Appl Microbiol Biotechnol (1998) 49: 277–284

ORIGINAL PAPER

S. Halldórsdóttir · E. T. Thórólfsdóttir · R. Spilliaert
M. Johansson · S. H. Thorbjarnardóttir · A. Palsdottir
G. Ó. Hreggvidsson · J. K. Kristjánsson · O. Holst
G. Eggertsson

Cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12

Received: 5 August 1997 / Received revision: 6 November 1997 / Accepted: 7 November 1997

Abstract A gene library from the thermophilic eubacterium *Rhodothermus marinus*, strain ITI 378, was constructed in pUC18 and transformed into *Escherichia coli*. Of 5400 transformants, 3 were active on carboxymethylcellulose. Three plasmids conferring cellulase activity were purified and were all found to contain the same cellulase gene, *celA*. The open reading frame for the *celA* gene is 780 base pairs and encodes a protein of 260 amino acids with a calculated molecular mass of 28.8 kDa. The amino acid sequence shows homology with cellulases in glycosyl hydrolase family 12. The *celA* gene was overexpressed in *E. coli* when the pET23, T7 phage RNA polymerase system was used. The enzyme showed activity on carboxymethylcellulose and lichenan, but not on birch xylan or laminarin. The expressed enzyme had six terminal histidine residues and was purified by using a nickel nitrilotriacetate column. The enzyme had a pH optimum of 6–7 and its highest measured initial activity at 100 °C. The heat stability of the enzyme was increased by removal of the histidine residues. It then retained 75% of its activity after 8 h at 90 °C.

S. Halldórsdóttir · E. T. Thórólfsdóttir · R. Spilliaert[1]
S. H. Thorbjarnardóttir · A. Palsdottir[2] · G. Eggertsson (✉)
Laboratory of Molecular Genetics, Institute of Biology,
University of Iceland, IS-108 Reykjavik, Iceland
Tel.: +354 525 4603
Fax: +354 525 4069
e-mail: gudmegg@rhi.hi.is

G. Ó. Hreggvidsson · J. K. Kristjánsson
Department of Biotechnology,
Technological Institute of Iceland,
Keldnaholt, IS-112 Reykjavik, Iceland

O. Holst · M. Johansson
Center for Chemistry and Chemical Engineering,
Lund University, S-22100 Lund, Sweden

Present address:
[1] Science Institute, University of Iceland,
IS-101 Reykjavik, Iceland
[2] Institute for Experimantal Pathology, Keldur, IS-112,
Reykjavik, Iceland

Introduction

Rhodothermus marinus is a marine thermophilic eubacterium isolated from alkaline submarine hot springs (Alfredsson et al. 1988). It is an obligate aerobe and grows optimally at 65 °C, pH 7.0 and 2% NaCl (Alfredsson et al. 1988). Sequence analysis of a 16S ribosomal RNA gene revealed that *R. marinus* is most closely allied to the Flexibacter-Cytophaga-Bacteroides group (Andrésson and Fridjónsson 1994). It has been shown to produce several thermostable glycosyl hydrolases, including amylases (Hreggvidsson et al. 1992), a xylanase (Nordberg Karlsson et al. 1997), a β-xylosidase (Manelius et al. 1994) and a cellulase (Hreggvidsson et al. 1996). A gene from *R. marinus* encoding a thermostable β-glucanase belonging to glycosyl hydrolase family 16 has been sequenced and the enzyme characterized (Spilliaert et al. 1994).

A classification of glycosyl hydrolases, in families based on amino acid sequence similarities, was proposed by Henrissat (1991) and, in 1993, 480 glycosyl hydrolases with 52 different EC numbers were classified in 45 families (Henrissat and Bairoch 1993).

There are many potential applications for cellulases in industry and biotechnology (Bayer et al. 1994; Béguin and Aubert 1994). Thermostable enzymes are attractive candidates for some of these applications (Herbert 1992; Kristjansson 1989). In this paper we describe the cloning, sequencing and overexpression of a *R. marinus* gene encoding a thermostable cellulase as well as comparison of the amino acid sequence of the *R. marinus* cellulase to that of its hyper-thermophilic and mesophilic counterparts.

Materials and methods

Bacteria, plasmids and culture conditions

R. marinus strain ITI 378 was obtained from the strain collection of the Technological Institute of Iceland. It was grown at 65 °C with

shaking in medium 162 (Degryse et al. 1978) with 2% NaCl added (Alfredsson et al. 1988). *Escherichia coli* TG1 *supE hsd*Δ5 *thi* Δ(*lac-proAB*), F′*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15 was used as a host for the cloning vectors pUC18 and pUC19 and for propagation of M13 phages mp18 and mp19. *E. coli* GE1731 (Thorbjarnardóttir et al. 1995) was used as the host strain for recombinant plasmid pET23 (Studier et al. 1990). *E. coli* was grown in Luria-Bertani (LB) broth containing 10 g tryptone, 5 g yeast extract and 5 g NaCl/l. Ampicillin was added to the LB/agar medium to a final concentration of 100 μg/ml for the selection of recombinants. X-gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside) and IPTG (isopropyl β-D-thiogalactopyranoside) were added at a final concentration of 25 μg/ml and 0.1 mM respectively. IPTG was added to LB medium to a final concentration of 0.4 mM for the induction of expression for 4 h of the *celA* gene in the pET23 expression vector system. For in situ enzyme assays (Wood et al. 1988), the media were supplemented with 0.5% polysaccharide substrate.

