This is to certify that the

thesis entitled

Molecular Characterization of a Conserved Arabidopsis Gene Involved in Cell Wall Synthesis

presented by

Zhaohong Wang

has been accepted towards fulfillment
of the requirements for

MS degree in Botany

Kenneth Keegstra
Major professor

Date 8/9/95

MOLECULAR CHARACTERIZATION OF A CONSERVED ARABIDOPSIS GENE INVOLVED IN CELL WALL SYNTHESIS

By

Zhaohong Wang

A THESIS

Submitted to

Michigan State University

in partial fulfilment of the requirements

for the degree of

MASTER OF SCIENCE

Department of Botany and Plant Pathology

1995

ABSTRACT

MOLECULAR CHARACTERIZATION OF A CONSERVED ARABIDOPSIS GENE INVOLVED IN CELL WALL SYNTHESIS

By

Zhaohong Wang

Although the plant cell wall plays a critical role in plant cells, very little is known about cell wall synthesis. Xyloglucan is the major hemicellulose in the primary cell wall of dicots, comprising about 20% of its dry weight. Xyloglucan consists of on extended chain of β-1,4-linked glucosyl residues to which xylosyl and xylosyl-galactosyl-fucosyl side chains are attached at regular intervals. Xyloglucan hydrogen bonds to and forms cross bridges between neighboring cellulose microfibrils. Xyloglucan cleavage is thought to allow cellulose microfibrils to separate during cell growth.

Xyloglucan biosynthesis is known to occur in Golgi membranes. Glucan synthase I is a glucosyl transferase also present in the Golgi; one of its possible functions is the synthesis of xyloglucan backbone. Dhugga et. al. (Dhugga KS Ulvskov P, Gallagher SR, Ray PM [1991] J. Biol. Chem. 266: 21977-21984) have purified to homogeneity a 40 kD protein reportedly involved in glucan synthase I function. The 40 kD protein can be reversibly glycosylated by UDP-Glc, UDP-Gal, UDP-Xyl, but not UDP-Man. Based on this evidence, they proposed that this protein could play a role in xyloglucan synthesis.We used

their peptide sequence information to search the dBEST and found 5 Arabidopsis cDNAs that encode these peptides. Restriction mapping of these clones revealed that the peptides are encoded by two similar, but not identical, genes. Two cDNA clones, ATA1 and ATB1 were sequenced. The complete nucleotide sequence of ATA1 was used to search dBEST. Fifteen Arabidopsis cDNA clones and seven rice cDNA clones were found which are similar to ATA1 as of the last search on June 15. Some of the Arabidopsis clones, were mapped with restriction enzymes and found to have restriction maps identical to either the ATA and ATB genes. Although the deduced amino acid sequences of the rice and Arabidopsis proteins are very similar, they have little similarity to other proteins in the databases. I postulate that these proteins play an important role that is unique to plants. Cell wall synthesis is one possibility.

Based on the work of Dhugga et. al. plus what we learned about this gene family, we propose two possible functions for this protein. A) It may function as an monosacchride intermediate. Monosaccharides are transferred from nucleotide sugar to the protein and then to the growing xyloglucan chain. B) It works as an oligosaccharide intermediate. Several sugar residues are transferred to the protein to form an oligosaccharide repeat unit, which is transferred to the growing xyloglucan chain. Further work will be needed to evaluate the possibilities.

ACKNOWLEDGMENTS

I wish to express my sincerest thanks to the people that have shared their knowledge, advice, experience, and time with me. My sincerest thanks to my advisor Dr. Ken Keegstra for his guidance, help and support for the experiments described in this thesis, and for his understanding and encouragment.

I thank Dr. Jon Walton and Dr. Michael Bagadasarian for serving on my guidance committee.

I also thank members of Keegstra lab, Pat, Amy, John, Eric, Jenny, Arun, Mitsru and Karen, and my friends at MSU for their kindly help.

And thanks my family for their forever love and support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The wall is an important and unique part of a plant cell. It not only provides mechanical support to help a plant to maintain form and structural integrity, but it also contains components involved in signaling, communication and pathogen defense. The cell walls of higher plants can be divided into two types: primary walls and secondary walls. Primary walls are produced by growing cells and can elongate during growth; secondary walls are less flexible and almost inelastic. Once the primary wall is formed, the wall continues to thicken after elongation stops, as layers of cellulose and lignin are deposited to form the secondary wall. This brief review will consider only the primary wall; it consists of cellulose, hemicellulose, pectins, and proteins.

Cellulose constitutes 20%-30% of the dry weight of primary walls. Chemically, cellulose is a linear β-1-4-linked D-glucan. The glucan chains of cellulose are aggregated together to form structures called microfibrils. Xyloglucan, a hemicellulose, is about 20% of the dry weight of the primary wall in dicots. Although it is present both in dicots and monocots, it is a minor component in most monocot cell walls. The basic structure of this cell wall polymer consists of a backbone of β-4-linked-D-glucosyl residues, with D-xylosyl side chains α-linked to O-6 of some the glucosyl residues. Some xylosyl side chains have D-galactosyl or L-fucosyl-α-2-D-galactosyl β-linked to the O-2 of the xylosyl residues (see Figure 1-1). In the wall, most of the xyloglucans are tightly hydrogen bonded to cellulose microfibrils. Pectins and proteins comprise the remainder (approximately 50%) of the

