BARLEY CELLULOSE SYNTHASES INVOLVED IN SECONDARY CELL WALL FORMATION AND STEM STRENGTH: GENERATION OF cDNA CONSTRUCTS FOR FUNCTIONAL ANALYSIS

by

Partha Varanashi

A thesis submitted for the partial fulfillment of the requirements of the Masters of Biotechnology (Plant Biotechnology)

The University of Adelaide

Faculty of Sciences

School of Agriculture, Food and Wine

Waite Campus

2008

DECLARATION

I declare that this thesis is a record of original work and contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text.

Partha Varanashi

Associate Professor Maria Hrmova

PREFACE

This research was performed over 10 months as part of a Masters in Biotechnology (Plant Biotechnology). The literature review was previously assessed in accordance with the correction suggested by the examiners. The main focus of the project remains very similar to that of the research proposal. However the goals were not achieved according to the time deadline stated in the research proposal. Hence protein purification was could not be carried out.

Although the research manuscript contained herein will provide the first draft of a future publication to be submitted to Plant journal, due to time constraint, all data relevant to that publication has not been collected. However, additional data which was not conclusive was collected and this is provided within the appendices. The research manuscript outlines stages involved in the construction and heterologus expression of barley CesA4 cDNA. While the appendices contain additional data from HvCesA4 protein structure prediction, media recipes, in-silico representation of the HvCesA4 constructs with respective vectors.

Title: BARLEY CELLULOSE SYNTHASES INVOLVED IN SECONDARY CELL WALL FORMATION AND STEM STRENGTH: GENERATION OF cDNA CONSTRUCTS FOR FUNCTIONAL ANALYSIS

Author:
Partha Varanashi
The University of Adelaide
Plant Genomics Centre
Glen Osmond 5064, Australia
Email partha.varanashi@student.adelaide.edu.au

Corresponding author:
Dr Maria Hrmova
Associate Professor
The University of Adelaide
Plant Genomics Centre
Glen Osmond 5064, Australia
T: +61 8 8303 7280
F: +61 8 8303 7102
Email: maria.hrmova@adelaide.edu.au

Table of Contents

PART I- LITERATURE REVIEW & RESEARCH PLAN

Introduction

Barley is ranked as the fourth largest crop grown in the world based on quantity of production. One of the major problems farmers face during barley cultivation is crop lodging. Owing to weak stems, barley crops cannot withstand heavy winds and hence fall over. Crop lodging may result in a loss of crop production or in difficulties during harvesting. The stem strength depends on a type of cells present in the stem region of the plant. Secondary plant cell walls present in each cell are known to provide the stem strength. These secondary cell walls are made up of cellulosic microfibrils embedded in a matrix containing complex non-cellulosic polysaccharides; usually these are pectic polysaccharides, xyloglucans and heteroxylans as major constituents, and heteromannans as minor constituents. In vascular plants the major constituent of cell walls are found to be cellulose, which constitutes (1-4)-β-D-glucan microfibrils. In recent discoveries candidate genes for biosynthesis of key components of secondary cell walls have been tracked down to cellulose synthase (CesA) and cellulose synthase like (Csl) gene families (Richmond 2000).

It is now well known that assembly of cellulose or (1-4)-β-D-glucan microfibrils are likely to be carried out by a certain set of enzymes (Doblin *et al.*, 2002). Among the cellulose synthase gene family, cellulose synthase A (CesA) genes in plants are thought to be responsible for encoding glycosyl transferases, which plays a key roles in biosynthesis of cellulose (Delmer, 1999). So far in barley nine HvCesA genes have been designated (Burton et al., 2004). These represent HvCesA1 to HvCesA8. In barley, however, the functional analyses of the HvCesA genes are incomplete. Thus functional analysis of

these proteins is important for understanding of a complete process of cellulose biosynthesis. Complete information generated on the enzymes encoded by the *HvCesA* gene would aid in manipulation of genes responsible for cellulose biosynthesis. Manipulations would be aimed at improving stem strength of barley plants, thus helping in reducing crop lodging and undoubtedly increasing the yields.

Literature Review

Introduction

One of the most important distinguishing features of plant cells is the presence of a cell wall on their surfaces. Plant cell walls are bound by rigid walls, which are mainly made up of cellulosic, non cellulosic and pectic compounds. The amount of each compound varies from plant to plant. These cell walls provide the exoskeletal structure for the plant and are crucial to the structure and growth of the plant. They also contribute to the shape and morphology of the plant. Cell walls also take part in regulating cell growth along with providing the structural and mechanical support for the plant (Bacic *et al.*, 1988). Cellulose holds the key role in the above activities of cell wall, thus attracting the interest of researchers. This literature review mainly focuses on the biosynthesis of cellulose, which involves cellulose synthase (Ces) genes. Few genes are found in the cellulose synthase gene family. They are very similar to each other and are conserved in most of the higher plants.

Cellulose

Among polymers, cellulose is found to be the most plentiful and is a major constituent of vascular plants (Brown, 2004). Cellulose also contributes to growth and cell division.

Plant cell walls are made up of cellulose and other non-cellulosic components. Cellulose microfibrils are insoluble, cable-like structures composed of 36 hydrogen bonded (1-4)-β-D-glucan chains (Fig. 1) (Finaev, 2007). These 36 chains are arranged in parallel arrays to form para-crystalline microfibrils. The microfibrils vary in diameter from primary cell wall to secondary cell wall. In primary cell wall the diameter of paracrystalline microfibrils is found to be 3 nm and in secondary cell walls the diameter is found to be 5-10 nm (Carpita and Gibeaut, 1993).

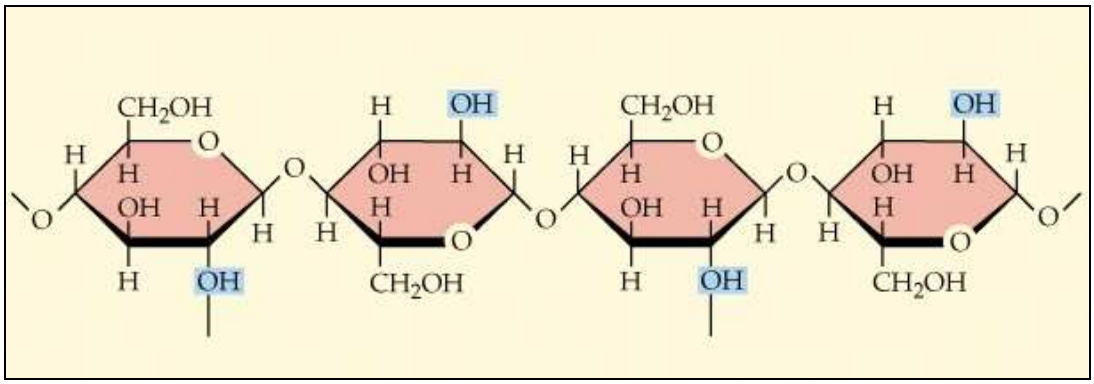


Figure 1. Structure of cellulose showing the (1-4)-β-D glucan chain.

In order to study the structure of cellulose, chemists have extracted and purified the polymer from plant cell walls. However, cellulose thus obtained may have different properties than native cellulose (Brett, 2000). Hence, studies on cellulose biosynthesis at molecular levels using purified or recombinant proteins are preferable, and the results obtained from these studies would aid to a better understanding of cellulose biosynthesis and structure.

Cellulose Biosynthesis

Cellulose biosynthesis in vascular plants is found to be co-localised within plasma membranes. Freeze fractured plasma membranes of vascular plants have shown that

cellulose synthase forms complexes and appear as a rosette of six globular particles arranged symmetrically. These six globular particles seems to produce (1-4)-β-D-glucan molecules, which co-crystallize into cellulosic microfibrils (Roelofsen,1958; Mueller and Brown 1980; Kimura at al. 1999; Kimura, 2002). This rosette terminal complex (Fig. 2) was also seen at the ends of cellulose microfibrils, which were synthesized *in-vitro* using membrane extracts of suspension-cultured cells of *Rubus fructicosus* (Lai-Kee-Him *et al.*, 2002). Once the cellulose microfibrils are ordered in a specific orientation, the direction of cell elongation is essentially decided. These microfibrils, which are long and inelastic are found in the secondary cell wall region and provide the mechanical strength for the cell

NOTE: This figure is included on page 10 of the print copy of the thesis held in the University of Adelaide Library.

Figure 2. Biosynthesis of cellulose in plama membrane of a vascular plant (Somerville, 2006).

Cellulose synthase gene family

A gene family responsible for the biosynthesis of (1-4)-β-D-glucan molecules, or cellulose, have been designated cellulose synthases (Ces) (Pear *et al.*, 1996). The

nucleotide sizes of barley *CesA* genes vary between 3.5 to 5.5 kb and these genes are known to have 8 to 9 introns (Finaev, 2007).

NOTE: This figure is included on page 11 of the print copy of the thesis held in the University of Adelaide Library.

Figure 3. A schematic model of CesA showing its organization within plasma membrane (http://www.unimuenster.de/Biologie. Botanik/agschaewen/ forschung/cesa/e_cesa.html).

Cellulose synthase genes from barley are translated into approximately 1,100 amino acid residues. The protein sequences of CesAs have a large central domain of approximately 530 amino acids (Pear *et al.*, 1996). This catalytic centre lies between the two regions bound by trans-membrane helices or domains and is thought to be localized in the cytoplasm. The trans-membrane domains on the other hand are thought to be positioned in membrane regions. *In-silico* studies of barley CesAs, using the protein topology prediction tools predict that their NH2-termini co-locate in cytoplasm (Fig. 3). Interestingly, the central catalytic domain or the catalytic centre among all CesA proteins, which have been studied so far, seems to be highly conserved. Further, in barley cellulose synthase proteins (HvCesAs) the motif QXXRW and catalytic residues SDD and TED

are highly conserved among all HvCesAs. The cellulose synthase proteins are classified within glycosyl transferases, GT2 group (Coutinho and Henrissat, 1999), where also chitin synthases are listed. Experiments on chitin synthase 2 of yeasts proved the QVLRW motif is required for chitin synthase activity (Nagahashi *et al.*, 1995). One of the barley CesA genes, designated *HvCesA4* is 3,129 kb long and the *HvCesA4* gene translates into 1,042 amino acid residues in length. Both NH_2- and COOH- termini are predicted to be localized in the cytoplasm.

Conclusions

The *HvCesA4* gene that is the subject of my study shows a high similarity to the other cellulose synthase genes in barley. Functional analyses of *HvCesA4* protein will take us one step closer in understanding the biosynthesis of cellulose in vascular plants.

The Research Plan

Overview

A goal of my project is to design and generate a series of DNA constructs of the *HvCesA4* isoenzyme through molecular cloning and express these DNA constructs in heterologous system of *Pichia pastoris*. The *vCesA4* gene has not yet been sequenced completely, however partial cDNAs of *HvCesA4* gene were obtained from the National Center for Biotechnology Information (NCBI) Database (Burton *et al.*, 2004). Further, a RNA transcript of *HvCesA4* gene was obtained (Burton et al. 2008, unpublished data). The introns were removed from these sequences to obtain an open-reading-frame of the *HvCesA4* gene. The gene thus obtained was analysed through BLAST searches provided

by the NCBI. These searches confirmed that the *HvCesA4* gene is very similar to the other members of cellulose synthase gene family in barley.

Bioinformatics

Information Technology has been playing a major role in modern research. Bioinformatics tools helps to improve the accuracy of research. During this project we have used various *in-silico* protein prediction tools in order to get insights into the organization of the *HvCesA4* gene. These tools predict the secondary structure and topology of proteins based on their amino acid sequences. Along with the above predictions we have also estimated potential *N*- and *O*-glycosylation and phosphorylation sites. The results obtained from these protein prediction tools aided us in designing the constructs of the *HvCesA4* gene for cloning and expression in *Pichia pastoris*. The sites to be used for prediction of protein topology and post-translational modifications include TOPRED, HMMTOP, SOSUI, DAS, etc. (the list of these sites is summarized in References). However, it is important to note that these results represent just predictions and often they vary from one tool to another. Thus, it is important to re-check the results with different sites and conclude a prediction consensus.

NOTE: This table is included on page 14 of the print copy of the thesis held in the University of Adelaide Library.

Table 1 Predictions of membrane helices for *HvCesA4* by SOSUI (http://bp.nuap.nagoyau.ac.jp/sosui/sosui_submit.html). Eight membrane helices were predicted.

Design of HvCesA4 constructs

The partial cDNA sequence of the barley *HvCesA4* gene was obtained from the NCBI database (Burton *et al.*, 2004). Rest of the sequence is obtained by removing introns from the open reading frame of *HvCesA4* gene. In order to study the functions of this gene it is essential to design a series of DNA constructs that could then suggest possible functions of individual segments of the gene. Hence, the entire *HvCesA4* gene needs to be truncated into segments and these truncations need to be very carefully designed. In this project four constructs of *HvCesA4* will be designed. According to the topology prediction the *HvCesA4,* the protein has its NH2-terminus inside the cytoplasm (Fig. 4). Eight transmembrane helices were predicted. According to several prediction tools the first helix

starts at the 197th amino acid, followed by the second trans-membrane helix at the 227th amino acid residue. A catalytic domain of 570 amino acid residues was predicted, and this domain extends until the third trans-membrane helix, which was predicted to start at the 818th amino acid residue. Five more trans-membrane helices were predicted after the third helix, whereby the sequence ends at 1042 amino acid residue (Table 1).