Polysaccharide substrates

The following substrates were used: Avicel (Fluka Chemica), Konjak mannan (glucomannan, from Megazyme International Ltd.), microgranular cellulose, lichenan, laminarin, birchwood xylan (all from Sigma) and carboxymethylcellulose (CMC). The CMC types used were CMC (Sigma), Blanose (Hercules) 7M31F (degree of substitution, DS, 0.7), here designated CMC1, and Blanose 12M31P (DS 1.3), here designated CMC2. CMC1 was, unless otherwise stated, used for enzyme assays and activity tests in gels. CMC (Sigma) was used for plate assays.

Construction of a genomic library and isolation of the cellulase gene

R. marinus chromosomal DNA was prepared and a gene library constructed in the pUC18 vector as described before (Spilliaert et al. 1994). *E. coli* TG1 cells were made competent by the method of Chung et al. (1989), transformed with the library and plated on LB/agar plates containing X-gal and IPTG. White colonies, with recombinant pUC18 plasmids, were transferred onto LB plates containing 0.5% CMC. Plates were incubated overnight at 37 °C and then for 5 h at 65 °C, and stained with 0.1% Congo red (Sigma) for 15 min according to Wood et al. (1988). A zone of hydrolysis on a red background was visible around colonies showing cellulase activity.

Plasmid DNA analysis

Plasmid DNA was prepared according to standard methods and the inserted DNA digested with restriction endonucleases in single or double digests for the construction of a restriction map. For Southern-blot analysis, restriction fragments were transferred onto Hybond N hybridization membranes (Amersham) and hybridized to a DNA fragment, previously gel-purified and labelled with digoxigenin, using a DIG labelling kit (Boehringer Mannheim). Restriction fragments were cloned into M13mp18 and M13mp19 and both strands of the DNA were sequenced by the dideoxy-DNA chain-termination method and using a Sequenase 2.0 sequencing kit (United States Biochemical Corp.) or a Thermo Sequenase cycle sequencing kit (United States Biochemical Corp.). A homology search and sequence analysis were performed by using the available nucleotide and/or protein sequence databases (GenBank, pir, swissprot) using the standard search algorithms. Compositional analysis of protein sequences was done by using the SAPS computer program (Brendel et al. 1992). The computer program ClustalW1.6 (Thompson et al. 1994) was used to generate pairwise sequence alignments.

Overexpression of the *celA* gene in pET23 and purification of the cellulase

The *celA* gene was cloned via the polymerase chain reaction (PCR) between the *Nde*I and *Hin*dIII sites of pET23b$^+$ in order to overexpress the gene. The primers used were forward 5′GGAATTCCATATGAACGTCATGCGTGCGG, and reverse 3′CCCAAGC TTCTGCACCGTTACGGAAA. The resulting recombinant plasmid was named pET23bAH. The GE1731 derivative containing the recombinant plasmid pET23bAH (GE2205) was grown overnight at 32 °C. The overnight culture was diluted 1:50 and grown at 32 °C until A_{600} reached 0.8. The culture was induced with 0.4 mM IPTG and aliquots of 100 μl were taken for sodium dodecyl sulphate/polyacrylamide gel electrophoresis (SDS/PAGE) analysis after 0, 1, 2 and 4 h incubation.

After 4 h of induction, the cells were harvested for protein purification. The cellulase was purified using a nickel nitrilotriacetate (Ni-NTA) column from QIAGEN according to the manufacturer's protocol. Purification was monitored by visualizing samples on SDS/PAGE and by in situ enzyme assay. Enzyme activity was measured with the standard assay and the protein concentration was measured with the Bio-Rad (Bradford) protein assay (Zaman and Verwilghen 1979).

Enzyme characterization

SDS/PAGE was performed in a 12% gel by the method of Laemmli (1970) and proteins were stained with Coomassie brilliant blue R-250 (Sigma). Enzyme activity was detected in situ in SDS/PAGE containing 0.13% CMC. The gel was incubated for 4 h at 70 °C and stained with 0.1% Congo red. For detecting activity in SDS/PAGE gels not containing substrate, the gels were covered with a 1% agarose layer containing 0.15% CMC.

The standard assay for cellulase activity was done by incubating the enzyme (15 mg/ml) at 70 °C for 15 min, with 0.5% (w/v) CMC as a substrate in 0.1 M phosphate buffer ($Na_2H_2PO_4$), pH 7.0. The reducing sugars released were detected by the dinitrosalicylic acid method (Sumner and Somers 1949) with glucose as a standard. One unit of enzyme activity is the amount that leads to the release of 1 μmol reducing sugars/min. All activity measurements were done in triplicate.

The temperature optimum was determined by running the standard assay at 60, 70, 75, 80, 90, 95 and 100 °C. Thermostability was measured by incubating the enzyme (15 mg/ml in phosphate buffer, pH 7.0) at 90 °C for 8 h. Samples were taken at intervals and residual activity was measured by the standard assay. All assays were done in triplicate.

The pH optimum was measured by running the standard assay at different pH values. The buffers used were citrate/phosphate buffer (0.2 M) for pH 4–5 and sodium phosphate buffer (0.2 M) for pH 6–10. For pH 9 and 10, which are outside the buffering range, the pH was monitored before and after the reaction. Significant changes in pH due to the reaction were not observed.