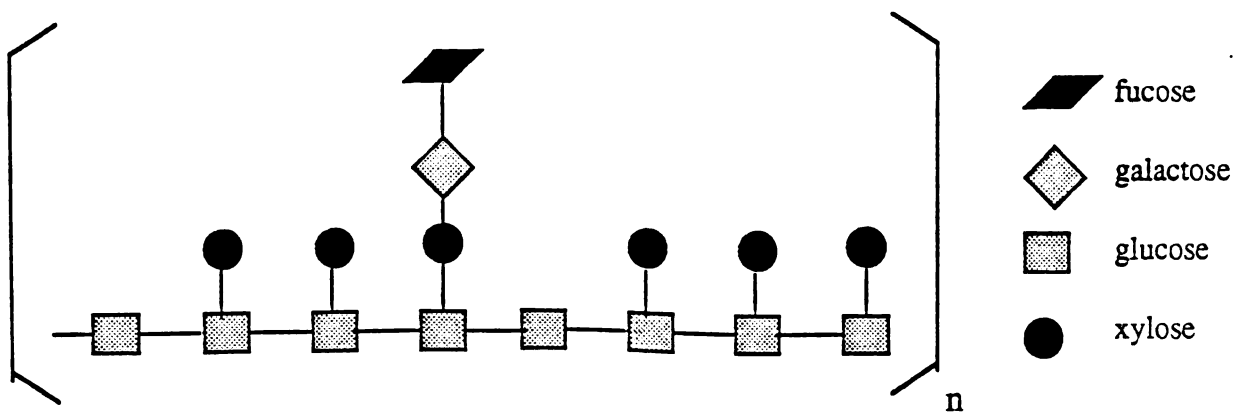


Figure 1-1. The basic structure of xyloglucan

primary cell wall (McNeil et al., 1984). Although they are important components of primary cell walls, they will not be considered further in this brief review.

How are these different polymers assembled and altered to form an extensible wall? The pioneering work of Albersheim's group provided an early comprehensive model. In their model, all non-cellulosic polymers are covalently linked except for hydrogen bonding of the xyloglucans to the cellulose microfibrils (Keegstra et al., 1973; Talmadge et al., 1973) The concept that primary wall coherence is based on a covalently cross-linked meshwork of matrix molecules has influenced thinking about wall growth for a long time. Another new model was later provided by Peter Ray's group (Talbott and Ray, 1992). They describe a model in which noncovalent associations, such as among pectic polyuronides, between them and hydroxyl proline-rich wall glycoproteins (extensins) and between XG and cellulose, are principally responsible for the mechanical coherence of the growing wall.

Considerable progress has been made in recent years in understanding the structure of wall components and the ways in which these components are arranged in the wall. However, there has been little progress in understanding the molecular details of the biosynthesis and assembly of the plant cell wall and its components. This information is also essential for a complete description of the physiology of plant growth.

Polysaccharides, which are the major components of cell wall, will be the focus of this brief review of the synthesis of wall components. Their synthesis can be divided into three steps: prepolymerization, polymerization,

post polymerization (Stoddart, 1984). The prepolymerization steps are synthesis and activation of monomers. The source of monosaccharide building units for wall polysaccharides is carbohydrate translocated from photosynthesis or storage tissues to the site of synthesis. The immediate donors of monosaccharides for polymerization are the nucleoside diphosphate monosaccharides (NDP-monosaccharides). These are generated from the cellular pool of hexose phosphates by soluble nucleoside triphosphate pyrophosphorylases. In tissues showing active wall synthesis, total NDP-monosaccharides are estimated to be present in up to millimolar concentrations in the cytoplasm with UDP-Glc (75%) being by far the most abundant, the next most abundant being UDP-Gal (Carpita and Delmer, 1981).

The enzymes generating nucleoside precursors have been found both in soluble and membrane fractions (Fincher and Stone, 1981). Although it's not clear in which subcellular compartment they occur, this is a crucial point to understanding the control of precursor supply to the sites of polysaccharide synthesis. Current available evidence points to the endomembrane system as the site of polymerization of all of the wall matrix, like polysaccharides and glycoproteins (Northcote, 1985). In higher plants, cellulose is simultaneously polymerized and deposited in the wall by a plasma-membrane-localized cellulose synthase complex. This latter important process will not be described here.

The process of polymerization can be divided into three steps: chain initiation, chain elongation and chain termination. It seems likely that the first sugar transfer must differ in mechanism from the succeeding transfers and may involve use of a lipid or protein as acceptor (Maclachlan, 1981; Stoddart, 1984). Phosphorylated polyprenols have a well-established role as carrier of mono-, di-, and oligosaccharides in the synthesis of bacterial wall polymers such as peptidoglycan, teichoic acids, and the O-antigen portion of lipopolysaccharide (Stoddart, 1984) and in the glycosylation of asparagine-linked glycoproteins in both plants and animals (Elbein, 1979; Kornfeld et al., 1985). Such carriers might well be expected to play a role in plant cell wall polysaccharide synthesis either as chain initiators or in subsequent elongation. However, to date, there is no solid evidence that any of these compounds are involved in wall polysaccharide biosynthesis.

Xyloglucan, a branched polysaccharide, has a regular repeat in its structure (Figure 1-1). However, the pathway whereby the repeat is synthesized is unknown. A Mn^{2+}-stimulated (1-4)-β-glucan synthase and an associated xylosyltransferase are thought to be involved in synthesis of xyloglucan; both enzymes are Golgi located (Hayashi and Matsuda, 1981; Delmer et al., 1985). The extension of the xyloglucan chain of soybean preparations involves cooperative transfer since the addition of xylosyl residues is apparently essential for extension of the glucosyl chain. Thus the soybean xyloglucan pentasaccharide unit is converted to the heptasaccharide unit and by addition of a nonreducing glucosyl residue and its substitution by a xylosyl residue and this conversion is dependent on the UDP-Xyl concentration

(Hayashi et al., 1984) . However in peas (Ray, 1980) the formation of the glucan backbone was not stimulated by UDP-Xyl but UDP-Glc stimulates xylose transfer.