Based on this topology prediction, we decided on a series of truncation sites that will produce three DNA constructs designated a truncated catalytic module (long and short), and a catalytic module with flanking membrane-anchoring helices. The following four constructs will be prepared for cloning into the yeast expression vectors:

1. Full length *HvCesA4*
2. Truncated catalytic module of *HvCesA4* with flanking membrane-anchoring helices
3. A short truncated catalytic module of *HvCesA4* without the QVLRW region
4. A long truncated catalytic module of *HvCesA4* with the QVLRW region.

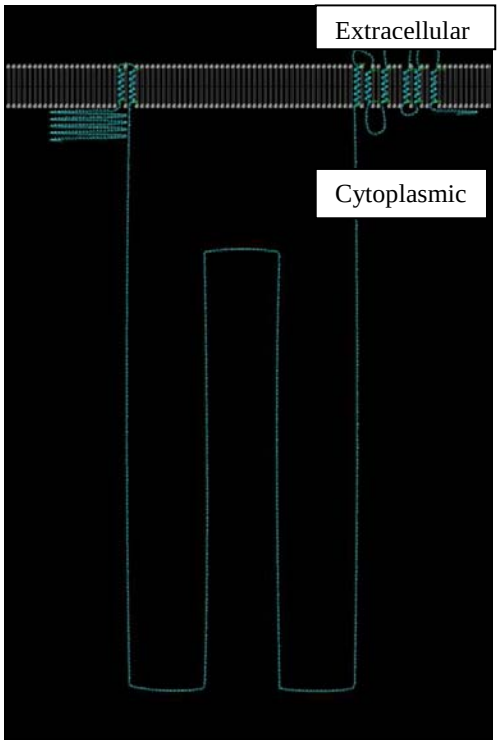


Figure 4. Topology of *HvCesA4* enzyme predicted with the HMMTOP algorithm reveals eight trans-membrane helices (Tusnady and Simon, 2001). The graphics was visualized with TMRpres2D software (Spyropoulos et al., 2004).

Primer design

Primers need to be designed for each of the four constructs, thus a total of eight primers is required. The full length construct and the construct with flanking membrane-anchoring helices would also include a His tag and TEV sites. Further, according to *Pichia* Expression Kit Manual, each primer should begin with A/YAA/TAATGTCT yeast consensus sequence, when the pPICZ vector is used. However, recent experiments with the pPICZ vector indicated that expression was also successful when CAAA sequence was used instead of A/YAA/TAATGTCT yeast consensus sequence . Each primer will beneeds to be validated using the Netprimer tool provided by Primer Biosoft International.

Cloning into pPICZ and pPICZα vectors:

The full length *HvCesA4* cDNA and the remaining three truncated cDNA segments will be cloned into pPICZ vector for expression in the yeast expression system *Pichia pastoris* (Figure 5). EcoRI is to be used as the restriction site at the 5' primers and Xba1 is to be used as the restriction site at 3' primers.

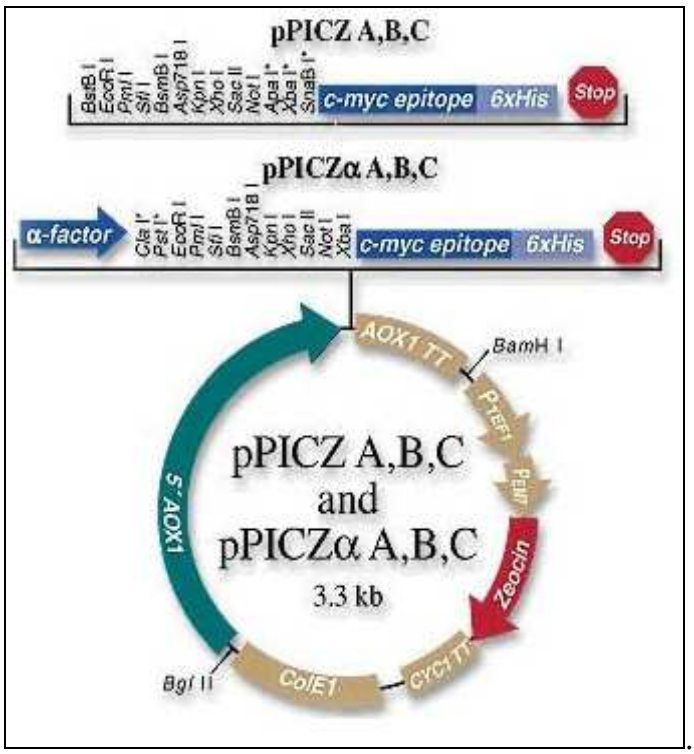


Figure 5. Representation of pPICZ vector with AOX region and multiple cloning sites

The two remaining primers, which are designed for long and short truncation segments of the catalytic module of the *HvCesA4* cDNA, will be cloned into pPICZα vector. The restriction sites to be used are again EcoRI and XbaI. The pPICZ vector contains a 942 bp fragment containing the AOX1 promoter that allows methanol-inducible, high-level expression in *Pichia* (Figure 4). Standard cloning procedures that will be used throughout my project are specified in the Easy Select Manual iew (Invitrogen Technologies).

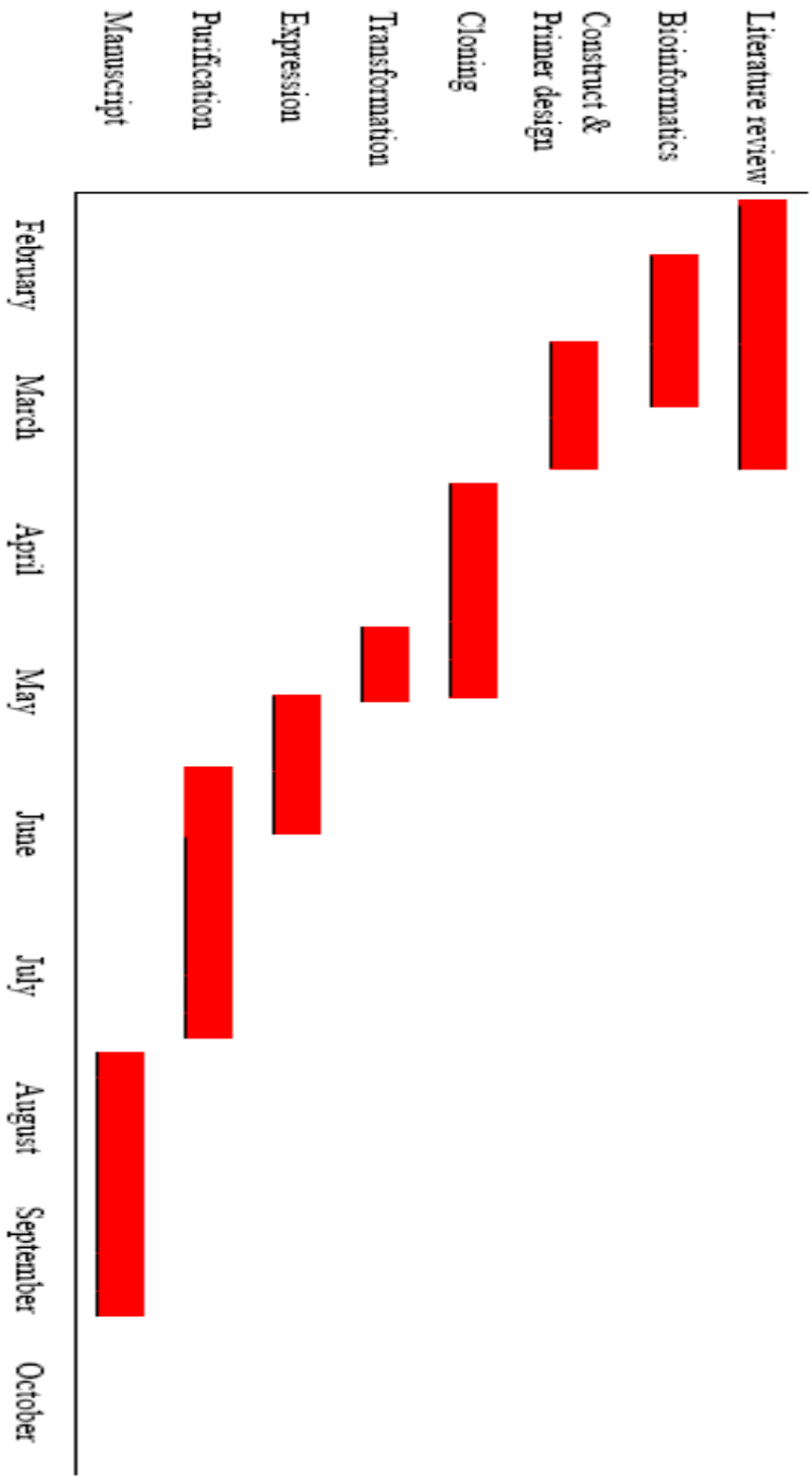
Transformation into *Pichia pastoris*

After the cDNA is cloned in the pICZ vectors, the vector needs to be linearized to stimulate recombination, when it is transformed into the *Pichia* cells. The pPICZ vectors along with four different gene constructs are then transformed into *Pichia pastoris* SMD1168H cells, and the ransformation will be performed by electroporation.

Pichia pastoris expression system

The *Pichia* Expression System Kit is based on the yeast *Pichia pastoris*. *Pichia pastoris* was developed into an expression system by Salk Institute Biotechnology/Industry Association (SIBIA) for high-level expression of recombinant proteins. As a eukaryote, *Pichia pastoris* has many of the advantages of higher eukaryotic expression system such as it allows protein folding, protein processing and post-translational modifications. Along with these advantages the *Pichia* expression system is easy to manipulate, less expensive and relatively faster than other eukaryotic expression systems. Also *Pichia* expression system is known to give higher expression levels. Good amount of HvCesA2 iso-enzyme was recently expressed in *Pichia* (Hrmova M, Peng C, Fincher G., unpublished data). The yeast *Pichia pastoris* shares the advantages of molecular and genetic manipulations with *Saccharomyces*, and it has the added advantages of producing10-100 fold higher protein expression levels. These features make *Pichia* one of the most suitable expression systems for eukaryotic genes.

Time Table


Literature review
Bioinformatics
Construct & Primer design
Cloning
Transformation
Expression
Purification
Manuscript
February
March
April
May
June
July
August
September
October

Conclusions

Study of cellulose synthesis has been one of the most interesting fields in bioscience industry today due to its promising involvement not only in agriculture but also in bio-ethanol production and a number of other industries.

My research project will be focusing on the functional analysis of *HvCesA4* enzyme and will contribute to a better understanding of cellulose biosynthesis in barley. The results will define the importance of specific *HvCesA4*protein motifs, which are found to be conserved in the cellulose synthase gene family. A successfully expressed and purified *HvCesA4* protein will help in determining two- and/or three-dimensional structure of the enzyme. Further, production of highly purified protein will be useful in production of antibodies that could determine sub-cellular localization of the *HvCesA4* protein by immuno-histochemistry and electron microscopy. Finally, the findings would help in characterizing the biochemical and biophysical properties of the *HvCesA4* protein.

Appendix

>*HvCesA4* translated amino acid sequence of 1042 amino acid residues:

MDTGEPKAKVCRACGDDVGTREDGSPFVACAECGFPVCRPCYEYERSDGT
QCCPQCNARYKRHKGCPRVEGDDEDGDMDDLEEEFQVK**S**PKKPHEPVPFD
VYSENGEQPPQKWRPGGPAMSSFGGSGQELEAEREMEGSMEWKDRIDKWK
TKQEKRGKLNRDNSDDDDDDKNDDEYMLLAEARQPLWRKLPVPSSQINPY
RIVIVLRLVVLCFFLRFRI**M**TPANDAIPLWLVSVICELWFALSWILDQLP
KWSPVTRETYLDR**LALRY**DREGEPSRLSPIDFFVSTVDPLKEPPIITANT
VLSILAVDYPVDRNSCYV**SDD**GASMLCFDTLSETAEFARRWVPFCKKFAI
EPRAPEFYFSQKIDYLKDKVQPTFVKERRAMKREYEEFKVRINGLVAKAE
KKPEEGWVMQDGTPWPGNNTRDHPGMIQVYLGSQGALDVEGHELPRLVYV
SREKRPGHNHHKKAGAMNALVRVSAVLTNAPFILNLDCDHYV**N**NSKAVRE
AMCFLMDPQLGKKLCYVQFPQGFDGIDLHDRYANRNVVFFDINMKGLDGI
QGPVYVGTGCVFNRQALYGYDPPRPEKRPKMTCDCWPSWCCCCCCFGGGK
HRKSSKDKKGGGGGDDEPRRGLLGFYKKRGKKDKLGGGPKKGSYRKRQRG
YELEEIEEGIEGYDELERS**SLMSQ**KSFQKRFGQSPVFIASTLVEDGGLPQ
GAAADPAGLIKEAIHVISCGYEGKTEWGKEIGWIYGSV**TED**ILTGFKMHC
RGWKSVYCTPTRPAFKGSAPI**N**LSDRLH**QVLRW**ALGSVEIFMSRHCPLW**Y**
AYGGRLKWLERFAYTNTIVYPFTSIPLIAYCTIPAVCLLT**G**KFIIPTLNN
LASIWFIALFMSIIATGVLELRWSGVSIEDWWRNEQFWVIGGVSAHLFAV
FQGFLKVLGGVDT**N**FTVTSKAGADEADAFGDLYLFKWTTLLIPPTTLIII
NMVGIVAGVSDAVNNGYGSWGPLFGKLFFSFWVIVHLYPFLKGLMGRQ**N**R
TPTIVVLWSVLLASIFSLVWVRIDPFIAKPKGPILKPCGVQC

- Eighth Membrane helices (underlined)
- Four *N*-glycosylation sites (bold and underlined)
- Two *O*-glycosylation sites (bold and underlined)
- Catalytic amino acid residues (bold and underlined): **SDD**, **TED**
- **QVLRW** highly conserved sequence (protein interactions)(bold)

References

Arioli, T., Peng, L.C., Betzner, A.S., Burn, J. and Wittke, W. (1998) Molecular analysis of cellulose biosynthesis in Arabidopsis. *Science*, **279**, 717-745.