Results

Cloning and sequencing of the *R. marinus celA* gene

Three clones among 5400 transformants were found to produce a clear halo on 0.5% CMC. Plasmid DNA was isolated from these clones, $pLHC_1$, $pLHC_2$ and $pLHC_3$, and the size of the inserts determined. The clones had inserts of 3 kb, 5 kb and 1.5 kb respectively. Plasmid $pLHC_3$ was selected for restriction endonuclease mapping and sequencing. An insert of 1560 bp was found to contain an open reading frame of 780 bp. Two potential methionine initation codons were found. Translation

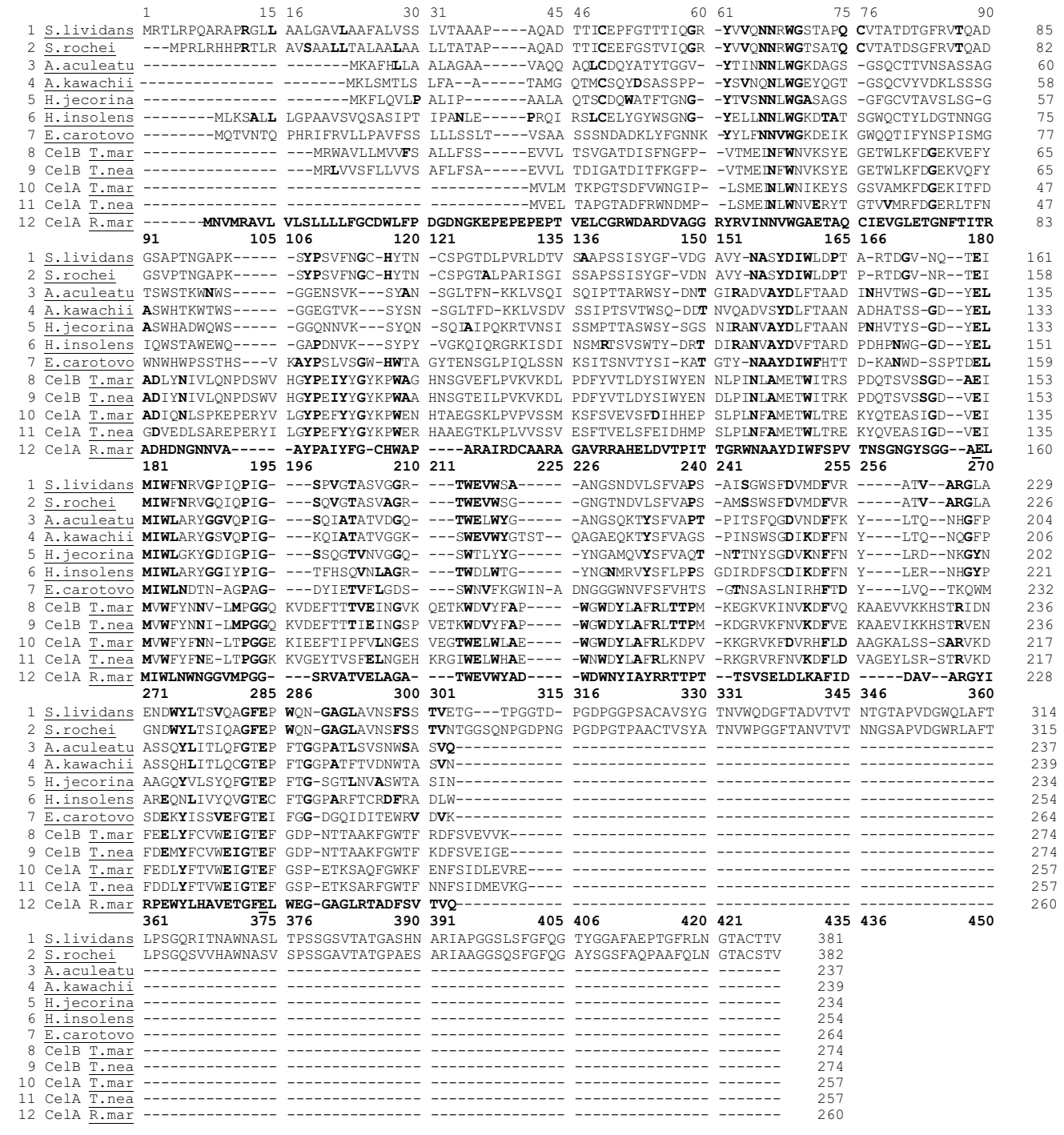

Fig. 1 Multiple alignment of the deduced amino acid sequences of the *celA* gene from *Rhodothermus marinus* and the 11 sequences that showed homology to it. Amino acid residues identical to those of the CelA protein are shown in bold. The putative catalytic residues (Glu^{159} and Glu^{242}) are underlined. Accession numbers for the deduced sequences are as follows: (*1*) *Streptomyces lividans*, *celB*, U04629; (*2*) *Streptomyces rochei*, *eglS*, X73953; (*3*) *Aspergillus aculeatus*, FI-CMCase mRNA, X52525; (*4*) *Aspergillus kawachii*, mRNA for an endoglucanse, D12901; (*5*) *Hypocrea jecorina*, endoglucanase gene, AB003694; (*6*) *Humicola insolens*, mRNA fragment, A22907; (*7*) *Erwinia carotovora*, *celS*, M32399, (*8, 9*) *Thermotoga maritima*, *celA* and *celB*, Z69341; (*10, 11*) Thermotoga neapolitana, *celA* and *celB*, Z86103, U93354; (*12*) *Rhodothermus marinus*, *celA* U72637

initation at the first codon would result in a 780-bp message corresponding to a polypeptide of 260 amino acids with a calculated molecular mass of 28.8 kDa, while translation initation of the second codon would result in a 771-bp message corresponding to a polypeptide of 257 amino acids. By using the computer program PSORT, the polypeptide of 260 amino acids was found to contain a putative signal peptide of 17 amino acids (Nakai and Kanehisa 1991). Assuming that the preprotein is 28.8 kDa, the predicted mature protein would be 243 amino acids with a calculated molecular mass of

26.9 kDa. No sequences resembling known eubacterial promoter sequences were identified upstream from the *celA* gene but promoter sequences in *R. marinus* have not been characterized. The cellulase gene has a G + C content of 65.4%, which corresponds to the 64.4% value found for *R. marinus* (Alfredsson et al. 1988). The cellulase gene was named *celA*. The nucleotide sequence data will appear in the GenBank/EMBL/DDBJ databases under the accession no. U72637. Partial nucleotide sequencing confirmed that plasmids $pLHC_1$ and $pLHC_2$ also contained the *celA* gene.

