not dependent on the presence of such a domain. The role of this domain may be to control the release and diffusion of Cel12A in its natural environment. The truncated form of the Cel12A of *Streptomyces* sp. strain 11AG8, showing favorable characteristics such as CMCase activity over a broad pH range, good stability at high pH, and elevated temperatures, indicates that this enzyme is a good candidate for industrial applications.

Acknowledgements We thank colleagues at Palo Alto, Christian Paech for his input in mass spectrometry analysis of the Cel12A gene products and Colin Mitchinson for biochemical support.

References

Béguin P (1983) Detection of cellulase activity in polyacrylamide gels using Congo red stained replicas. Anal Biochem 131:333–336

Béguin P, Aubert JP (1994) The biological degradation of cellulose. FEMS Microbiol Rev 13:25–58

Duckworth AW, Grant S, Grant WD, Jones BE, Meijer D (1998) *Dietzia natronolimnaios* sp. nov., a new member of the genus *Dietzia* from an East African soda lake. Extremophiles 2:359–366

Gryczan TJ, Contente S, Dubnau D (1978) Characterisation of *Streptococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. J Bacteriol 134:318–329

Henrissat B, Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino sequence similarities. Biochem J 293:781–788

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, vol III. Academic Press, New York

Kawamura F, Doi RH (1984) Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. J Bacteriol 160:442–444

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA 82:6955–6859.

Lange NK (1993) Application of cellulases in the textile industry. In: Souminen P, Reinikainen T (eds) Proceedings of the second TRICEL symposium on *Trichoderma reesei* cellulases and other hydrolases, Espoo, Finland. Found Biotech Ind Ferment Res 8:263–272

Lao G, Ghangas GS, Jung ED, Wilson DB (1991) DNA sequences of three beta-1,4-endoglucanase genes from *Thermomonospora fusca*. J Bacteriol 173:3397–3407

Lever M (1972) A new reaction for colorimetric determination of carbohydrates. Anal Biochem 47:273–279

Maurer K-H (1997) Development of new cellulases. In: van Ee JH, Misset O, Baas EJ (eds) Enzymes in detergency. Marcel Dekker, New York, pp 175–202

Miller JH (1978) The *lac*I gene: its role in *lac* operon control and its use as a genetic system. In: Miller JH, Reznikoff WS (eds) The operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp 31–88

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Saito H, Miura K-I (1963) Preparation of transforming deoxyribonucleic acid by phenol treatment. Biochim Biophys Acta 72:619–629

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Schlochtermeier A, Walter S, Schroder J, Moorman M, Schrempf H (1992) The gene encoding the cellulase (Avicelase) Cel1 from *Streptomyces reticuli* and analysis of protein domains. Mol Microbiol 6:3611–3621

Shyamala V, Ames GF (1989) Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. Gene (Amst) 84:1–8

Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC (1985) Measurement of protein using bicinchoninic acid. Anal Biochem 150:76–85

Staley JT, Fuerst JA, Giovannoni S, Schlesner H (1992) The order *Planctomycetales* and the genera *Planctomyces*, *Pirellula*, *Gemmata* and *Isosphaera*. In: Barlows HG, Truper MG, Dworkin M, Harder W, Scheifer KH (eds) The prokaryotes: handbook of the biology of bacteria, vol 4 (203), 2nd edn. Springer, New York, pp 3722–3723

Teather RM, Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterisation of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol 43:777–778

Tomme P, Warren RAJ, Gilkes NR (1995a) Cellulose hydrolysis by bacteria and fungi. Adv Microb Physiol 37:1–81

Tomme P, Warren RAJ, Miller RC, Kilburn DG, Gilkes NR (1995b) Cellulose-binding domains: classification and properties. In: Saddler JN, Penner MH (eds) Enzymatic degradation of insoluble carbohydrates. ACS symposium series, vol 618. American Chemical Society, Washington, DC, pp 142–163

Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci 10:569–570

van Ee JH (1991) Regulatory region cloning and analysis plasmid for *Bacillus*. US Patent no. 5024943

Vehmaanperä J (1988) Transformation of *Bacillus amyloliquefaciens* protoplasts with plasmid DNA. FEMS Microbiol Lett 49:101–105

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

Walter S, Schrempf H (1995) Studies of *Streptomyces reticuli* cel-1 (cellulase) gene expression in *Streptomyces* strains, *Escherichia coli*, and *Bacillus subtilis*. Appl Environ Microbiol 61:487–494

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