Bacic A, Harris PJ, Stone BA (1998) Structure and function of plant cell walls. *The biochemistry of plants,* **297**,371-398.

Brett, C.T. (2000) Cellulose microfibrils in plants: biosynthesis, deposition, and integration into the cell wall. *Int. Rev. Cytol*, **199**, 161-189.

Brown, R.M. (2004) Cellulose structure and biosynthesis: What is in store for the 21st century. *J. Polymer. Sci. A*, **42**, 487-511.

Burton, R.A., Shirley, N.J., King, B.J., Harvey, A.J. and Fincher, G.B. (2004) The CesA Gene Family of Barley. Quantitative Analysis of Transcripts Reveals Two Groups of Co-Expressed Genes. *Plant Physiol.*, **134**, 224-236.

Carpita, N.C. and Gibeaut, D.M. (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *The Plant Journal*, **3**, 1-30.

Coutinho, P. and Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated database approach. *Royal Society of Chemistry.*,. 3–12.

Delmer, D.P. (1999) Cellulose biosynthesis: exciting times for a difficult field of study. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **50**, 245-265.

Doblin, M.S., Kurek, I., Jacob-Wilk, D. and Delmer, D.P. (2002) Cellulose Biosynthesis in Plants: from Genes to Rosettes. *Plant Cell Physiol.*, **43**, 1407-1420.

Finaev, D. (2007) Some aspects of cellulose biosynthesis. *Biologia Plantarum*, **51**, 407-413.

Herth, W. (1983) Arrays of plasma membrane rosettes in cellulose microfibril formation of Spirogyra. *Planta*, **159**, 347-360.

Kimura, S., Laosinchai, W., Itoh, T., Cui, X.J., Linder, C.R. and Brown, R.M. (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant Vigna angularis. *Plant Cell*, **11**, 2075-2087.

Lai-Kee-Him, J., Chanzy. H., Muller. M, Putaux, J.L., Imai, T. and Bulone, V. (2002) In vitro versus in vivo cellulose microfibrils from plant primary wall synthases: structural differences. *J. Biol. Chem.*, **277**, 36931-36955.

Mueller. S. C, Brown RM Jr (1980) Evidence for an intramembranous component associated

with a cellulose microfibril synthesizing complex in higher plants. *J Cell Biol* **84:** 315–326.

Nagahashi, S., Sudoh, M., Ono, N., Sawada, R. and Yamaguchi, E. (1995) Characterization of chitin synthase 2 of Saccharomyces cerevisiae: implication of two

highly conserved domains as possible catalytic sites. *J. Biol. Chem.*, **270**, 13961-13977.

Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P. and Stalker, D.M. (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. *PNAS*, **93**, 12637-12642.

Roelofsen A (1958) Cell wall structure as related to surface growth. *Acta Bot Neerl*, **7,** 77–89.

Richmond, T. (2000) Higher plant cellulose synthases. *Genome Biol.*, **4**, 30011-30033.

Somerville C (2006) Cellulose Synthesis in Higher Plants. Annual Review of Cell and Developmental Biology, **22,** 53-78

Spyropoulos, I.C., Liakopoulos, T.D., Bagos, P.G. and Hamodrakas, S.J. (2004) TMRPres2D: high quality visual representation of transmembrane protein models. *Bioinformatics*, **20**, 3258-3260.

Tusnady, G.E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics*, **17**, 849-850.

World Wide Web sites:

NCBI Nucleotide database: http://www.ncbi.nlm.nih.gov/

Expasy prediction: http://expasy.org/

HMMTOP prediction: http://www.enzim.hu/hmmtop/html/submit.html

NetNgly prediction: http://www.cbs.dtu.dk/services/NetNGlyc/

PSIPRED prediction: http://bioinf.cs.ucl.ac.uk/psipred/psiform.html

TopPred prediction: http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html

SOSUI prediction: http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html

Acknowledgements

I thank Associate Professor Maria Hrmova for supervision and Professor Geoffrey B. Fincher for being a co-supervisor and for giving me the opportunity to work on the project in his laboratory. I also thank Dr. Andrew J. Harvey for his guidance and for being my internal examiner. Finally, I thank Professor Tony Bacic for being my external examiner. I would also like to extend my thanks to all the ACPFG colleagues for their support.

PART II- THE MANUSCRIPT

Summary

Cellulose is the most abundant polysaccharide polymer on earth that contributes significantly to the structure and development of flora. A part of the biosynthesis of cellulose has been tracked to two gene families, namely cellulose synthase (CesA) and cellulose synthase like (Csl) gene families. CesA genes are known to encode a catalytic subunit of cellulose synthase. It has been suggested that cellulose synthase forms a protein complex, which is believed to be located in the plasma membrane of plant cells and to contribute to the production of cellulosic micro-fibrils. These micro-fibrils co-crystallize to form cellulose. Up to now, nine CesA genes have been assigned in barley, which are designated HvCesA1 to HvCesA9 (Rachel Burton unpublished data). Out of these nine genes, it was found that *HvCesA4*, HvCesA7 and HvCesA8 were expressed abundantly during stem and root development in barley. To our knowledge, none of these protein sequences have so far expressed heterologously. We believe that heterologous expression of these proteins would be beneficial to our understanding of protein properties, and to structural studies that would lead to our better understanding of cellulose synthesis. In this study four constructs of the *HvCesA4* isoenzyme from barley have been cloned into the pPICZ vector system. In addition, protein expression of *HvCesA4* constructs was attempted in a heterologous host of *Pichia pastoris*.

Introduction

The majority of plant cell walls contain cellulose as a major structural component, where cellulosic microfibrils with xyloglucan represent a major load-bearing components (Pear

et al., 1996). Cellulose is the most abundant polysaccharide in the biosphere, and is involved in key cellular and physiological processes of plants, but the detailed mechanism of its biosyntheses at molecular level is not understood. Cellulose is a linear β-D-glucan polymer composed of β-linked glucosyl residues, and linkage configuration between the β-glucosyl residues in cellulose has decisive effects on its conformation and hence biological function (Stone et al., 2008). Cellulose synthase occurs in large protein families with more than 9,000 members and are classified in the glycosyl transferase GT2 group of enzymes (CAZy; http://www.cazy.org/fam/acc_GH.html; Coutinho and Henrissat, 1999). Their cDNAs, e.g. barley cellulose synthase *HvCesA4*, are translated into proteins with 1082 amino acid residues (Burton et al., 2004). The topology predictions show 7-8 membrane-spanning helices positioned at *HvCesA4* NH_2- and COOH-termini, while single large cytoplasmic regions encompass the catalytic module.

CesA genes were found to be involved in synthesis of polysaccharides that form primary and secondary cell walls (Arioli et al., 1998; Fagard et al., 2000; Burn et al., 2002; Somerville, 2006). It has been shown in Arabidopsis that a CesA gene knockout mutant showed cellulose deficiency (Arioli et al., 1998; Taylor et al., 1999; Fagard et al., 2000; Taylor et al., 2000; Scheible et al., 2001; Beeckman et al., 2002; Burn et al., 2002; Gardiner et al., 2003; Taylor et al., 2003). Out of eight CesA gene so far identified in barley (Burton et al., 2004), few have been studied.

The plant glucan synthases are predicted to form oligomeric assemblies in the plasma membrane (hexameric rosettes) (Somerville, 2006), and when purified, they are rapidly

inactivated (Brown, 1996; Kimura et al., 1999; Lai-Kee-Him et al., 2002). The only known three-dimensional (3D) structure of a GT2 enzyme is for a 256-residue spore-coat protein from *Bacillus subtilis*; this offers insights into the geometry of the active site of these GTs (Charnock and Davies, 1999), but the function of the protein is unknown.

The focus of this project is on cellulose synthase that mediates synthesis of cellulose or (1,4)-β-D-glucan, which represents the key component of plant cell walls of economically important cereals and grasses (Stone et al., 2008). In this project effort was put towards constructing expression vectors containing cDNA sequence of *HvCesA4* and towards expressing some of these proteins heterologously. It has been reported that CesA genes have unique and conserved motifs such as QXRLW and TED (Burton et al., 2004(Holland et al., 2000). The *HvCesA4* isoenzyme also contains QVLRW and TED motifs in its catalytic region. For the better understanding of the *HvCesA4* protein structure, we have conducted *in-silico* protein sequence investigations, before attempting expressing the protein in *Pichia pastoris*. In summary, three types of vectors have been successfully constructed and tested for protein expression

Results

Prediction of membrane helices and catalytic module of HvCesA4

Several web-based applications were used for predicting the topology, secondary structure dispositions and locations of the membrane helices in the *HvCesA4* protein. These tools revealed that several locations of these helices were possible. In the end the consensus was reached that *HvCesA4* contains eight membrane helices (Fig. 1). The N-

terminus of *HvCesA4* was identified to be located on the intracellular side of the plasma membrane, from where the protein sequence traverses the membrane to extra-cellular space, before the sequence enters back to the cytoplasm and forms a large catalytic region (Fig. 1). According to the topology prediction server TopPred, two membrane helices were found at the N-terminus, while six membrane helices were found at the C-terminus (Fig. 1). A 570 amino acid residues long catalytic region was predicted to be located inside the cytoplasm, which contains TED, LALRY and QVLRW conserved motifs. The eight membrane helices were predicted between 197-216, 226-248, 818-840, 849-871, 888-910, 938-960, 971-993, and 1003-1025 amino acid residues. Surprisingly only four N-glycosilation site and two O-glycosylation sites were predicted in the sequence. It is of interest that various protein topology and secondary structure prediction servers revealed slightly different possibilities for positions of membrane helices (data not shown). However, the consensus from these predictions supports the view that *HvCesA4* contains 8 membrane helices (SOSUI server at http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html).

Cloning into Entry Vector Systems

Ligation was carried out with four constructs of *HvCesA4*. As an entry vector, primarily pGEM-T vector system was used throughout. However, when the clones were transformed into the DH5α competent cells, no colonies were observed. For this reason, the cloning procedure was repeated with pCR8-TOPO vector system. In this case, the colonies were observed, which were screened to be positive by colony PCR screening technique.

Cloning into destination vector system

The three of four constructs (constructs 2-4) were successfully ligated into the destination pPICZ vector system that was in-frame with the yeast α-factor secretion sequence and under the transcriptional control of the alcohol oxidase promoter AOX1. All three constructs were sequenced in both directions and have not contained any errors in their open reading frames.

Protein expression

Protein expression was attempted with the three *HvCesA4* constructs (constructs 2-4) in *Pichia pastoris* SMD1168H cells and the expression was monitored by SDS-PAGE and Western blot analyses. The preliminary experiment revealed that catalytic module of *HvCesA4* with flanking helices has been successfully expressed, where 70 kDa His-tagged protein was clearly visible (Fig.2).

As for the expression of the two truncated sequences (constructs 2 and 3) corresponding to the two versions of the catalytic module of *HvCesA4*, the dot blot analyses have not revealed the presence of these proteins in soluble forms in the cultivation media (data shown in supplementary material). The most likely explanation for this observation is the fact that protein concentration might have been low for SDS-PAGE and Western blot analyses.

Discussion

It has been concluded that at least three barley genes, encoding cellulose-synthases, are coordinately expressed out of nine barley *HvCesA* genes, during cellulose synthesis (Appenzeller et al., 2004; Burton et al., 2004). It has been suggested that at least three different genes are essential for making a planar rosette complex that typically contains up to 30-36 subunits (Perrin, 2001; Scheible et al., 2001). The studies have also shown that HvCesA1, HvCesA2 and HvCesA6 isoenzymes are predominantly and coordinately transcribed mostly in young tissues (Burton et al., 2004). Another group of HvCesAs contains HvCesA4, HvCesA7 and HvCesA8 isoenzymes and these were found to be coordinately transcribed in the mature tissues. With this information it could be concluded that the group containing HvCesA1, HvCesA2 and HvCesA6 is most likely to contribute for primary cell wall synthesis and the second group containing HvCesA4, HvCesA7 and HvCesA8, is likely to take part in secondary cell wall synthesis. For this reason, as outlined in previous sections of my thesis, my aim was to generate DNA vectors of HvCesA genes that are involved in secondary cell wall formation, in particular of the HvCesA8 isoform.

Bioinformatical analysis of the 1042-amino acid residue long HvCesA4 protein sequence (Figure 1; this Figure should be the sequence of HvCesA4) has shown that the full-length protein contains eighth membrane helices (Figure 2) (and a long catalytic centre region, which was about 570 amino acid residue long. A highly conserved motif QVLRW was found within this catalytic region, alongside with four *N*-glycosylation sites and two *O*-glycosylation sites (Figure 1). Apart of catalytic motifs NED and SDD (Figure 1), all the

above HvCesA genes carry a common sequence such as QXXLRW motif, which is highly conserved in plant cellulose synthases (Pear et al., 1996; Delmer, 1999). It is of interest that the QXXLRW motif is also known to be associated with bacterial cellulose synthases and other processive glycosyltransferases such as chitin and hyaluronan synthases (Marks et al., 2001). Thus, we were interested in constructing expression vectors encoding soluble catalytic modules of HvCesA4 with and without the QXXLRW motifs, in addition to constructing vectors containing its full-length sequence and the sequences of the catalytic module with flanking helices on each side of the catalytic module.