Homology with cellulases from other microorganisms

A search for homology between the CelA protein and cellulases from other microorganisms was done by using the available nucleotide and protein databases with the BLAST algorithm (Altschul et al. 1990). The CelA protein showed homology to 11 cellulases from 9 species. Eight of these enzymes have been assigned to cellulase family H, which corresponds to glycosyl hydrolase family 12 (Liebl et al. 1996; Tomme et al. 1995). It is concluded that CelA of *R. marinus* belongs to this family. The homology between CelA and the other enzymes is low, identity ranging from 17% to 27%. Multiple alignment of the deduced amino acid sequence of CelA of *R. marinus* and the other cellulases is shown in Fig. 1. The accession number of each sequence is given in the figure legend. The amino acid composition of the 12 cellulases is given in Table 1.

An open reading frame of about 500 bp was found about 1 kb upstream of the *celA* gene in the $pLHC_2$ insert. The derived amino acid sequence showed homology to argininosuccinase from both eukaryotes and prokaryotes.

Overexpression of the *celA* gene in *E. coli*, and SDS/PAGE analysis

In order to overexpress the *celA* gene in *E. coli*, a PCR copy of the gene with *Nde*I-*Hin*dIII restriction sites was ligated into the expression vector pET23b(+) cut with the same enzymes. The resulting plasmid, named pET23bAH, encodes a cellulase protein with six terminal His residues.

Synthesis of the enzyme was monitored by removing aliquots for SDS/PAGE analysis during 4 h induction with 0.4 mM IPTG. Coomassie-blue staining of SDS/PAGE gels revealed two bands in induced samples. One band had an apparent molecular mass near 50 kDa, but the other had an apparent molecular mass of 28 kDa, which corresponds to the expected calculated value for the *R. marinus* cellulase (data not shown).

The cellulase was purified by using the Ni-NTA spin column. Purification was monitored by removing aliquots for SDS/PAGE analysis and in situ assays, after column loading, washing and elution. Coomassie blue staining of the purified cellulase protein revealed a strong band of apparent molecular mass 28 kDa, which is in good agreement with the calculated value of 28.8 kDa (Fig. 2).

The purified protein was treated with proteinase K (0.01 mg/ml) for 4 h at 37 °C. After this treatment the protein did not adsorb to the Ni-NTA column. This implies that the histidine tail and possibly linking amino acids were removed by the treatment. No reduction in

Table 1 Amino acid composition of 12 cellulases from glycosyl hydrolase family 12. For the full names of the organisms see Fig. 1. The percentages for Glu from thermophilic organisms are shown in bold

Amino acid	Composition (%)											
	A. acul	*A. kaw*	*E. caro*	*H. inso*	*H. jeco*	*S. livid*	*S. roch*	*T. mar A*	*T. mar B*	*T. nea A*	*T. nea B*	*R. mar*
A	11.0	7.9	5.3	6.3	9.4	12.6	13.1	4.7	4.7	4.3	4.7	11.9
R	1.3	0.4	1.5	6.7	1.3	4.2	4.2	2.7	1.8	6.2	1.8	6.9
N	5.9	4.2	8.0	6.3	9.0	5.0	6.0	3.9	5.5	4.7	4.7	5.4
D	4.2	5.4	6.1	6.7	3.0	4.7	3.7	5.1	5.8	5.4	6.9	6.2
C	0.8	1.3	0.0	2.0	0.9	1.6	1.6	0.0	0.4	0.0	0.4	1.9
E	1.7	2.5	3.0	3.1	0.9	1.8	1.6	**10.9**	**7.7**	**10.9**	**7.7**	**6.5**
Q	6.8	6.3	3.0	3.9	6.0	3.7	4.2	1.2	1.8	0.4	1.5	0.8
G	9.7	9.2	7.6	9.8	11.5	11.8	11.0	7.0	6.2	7.4	6.2	10.4
H	1.3	1.3	2.7	1.2	0.9	0.5	1.0	1.6	1.1	1.6	1.1	1.5
I	4.2	2.9	6.1	5.5	3.8	2.9	3.1	4.3	4.0	2.7	5.8	4.2
L	6.3	5.4	6.1	7.5	5.6	5.0	4.5	7.8	6.6	8.2	5.8	6.9
K	3.4	4.2	4.5	2.4	2.6	0.3	0.3	7.0	6.6	4.7	6.6	0.8
M	0.8	2.1	1.5	1.6	1.7	0.8	1.0	2.7	2.6	3.1	2.6	1.5
F	4.2	3.8	4.9	3.5	3.8	4.5	4.2	7.8	7.3	7.4	7.7	3.1
P	3.0	3.3	3.0	5.1	3.0	6.8	6.5	5.4	4.4	4.3	4.7	4.6
S	10.5	14.2	12.1	6.7	11.5	8.9	11.0	7.8	5.5	5.4	4.7	3.1
T	10.1	10.5	8.3	6.3	8.1	11.3	9.4	5.1	6.9	6.2	7.7	6.9
W	4.2	4.2	5.7	4.7	3.8	2.9	2.9	4.3	4.7	4.3	4.4	5.4
Y	5.1	4.6	4.5	6.3	6.8	2.4	2.4	3.9	5.1	4.7	5.1	3.8
V	5.5	6.3	6.1	4.3	6.4	8.4	8.4	7.0	11.3	8.2	9.9	8.1