Completion of the soybean xyloglucan structure by addition of galactose and fucose residues to the xylose substituents on the xyloglucan appears to be independent of chain growth since enzymically synthesized xyloglucan which lacks these residues has a similar molecular weight to the xyloglucan from soybean walls (Hayashi et al. 1984). A considerable amount of work on the pea fucosyl transferase has been done by Maclachalan's group (Maclachalan 1985).

Construction of the cell wall requires that substances produced inside the cell migrate through the plasmalemma to the wall. The matrix wall components are delivered to the plasmalemma by a process similar to secretion in animal cells. The cell wall proteins, which are synthesized on the rough endoplasmic reticulum and cell wall polysaccharides, hemicelluloses, and pectic substances produced in the Golgi apparatus, seem to be incorporated into secretory vesicles that "bud" of the Golgi apparatus or the endoplasmic reticulum and then fuse with the plasmalemma in such a manner that the products are released to the wall area (Devin and Witham, 1983).

Considering the complex composition and organization of the plant cell wall, its synthesis should be highly regulated. It could be controlled at the level of the supply of monosaccharide, at the level of their activation and transport, at polymerization steps, by control of the amounts of activity of the polymerase and glycosyltransferases, and by control of precusor

polysaccharide transport to the cell wall surface and their secretion into the wall. The phytohormones, auxin, gibberellins, cytokinins, abscisic acid and ethylene also have direct or indirect effects on wall metabolism and could regulate some or all of these steps..

During their study of cell wall biosynthesis enzymes glucan synthase I (GS-I) and glucan synthase II (GS-II), Dhugga et. al. found that two closely spaced polypeptides (doublet) of ~40 kD became reversibly glycosylated by UDP-Glc under glucan synthase-I (GS-I) assay conditions (Dhugga et al., 1991) . In addition to being Golgi-associated, these polypeptides also occur in a soluble form. The 40 kD polypeptides, which can be labeled with [^{14}C]UDP-Glc or [^{14}C]UDP-Xyl under GS-I assay conditions. The labeling can be inhibited by including unlabeled UDP-Glc or UDP-Xyl. These polypeptides do not possess GS-I activity by themselves, but rather appear to be primary acceptors for sugars (Glc, Gal, Xyl, but not Man) from UDP-sugars (Dhugga et al., 1991) . In our work, we refer to this protein as polysaccharide synthesis intermediate (PSI) protein.

Based on this evidence, they proposed that this protein could play a role in xyloglucan synthesis. Subsequently they have purified the soluble 40 kD polypeptides to homogeneity by ammonium sulfate precipitation and affinity chromatography on UDP-Agarose. After passing the ammonium sulfate concentrate through the affinity column in the presence of Mn^{2+}, the bound polypeptides were eluted with EDTA. They also sequenced 3 peptide fragments from this 40 kD protein (Dhugga et al., 1994). We compared these sequences with dBEST database and identified 5 Arabidopsis cDNAs that

could encodes these peptides. Within these 5 clones, there are 2 related but not identical cDNAs. I characterized these clones, and used ATA1 clone, which we believe is full length to further search dBEST. This search yielded several more Arabidopsis and rice clones. All these clones are highly similar. But searching the protein data bases, we don't find any similar proteins. All this suggests that this kind of protein may play an important and essential role unique to plants, though its function is not known now. And my work is to do some initial molecular characterization of these Arabidopsis clones.

CHAPTER 2

MATERIALS AND METHODS

Restriction Mapping

All of the cDNA clones described here were from the Columbia wild type of *Arabidopsis thaliana* (L.) Heynh. They were constructed in a λZipLox library designated PRL2. For the PRL2 library, equal amounts of poly-A(+) mRNA from the four tissue types (seedlings, roots, stems, flowers) were converted to cDNA using a SuperScript Kit from Gibco BRL. The cDNAs were ligated into the SalI-NotI sites of λZipLox, packaged, and plated on *Escherichia coli* Y1090 (pZIP) (Newman et al. 1994) . All of the cDNA clones (38C4T7, 105K4T7, 109C5T7, 86B10T7, 108D6T7, 114H5T7, 139A8T7)were kindly provided by Arabidopsis Biological Resource Center at Ohio State University.

The *E. coil* Y1090 containing the cDNA clones were grown at 37 ^{o}C in 5ml cultures of LB medium plus Ampicillin (100mg/ml) and plasmids were extracted using Magic Minipreps from Promega (Madison, WI).

The plasmids were digested with restriction enzymes SalI, EcoRI, AvaII, HindIII, BamHI, XbaI, NciI and their combinations at 37^{o}C for 4 hours. The DNA amount is 1.5 μg per aliquot. The result of restriction mapping is shown in Figure 3-1.

