Molecular Cloning of a Cellulase Gene from Abalone *Haliotis discus hannai* and Its Expression in *E. coli*

Eun-Mi Park, Yun-Hee Han, In-Suk Park, Bo-Hye Nam, Hee Jeong Kong, Woo-Jin Kim, Sang-Jun Lee and Young-Ok Kim*

Biotechnology Research Institute, National Fisheries Research & Development Institute, 408-1 Sirang-ri, Gijang-eup, Gijang-gun, Busan 619-705, Republic of Korea

Abstract A cellulase (endo-ß-1,4-D-glucanase(E.C.3.2.1.4)) was isolated from the hepatopancreas of abalone *Haliotis discus hannai* by EST analysis. The abalone cellulase named HdEG compassed 1977 bp, including 195 bp in the 5'untranslated region, 1680 bp in the open reading frame which encodes 560 amino acid residues, and 92 bp in the 3'-untranslated region. The C-terminal region of the HdEG showed 44-52% identity to the catalytic domains of glycoside hydrolase family 9 (GHF9)-cellulases from arthropods and bacteria. The recombinant cellulase, pEHdEG was produced in *E. coli* with being fused with C-terminal His-tag. The expressed protein showed a single band (~62 kDa) on Western blotting which was consistent with the value (61,878 Da) calculated from the DNA sequence.

Key words : Cellulase, abalone, cDNA cloning, recombinant DNA

Introduction

Cellulose is the most abundant organic compound produced by terrestrial plants, where it forms the main structural component of cell wells [2]. The cellulose molecule is essentially a linear homopolymer, consisting of glucose units linked by ß-1,4 bonds. These polymers bond together to form highly ordered (crystalline) structures, and less ordered (amorphous) regions, which are more prone to enzyme degradation [3].

Endoglucanase (endo-1,4-ß-D-glucanase, EC 3.2.1.4), cellobiohydrolase (CBH, exo-1,4-ß-D-glucanase, EC 3.2.1.91), and ß-glucosidase (EC 3.2.1.21) are three major types of cellulolytic enzymes. Endoglucanases randomly hydrolyze ß-1,4 bonds along the interior of the cellulose chain. Cellobiohydrolases cleave cellobiosyl units from non-reducing ends of the cellulose chains. ß-Glucosidases cleave glucosyl units from non-reducing ends of cello-oligosaccharides. The cellulase has been shown to exist not only in plants [5], molds [4], fungi [18], bacteria [18] and protista [9] but also in herbivorous invertebrates such as arthropods [16,19,23], nematodes [14] and mollusks [1,8,9,21,25]. Most cellulases from microorganisms are composed of a catalytic domain and ancillary domains such as CBMs (carbohydrate-binding module) and linkers, while the invertebrate cellulases except for two nematode enzymes have just a catalytic domain [18, 15]. The origin of the invertebrate cellulases was initially explained as products of symbiotic microorganisms in the intestine [7]. However, those cellulases have become considered to be the products of invertebrates themselves [10] and cellulase genes were cloned from termite [20], crayfish [6], nematode [24], and mussel [22].

In this study, we cloned a cellulase gene from abalone *Haliotis discus hannai* which is one of the most common and valuable herbivorous molluscs in Korea, and determined its primary structure. In addition, recombinant cellulase was produced in *Escherichia coli.*

* Corresponding author
Phone: +82-51-720-2455, Fax: +82-51-720-2456
E-mail: yokim@momaf.go.kr

Materials and Methods

Cloning and sequencing of HdEG gene

Previously, we constructed a cDNA library from the hepatopancreas of 3-yr-old abalones (*Haliotis discus hannai*) and analyzed expressed sequence tag (EST) of 110 clones [11]. The EST clone VHP-059, which carries a 780-bp insert, showed significant similarity to cellulase. The full-length cDNAs, which was named as HdEG were amplified from the first-strand cDNAs using 5'race and 3'race using primers based on the sequences using the SMART RACE Amplification Kit (Clontech).

Sequence analysis

Sequencing was performed using the ABI 3100 automatic DNA sequencer (PE Applied Biosystem, CA, USA) and the ABI Prism Big Dye Terminatior Cycle Sequencing Ready Reaction kit (PE Applied Biosystems). The nucleotide sequences and deduced amino acid sequences were analyzed using the Genetyx-Win program ver. 4.0 (Genetyx Co, Japan). Signal peptides were predicted using the Signal P program. Multiple alignments of the cellulase proteins were constructed using the ClustalW program and visualized with MEGA (ver. 3).