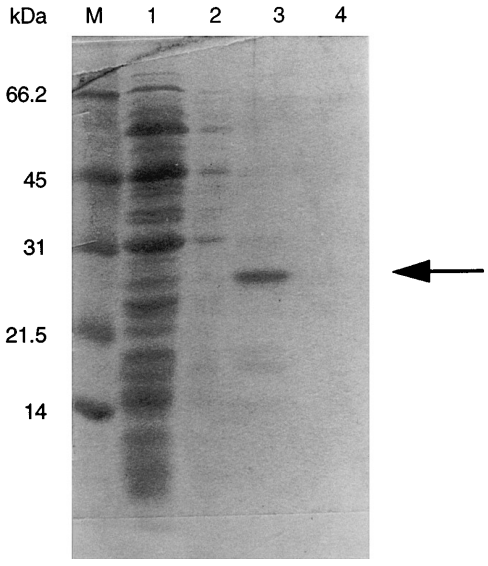

Fig. 2 Sodium dodecyl sulphate (SDS)/polyacrylamide gel of purification of the CelA protein. *M* molecular size marker (kDa), *1* nickel nitrilotriacetase column flow-through after sample loading, *2* flow-through after column washing, *3* flow-through after column elution with 1 M imidazole, *4* flow-through after second elution of column with 1 M imidazole. *Arrow* the purified CelA protein band in *3*

enzyme activity was observed, as measured by the dinitrosalicylic acid method. Samples of the heated enzyme were boiled for different lengths of time, run on a SDS/PAGE gel containing CMC and stained both for activity and for protein (Fig. 3). The unboiled control showed three activity bands with an apparent molecular mass of 35–46 kDa. After boiling, activity disappeared and a protein band of about 26 kDa appeared (Fig. 3). By running proteinase-treated and untreated samples in parallel in a SDS/PAGE Phast gel it was confirmed that the treated protein is 1–2 kDa smaller than the untreated one (data not shown). No enzyme activity was connected with the protein monomer.

Samples of the recombinant enzyme and a previously described cellulase partially purified from *R. marinus* (Hreggvidsson et al. 1996) were run together on a gel containing CMC (Fig. 4). The native activity appeared in a position corresponding to the smallest activity band of the recombinant protein. In order to see if CMC was affecting the mobility of the active enzyme, samples identical to those shown in Fig. 4 were run on a SDS/PAGE gel without CMC. Activity was detected in an agarose overlay with CMC. The same bands were seen and no difference in mobility was observed (data not shown). It is noted that the same activity bands were seen whether the enzyme had been heated with protease or not (Figs. 3, 4). The nature of the apparent aggregates that represent the active protein is not known.

Some properties of the *R. marinus* cellulase

The activity of purified cellulase was investigated at different temperatures and pH values.

The temperature for maximal activity was determined by incubating the enzyme at different temperatures according to the standard assay. As shown in Fig. 5, the initial activity was still increasing between 90 °C and 100 °C. Thermal stability was investigated by incubating the enzyme for up to 8 h at 90 °C (Fig. 6). Both prote-

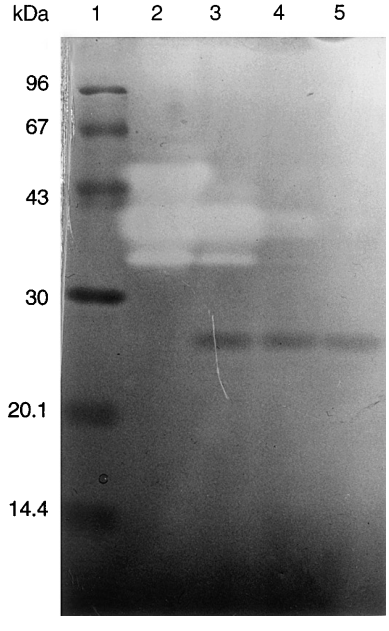

Fig. 3 SDS/polyacrylamide gel electrophoresis of recombinant cellulase treated with proteinase K. Lanes: *1* low-molecular-mass markers from Pharmacia, *2* unboiled sample, *3* sample boiled for 2 min in mercaptoethanol, *4* sample boiled for 5 min, 5 sample boiled for 10 min

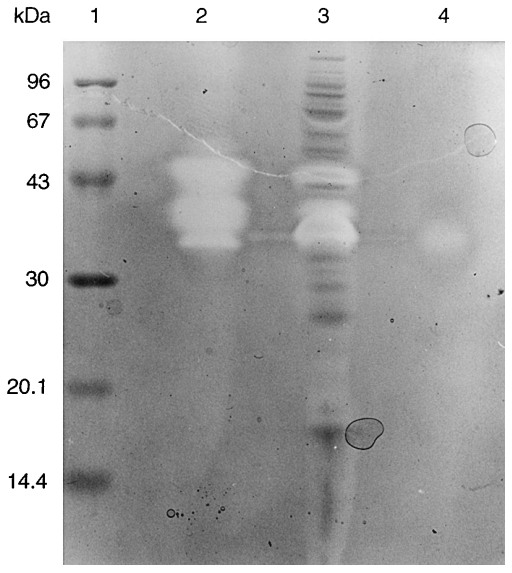

Fig. 4 SDS/polyacrylamide gel electrophoresis of native and recombinant cellulase from *R. marinus*. The gel contained carboxymethylcellulose and was stained for activity and protein. Lanes: *1* low-molecular-mass markers from Pharmacia; *2* recombinant cellulase treated with proteinase K, unboiled sample; *3* supernatant from *E. coli*-expressed recombinant clone, unboiled; *4* partially purified cellulase from *R. marinus*, unboiled