Sequencing and Data analysis

The ATA1 (38C4T7) and ATB1 (86B10T7) were digested with restriction enzymes and the fragments were subcloned into pBluescript II SK(+) vector (Stratagene) to make constructs for sequencing. The automated fluorescent sequencing reactions were performed by our Plant Biochemistry Facility at Michigan State University by T3, T7, SP6 dye primers using the ABI catalyst 800 for Taq cycle sequencing and the ABI 373A Sequencer for the analysis of products. Both strands of ATA1 and ATB1 were sequenced to generate the results shown in Figure 3-2 and Figure 3-3. The complete nucleotide sequence of pea clone was kindly provided by Dr. Dhugga. The complete nucleotide sequence of ATA1 was used to search dBEST and GenBank. The deduced amino acid sequence of ATA1 was used to search GenBank data. The GCG, Sequencer, DNASIS, PROSIS program were used to analysis data. The result were shown in Figure 3-4, Figure 3-5, Figure 3-6, Figure 3-7, Figure 3-8, Figure 3-9, Figure 3-10.

The chimeric protein overexpressed in *E. coli* and purification

An BamHI/HindIII fragment containing the last two thirds of the coding sequence from the cDNA clone ATA1 was inserted into corresponding sites of the pET22b(+) expression vector (Novagen) (see Figure 3-11). A single colony of *E.coli* strain BL21(DE3), harbouring the expression plasmid was grown in 50ml LB medium containing 50μg/ml ampicillin and 34μg/ml chloramphenicol at 37°C for about 3 hours until the OD_{600} reached 0.6-1. Then IPTG was added to a final concentration of 0.4mM and the expression was achieved under the control of the T7 promoter for 2-3 hours. The bacterial was harvested, resuspended in 50mM Tris-HCl, 2mM EDTA pH 8.0, and disrupted by several cycles of sonication, each 20-30 s with an interval on ice. The total lysate was separated into soluble and "inclusion body" fractions by centrifugation for 20 min at 4°C and at 10,000xg.

The samples were prepared in SDS gel loading buffer containing 0.2% 2-mer-captoethanol and denatured by heating for 10 min at 100°C. Equal volumes of extracts were loaded on SDS gel to check the location of fusion protein. The ATA1 fusion protein was founded in "inclusion body". The His-Tag affinity column (Novagen) was uesd to purify the fusion protein under denaturing condition suggested by the supplier.

CHAPTER 3

RESULTS

Restriction Mapping indicates at least two genes

The restriction map (see Figure3-1) of seven different Arabidopsis cDNA clones indicated that they can be divided into two gene groups: ATA and ATB. The ATA group includes ATA1(38C4T7), ATA3(109C5T7). ATA1 has AvaII, BamHI, NciI, HincII and HindIII restriction sites found by mapping. ATA3 also has AvaII, NciI and HindIII sites. Its restriction pattern is same as ATA1, suggesting it may be a partial clone of ATA1. The ATB group includes ATB1(86B10T7), ATB2(139A8T7), ATB3(114H5T7), ATB4(105K4T7), ATB5(108D6T7). The ATB1 has EcoRI, AvaII, BamHI, AvaII, HindIII sites found by mapping. The ATB2, ATB3, ATB4 and ATB5 has the same restriction pattern as ATB1.

Structures of the ATA1 and ATB1 genes

The ATA1 was sequenced, showing that it is 1405bp long with ATG start codon in front of a putative open reading frame and Poly(A) tail. Figure 3-2 shows the nucleotide and deduced amino acid sequence of ATA1. ATA1 potentially encodes a 364 amino acid polypeptide. Approximately 400bp 3′ from the coding region have been sequenced, allowing for the identification of poly(A) signal AATAAA, located at 27bp upstream of poly(A) addition site of the cDNA clone. The calculated molecular mass of the ATA1 polypeptide is 40kD. The amino acid sequence does not contain a typical signal sequence,

Legend to Figure 3-1. The restriction map of Arabidopsis ATA and ATB genes. Seven different Arabidopsis cDNA clones can be divided into two groups: ATA and ATB. The ATA group includes ATA1 (38C4T7), ATA3 (109C5T7). ATA1 has AvaII, BamHI, NciI, HincII, HindIII restriction sites. The ATA3 also has AvaII, NciI, HindIII sites. Its restriction pattern is same as ATA1. The ATB group has ATB1 (86B10T7), ATB2 (139A8T7), ATB3 (114H5T7), ATB4 (105K4T7), ATB5 (108D6T7). The ATB1 has EcoRI, AvaII, BamHI, AvaII, HindIII sites. The ATB2, ATB3, ATB4, ATB5 have the same restriction pattern as ATB1. A stands for AvaII. H stands for Hind III.