Due to a high GC content of the HvCesA4 sequence, the primer design and amplification of DNA by PCR proved to be challenging. The amounts of PCR products obtained specific primers and HvCesA4 template were low and these amounts were not sufficient for further cloning steps. To overcome these problem three new primers were designed for a full-length sequence of HvCesA4 with longer priming sites that would avoid GC rich segments in the sequence. Using this approach, all four HvCesA4 constructs were successfully cloned into the entry vector pCR8. An attempt was also made to clone the four HvCesA4 constructs into the entry vector pGEM-Teasy, however they were unsuccessful. Further cloning of DNA sequences into the destination vector pPICZ was carried out successfully for three HvCesA4 constructs containing truncated sequences. The cloning of the full-length sequence of HvCesA4 into pPICZ was unsuccessful. The 3164 bp long sequence of HvCesA4 could not be cloned into pPICZ vector for a variety

of reasons, such as e.g. forming secondary DNA structures that could occlude a steric availability of restriction sites.

Pichia pastoris expression system was chosen over the *E.coli* expression system for expressing the HvCesA4 proteins, because the HvCesA4 sequence showed the presence of several *N*- and *O*-glycosylation sites, as specified above. As an eukaryote, *Pichia pastoris* possesses many advantages of higher eukaryotic expression systems, such as protein processing, protein folding, and the availability of post-translational modifications, while it could not be manipulated as easily as the *E. coli* expression system. Due to all these advantages, *Pichia pastoris* expression system was chosen over the *E.coli* one. From the three successfully cloned sequences, HvCesA4_T1, a truncated construct containing the central catalytic module flanked by two membrane helices on each side, was chosen to be tested for protein expression. The HvCesA4-pPICZ vector was transformed into the Pichia cells and successfully expressed. A 70 kDa His-tagged protein was observed by Western analysis by blotting the protein produced by the transformed *Pichia* cells with His-tag antibodies. Based on this detection it can be concluded that the expressed protein was encoded by HvCesA4_T1 construct carrying the 6x-His epitope. On the other hand, no expressed protein was observed with the two other successfully cloned constructs. Various reasons were implicated including probable contamination of *Pichia* cell cultures with bacteria. Due to the lack of time, the experiments could not be repeated hence further conclusions cannot be drawn. Yet, these experiments back up the idea of successful use of *Pichia* expression system for production the HvCesA proteins in general. It is very likely that all the constructs

designed for HvCesA4 production could be expressed successfully in *Pichia pastoris* expression system.

As for the future prospects of these studies to investigate HvCes4 proteins, the heterologus expression of the above gene followed by purification and characterization could yield significant data for understanding of secondary cell wall synthesis and stem strength in barley. This project has taken us a step closer in the above process, where the three constructs of HvCesA4 were cloned into pPICZ vector system for expression in *Pichia pastoris*. In the future, the full-length sequence of HvCesA4 needs to be cloned into pPICZ vector using different techniques such as a direct PCR product cloning into destination vectors. It is projected that successful expressions of these HvCesA4 proteins should be followed be purification of the proteins and subjecting them These to activity assays, followed by complete bio-chemical, bio-physical and structural studies needs

Experimental Procedures

Bioinformatics and predictions of membrane helices

Barley *HvCesA4* cDNA (GenBank accession number AY483154) ((Burton et al., 2004) was retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/). *HvCesA4* cDNA was translated and the amino acid sequence was fed into the following protein secondary structure and topology prediction web-based applications (Expasy http://expasy.org/; HMMTOP http://www.enzim.hu/hmmtop/html/submit.html; NetNgly - ttp://www.cbs.dtu.dk/services/NetNGlyc/;PSIPRED,http://bioinf.cs.ucl.ac.uk/psipred/psiform.html, TopPred, http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html, SOSUI, http://bp.nuap.nagoyau.ac.jp/sosui/sosui_submit.html). Based on the bioinformatical

analyses, the four constructs were designed and designated as *HvCesA4* constructs 1-4. The first construct (*HvCesA4* construct 1) represents the full-length sequence. The second construct represents the catalytic subunit and two flanking membrane helices on each side (*HvCesA4* construct 2). The third construct (*HvCesA4* construct 3) contained the central catalytic subunit with QVLRW motif. The forth construct (*HvCesA4* construct 4) is the shortest of all, and contained the partial catalytic module excluding the QVLRW motif.

Molecular cloning and primers for HvCesA4 constructs

Initially seven primers were designed based on the *HvCesA4* protein topology and membrane helical predictions. For all the constructs, the primers with restriction sites, yeast consensus sequence and poly-histidine sequences were designed. For all 5' primers EcoR1 restriction digestion sites were chosen and for all 3' primers Xba1 restriction sites were incorporated. The sequences of the primers are as follows.

HvCesA4 construct 1:

Forward primer	GTCAGAATTCAAACATCATCATCATCATCATGAAAATCTGTACTT TCAAGGTGACACCGGCGAGCCCAAGGCC
Reverse primer	GCCAAGCCCAAGGGACCCATTCTTAAGCCGTGTGGAGTACAGTGC TGAGGATAATCTAGA

HvCesA4 construct 2:

Forward primer	AGTCGAATTCAAAATGGGTCATCATCATCATCATCATGAAAATCTGT ACTTTCAAGGTACCCCGGCCAACGACGCCATC
Reverse primer	GGGGTGCTGGAGCTGCGGTGGAGCGGGGTGAGCTGAGGATAATCTA GA

For *HvCesA4* construct 3 and 4, primers were designed with restriction sites EcoR1 and Xba1 on 5’ and 3’ ends respectively, where also a TEV (Tobacco Etched Virus) site was added at the 3’ end.

HvCesA4 constructs 3 and 4:

Forward primer	CCGGAATTCAGACTCGCCCTGCGCTACGACCGCGAG

HvCesA4 construct 3:

Reverse primer	TCTGGTACGCCTACGGCGGCCACCTTGAAAGTACAGATTTTCTCT AGA

HvCesA4 construct 4:

Reverse primer	CGCTCCTCGCTCATGTCACAGACCTTGAAAGTACAGATTTTCTCTAG AAGG

Due to high GC content in the primers designed for *HvCesA4* construct 1, two additional forward primers and one additional reverse primer were designed.

Forward primer 2	GAATTCAAACATCATCATCATCATCATGAAAATCTGTACTTTCAAG GTGACACCGGCGAGCCCAA
Forward primer 3	GAATTCAAACATCATCATCATCATCATGAAAATCTGTACTTTCAAG GTGAGCCCAAGGCCAAGGT
Reverse primer 2	ATTCTTAAGCCGTGTGGAGTACAGTGCTGAGGATAATCTAGA.

The expected size of PCR products with the primers for *HvCesA4*_FL is 3162bp, *HvCesA4*_T1 is 2044bp, *HvCesA4*_T2 is 1669bp and for *HvCesA4*_T3 is 1273bp.

Cloning into entry vector

The constructs are produced and amplified with respective primers using full length *HvCesA4* (kindly donated by Dr. Rachel Burton).The sticky end PCR products of the four constructs *HvCesA4* (constructs 1-4) were purified using nucleospin extract II kit and the fresh DNA were ligated into pCR8/GW-TOPo vector (Invitrogen), which served as an entry vector. Cloned vector was transformed into One Shot® Mach1™ T1 phage-resistant chemically competent *E. coli* cells (Invitrogen). Antibiotic zeocin was used for clone selection at 100 µg/ml. Successfully transformed colonies were picked and grown overnight on Luria-Bretani medium along with the zeocin at 30 µg/ml. DNA was extracted from the cells using alkaline lysis extraction method. Extracted DNA was subjected to digestion with *Xba1* and *EcoR1* restriction enzymes. The constructs from the entry vector pCR8 were gel purified and stored at -20 ◦C for further cloning into the destination vector.

Preparation of destination vectors

Glycerol stocks of *E. coli* cells containing the destination vector pPICZB and pPICZαA were obtained from Mrs. Margaret Buchanan. The cells were inoculated into the Luria-Bertani agar media plates with zeocin at 30 µg/ml. The plates were incubated overnight at 37º C. The positive colonies are inoculated on to LB medium with Zeocin (30µg/ml). Cell culture is incubated overnight at 37ºC. Plasmids from the cell suspensions were

extracted using alkaline lysis DNA-extraction method. The plasmids were subjected to restriction digestion with Xba1 and EcoR1 restriction enzymes. The digested products were run on 1% (w/v) agarose gels and the small fragments released by restriction digestion were eliminated from the vectors. Linearised vectors were cut off the gels, purified by Nucleo-spin DNA purification kit (Macherey-Nagel) and stored at -20º C.

Cloning into destination vectors

The four constructs with EcoR1 restriction sites on 5'end and Xba1 restriction sites on 3' end were ligated into the destination pPICZ vector that was in-frame with the yeast α-factor secretion sequence and under the transcriptional control of the alcohol oxidase promoter AOX1. Two different variants of pPICZ were used. Constructs 1 and 2 and were cloned on to pPICZB, while constructs 3 and 4 were ligated into pPICZαA. The cloned vectors were transformed into One Shot® Mach1™ T1 phage-resistant chemically competent *E. coli* cells (Invitrogen) as recommended by in the manufacturer (EasySelect ™ Pichia Expression Kit, version G). The transformed cells were incubated overnight at 37◦ C. DNA was extracted from the *E. coli* cells using alkaline lysis method.

DNA sequencing

All constructs were sequenced using AOX1 forward and reverse primers (EasySelect ™ Pichia Expression Kit, version G), and by using *HvCesA4* gene specific internal primers kindly provided by Dr. Rachel Burton.

Transformation into *Pichia pastoris*

About 10 ml of competent *Pichia pastoris* cells were prepared using the protocol from the EasySelect ™ Pichia Expression Kit version G. The *HvCesA4* constructs 2-4 were linearised with the restriction enzyme Pme1. The enzyme mixture was heat inactivated at 60º C for 20min. The linearised plasmids were purified using NucleoSpin® Extract II Kit from Macherey-Nagel kit. About 5-10 µg of the purified DNA was used for transformation of *Pichia pastoris* SMD1168H competent cells, using electroporation technique. Approximately 80 µl of *Pichia pastoris* SMD1168H competent cells were deposited in a 0.2cm electroporation cell and 10 µl of purified and linearised plasmid was added. The electroporation was carried out with 1500 V, 400Ω and 25µF. The transformed mixture was platted onto the YPD media with zeocin antibiotic at 100 µg/ml. Plates were incubated at 28ºC for 4-6 days, after which colonies were observed.

Protein expression

Successfully transformed *Pichia pastoris* cells were inoculated into the BMGY media after 4-6 days of growth on YPD media at 28ºC. Cells were incubated at 28ºC overnight. Cells were washed by centrifugation (1500g, 10 min) and transferred to the BMMY medium, where they were incubated for 5 days, during which 0.5% (v/v) methanol concentration was maintained. The cells were incubated with constant shaking at 220 rpm, and harvested after 5 days of growth.

Protein extraction, dot blot, Western blot and SDS-PAGE analyses

After 5 days of incubation of cell expressing construct 1, proteins were extracted from the cell suspension using alkaline lysis method (Tobais von der Haar, 2007) and fractionated into three fractions. The samples were spotted onto a nitrocellulose membrane, the membrane was blocked with non-fat milk for 3 hours, washed with phosphate buffered saline and again blocked with 1:2000 dilution of monoclonal anti-polyhistidine, alkaline phosphatase conjugate clone His-1 antibodies overnight at 4 oC. The following day the nitrocellulose membrane was developed with BCIP\NBT reagent. For the constructs 3 and 4, the cell cultures after 5 days of incubation were spun (1500g, 10 min) and the culture media were analysed for the protein expression. SDS-PAGE analyses proceeded according to the standard protocol (Hrmova et al., 1996)

Acknowledgements

I thank Associate Professor Maria Hrmova for supervision and Professor Geoffrey B. Fincher for giving me the opportunity to work on the project in his laboratory. I also thank Dr. Andrew J. Harvey for being my internal examiner and to Professor Tony Bacic for being my external examiner. I would also like to extend my thanks to Dr. Rachel Burton, Dr. Andrew Jacob for their support. Also I thank Ms. Jessica Smith, Ms.Maria Lombardi and Mrs. Jodie Kretschmer Jodie and all my colleagues at Plant Genomics Centre for their support.