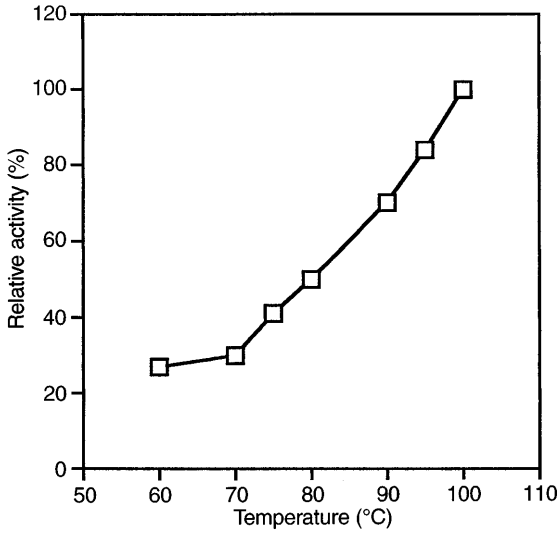

Fig. 5 Influence of temperature on the activity of CelA from *R. marinus*. Enzyme was measured according to the standard assay at 60, 70, 75, 80, 90, 95 and 100 °C. The highest level of activity, 3.44 U/mg, was obtained at 100 °C and was set at 100%

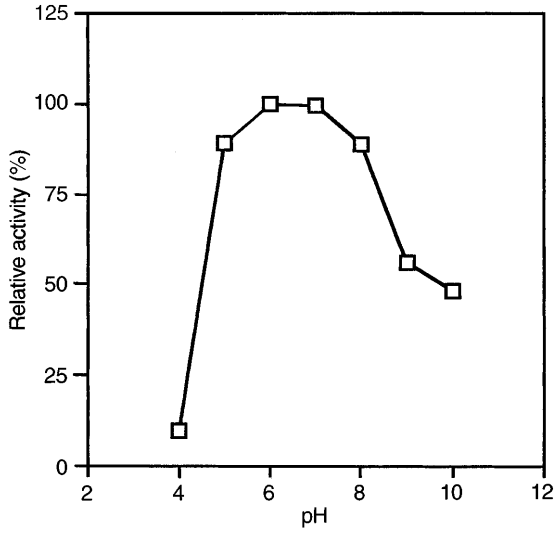

Fig. 7 Influence of pH on the activity of CelA from *R. marinus*. Enzyme activity was measured at pH 4, 5, 6, 7, 8, 9 and 10. The highest level of activity, 0.77 U/mg, was obtained at pH 6 and was set at 100%

ase-treated and untreated enzyme was tested. The protease-treated enzyme showed more heat stability than the untreated one. After 8 h incubation the treated and untreated enzyme retain 75% and 45% of their activity respectively.

The pH optimum of the purified cellulase was determined by incubating the enzyme at different pH values according to the standard assay (Fig. 7). The enzyme showed a pH optimum of 6–7.

The enzyme was found to hydrolyse polysaccharides with β1-4 and β1-3–1-4 linkages. The specific activity of the purified enzyme at 70 °C in CMC, lichenan and glucosomannan was 1.0, 1.7 and 2.3 U/mg, respectively.

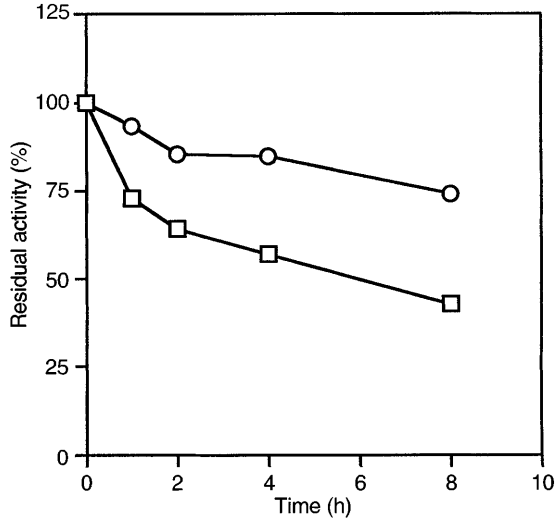

Fig. 6 Thermal stability of CelA from *R. marinus*. Approximately 3.5 U enzyme was preincubated for 1,2, 4 and 8 h at 90 °C. Residual activity was measured according to the standard assay. □ Untreated cellulase, ○ cellulase treated with proteinase K with the histidine tail removed

The extent of hydrolysis measured after 48 h at 70 °C was 13% for CMC, 51% for lichenan and 34% for glucosomannan. For the highly substituted CMC2 the extent of hydrolysis was only 4% (specific activity 0.12 U/mg). No activity was found with Avicel, microgranular cellulose, birch xylan or laminarin (plate assays).

Discussion

In this paper we describe the cloning, sequencing and overexpression of a gene, *celA*, from the thermophilic eubacterium *R. marinus*. The gene encodes a thermostable cellulase. It has a coding sequence corresponding to a polypeptide of 28.8 kDa, including a putative signal peptide of 17 amino acids. The CelA enzyme is able to hydrolyse β1-4 linkages in carboxymethyl cellulose and mixed β-glucans (lichenan) but does not cleave β1-3 linkages, as shown by its inability to hydrolyse laminarin, a homopolymer of β1-3-linked glucose residues. Activity towards birch xylan or crystalline cellulose was not detected. On the basis of these observations the CelA enzyme can be classified as an endo-1,4-β-glucanase (EC 3.2.1.4).