Construction of HdEG cellulase expression plasmid

For construction of *E. coli* expression vector, the cellulase gene containing signal peptide was amplified using a sense *Nde*I-linker primer (5′-AGGCATATGAAG CTGAGTCAG-3′) and an antisense *Xho*I linker primer (5′-TACAAGGACATCTGCCCTCTCGAGCAC-3′) by PCR. The amplified product was cloned into *Nde*I and *Xho*I sites of expression vector pET22b (Novagen). The construction plasmid was designated as pEHdEG.

Expression of recombinant protein in *E. coli*

E. coli BL21(DE3) cells harboring the pEHdEG was isolated and inoculated into 100 ml LB medium containing ampicillin (100 μg/ml). For induction of the recombinant protein, isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 1 mM when the cell density achieved at OD_{600}=0.6~0.8. The cells were cultivated for further 12-14 hr at 19°C after induction and the culture supernatant and pellet were collected for the protein analysis.

Analytical method

For protein analysis, the proteins were separated by SDS-PAGE on a 12% polyacrylamide gel. The gels were stained with Coomassie brilliant blue R-250. Recombinant HdEG proteins containing a His-tag were transferred to a nitrocellulose membrane (Roche). After locking with 5% skimmed milk (Difco), the membrane was incubated with anti-His6 mAb for 2h at 37°C. HRP-conjugated rabbit antimouse IgG was used as secondary antibody, and 3,3'-diaminobenzidine tetrahydrochloride (DAB; Sigma-Aldrich) as the staining substrate.

Results and Discussion

Isolation and analysis of abalone cellulase cDNA

Among the 110 of EST clones, the VHP-059 was revealed to be homologous to known cellulase genes of other species by computer homology search of the public database. The nucleotide and deduced amino acid sequences of HdEG were shown in Fig. 1. The full cDNA sequence of HdEG spans 1977 bp and comprises a 5'-UTR of 195 bp, a 3'-UTR of 92 bp and an ORF of 1680 bp that encodes a polypeptide of 560 amino acids. The deduced polypeptide has a molecular weight of 62 kDa and an expected isoelectric point of 5.39. In the 3'-terminal region, a putative polyadenylation signal sequence AATAAA and a poly (A+) tail were found. These structural characteristics indicate that the HdEG cDNA is not derived from prokaryote like intestinal bacteria. By the comparison with consensus sequence for signal peptides of eukaryote secretory proteins, we could not find putative signal peptide region in the HdEG sequence. Accordingly, this abalone cellulase is considered as a non-secretory protein, which may be due to differences in cellulase function throughout their evolution.

Alignments and phylogenetic tree construction

By sequence comparison with other invertebrate and bacterial cellulases, the C-terminal region of 400 residues in the HdEG was regarded as the glycoside hydrolase family (GHF)9-type catalytic domain i.e. it showed 44-52% identity with the corresponding regions of abalone HdEG [15], termite RsEG [20], crayfish CqEG [6], and *Paenibacillus* sp. BP-23 [12] cellulases, respectively (Fig. 2). Further, the catalytically important

residues in GHF9 cellulases [13, 17], i.e. Asp169, Asp172, His472, Asp517 and Glu526 in the HdEG sequence were all conserved (Fig. 2). Based on these results, we conclude that HdEG is classified into GHF9.

A phylogenetic tree was generated using MEGA. The HdEG showed the closest relationship with another abalone cellulase [15] and they formed a cluster with the GHF9 proteins maintaining a high bootstrap value (Fig. 3).

Heterologous expression of abalone cellulase

We constructed the expression plasmid, pEHdEG as described in materials and methods. Because the most of pEHdEG expressed in *E. coli* was accumulated as inclusion body in the cells at 37°C induction (data not shown), induction temperature was reduced to 19°C. Induction of the transformant resulted in slow bacterial growth and western blot analysis of cell lysates showed that the recombinant protein was expressed (Fig. 4). When both pellets and supernatants from bacterial lysates were analyzed, results showed that the HdEG protein was located in the cytoplasm somewhat and mostly present in the insoluble fraction. The expressed protein has a molecular mass of ~62 kDa, a consistent with a monomer size from deduced amino acid sequence. Further studies are needed to clarify the biochemical properties of the expressed HdEG protein and also the tissue-specific expression and the function of the cellulase in *Haliotis discus hannai.* In the long term, application of such studies may assist the development of improved artificial diets for cultured aquatic species through the incorporation of low-cost feed components sourced from plant material.