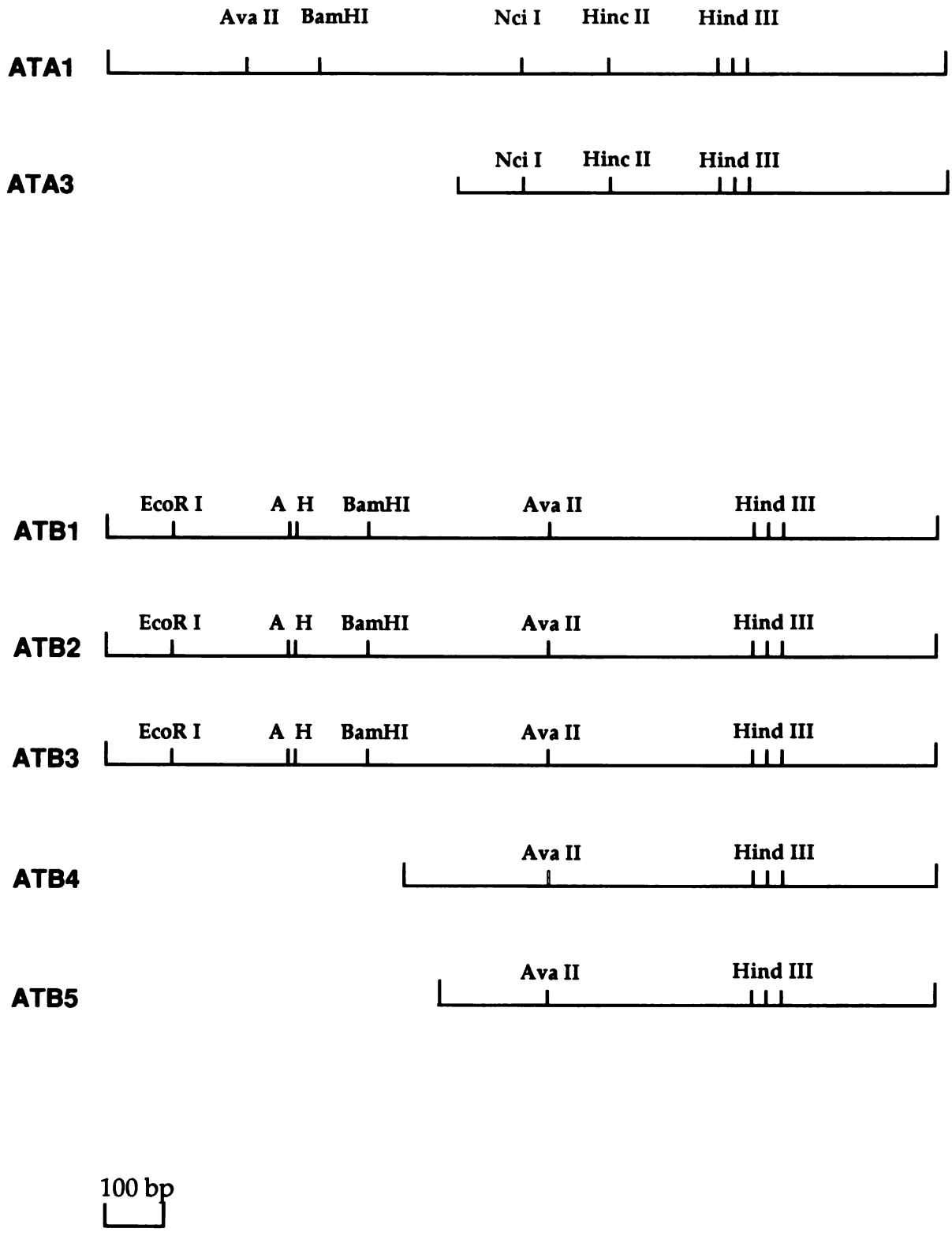


Figure 3-1. The restriction map of Arabidopsis ATA and ATB genes

Legend to Figure 3-2. The nucleotide and deduced amino acid sequences of the ATA1 gene. Numbers on the right refer to the amino acid sequence. ATA1 is 1405bp long with ATG start codon and Poly(A) tail. It potentially encodes a 364 amino acid polypeptide. The amino acid sequences matching Dhugga's pea peptide fragments are underlined. The DNASIS program was used to generate this figure.

```
GAC CCA CGC GTC CGG TAA ACC ATG GTT GAG CCG GCG AAT ACT GTT GGT CTT CCG GTG AAC
                            M   V   E   P   A   N   T   V   G   L   P   V   N      13

CCG ACT CCG TTG CTG AAA GAT GAG CTC GAT ATC GTG ATT CCG ACT ATC AGA AAC CTC GAT
P   T   P   L   L   K   D   E   L   D   I   V   I   P   T   I   R   N   L   D      33

TTC CTC GAG ATG TGG AGG CCT TTT CTT CAG CCT TAC CAT CTG ATC ATC GTC CAG GAC GGA
F   L   E   M   W   R   P   F   L   Q   P   Y   H   L   I   I   V   Q   D   G      53

GAT CCA TCG AAG AAG ATC CAT GTC CCT GAA GGT TAC GAC TAC GAG CTC TAC AAC AGG AAC
D   P   S   K   K   I   H   V   P   E   G   Y   D   Y   E   L   Y   N   R   N      73

GAC ATT AAC CGA ATC CTC GGA CCT AAG GCT TCT TGT ATC TCG TTT AAG GAT TCT GCT TGT
D   I   N   R   I   L   G   P   K   A   S   C   I   S   F   K   D   S   A   C      93

CGA TGC TTT GGG TAC ATG GTG TCT AAG AAG AAG TAT ATC TTC ACC ATT GAT GAC GAT TGC
R   C   F   G   Y   M   V   S   K   K   K   Y   I   F   T   I   D   D   D   C     113

TTC GTT GCC AAG GAT CCA TCA GGC AAA GCA GTG AAC GCT CTT GAG CAA CAC ATC AAG AAC
F   V   A   K   D   P   S   G   K   A   V   N   A   L   E   Q   H   I   K   N     133

CTT CTC TGC CCA TCG TCT CCC TTT TTC TTC AAC ACC TTG TAT GAT CCT TAC CGT GAA GGT
L   L   C   P   S   S   P   F   F   F   N   T   L   Y   D   P   Y   R   E   G     153

GCT GAT TTC GTC CGT GGA TAC CCT TTC AGT CTC CGT GAA GGT GTT TCC ACT GCT GTT TCC
A   D   F   V   R   G   Y   P   F   S   L   R   E   G   V   S   T   A   V   S     173

CAT GGT CTT TGG CTC AAC ATC CCT GAC TAC GAT GCC CCG ACC CAA CTC GTG AAG CCT AAG
H   G   L   W   L   N   I   P   D   Y   D   A   P   T   Q   L   V   K   P   K     193

GAG AGG AAC ACC AGG TAT GTG GAT GCT GTC ATG ACC AAC CCA AAG GGA ACA CTT TTC CCA
E   R   N   T   R   Y   V   D   A   V   M   T   N   P   K   G   T   L   F   P     213

ATG TGT GGT ATG AAC TTG GCT TTT GAC CGT GAT TTG ATT GGC CCG GCT ATG TAC TTT GTT
M   C   G   M   N   L   A   F   D   R   D   L   I   G   P   A   M   Y   F   V     233

CTC ATG GGT GAT GGT CAG CCT ATT GGT CGT TAC GAC GAT ATG TGG GCT GGT TGG TGC ATC
L   M   G   D   G   Q   P   I   G   R   Y   D   D   M   W   A   G   W   C   I     253

AAG GTG ATC TGT GAC CAC TTG AGC TTG GGA GTG AAG ACC GGT TTA CCG TAT ATC TAC CAC
K   V   I   C   D   H   L   S   L   G   V   K   T   G   L   P   Y   I   Y   H     273