A thermostable cellulase from *R. marinus* was recently described. In our study the apparent molecular mass of this enzyme on SDS/PAGE gels was found to correspond to one of three bands of activity found for the recombinant enzyme. Comparing the characteristics of the recombinant and native cellulases (Hreggvidsson et al. 1996, and the present study) we find no significant differences. The highest measured activity on CMC was at 100 °C for both enzymes, the pH optimum was 7 for the native and 6–7 for the recombinant enzyme and the specific activity at 70 °C was 1.4 U/mg and 1.0 U/mg for the native and recombinant enzyme respectively. No

differences have been found in substrate specificity. It is important to note that only one band of cellulase activity was found in the crude supernatant on SDS gels (Hreggvidsson et al. 1996). We therefore think that most likely the native enzyme is encoded by the *celA* gene. Enzyme activity in SDS/PAGE gels was associated with apparent aggregates of the recombinant protein. Nothing can be concluded about the nature of these aggregates.

A thermostable *R. marinus* glucanase, estimated to be 29.7 kDa, has also been characterized (Spilliaert et al. 1994). This enzyme had no activity on CMC cellulose but high activity on lichenan and laminarin. It therefore differs considerably in specificity from the recombinant and native enzymes of this study. It belongs to glycosyl hydrolase family 16.

On the basis of sequence comparisons, the CelA enzyme was assigned to glycosyl hydrolase family 12 (Fig. 1). The overall homology between CelA from *R. marinus* and the 11 cellulases shown in Fig. 1 is low. There is, however, a box of three conserved amino acids to be found among the 11 cellulases, starting at Glu^{159} in the *R. marinus* enzyme (E-M-W) as well as eight other conserved amino acids: $Asn^{60,}$ Trp^{62}, Pro^{172}, Gly^{174}, Trp^{187}, Phe^{218}, Gly^{240} and Glu^{242}. In spite of low overall identity between the protein sequences, it is interesting to compare the thermostable candidates (*R. marinus* CelA and *T. maritima* and *T. neapolitana* CelA and CelB) to their mesophilic counterparts. By comparing small proteins that comprise only a single domain, the search for determinants of protein stability is somewhat simplified. Liebl et al. (1996) noted an increased percentage of the charged residue Glu and a slight increase in Pro in CelA and CelB from *T. maritima* compared to the less thermostable cellulases of family 12. By analysing differences in amino acid composition between the thermophilic and mesophilic enzymes in family 12 given in Table 1, it is evident that the percentage of Glu is higher in the thermophilic candidates, but an increase in Pro content is not observed. The increased percentage of Glu among the thermostable enzymes is not surprising since charged residues play a role in structure stabilization by salt bridge formation and by forming hydrogen bonds with neutral residues. Correlation between thermostability and the number of salt bridges and hydrogen bonds in proteins has indeed been observed (Macedo-Ribeiro et al. 1996; Szilagyi and Zavodszky 1995; Tanner et al. 1996).

Glycosyl hydrolases hydrolyse glycosidic bonds via general acid catalysis that requires two amino acid residues: a proton donor and a nucleophile/base. In most glycosyl hydrolases studied, only Asp and/or Glu residues have been found to perform catalysis (Davies and Henrissat 1995). Indeed, two conserved Glu residues are found in *celA* from *R. marinus*, Glu^{159} and Glu^{242}. These amino acid residues correspond to Glu^{158} and Glu^{246} in *E. carotovora* CelS, and Glu^{134} and Glu^{218} in *A. aculeatus* FI-CMCase. Törronen et al. (1993) have compared the *E. carotovora* and *A. aculeatus* enzymes to 17 enzymes from glycosyl hydrolase family 11, where the catalytic role of Glu residues has been estabilished (Ko et al. 1992). They predicted that the two conserved Glu residues in the family 12 enzymes are essential for catalytic activity.

Although a variety of factors have been implicated in the origins of thermostability, no general conclusion has emerged (Danson et al. 1996). It has been suggested that both thermophilic and mesophilic proteins attain a similar degree of flexibility at their respective optimum temperatures (Jaenicke 1996). Furthermore, comparisons of the three-dimensional structures of thermostable proteins to those of their corresponding mesophilic counterparts show that high thermostability can be achieved without large changes of overall protein structure (Macedo-Ribeiro et al. 1996; Delboni et al. 1995) It is likely that the same would apply for the thermostable cellulases of glycosyl hydrolase family 12. However, it is clear that this question can only be resolved by determining the three-dimensional structure of thermophilic and mesophilic enzymes of this family.

Acknowledgements This work was supported by grants from the Icelandic National Research Council and the EC generic project Biotechnology of Extremophiles ERB BI04 CT96 0488.