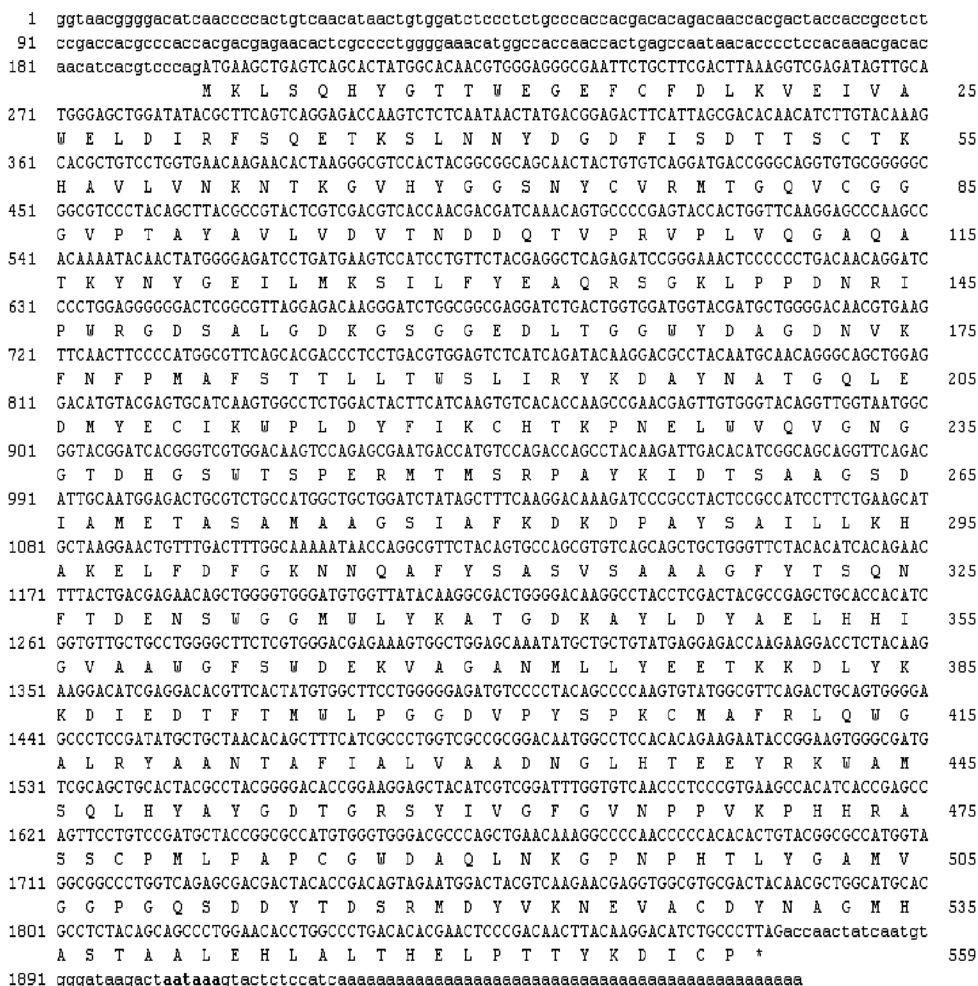

Fig. 1. The nucleotide and deduced amino-acid sequences of the HdEG.. Residue numbers for nucleotide and amino acid are indicated in the left and right of each row, respectively. 5' and 3' untranslated region are given in lowercase letters.

```
HdEG    148 RGDSALGDKGSGGEDLTGGWYDAGDNVKFNFPMAFSTTLLTWSLIRYKDAYNATGQLEDM 207
HdEG66  179 RGDSALGDKGDNGEDLTGGWYDAGDHVKFSLPMSSTSTVLLWGYLQWKDAYATTKQTDMF 238
RsEG    48  RKDSALNDKGQKGEDLTGGYYDAGDFVKFGFPMAYTVTVLAWGVIDYESAYSAAGALDSG 107
CqEG    60  RGDSALNDGSDVGHDLTGGYYDAGDHVKFGFPMAFTATMLAWGQIDFANGHSKAGQTSYG 119
BP-23   69  RGDSGMQDGADVGVDLTGGWYDAGDHVKFGFPMAASATMLAWSVVEYADGYEQAGQLEEI 128

HdEG    208 YECIKWPLDYFIKCHT-KPNELWVQVGNGGTDHGSWTSPERMTMSRPAYKIDTSAAGSDI 266
HdEG66  239 FDMIKWPLDYFLKCWIPKSQTLYAQVGEGNDDHHFWGRAEDMKMARPAYKLTPSKPGSDV 298
RsEG    108 RKALKYGTDYFLKAHT-AANEFYGQVGQGDVDHAYWGRPEDMTMSRPAYKIDTSKPGSDL 166
CqEG    120 HAALKWATDYFLKAHT-ATNEFYGQVGEGNTDHAYWGRPEDMTMSRPSYKIDQSRPGSDL 178
BP-23   129 KDNIRWATDYFMKAHT-KPNELWGQVGAGNTDHAWWGPAEVMQMNRPSFKIDASCPGSDL 187

HdEG    267 AMETASAMAAGSIAFKDKDPAYSAILLKHAKELFDFGKNNQAFYSASVSAAAGFYTSQ-N 325
HdEG66  299 AGEIAASLAAGYLAFKQRDAKYAATLLSTSKEIYEFGKKYPGIYSSSIQDAGQFYSSS-G 357
RsEG    167 AAETAAALAATAIAYKSADATYSNNLITHAKQLFDFANNYRGKYSDSITDAQNFYASG-D 225
CqEG    179 AGETAAALAAASIVFKSSDSSYSSQVLSVAKELYEFADQHRDIYTNAITDAASFYNSWSG 238
BP-23   188 AAETAAALAASSIVFADSDPVYSAKLLQHAKELYNFADTYRGKYTDCITDAAAFYNSWTG 247

HdEG    326 FTDENSWGGMWLYKATGDKAYLDYAELHHIGVAAW----------GFSWDEKVAGANMLL 375
HdEG66  358 YKDEMCEGAMWLYKATGDKSYLADAKGYHENAWAW----------ALGWDDKKIACQLLL 407
RsEG    226 YKDELVWAAAWLYRATNDNTYLTKAESLYNEFGLGNWNG------AFNWDNKISGVQVLL 279