AGC AAA GCG AGC AAC CCT TTT GTT AAC CTG AAG AAG GAA TAC AAG GGA ATC TTC TGG CAG
S   K   A   S   N   P   F   V   N   L   K   K   E   Y   K   G   I   F   W   Q     293

GAG GAG ATC ATT CCG TTC TTC CAG AAC GCA AAG CTA TCG AAA GAA GCA GTA ACT GTT CAG
E   E   I   I   P   F   F   Q   N   A   K   L   S   K   E   A   V   T   V   Q     313

CAA TGC TAC ATT GAG CTC TCA AAG ATG GTC AAG GAG AAG TTG AGC TCC TTA GAC CCG TAC
Q   C   Y   I   E   L   S   K   M   V   K   E   K   L   S   S   L   D   P   Y     333

TTT GAC AAG CTT GCA GAT GCC ATG GTT ACA TGG ATT GAA GCT TGG GAT GAG CTT AAC CCA
F   D   K   L   A   D   A   M   V   T   W   I   E   A   W   D   E   L   N   P     353

CCA GCA GCC AGT GGC AAA AGC TTG AGA GCA GTA TGA GCC AAA AAG AAA AAG CCA CCA AAG
P   A   A   S   G   K   S   L   R   A   V   *                                     365
TTT TGG TTA TTT TTA GCT CAA ATT ATC GTT ACT TTT AAA TTT CTG ATT TTA CGA ACC TTT
CTT GCT TTT TTT ACA CAT TTG AGT AGT TTT CAT CAT CAG TAC TTT CTC ATT GTC CGG TTA
TGG TTT TTG CAT TTG GTT TAA ATA TCA CCG GTT TAT TTA TAA ACA GTG GTG GAT TAG TAG
TAC TAT TTT CTG AGT TTT TTT CTT TGT TTC ATT AAT AAA AAG GCC TTT TCA TAG GTG TTT
GCA ATT AAA AAA AAA AAA AAA GGG C
```

Figure 3-2. Nucleotide and Predicted Amino Acid Sequences of ATA1

ER retention signal, or N-glycosylation signal, suggesting that the ATA1 polypeptide does not enter the secretory pathway.

The sequence of ATB1 shows that it is 1398bp long and also has an ATG start codon and Poly(A) tail. The 3′ end of ATB4 has been sequenced and it matches the 3′ end of ATB1. The 5′ ends of ATB2 and ATB3 are the same as ATB1. Figure 3-3 shows the nucleotide and deduced amino acid sequence of ATB1. Similar to ATA1, ATB1 potentially encodes a 357 amino acid amino acid polypeptide. Also the polyA signal AATAAA can be identified at 17bp upstream of poly(A) addtition site of the cDNA clone. The calculated molecular mass of ATB1 is 39.3 kD. The amino acid sequence of ATB1 also does not contain a typical signal sequence.

The nucleotide sequences of ATA1 and ATB1 were aligned as shown in Figure 3-4. Comparison at the nucleotide level, demonstrated that the ATA1 and ATB1 sequences are highly similar, are about 85% identical. The 3′ non-coding region of ATA1 and ATB1 gene show little similarity. In contrast, the sequences are highly similar within the coding region. As Shown in Figure 3-5, the deduced amino acid sequences are 93% identical. The peptide fragment "NLDFLEMXRPFFEQY" and "EGVPTAVSHGLXLNI" of the 40 kD pea protein purified by Dhugga et al. (Dhugga et al., 1994) were found in both ATA1 and ATB1 with a little variation. For the first fragment, F, E, Q were substituted by L, Q ,P in ATA1; E, Q were substituted by Q, P in ATB1. For the second fragment, P was substitued by S, both in ATA1 and ATB1. And comparing the size of the deduced amino acid peptide of ATA1 and ATB1