References

Appenzeller L, Doblin M, Barreiro R, Wang HY, Niu XM (2004) Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. Cellulose **11:** 287

Arioli T, Peng LC, Betzner AS, Burn J, Wittke W (1998) Molecular analysis of cellulose biosynthesis in Arabidopsis. Science **279:** 717

Beeckman T, Przemeck GKH, Stamatiou G, Lau R, Terryn N (2002) Genetic complexity of cellulose synthase A gene function in Arabidopsis embryogenesis. Plant Physiol. **130:** 1883

Brett CT (2000) Cellulose microfibrils in plants: biosynthesis, deposition, and integration into the cell wall. Int. Rev. Cytol **199:** 161

Brown RM (1996) The biosynthesis of cellulose. J. Macromol. Sci. A **33:** 1345

Brown RM (2004) Cellulose structure and biosynthesis: What is in store for the 21st century. J. Polymer. Sci. A **42:** 487

Burn JE, Hurley UA, Birch RJ, Arioli T, Cork A, Williamson RE (2002) The cellulose-deficient Arabidopsis mutant rsw3 is defective in a gene encoding a putative glucosidase II, an enzyme processing N-glycans during ER quality control. Plant J. **32:** 949

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The CesA Gene Family of Barley. Quantitative Analysis of Transcripts Reveals Two Groups of Co-Expressed Genes. Plant Physiol. **134:** 224-236

Burton RA, Shirley NJ, King BJ, Harvey AJ, Fincher GB (2004) The cesa gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. **134:** 224

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. The Plant Journal **3:** 1-30

Delmer DP (1999) Cellulose biosynthesis: exciting times for a difficult field of study. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50:** 245

Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP (2002) Cellulose Biosynthesis in Plants: from Genes to Rosettes. Plant Cell Physiol. **43:** 1407-1420

Fagard M, Desnos T, Desprez T, Goubet F, Refregier G (2000) PROCUSTE1 encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of Arabidopsis. Plant Cell **12:** 2409

Finaev D (2007) Some aspects of cellulose biosynthesis. Biologia Plantarum **51:** 407-413

Gardiner JC, Taylor NG, Turner SR (2003) Control of cellulose synthase complex localization in developing xylem. Plant Cell **15:** 1740

Herth W (1983) Arrays of plasma membrane rosettes in cellulose microfibril formation of Spirogyra. Planta **159:** 347

Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP (2000) A Comparative Analysis of the Plant Cellulose Synthase (CesA) Gene Family. Plant Physiol. **123:** 1313-1324

Kimura S, Laosinchai W, Itoh T, Cui XJ, Linder CR, Brown RM (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant Vigna angularis. Plant Cell **11:** 2075

Lai-Kee-Him J, Chanzy H, Muller M, Putaux JL, Imai T, Bulone V (2002) In vitro versus in vivo cellulose microfibrils from plant primary wall synthases: structural differences. J. Biol. Chem. **277:** 36931

Marks DL, Dominguez M, Wu KJ, Pagano RE (2001) Identification of active site residues in glucosylceramide synthase: a nucleotide-binding/catalytic motif conserved with processive beta-glycosyltransferases. J. Biol. Chem. **276:** 26492

Nagahashi S, Sudoh M, Ono N, Sawada R, Yamaguchi E (1995) Characterization of chitin synthase 2 of Saccharomyces cerevisiae: implication of two highly conserved domains as possible catalytic sites. J. Biol. Chem. **270:** 13961

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA **93:** 12637

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. PNAS **93:** 12637-12642

Perrin RM (2001) Cellulose: how many cellulose synthases to make a plant. Curr. Biol. **11:** R213

Scheible WR, Eshed R, Richmond T, Delmer D, Somerville CR (2001) Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in Arabidopsis Ixr1 mutants. Proc. Natl. Acad. Sci. USA **98:** 10079

Somerville C (2006) Cellulose Synthesis in Higher Plants. Annual Review of Cell and Developmental Biology **22:** 53-78

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. U.S.A **100:** 1450

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. Plant Cell **12:** 2529

Taylor NG, Scheible WR, Cutler S, Somerville CR, Turner SR (1999) The irregular xylem3 locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell **11:** 769

Table 1. SOSUI protein topology prediction of *HvCesA4* protein sequence. The presence of eight membrane helices is indicated, where the first two helices are positioned at the N-terminus, while the next six helices are located at the C-terminus of the protein. The average lengths of the membrane are 20-23 residues, which are ideal as anchor helices.

No.	N terminal	transmembrane region	C terminal	type	length
1	197	INPYRIVIVLRLVVLCFFLR	216	PRIMARY	20
2	226	AIPLWLVSVICELWFALSWILDQ	248	PRIMARY	23
3	818	IVYPFTSIPLIAYCTIPAVCLLT	840	PRIMARY	23
4	849	NNLASIWFIALFMSIIATGVLEL	871	PRIMARY	23
5	888	WVIGGVSAHLFAVFQGFLKVLGG	910	SECONDARY	23
6	938	TTLLIPPTTLIIINMVGIVAGVS	960	PRIMARY	23
7	971	GPLFGKLFFSFWVIVHLYPFLKG	993	SECONDARY	23
8	1003	TIVVLWSVLLASIFSLVWVRIDP	1025	PRIMARY	23

Table 2. The number of successful clones and transformants (constructs 1-4). The two stages of cloning are indicated, such as cloning into the entry vector and cloning into destination vector. The statistics of cell transformation into P*ichia pastoris* is also summarized.

Construct (number)	Vector	Competent cells	Number of colonies
1	pGEM-T(Invitrogen)	*E.coli* DH5α	-
2	pGEM-T(Invitrogen)	*E.coli* DH5α	-
3	pGEM-T(Invitrogen)	*E.coli* DH5α	-
4	pGEM-T(Invitrogen)	*E.coli* DH5α	-
1	pCR8 TOPO (Invitrogen)	One shot Mach1 *E. coli* *	9
2	pCR8 TOPO (Invitrogen)	One shot Mach1 *E. coli* *	15
3	pCR8 TOPO (Invitrogen)	One shot Mach1 *E. coli* *	27
4	pCR8 TOPO (Invitrogen)	One shot Mach1 *E. coli* *	23
1	pPICZB (Invitrogen)	One shot Mach1 *E. coli* *	-
2	pPICZB (Invitrogen)	One shot Mach1 *E. coli* *	13
3	pPICZ α A (Invitrogen)	One shot Mach1 *E. coli* *	18

4	pPICZ α A (Invitrogen)	One shot Mach1 *E .coli* *	24
2	pPICZ B (Invitrogen)	*Pichia pastoris* SMD1168H	22
3	pPICZ α A (Invitrogen)	*Pichia pastoris* SMD1168H	12
4	pPICZ α A (Invitrogen)	*Pichia pastoris* SMD1168H	9

*One Shot® Mach1™ T1 Phage-Resistant Chemically Competent E. coli from Invitrogen.

Figures Legends

Figure 1. Amino acid sequence of HvCesA4 from barley representing 1042 residues. Eight predicted membrane anchoring are color coded, *N*- and *O*-glycosylation sites are highlighted. The conserved motifs QXXLRW, TED, SDD are also shown.

Figure 2. Hydrophobicity plot of the *HvCesA4* protein sequence. The wedges near and above the green line indicate the positions of membrane helices.

Figure 3. Position of membrane helices predicted by TOPRED web-based protein topology prediction tool. The figure indicates the position of a central catalytic module flanked by the two membrane helices on the N-terminus and the six membrane helices on the C-terminus. The catalytic module comprises of the conserved QVLRW, TED and SDD motifs.

Figure 4. Western blot (left) and SDS-PAGE (right) analyses of construct 2 reveal the presence of the 70 kDa His-tagged proteins that was expressed in P*ichia pastoris* cells. Lanes 1 and 2 indicate protein that was isolated from colonies number 7 and 9 (fraction 3 after alkaline lysis). Lane 3 contains a crude total protein from the colony number 7. St indicates protein standards ranging from 220 to 20 kDa.

Figure 1

MDTGEPKAKVCRACGDDVGTREDGSPFVACAECGFPVCRPCYEYERSDGTQCCPQCNAR
YKRHKGCPRVEGDDEDGDMDDLEEEFQVKSPKKPHEPVPFDVYSENGEQPPQKWRPGGP
AMSSFGGSGQELEAEREMEGSMEWKDRIDKWKTKQEKRGKLNRDNSDDDDDDKNDDEYM
LLAEARQPLWRKLPVPSSQINPYRIVIVLRLVVLCFFLRFRIMTPANDAIPLWLVSVIC
ELWFALSWILDQLPKWSPVTRETYLDRLALRYDREGEPSRLSPIDFFVSTVDPLKEPPI
ITANTVLSILAVDYPVDRNSCYVSDDGASMLCFDTLSETAEFARRWVPFCKKFAIEPRA
PEFYFSQKIDYLKDKVQPTFVKERRAMKREYEEFKVRINGLVAKAEKKPEEGWVMQDGT
PWPGNNTRDHPGMIQVYLGSQGALDVEGHELPRLVYVSREKRPGHNHHKKAGAMNALVR
VSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLMDPQLGKKLCYVQFPQGFDGIDLHDR
YANRNVVFFDINMKGLDGIQGPVYVGTGCVFNRQALYGYDPPRPEKRPKMTCDCWPSWC
CCCCCFGGGKHRKSSKDKKGGGGGDDEPRRGLLGFYKKRGKKDKLGGGPKKGSYRKRQR
GYELEEIEEGIEGYDELERSSLMSQKSFQKRFGQSPVFIASTLVEDGGLPQGAAADPAG
LIKEAIHVISCGYEGKTEWGKEIGWIYGSVTEDILTGFKMHCRGWKSVYCTPTRPAFKG
SAPINLSDRLHQVLRWALGSVEIFMSRHCPLWYAYGGRLKWLERFAYTNTIVYPFTSIP
LIAYCTIPAVCLLTGKFIIPTLNNLASIWFIALFMSIIATGVLELRWSGVSIEDWWRNE
QFWVIGGVSAHLFAVFQGFLKVLGGVDTNFTVTSKAGADEADAFGDLYLFKWTTLLIPP
TTLIIINMVGIVAGVSDAVNNGYGSWGPLFGKLFFSFWVIVHLYPFLKGLMGRQNRTPT
IVVLWSVLLASIFSLVWVRIDPFIAKPKGPILKPCGVQC

Colour coding:
Eight membrane helices (gray)
Four *N*- (green) and two *O*-glycosylation (magenta) sites
Catalytic amino acid residues SDD, TED (red)
QVLRW highly conserved sequence (yellow).

Figure 2


hydrophobicity value
2
1
0
-1
-2
-3
-4
-5
0
100
200
300
400
500
Upper cutoff
Lower cutoff
hydrophobicity

Figure 3

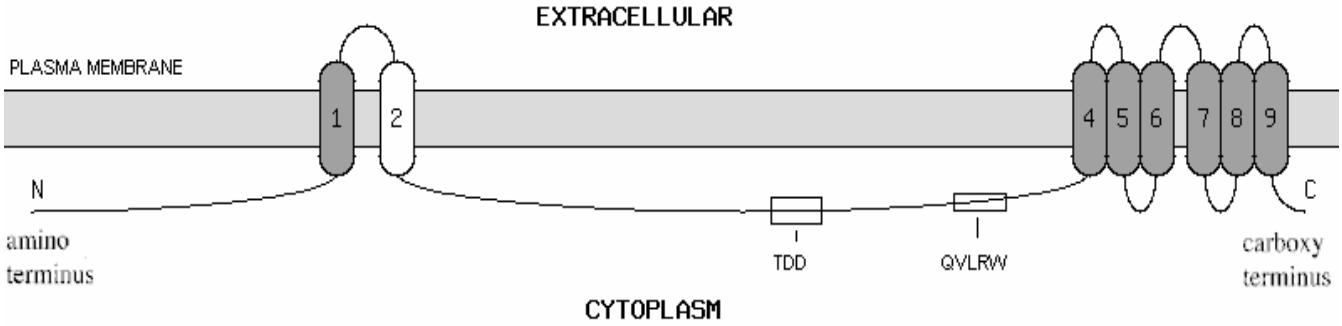
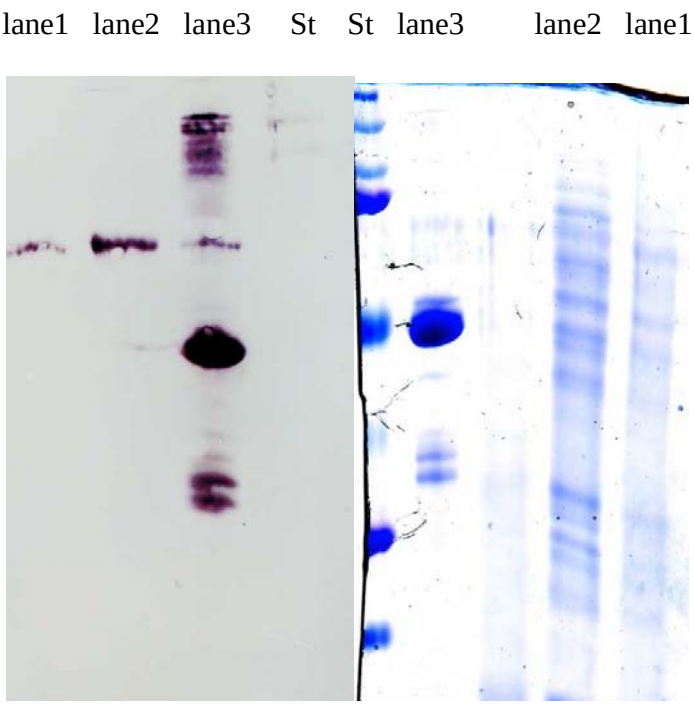

EXTRACELLULAR
PLASMA MEMBRANE
1
2
4
5
6
7
8
9
N
C
amino
terminus
TDD
QVLRW
carboxy
terminus
CYTOPLASM

Figure 4



Supplementary data

Standard FL T1 T1 T3 T2 T2

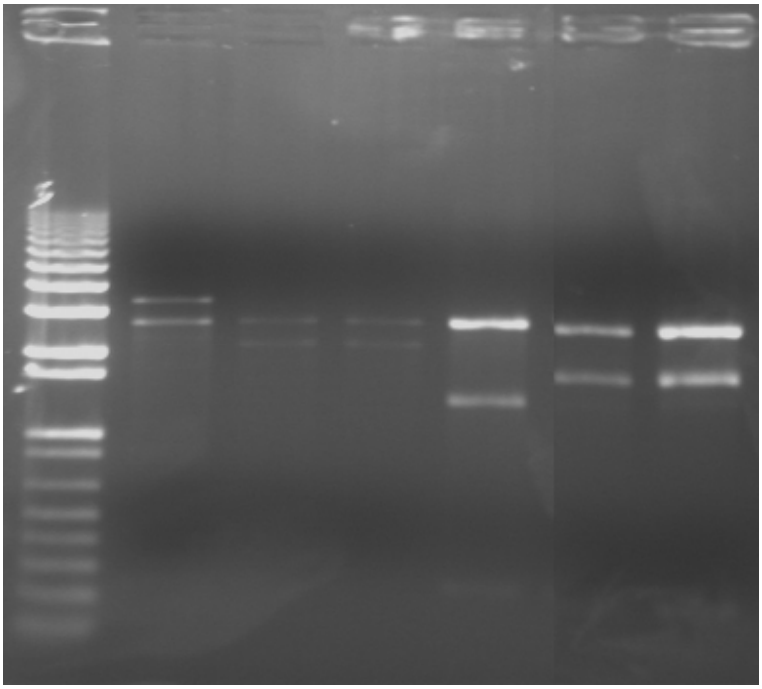


Figure 1. The plasmid pCR8+HvCesA4 constructs subjected to restriction digestion. The HvCes4 constructs are extracted from the gel, purified and cloned in to the respective destination vector.