CqEG    239 YGDELAWAALWLARATGDNSYLDRAKGHVSEFNLLGTPS------QFGWDDKKAGVQALL 292
BP-23   248 YEDELAWGGAWLYLATNDNAYLSKALSAADRWSTSGGSANWPYTWTQGWDSKHYGAQILL 307

HdEG    376 YEETK------KDLYKKDIEDTFTMWLPG---GDVPYSPKCMAFRLQWGALRYAANTAFI 426
HdEG66  408 YEATK------DTAYKTEVEGFFKGWLPG---GSITYTPCGQAWRDKWGSNRYAANSAFA 458
RsEG    280 AKLTS------KQAYKDKVQGYVDYLIS-----SQKKTPKGLVYIDQWGTLRHAANSALI 328
CqEG    293 VLLDG------SSEYTNALNQFLNFVRN-----QAPYTPEGLVFLDAWGSNRHAANVAFI 341
BP-23   308 ARITSNLNMPEATKFIQSTERNLDYWTVGTNGGRVKYTPGGLAWLDQWGSLRYAANAAFI 367

HdEG    427 ALVAADNGLHTEEY---RKWAMSQLHYAYGDT--GRSYIVGFGVNPPVKPHHRASSCPML 481
HdEG66  459 ALVAADAGIDTVTY---RKWAVEQMNYILGDNKYGISYQIGFGTKYPRNPHHRSASCPDI 515
RsEG    329 ALQAADLGINAATY---RAYAKKQIDYALGDGGR--SYVIGFGTNPPVRPHHRSSSCPDA 383
CqEG    342 ALYAAKLGIDAGTN---QQWARGQIGQLLGDNSRYQSFVVGFGVNPPTRPHHRSSSCPDR 398
BP-23   368 SFVYSDWVSDPVKKSRYQNFATSQINYILGDNPRQSSYVVGYGQNSPQHPHHRTAHSS-- 425

HdEG    482 PAPCGWDAQLNKGPNPHTLYGAMVGGPGQSDDYTDSRMDYVKNEVACDYXAGMHASTAAL 541
HdEG66  516 PAPCSETNLHTAGPSPHILVGAIVGGPDNDDSYKDNREDYVHNEVACDYNSGFQSALAGL 575
RsEG    384 PAVCDWNTYNSAGPNAHVLTGALVGGPDSNDSYTDARSDYISNEVATDYNAGFQSAVAGL 443
CqEG    399 PADCSNGLTNSG-PNPQTLWGALVGGPAQDGSYNDDRQDYQHNEVACDYNAAYTGALAAL 457
BP-23   425 ---WMNNEDIPA-NHRHILYGAMVGGPNASDQYTDDIGDYVSNEVATDYNAGFTGALAKM 481
```

Fig. 2. Comparison of amino acid sequences for the HdEG and other cellulases. The sequence of C-terminal 400 residues fro the HdEG was aligned with the sequences for catalytic domains of HdEG66 (*Haliotis discus hannai),* RsEG (*Reticulitermes speratus*), CqEG (*Cherax quadricarinatus*), and BP-23 (*Paenibacillus* sp. BP-23). Catalytically important residues in GHF9 are boxed.