References

Alfredsson GA, Kristjansson JK, Hjorleifsdottir S, Stetter KO (1988) *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in Iceland. J Gen Microbiol 134: 299–306

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Andrésson ÓS, Fridjónsson ÓH (1994) The sequence of the single 16S rRNA gene of the thermophilic eubacterium *Rhodothermus marinus* reveals a distant relationship to the group containing *Flexibacter*, *Bacteroides* and *Cytophaga* species. J Bacteriol 176: 6165–6169

Bayer EA, Morag E, Lamed R (1994) The cellulosome – a treasure-trove for biotechnology. Trends Biotechnol 12: 379–386

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

Brendel V, Bucher P, Nourbakhsh I, Blaisdell BE, Karlin S (1992) Methods and algorithms for statistical analysis of protein sequences. Proc Natl Acad Sci USA 89: 2002–2006

Chung CT, Niemeda SL, Miller RH (1989) One step preparation of competent *Escherichia coli*: transformation and storage in the same solution. Proc Natl Acad Sci USA 82: 2172–2175

Danson MJ, Hough DW, Russell RJM, Taylor GL, Pearl L (1996) Enzyme thermostability and thermoactivity. Prot Eng 9: 629–630

Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. Structure 3: 853–859

Degryse E, Glansdorff N, Piérard A (1978) A comparative analysis of extreme thermophilic bacteria belonging to the genus *Thermus*. Arch Microbiol 117: 189–196

Delboni LF, Mande SC, Rentier-Delrue F, Mainfroid V, Turley S, Vellieux FM, Martial JA, Hol WG (1995) Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. Prot Sci 4: 2594–2604

Henrissat B (1991) A classification of glycosyl hydrolases based on amino-acid sequence similarities. Biochem J 280: 309–316

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

Herbert RA (1992) A perspective on the biotechnological potential of extremophiles. Trends Biotechnol 10: 395–402

Hreggvidsson GO, Helgason E, Petursdottir S, Kristjansson JK (1992) Comparison of α-amylases from three phylogenetically distant genera of thermophilic bacteria: *Bacillus*, *Thermus*, and *Rhodothermus*. *In:* Geirsdóttir AM, Brown HP, Skjenstad, T (eds) Conference programme and abstracts of the International Conference on Thermophiles. Science and technology, Reykjavik, Iceland Ice Tech, Reykjavik p 26

Hreggvidsson GO, Kaiste E, Holst O, Eggertsson G, Palsdottir A, Kristjansson JK (1996) An extremely thermostable cellulase from the thermophilic eubacterium *Rhodothermus marinus*. Appl Environ Microbiol 62: 3047–3049

Jaenicke R (1996) Stability and folding of ultrastable proteins: eye lens crystallins and enzymes from thermophiles. FASEB J 10: 84–92

Ko EP, Akatsuka H, Moriyama H, Shinmyo A, Hata Y, Katsube Y, Urabe I, Okada H (1992) Site-directed mutagenesis at aspartate and glutamate residues of xylanase from *Bacillus pumilus*. Biochem J 288: 117–121

Kristjansson JK (1989) Thermophilic organisms as sources of thermostable enzymes. Trends Biotechnol 7: 349–353

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685

Liebl W, Ruile P, Bronnenmeier K, Riedel K, Lottspeich F, Greif I (1996) Analysis of a *Thermotoga maritima* DNA fragment encoding similar thermostable cellulases, CelA and CelB, and characterization of the recombinant enzymes. Microbiology 142: 2533–2542

Macedo-Ribeiro S, Darimont B, Sterner R, Huber R (1996) Small structural changes account for the high thermostability of 1[4Fe-4S] ferridoxin from the hyperthermophilic bacterium *Thermotoga maritima*. Structure 4: 1291–1301

Manelius Á, Dahlberg L, Holst O (1994) Some properties of a thermostable β-xylosidase from *Rhodothermus marinus*. Appl Biochem Biothechnol 44: 39–48

Nakai K, Kanehisa M (1991) Expert system for predicting protein localization sites in gram-negative bacteria. Proteins 11: 95–110

Nordberg Karlsson E, Bartonek-Roxa E, Holst O (1997) Cloning and sequence analysis of a thermostable multidomain xylanase from the bacterium *Rhodothermus marinus*. Biochim Biophys Acta 1353: 118–124

Spilliaert R, Hreggvidsson GO, Kristjansson JK, Eggertsson G, Palsdottir A (1994) Cloning and sequencing of a *Rhodothermus marinus* gene, *bglA*, coding for a thermostable β-glucanase and its expression in *Escherichia coli*. Eur J Biochem 224: 923–930

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 185: 60–89

Sumner JB, Somers GF (1949) Dinitrosalisylic method for glucose. In: Laboratory experiments in biological chemistry. Academic Press, New York

Szilagyi A, Zavodszky P (1995) Structural basis for the extreme thermostability of D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: analysis based on homology modelling. Prot Eng 8: 779–789

Tanner JJ, Hecht RM, Krause KL (1996) Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 25 angstroms resolution. Biochemstry 35: 2597–2609

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

Thorbjarnardottir SH, Jónsson ZO, Andrésson ÓS, Kristjánsson JK, Eggertsson G, Palsdottir A (1995) Cloning and sequence analysis of the DNA ligase-encoding gene of *Rhodothermus marinus*, and overproduction, purification and characterization of two thermophilic ligases. Gene 161: 1–6.

Tomme P, Warren RAJ, Gilkes NR (1995) Cellulose hydrolysis by bacteria and fungi. Adv Microbial Physiol 37: 1–18

Törrönen A, Kubicek CP Henrissat B (1993) Amino acid sequence similarities between low molecular weight endo-1,4-β-xylanases and family H cellulases revealed by clustering analysis. FEBS Lett 321: 135–139

Wood PJ, Erfle JD, Teather RM (1988) Use of complex formation between Congo red and polysaccharides in detection and assay of polysaccharide hydrolases. Methods Enzymol 160: 56–74

Zaman Z, Verwilghen RL (1979) Quantitation of proteins solubilized in sodium dodecyl sulfate-mercaptoethanol-tris electrophoresis buffer Anal Biochem 100: 64–69