Legend to Figure 3-3. The nucleotide and deduced amino acid sequences of the ATB1 gene. Numbers on the right refer to the amino acid sequence. ATB1 is 1398bp long with ATG start codon and Poly(A) tail. It potentially encodes a 357 amino acid polypeptide. The amino acid sequences matching Dhugga's pea peptide fragments are underlined. The DNASIS program was used to generate this figure.

```
 GA CCC ACG CGT CCG CTC AAA TCT CCA CAA TCT TTT CTC TCC AAA TTC TCT TCT CTC TGA
ATC AAT CCA ATC ATG GTT GAG CCG GCG AAC ACC GTT GGA ATT CCG GTG AAC CAC ATC CCA
                M   V   E   P   A   N   T   V   G   I   P   V   N   H   I   P      16

TTG TTG AAG GAT GAG CTC GAT ATC GTG ATC CCC ACG ATC CGT AAC CTC GAT TTC CTG GAG
L   L   K   D   E   L   D   I   V   I   P   T   I   R   N   L   D   F   L   E      36

ATG TGG AGG CCT TTT TTT CAG CCT TAC CAT CTG ATT ATC GTC CAA GAT GGA GAT CCA TCG
M   W   R   P   F   F   Q   P   Y   H   L   I   I   V   Q   D   G   D   P   S      56

AAG ACC ATT GCT GTC CCT GAA GGG TTC GAT TAC GAA CTC TAC AAC AGG AAC GAC ATC AAC
K   T   I   A   V   P   E   G   F   D   Y   E   L   Y   N   R   N   D   I   N      76

CGT ATC CTT GGT CCT AAA GCT TCC TGC ATT TCC TTC AAG GAC TCT GCT TGT CGT TGC TTC
R   I   L   G   P   K   A   S   C   I   S   F   K   D   S   A   C   R   C   F      96

GGC TAC ATG GTC TCC AAG AAG AAG TAC ATC TTC ACT ATT GAT GAC GAT TGC TTC GTT GCT
G   Y   M   V   S   K   K   K   Y   I   F   T   I   D   D   D   C   F   V   A     116

AAG GAT CCA TCT GGA AAA GCT GTG AAC GCT CTT GAG CAA CAC ATC AAG AAC CTT CTC TGC
K   D   P   S   G   K   A   V   N   A   L   E   Q   H   I   K   N   L   L   C     136

CCA TCA ACT CCA TTT TTC TTC AAC ACC TTG TAC GAC CCA TAC CGT GAA GGT GCT GAC TTC
P   S   T   P   F   F   F   N   T   L   Y   D   P   Y   R   E   G   A   D   F     156

GTC CGT GGA TAC CCT TTC AGT CTC CGT GAG GGT GTT TCC ACC GCT GTT TCC CAC GGT CTG
V   R   G   Y   P   F   S   L   R   E   G   V   S   T   A   V   S   H   G   L     176

TGG CTC AAC ATC CCT GAT TAC GAT GCC CCT ACC CAA CTT GTG AAG CCT AAG GAA AGG AAC
W   L   N   I   P   D   Y   D   A   P   T   Q   L   V   K   P   K   E   R   N     196

ACA AGG TAT GTG GAT GCT GTC ATG ACC ATC CCA AAG GGG ACT CTT TTC CCT ATG TGT GGT
T   R   Y   V   D   A   V   M   T   I   P   K   G   T   L   F   P   M   C   G     216

ATG AAC TTG GCC TAT GAC CGT GAG CTC ATT GGT CCG GCT ATG TAC TTT GGT CTC ATG GGT
M   N   L   A   Y   D   R   E   L   I   G   P   A   M   Y   F   G   L   M   G     236

GAT GGT CAG CCT ATT GGT CGC TAC GAC GAT ATG TGG GCT GGA TGG TGT ATC AAG GTG ATC
D   G   Q   P   I   G   R   Y   D   D   M   W   A   G   W   C   I   K   V   I     256

TGT GAC CAT TTG GGA TTG GGA GTG AAG ACA GGT TTG CCC TAC ATT TAC CAC AGC AAA GCC
C   D   H   L   G   L   G   V   K   T   G   L   P   Y   I   Y   H   S   K   A     276

AGC AAC CCG TTT GTG AAC TTG AAG AAG GAG TAC AAG GGA ATC TTC TGG CAG GAG GAT ATC
S   N   P   F   V   N   L   K   K   E   Y   K   G   I   F   W   Q   E   D   I     296

ATT CCT TTC TTC CAG AGC CCA AAG CTC ACG AAA GAA GCT GTG ACA GTT CAA CAA TGC TAC
I   P   F   F   Q   S   P   K   L   T   K   E   A   V   T   V   Q   Q   C   Y     316

ATG GAG CTG TCC AAG TTG GTG AAG GAG AAG CTA AGC CCC ATT GAT CCT TAC TTT GAC AAG
M   E   L   S   K   L   V   K   E   K   L   S   P   I   D   P   Y   F   D   K     336

CTT GCA GAT GCT ATG GTC ACT TGG ATT GAA GCT TGG GAT GAG CTT AAC CCA CCC ACT AAA
L   A   D   A   M   V   T   W   I   E   A   W   D   E   L   N   P   P   T   K     356

GCT TGA GCA GCA AAA AAA CCA CCA CCG CAG TTT TGG TTA TTA GCT CAA CAT ATC ATC TAT
A   *                                                                             358
CTT CTC CGT TTT GTT TTT GTT TTG TCT TTT CTC AAA TTT TCC GGC GAT TCT TCA GTT CGT
TTT TAA GTT GCC GGA GTT TAT TTA AAT AGT GAA TGG TGT AGT TCT ATT TAT GAC CGA GAC
AAT CTC TCT TTT GAG GTT TTT CGT TTT TGT TTC AAT AAT AAA AAA TCA TAT CCC TTA AAA
AAA AAA AAA AAA AAG GGC
```