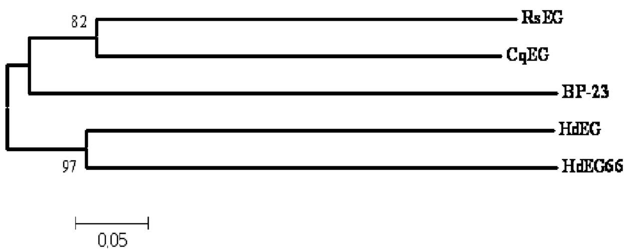

Fig. 3. Phylogenetic relationship based on deduced amino acid sequences of other cellulases. A phylogenetic tree of the aligned sequences was constructed using the Neighbour-Joining algorithm within MEGA (version 3.0). The degree of confidence for each branch point was determined by bootstrap analysis (1000 repetitions).The scale bar is 0.05, which refers to percentage divergence.

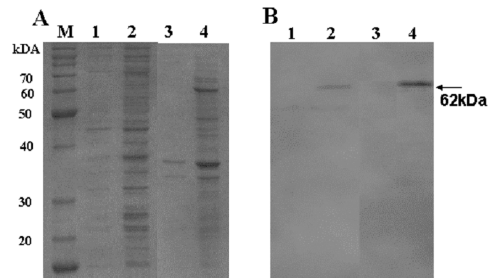

Fig. 4. SDS-PAGE (A) and western blotting (B) analysis of pEHdEG expressed in *E. coli*. *E. coli* BL21(DE3)/pEHdEG cells were induced with 1 mM IPTG at 19°C. Pelleted cells were lysed and separated into pellet and supernatant. Lane 1-2, sonicated supernatant of E. coli BL21 (DE3)/pEHdEG cultured for 0h and 12h after 1 mM IPTG induction. Lane3-4, sonicated pellet of *E. coli* BL21 (DE3)/pEHdEG cultured for 0h and 12h after 1 mM IPTG induction. Standards are shown to the left.

Acknowledgments

This work was supported by a grant from the National Fisheries Research and Development Institute (NFRDI), Republic of Korea.