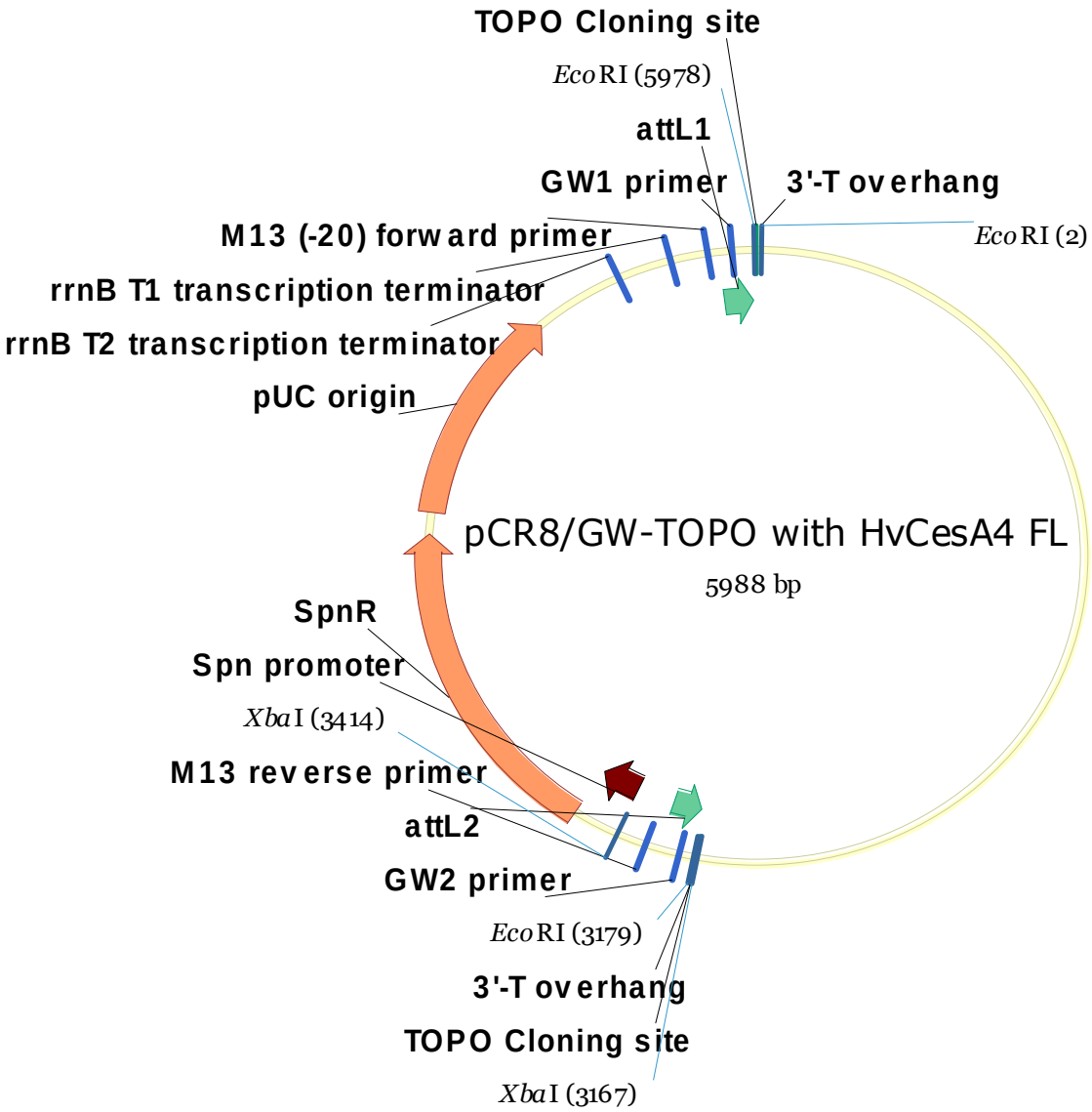


Figure 2. Entry vector pCR8/GW-TOPO containing HvCesA4_FL DNA.

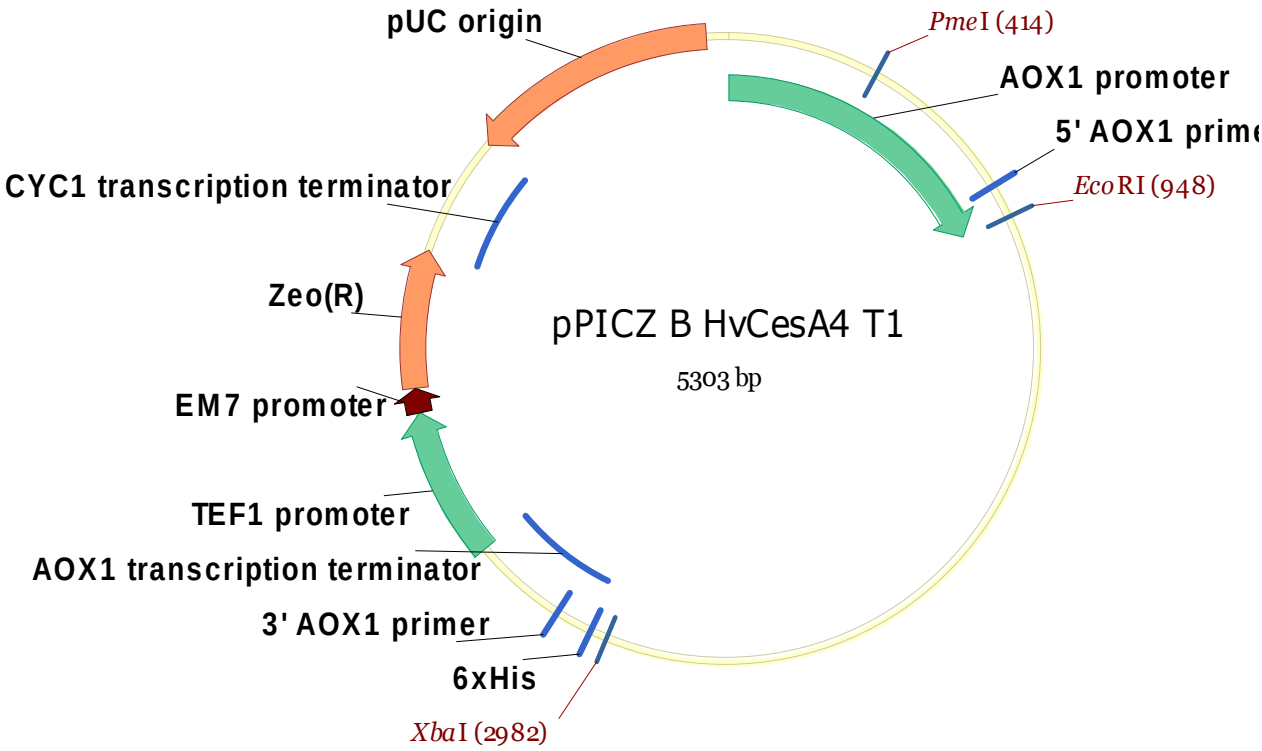


Figure 3. Destination vector pPICZB containing HvCesA4_T1 DNA.

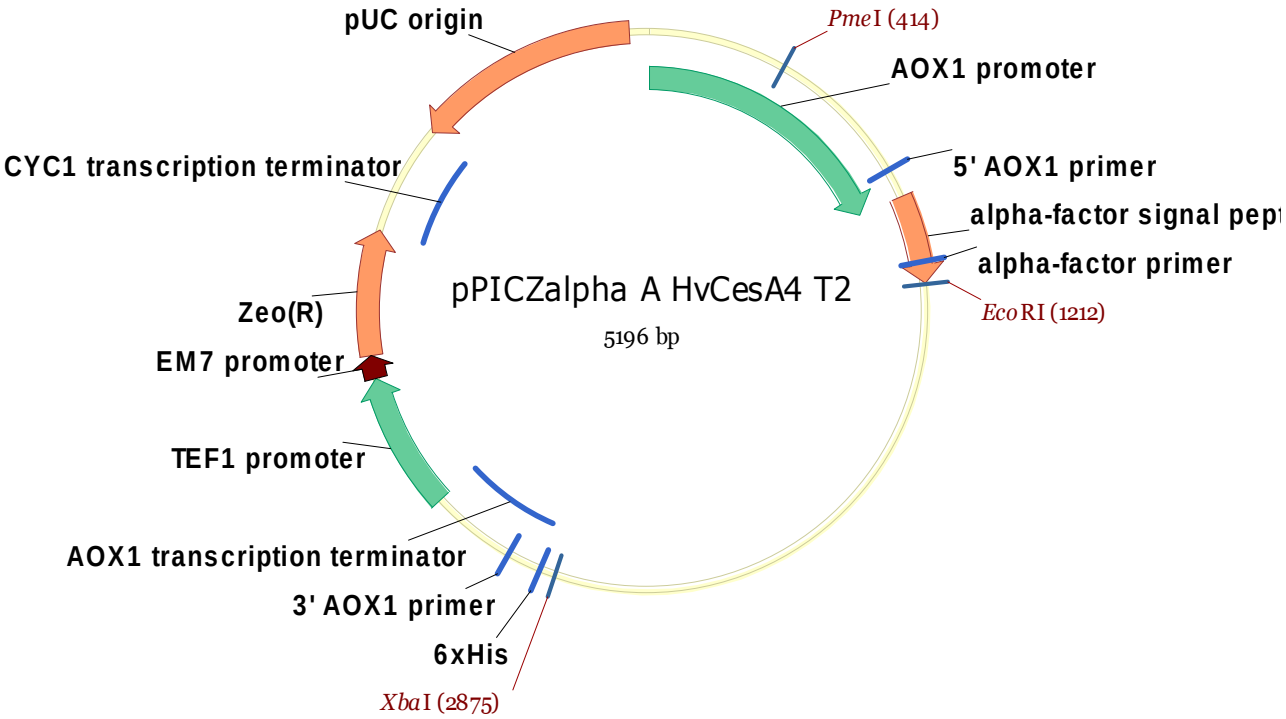


Figure 4. Destination vector pPICZα A containing HvCesA4_T2 DNA.

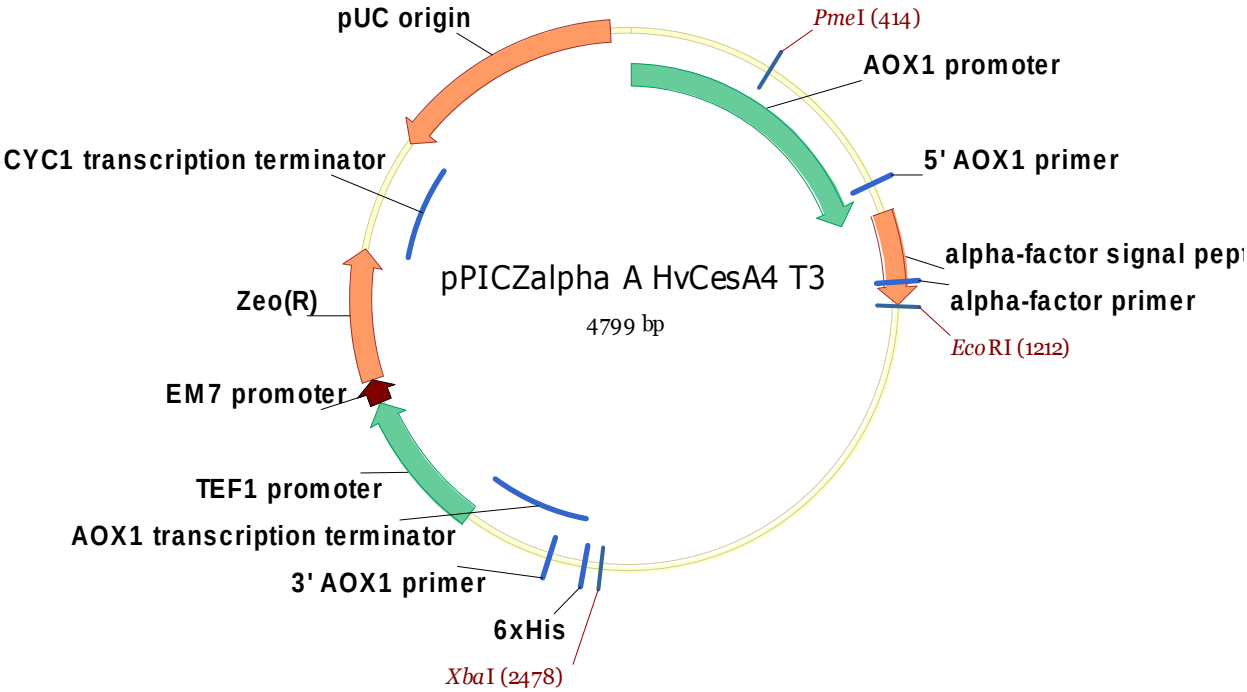


Figure 5. Destination vector pPICZα A containing HvCesA4_T3 DNA.