References

1. Anzai, H., Nisizawa, K. & Matsuda, K. 1984. Purification and characterization of a cellulose from *Dolabella auricularia*. *J. Biochem.* **96**, 1381-1390.
2. Aspinall, G..O., 1980. Chemistry of cell wall polysaccarides. In: Preiss, J. (Ed.), Carbohydrates: Structure and fuction, The Biochemistry of plants vol.3. Academic Press, NY, pp.473-500.
3. Berguin, P. and Aubert, J.P. 1994. The biological degradation of cellulose. *FEMS Microbiol. Rev.* 13, 25-28.
4. Blume, J.E. and Ennis, H.L. 1991. A Dictyostelium discoideum cellulase is a member of a spore germination-specific gene family. *J. Biol. Chem.* **266**, 15432-15437.
5. Brummell, D.A., Lashbrood, C.C. and Bennett, A.B. 1994. Plant endo-1,4-β-D-glucanases. *ACS Symp. Series* **566**, 100-129.
6. Byrne, K.A., Lehnert, S.A., Johnson, S.E. and Moore, S.S. 1999. Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. *Gene* **239**, 317-324.
7. Cleveland, L.R. 1924. The physiological and symbiotic relationships between the intestinal protozoa of temites and their host with special reference to *Reticulitermes flavipes* Kollar. *Biol. Bull. Mar. Biol. Lab.* **46**, 117-227.
8. Marshall, J.J. 1973. Purification of a β-1,4-glucan hydrolase (cellulase) from the snail, *Helix pomatic*. *Camp. Biochem. Physiol.* **33B**, 981-988.
9. Moriya, S., Ohkuma, M. and Kudo, T. 1998. Phylogenetic position of symbiotic protest *Dinemympha exilis* in the hindgut of the termite *Reticulitermes speratus* inferred from the protein phylogeny of elongation factor 1α. **Gene** **210**, 221-227.
10. O'Brien, G.W., Veivers, P.C., McEwen, S.E., Slaytor, M. and O'Brien, R.W. 1979. The origin and distribution of cellulase in the termites, *Nasutitermes exitiosus* and *Coptotermes lacteus*. *Insect Biochem.* **9**, 619-625.
11. Park, E.M., Nam, B.H., Kim, Y.O., Kong, H.J., Kim, W.J., Lee, S.J., Kong, I.S., Choi, T.J. 2007. EST-based survey of gene expression in seven tissue types from the abalone *Haliotis discus hannai*. *J. Fish. Sci. Technol.* **10**, 119-126.
12. Pastor, F.I., Pujol, X., Blanco, A., Vidal, T., Torres, A.L. and Diaz, P. 2001. Molecular cloning and characterization of a multidomain endoglucanase from *Paenibacillus* sp BP-23: evaluation of its performance in pulp refining. *Appl. Microbiol. Biotechnol.* **55**, 61-68.
13. Sakon, J., Irwin, D., Wilson, D.B. and Karplus, P.A. 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat. Struct. Biol.* **4**, 810-818.
14. Smant, G., Stokkermans, J.P.W.G., Yan, Y., De Boer, J.M., Baum, T.J., Wang, X., Hussey, P.S., Gommers, F.J., Henrissat, B., Davis, E.L., Helder, J., Schots, A. and Bakker, J. 1998. Endogenous cellulases in animals isolation of β-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc. Natl Acad. Sci. USA* **95**, 4906-4911.
15. Suzuki K.I., Ojima, T. and Nishita K. 2003. Purification and cDNA cloning of a cellulase from abalone *Haliotis discus hannai*. *Eur. J. Biochem.* **270**, 771-778.
16. Tokuda, G., Watanabe, H., Matsumoto, T. and Nodal, H. 1997. Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis* (Shiraki): distribution of cellulases and properties of endo-β-1,4-glucanase. *Zool. Sci.* **14**, 83-98.
17. Tomme, P., Chauvaux, S., Beguin, P., Millet, J., Aubert, J.-P. and Claeyssens, M. 1991. Identification of a histidyl residue in the active center of endoglucanase D from *Clostridium themocellum*. *J. Biol. Chem.* **266**, 10313-10318.
18. Tomme, P., Waren, R.A.J. and Gilkes, N.R. 1995. Cellulose hydrolysis by bacteria and fungi. *Adv. Microbiol. Physiol.* **37**, 1-81.
19. Watanabe, H., Nakamura, M., Tokuda, G., Yamaoka, I., Scrivener, A.M. and Noda, H. 1997. Site of secretion and properties of endogenous endo-β-1,4-glucanase components from *Reitculitermes speratus* (Kolbe), a Japanese subterranean termite. *Insect Biochem. Mol. Biol.* **27**, 305-313.
20. Watanabe, H., Noda, H., Tokuda, G. and Lo, N. 1998. A cellulase gene of termite origin. *Nature* **394**, 330-331.
21. Xu, B., Hellman, U., Ersson, B. and Janson, J.C. 2000. Purification, characterization and amino-acid sequence analysis of a thermostable, low molecular mass endo-β -1,4-glucanase from blue mussel, *Mytilus edulis*. *Eur. J. Biochem.* **267**, 4970-4977.
22. Xu, B., Janson, J.-C. and Sellos, D. 2001. Cloning and sequencing of a molluscan endo-β-1.4-glucanase gene from the blue mussel, *Mytilus edulis*. *Eur. J. Biochem.* **268**, 3718-3727.
23. Xue, X.M., Anderson, A.J., Richardson, N.A., Anderson, A.J., Xue, G.P. and Mather, P.B. 1999. Characterization of cellulase activity in the digestive system of the redclaw crayfish (*Cherax quadricarinatus*). *Aquaculture* **180**, 373-386.
24. Yan, Y., Smant, G., Stokkermans, J., Qin, L., Helder, J., Baum, T., Shots, A. and Davis, E. 1998. Genomic organization of four β-1,4-endoglucanase genes in plant-parasitic cyst menatodes and its evolutionary implications. *Gene* **220**, 61-70.
25. Yokoe, Y. and Yasumasu, I. 1964. The distribution of cellulase in invertebrates. *Comp. Biochem. Physiol.* **13**, 323-338.