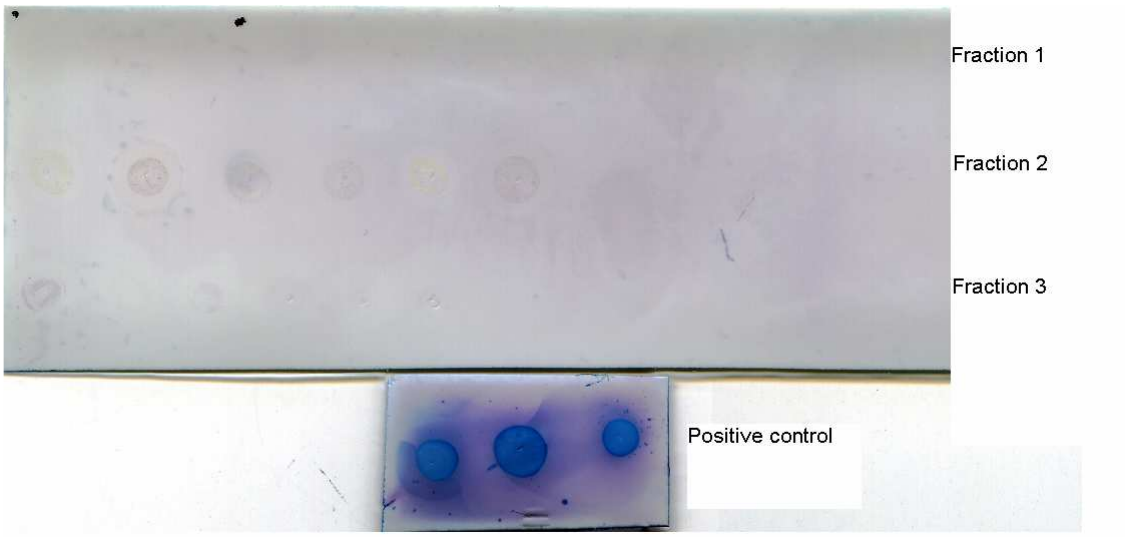


Figure 6. Dot blot of HvCesA4_T2 and HvCesA4_T3 reveals the presence of 6x-His-tagged protein in only fraction 2 and fraction 3 of the samples.

Appendices

HvCesA4 cDNA constructs without restriction sites

>*HvCesA4* full-length sequence (3128 bp)

ATGGACACCGGCGAGCCCAAGGCCAAGGTGTGCCGCGCGTGCGGCGACGATGTCGGGACGCGGGA
GGATGGCAGCCCCTTCGTGGCCTGCGCCGAGTGCGGGTTCCCGGTGTGCCGGCCGTGCTACGAGT
ACGAGCGCAGCGACGGCACGCAGTGCTGCCCCCAGTGCAACGCCCGCTACAAGCGCCACAAAGGG
TGCCCGAGGGTGGAAGGGGACGACGAGGACGGCGACATGGACGACTTAGAAGAGGAGTTCCAGGT
CAAGAGCCCCAAGAAGCCTCACGAGCCCGTTCCCTTCGACGTCTACTCGGAGAACGGCGAGCAGC
CGCCGCAGAAGTGGCGCCCCGGTGGCCCGGCCATGTCCTCCTTCGGTGGAAGCGGTCAGGAGCTT
GAGGCGGAGCGGGAGATGGAGGGGAGCATGGAGTGGAAGGACAGGATCGACAAGTGGAAGACCAA
GCAGGAGAAGAGGGGCAAGCTCAACCGCGACAACAGCGACGACGACGACGACGACAAGAACGACG
ACGAGTACATGCTGCTGGCTGAGGCGCGGCAGCCGCTGTGGCGCAAGCTGCCGGTGCCGTCGAGC
CAGATCAACCCGTACCGCATCGTCATCGTGCTCCGCCTGGTGGTGCTCTGCTTCTTCCTCCGCTT
CCGGATCATGACCCCGGCCAACGACGCCATCCCGCTGTGGCTGGTGTCCGTCATCTGCGAGCTCT
GGTTCGCGCTCTCCTGGATCCTGGACCAGCTGCCCAAGTGGTCGCCGGTGACCCGGGAGACGTAC
CTGGACCGCCTCGCCCTGCGCTACGACCGCGAGGGCGAGCCTTCCCGGCTGTCCCCCATCGACTT
CTTCGTGAGCACGGTGGACCCGCTCAAGGAGCCCCCCATCATCACGGCCAACACCGTGCTGTCCA
TCCTCGCCGTCGACTACCCCGTGGACCGCAACAGCTGCTACGTCTCCGACGACGGCGCCTCCATG
CTCTGTTTCGACACCCTCTCCGAGACGGCCGAGTTCGCGCGCCGCTGGGTGCCCTTCTGCAAGAA
GTTCGCCATCGAGCCCCGTGCCCCAGAGTTCTACTTCTCACAGAAGATCGACTACCTCAAGGACA
AGGTGCAGCCGACGTTCGTCAAGGAGCGGCGCGCCATGAAGCGCGAGTACGAGGAGTTCAAGGTG
CGGATCAACGGGCTGGTGGCCAAGCCGAGAAGAAGCCCGAGGAAGGGTGGGTGATGCAGGACGGC
ACGCCGTGGCCCGGGAACAACACCAGAGACCACCCCGGGATGATACAGGTGTATTTGGGCAGCCA
GGGCGCGCTGGACGTGGAGGGCCACGAGCTGCCGCGGTTGGTGTACGTGTCCCGAGAGAAGAGGC
CAGGACACAACCACCACAAGAAGGCCGGCGCCATGAACGCGCTGGTGCGGGTGTCGGCGGTGCTC
ACCAACGCGCCCTTCATCCTCAACCTCGACTGCGACCACTACGTGAACAACAGCAAGGCCGTCCG
CGAGGCCATGTGCTTCCTCATGGACCCCCAGCTCGGCAAGAAACTCTGCTACGTCCAGTTCCCGC
AGGGCTTCGACGGCATCGACCTCCACGATCGATACGCCAACCGCAACGTCGTCTTCTTCGACATC
AACATGAAGGGGCTAGACGGCATACAGGGCCCGGTGTACGTGGGCACGGGGTGCGTGTTCAACAG
GCAGGCGCTGTACGGGTACGACCCGCCGCGGCCGGAGAAGAGGCCCAAGATGACGTGCGACTGCT
GGCCATCGTGGTGCTGCTGCTGCTGCTGCTTCGGCGGAGGGAAGCACCGCAAATCGAGCAAGGAC
AAGAAGGGCGGCGGCGGCGGCGACGACGAGCCCCGGCGCGGGCTCCTTGGGTTCTACAAGAAGCG
GGGCAAGAAGGACAAGCTCGGCGGCGGGCCGAAGAAGGGGTCGTACAGGAAGCGGCAGCGCGGGT
ACGAGCTGGAGGAGATCGAGGAGGGCATCGAGGGGTACGACGAGCTGGAGCGCTCCTCGCTCATG
TCACAGAAGAGCTTCCAGAAGCGGTTCGGCCAGTCGCCGGTGTTCATCGCCTCCACGCTCGTCGA
GGACGGCGGGCTGCCGCAGGGCGCTGCCGCCGACCCCGCAGGCCTCATCAAGGAGGCTATCCACG
TCATCAGCTGTGGGTACGAAGGGAAGACCGAGTGGGGCAAGGAGATTGGGTGGATCTACGGGTCG
GTGACAGAGGACATCCTGACGGGCTTCAAGATGCACTGCCGTGGGTGGAAGTCCGTCTACTGCAC
GCCCACACGGCCAGCGTTCAAGGGATCGGCGCCCATCAACTTGTCGGACAGGCTTCACCAGGTGC
TTCGTTGGGCGCTCGGTTCCGTCGAGATCTTCATGAGCCGTCATTGCCCGCTCTGGTACGCCTAC
GGCGGCCGTCTCAAGTGGCTCGAGCGCTTCGCCTACACCAACACCATCGTCTACCCCTTCACCTC
CATTCCCCTCATCGCCTACTGCACCATCCCCGCCGTCTGCCTCCTCACCGGCAAGTTCATAATCC
CCACGCTCAACAACCTGGCGAGCATTTGGTTCATCGCCTTGTTCATGTCCATCATCGCGACGGGG
GTGCTGGAGCTGCGGTGGAGCGGGGTGAGCATCGAGGACTGGTGGCGTAACGAGCAGTTCTGGGT
CATAGGTGGCGTGTCCGCGCATCTGTTCGCCGTCTTCCAGGGCTTCCTCAAGGTGTTGGGCGGAG
TGGACACCAACTTCACTGTCACCTCCAAGGCGGGCGCTGATGAGGCCGATGCATTCGGTGACCTC
TACCTCTTCAAGTGGACAACCCTGCTGATCCCGCCCACCACGCTCATCATTATCAACATGGTCGG
CATCGTCGCCGGCGTGTCCGACGCCGTCAACAATGGGTACGGGTCCTGGGGCCCGCTCTTCGGGA

AGCTCTTCTTCTCCTTCTGGGTCATCGTTCACCTCTACCCGTTCCTCAAGGGGCTCATGGGGAGG
CAGAACCGGACGCCCACCATCGTTGTGCTCTGGTCTGTCCTGCTCGCCTCCATCTTCTCCCTTGT
GTGGGTCAGGATCGACCCCTTCATTGCCAAGCCCAAGGGACCCATTCTTAAGCCGTGTGGAGTAC
AGTGCTGA

>*HvCesA4_T1* (2044 bp)
AGTCGAATTCAAAATGGGTCATCATCATCATCATCATGAAAATCTGTACTTTCAAGGTACCCCGG
CCAACGACGCCATCCCGCTGTGGCTGGTGTCCGTCATCTGCGAGCTCTGGTTCGCGCTCTCCTGG
ATCCTGGACCAGCTGCCCAAGTGGTCGCCGGTGACCCGGGAGACGTACCTGGACCGCCTCGCCCT
GCGCTACGACCGCGAGGGCGAGCCTTCCCGGCTGTCCCCCATCGACTTCTTCGTGAGCACGGTGG
ACCCGCTCAAGGAGCCCCCCATCATCACGGCCAACACCGTGCTGTCCATCCTCGCCGTCGACTAC
CCCGTGGACCGCAACAGCTGCTACGTCTCCGACGACGGCGCCTCCATGCTCTGTTTCGACACCCT
CTCCGAGACGGCCGAGTTCGCGCGCCGCTGGGTGCCCTTCTGCAAGAAGTTCGCCATCGAGCCCC
GTGCCCCAGAGTTCTACTTCTCACAGAAGATCGACTACCTCAAGGACAAGGTGCAGCCGACGTTC
GTCAAGGAGCGGCGCGCCATGAAGCGCGAGTACGAGGAGTTCAAGGTGCGGATCAACGGGCTGGT
GGCCAAGGCCGAGAAGAAGCCCGAGGAAGGGTGGGTGATGCAGGACGGCACGCCGTGGCCCGGGA
ACAACACCAGAGACCACCCCGGGATGATACAGGTGTATTTGGGCAGCCAGGGCGCGCTGGACGTG
GAGGGCCACGAGCTGCCGCGGTTGGTGTACGTGTCCCGAGAGAAGAGGCCAGGACACAACCACCA
CAAGAAGGCCGGCGCCATGAACGCGCTGGTGCGGGTGTCGGCGGTGCTCACCAACGCGCCCTTCA
TCCTCAACCTCGACTGCGACCACTACGTGAACAACAGCAAGGCCGTCCGCGAGGCCATGTGCTTC
CTCATGGACCCCCAGCTCGGCAAGAAACTCTGCTACGTCCAGTTCCCGCAGGGCTTCGACGGCAT
CGACCTCCACGATCGATACGCCAACCGCAACGTCGTCTTCTTCGACATCAACATGAAGGGGCTAG
ACGGCATACAGGGCCCGGTGTACGTGGGCACGGGGTGCGTGTTCAACAGGCAGGCGCTGTACGGG
TACGACCCGCCGCGGCCGGAGAAGAGGCCCAAGATGACGTGCGACTGCTGGCCATCGTGGTGCTG
CTGCTGCTGCTGCTTCGGCGGAGGGAAGCACCGCAAATCGAGCAAGGACAAGAAGGGCGGCGGCG
GCGGCGACGACGAGCCCCGGCGCGGGCTCCTTGGGTTCTACAAGAAGCGGGGCAAGAAGGACAAG
CTCGGCGGCGGGCCGAAGAAGGGGTCGTACAGGAAGCGGCAGCGCGGGTACGAGCTGGAGGAGAT
CGAGGAGGGCATCGAGGGGTACGACGAGCTGGAGCGCTCCTCGCTCATGTCACAGAAGAGCTTCC
AGAAGCGGTTCGGCCAGTCGCCGGTGTTCATCGCCTCCACGCTCGTCGAGGACGGCGGGCTGCCG
CAGGGCGCTGCCGCCGACCCCGCAGGCCTCATCAAGGAGGCTATCCACGTCATCAGCTGTGGGTA
CGAAGGGAAGACCGAGTGGGGCAAGGAGATTGGGTGGATCTACGGGTCGGTGACAGAGGACATCC
TGACGGGCTTCAAGATGCACTGCCGTGGGTGGAAGTCCGTCTACTGCACGCCCACACGGCCAGCG
TTCAAGGGATCGGCGCCCATCAACTTGTCGGACAGGCTTCACCAGGTGCTTCGTTGGGCGCTCGG
TTCCGTCGAGATCTTCATGAGCCGTCATTGCCCGCTCTGGTACGCCTACGGCGGCCGTCTCAAGT
GGCTCGAGCGCTTCGCCTACACCAACACCATCGTCTACCCCTTCACCTCCATTCCCCTCATCGCC
TACTGCACCATCCCCGCCGTCTGCCTCCTCACCGGCAAGTTCATAATCCCCACGCTCAACAACCT
GGCGAGCATTTGGTTCATCGCCTTGTTCATGTCCATCATCGCGACGGGGGTGCTGGAGCTGCGGT
GGAGCGGGGTGAGCTGAGGATAATCTAGA

>*HvCesA4_T2* (1631 bp)
CTCGCCCTGCGCTACGACCGCGAGGGCGAGCCTTCCCGGCTGTCCCCCATCGACTTCTTCGTGAG
CACGGTGGACCCGCTCAAGGAGCCCCCCATCATCACGGCCAACACCGTGCTGTCCATCCTCGCCG
TCGACTACCCCGTGGACCGCAACAGCTGCTACGTCTCCGACGACGGCGCCTCCATGCTCTGTTTC
GACACCCTCTCCGAGACGGCCGAGTTCGCGCGCCGCTGGGTGCCCTTCTGCAAGAAGTTCGCCAT
CGAGCCCCGTGCCCCAGAGTTCTACTTCTCACAGAAGATCGACTACCTCAAGGACAAGGTGCAGC
CGACGTTCGTCAAGGAGCGGCGCGCCATGAAGCGCGAGTACGAGGAGTTCAAGGTGCGGATCAAC
GGGCTGGTGGCCAAGGCCGAGAAGAAGCCCGAGGAAGGGTGGGTGATGCAGGACGGCACGCCGTG
GCCCGGGAACAACACCAGAGACCACCCCGGGATGATACAGGTGTATTTGGGCAGCCAGGGCGCGC
TGGACGTGGAGGGCCACGAGCTGCCGCGGTTGGTGTACGTGTCCCGAGAGAAGAGGCCAGGACAC
AACCACCACAAGAAGGCCGGCGCCATGAACGCGCTGGTGCGGGTGTCGGCGGTGCTCACCAACGC
GCCCTTCATCCTCAACCTCGACTGCGACCACTACGTGAACAACAGCAAGGCCGTCCGCGAGGCCA

TGTGCTTCCTCATGGACCCCCAGCTCGGCAAGAAACTCTGCTACGTCCAGTTCCCGCAGGGCTTC
GACGGCATCGACCTCCACGATCGATACGCCAACCGCAACGTCGTCTTCTTCGACATCAACATGAA
GGGGCTAGACGGCATACAGGGCCCGGTGTACGTGGGCACGGGGTGCGTGTTCAACAGGCAGGCGC
TGTACGGGTACGACCCGCCGCGGCCGGAGAAGAGGCCCAAGATGACGTGCGACTGCTGGCCATCG
TGGTGCTGCTGCTGCTGCTGCTTCGGCGGAGGGAAGCACCGCAAATCGAGCAAGGACAAGAAGGG
CGGCGGCGGCGGCGACGACGAGCCCCGGCGCGGGCTCCTTGGGTTCTACAAGAAGCGGGGCAAGA
AGGACAAGCTCGGCGGCGGGCCGAAGAAGGGGTCGTACAGGAAGCGGCAGCGCGGGTACGAGCTG
GAGGAGATCGAGGAGGGCATCGAGGGGTACGACGAGCTGGAGCGCTCCTCGCTCATGTCACAGAA
GAGCTTCCAGAAGCGGTTCGGCCAGTCGCCGGTGTTCATCGCCTCCACGCTCGTCGAGGACGGCG
GGCTGCCGCAGGGCGCTGCCGCCGACCCCGCAGGCCTCATCAAGGAGGCTATCCACGTCATCAGC
TGTGGGTACGAAGGGAAGACCGAGTGGGGCAAGGAGATTGGGTGGATCTACGGGTCGGTGACAGA
GGACATCCTGACGGGCTTCAAGATGCACTGCCGTGGGTGGAAGTCCGTCTACTGCACGCCCACAC
GGCCAGCGTTCAAGGGATCGCGCCCATCAACTTGTCGGACAGGCTTCACCAGGTGCTTCGTTGGG
CGCTCGGTTCCGTCGAGATCTTCATGAGCCGTCATTGCCCGCTCTGGTACGCCTACGGCGGCCGT
CTCAAG

>*HvCesA4_T3* (1233 bp)
CTCGCCCTGCGCTACGACCGCGAGGGCGAGCCTTCCCGGCTGTCCCCCATCGACTTCTTCGTGAG
CACGGTGGACCCGCTCAAGGAGCCCCCCATCATCACGGCCAACACCGTGCTGTCCATCCTCGCCG
TCGACTACCCCGTGGACCGCAACAGCTGCTACGTCTCCGACGACGGCGCCTCCATGCTCTGTTTC
GACACCCTCTCCGAGACGGCCGAGTTCGCGCGCCGCTGGGTGCCCTTCTGCAAGAAGTTCGCCAT
CGAGCCCCGTGCCCCAGAGTTCTACTTCTCACAGAAGATCGACTACCTCAAGGACAAGGTGCAGC
CGACGTTCGTCAAGGAGCGGCGCGCCATGAAGCGCGAGTACGAGGAGTTCAAGGTGCGGATCAAC
GGGCTGGTGGCCAAGGCCGAGAAGAAGCCCGAGGAAGGGTGGGTGATGCAGGACGGCACGCCGTG
GCCCGGGAACAACACCAGAGACCACCCCGGGATGATACAGGTGTATTTGGGCAGCCAGGGCGCGC
TGGACGTGGAGGGCCACGAGCTGCCGCGGTTGGTGTACGTGTCCCGAGAGAAGAGGCCAGGACAC
AACCACCACAAGAAGGCCGGCGCCATGAACGCGCTGGTGCGGGTGTCGGCGGTGCTCACCAACGC
GCCCTTCATCCTCAACCTCGACTGCGACCACTACGTGAACAACAGCAAGGCCGTCCGCGAGGCCA
TGTGCTTCCTCATGGACCCCCAGCTCGGCAAGAAACTCTGCTACGTCCAGTTCCCGCAGGGCTTC
GACGGCATCGACCTCCACGATCGATACGCCAACCGCAACGTCGTCTTCTTCGACATCAACATGAA
GGGGCTAGACGGCATACAGGGCCCGGTGTACGTGGGCACGGGGTGCGTGTTCAACAGGCAGGCGC
TGTACGGGTACGACCCGCCGCGGCCGGAGAAGAGGCCCAAGATGACGTGCGACTGCTGGCCATCG
TGGTGCTGCTGCTGCTGCTGCTTCGGCGGAGGGAAGCACCGCAAATCGAGCAAGGACAAGAAGGG
CGGCGGCGGCGGCGACGACGAGCCCCGGCGCGGGCTCCTTGGGTTCTACAAGAAGCGGGGCAAGA
AGGACAAGCTCGGCGGCGGGCCGAAGAAGGGGTCGTACAGGAAGCGGCAGCGCGGGTACGAGCTG
GAGGAGATCGAGGAGGGCATCGAGGGGTACGACGAGCTGGAGCGCTCCTCGCTCATGTCACAG

Protein topology and secondary structures

>HvCesA4, amino acid sequence of 1082 residues, bold characters show conserved motifs QXXLRW, TED and SDD.

MDTGEPKAKVCRACGDDVGTREDGSPFVACAECGFPVCRPCYEYERSDGT
QCCPQCNARYKRHKGCPRVEGDDEDGDMDDLEEEFQVKSPKKPHEPVPFD
VYSENGEQPPQKWRPGGPAMSSFGGSGQELEAEREMEGSMEWKDRIDKWK
TKQEKRGKLNRDN**SDD**DDDDKNDDEYMLLAEARQPLWRKLPVPSSQINPY
RIVIVLRLVVLCFFLRFRIMTPANDAIPLWLVSVICELWFALSWILDQLP
KWSPVTRETYLDRLALRYDREGEPSRLSPIDFFVSTVDPLKEPPIITANT
VLSILAVDYPVDRNSCYVSDDGASMLCFDTLSETAEFARRWVPFCKKFAI
EPRAPEFYFSQKIDYLKDKVQPTFVKERRAMKREYEEFKVRINGLVAKAE
KKPEEGWVMQDGTPWPGNNTRDHPGMIQVYLGSQGALDVEGHELPRLVYV
SREKRPGHNHHKKAGAMNALVRVSAVLTNAPFILNLDCDHYVNNSKAVRE
AMCFLMDPQLGKKLCYVQFPQGFDGIDLHDRYANRNVVFFDINMKGLDGI
QGPVYVGTGCVFNRQALYGYDPPRPEKRPKMTCDCWPSWCCCCCFGGGK
HRKSSKDKKGGGGGDDEPRRGLLGFYKKRGKKDKLGGGPKKGSYRKRQRG
YELEEIEEGIEGYDELERSSLMSQKSFQKRFGQSPVFIASTLVEDGGLPQ
GAAADPAGLIKEAIHVISCGYEGKTEWGKEIGWIYGSV**TED**ILTGFKMHC
RGWKSVYCTPTRPAFKGSAPINLSDRLH**QVLRW**ALGSVEIFMSRHCPLWY
AYGGRLKWLERFAYTNTIVYPFTSIPLIAYCTIPAVCLLTGKFIIPTLNN
LASIWFIALFMSIIATGVLELRWSGVSIEDWWRNEQFWVIGGVSAHLFAV
FQGFLKVLGGVDTNFTVTSKAGADEADAFGDLYLFKWTTLLIPPTTLIII
NMVGIVAGVSDAVNNGYGSWGPLFGKLFFSFWVIVHLYPFLKGLMGRQNR
TPTIVVLWSVLLASIFSLVWVRIDPFIAKPKGPILKPCGVQC

Predicted membrane helices
TMpred(http://www.ch.embnet.org/software/TMPRED_form.html)

The sequence positions in brackets denote the positions of helices. Only scores above 500 are considered to be significant.
Inside to outside helices: 10 found

From	to	score	center
200 (204)	223 (220)	1187	212
226 (226)	246 (244)	1471	236
466 (466)	484 (482)	643	474
581 (581)	597 (597)	922	589
812 (820)	839 (836)	1310	828
851 (851)	869 (869)	2435	859
886 (886)	905 (905)	1331	895
939 (939)	959 (957)	1339	949
973 (973)	991 (989)	1727	981
1001 (1002)	1020 (1020)	2340	1011

Outside to inside helices: 13 found

from	to	score	center
199 (199)	215 (215)	1562	207
226 (228)	246 (244)	1584	236
292 (292)	310 (310)	109	300
469 (469)	486 (486)	563	478
546 (549)	570 (567)	396	559
581 (581)	600 (600)	285	591
777 (784)	803 (803)	161	793
822 (822)	840 (840)	1421	830
851 (851)	869 (869)	2364	861
886 (889)	905 (905)	1401	897
942 (942)	960 (960)	1511	950
970 (970)	991 (988)	1437	978
1005 (1005)	1021 (1021)	2285	1013

Toppred (http://cbi.labri.fr/outils/Pise/toppred.html)

Helix	Begin - End	Score	Certainity
1	226 – 246	1.583	Certain
2	291 – 311	0.625	Putative
3	581 – 601	0.998	Putative
4	821 - 841	1.699	Certain
5	850 - 870	2.191	Certain
6	885 - 905	1.419	Certain
7	940 - 960	1.697	Certain
8	968 - 988	1.447	Certain
9	1001 – 1021	2.175	Certain

DAS(http://molbioltools.ca/Protein_secondary_structure.htm)
Potential transmembrane segments
Start Stop Length ~ Cutoff
200 219 20 ~ 1.7
202 217 16 ~ 2.2
227 246 20 ~ 1.7
228 243 16 ~ 2.2
282 284 3 ~ 1.7
297 307 11 ~ 1.7
299 306 8 ~ 2.2
470 483 14 ~ 1.7
473 478 6 ~ 2.2
555 556 2 ~ 1.7
687 691 5 ~ 1.7
784 786 3 ~ 1.7
818 870 53 ~ 1.7
822 845 24 ~ 2.2
851 868 18 ~ 2.2
890 909 20 ~ 1.7
892 905 14 ~ 2.2
934 959 26 ~ 1.7
939 958 20 ~ 2.2
974 991 18 ~ 1.7
976 990 15 ~ 2.2
1003 1023 21 ~ 1.7
1004 1021 18 ~ 2.2