Figure 3-3. Nucleotide and Predicted Amino Acid Sequences of ATB1

Legend to Figure 3-4. Aligment of ATA1 nucleotide sequence with that of ATB1. The ATA1 and ATB1 sequences are highly similar, are about 85% identical. The 3' non-coding region of ATA1 and ATB1 gene show little similarity. The two sequences are highly similar within the coding region. ATG start codons and TGA stop codons are underlined. The GCG GAP program was used to generate this figure.

```
ATA1     18 AACCATGGTTGAGCCGGCGAATACTGTTGGTCTTCCGGTGAACCCGACTC 67
ATB1     68 AATCATGGTTGAGCCGGCGAACACCGTTGGAATTCCGGTGAACCACATCC 117

ATA1     68 CGTTGCTGAAAGATGAGCTCGATATCGTGATTCCGACTATCAGAAACCTC 117
ATB1    118 CATTGTTGAAGGATGAGCTCGATATCGTGATCCCCACGATCCGTAACCTC 167

ATA1    118 GATTTCCTCGAGATGTGGAGGCCTTTTCTTCAGCCTTACCATCTGATCAT 167
ATB1    168 GATTTCCTGGAGATGTGGAGGCCTTTTTTTCAGCCTTACCATCTGATTAT 217

ATA1    168 CGTCCAGGACGGAGATCCATCGAAGAAGATCCATGTCCCTGAAGGTTACG 217
ATB1    218 CGTCCAAGATGGAGATCCATCGAAGACCATTGCTGTCCCTGAAGGGTTCG 267

ATA1    218 ACTACGAGCTCTACAACAGGAACGACATTAACCGAATCCTCGGACCTAAG 267
ATB1    268 ATTACGAACTCTACAACAGGAACGACATCAACCGTATCCTTGGTCCTAAA 317

ATA1    268 GCTTCTTGTATCTCGTTTAAGGATTCTGCTTGTCGATGCTTTGGGTACAT 317
ATB1    318 GCTTCCTGCATTTCCTTCAAGGACTCTGCTTGTCGTTGCTTCGGCTACAT 367

ATA1    318 GGTGTCTAAGAAGAAGTATATCTTCACCATTGATGACGATTGCTTCGTTG 367
ATB1    368 GGTCTCCAAGAAGAAGTACATCTTCACTATTGATGACGATTGCTTCGTTG 417

ATA1    368 CCAAGGATCCATCAGGCAAAGCAGTGAACGCTCTTGAGCAACACATCAAG 417
ATB1    418 CTAAGGATCCATCTGGAAAAGCTGTGAACGCTCTTGAGCAACACATCAAG 467

ATA1    418 AACCTTCTCTGCCCATCGTCTCCCTTTTTCTTCAACACCTTGTATGATCC 467
ATB1    468 AACCTTCTCTGCCCATCAACTCCATTTTTCTTCAACACCTTGTACGACCC 517

ATA1    468 TTACCGTGAAGGTGCTGATTTCGTCCGTGGATACCCTTTCAGTCTCCGTG 517
ATB1    518 ATACCGTGAAGGTGCTGACTTCGTCCGTGGATACCCTTTCAGTCTCCGTG 567

ATA1    518 AAGGTGTTTCCACTGCTGTTTCCCATGGTCTTTGGCTCAACATCCCTGAC 567
ATB1    568 AGGGTGTTTCCACCGCTGTTTCCCACGGTCTGTGGCTCAACATCCCTGAT 617

ATA1    568 TACGATGCCCCGACCCAACTCGTGAAGCCTAAGGAGAGGAACACCAGGTA 617
ATB1    618 TACGATGCCCCTACCCAACTTGTGAAGCCTAAGGAAAGGAACACAAGGTA 667

ATA1    618 TGTGGATGCTGTCATGACCAACCCAAAGGGAACACTTTTCCCAATGTGTG 667
ATB1    668 TGTGGATGCTGTCATGACCATCCCAAAGGGGACTCTTTTCCCTATGTGTG 717

ATA1    668 GTATGAACTTGGCTTTTGACCGTGATTTGATTGGCCCGGCTATGTACTTT 717
ATB1    718 GTATGAACTTGGCCTATGACCGTGAGCTCATTGGTCCGGCTATGTACTTT 767

ATA1    718 GTTCTCATGGGTGATGGTCAGCCTATTGGTCGTTACGACGATATGTGGGC 767
ATB1    768 GGTCTCATGGGTGATGGTCAGCCTATTGGTCGCTACGACGATATGTGGGC 817

ATA1    768 TGGTTGGTGCATCAAGGTGATCTGTGACCACTTGAGCTTGGGAGTGAAGA 817
ATB1    818 TGGATGGTGTATCAAGGTGATCTGTGACCATTTGGGATTGGGAGTGAAGA 867
```

```
ATA1     818 CCGGTTTACCGTATATCTACCACAGCAAAGCGAGCAACCCTTTTGTTAAC 867
             | ||||| || || || ||||||||||||| |||||||| ||||| |||
ATB1     868 CAGGTTTGCCCTACATTTACCACAGCAAAGCCAGCAACCCGTTTGTGAAC 917

ATA1     868 CTGAAGAAGGAATACAAGGGAATCTTCTGGCAGGAGGAGATCATTCCGTT 917
             |||||||||| |||||||||||||||||||||||| |||||||| ||
ATB1     918 TTGAAGAAGGAGTACAAGGGAATCTTCTGGCAGGAGGATATCATTCCTTT 967

ATA1     918 CTTCCAGAACGCAAAGCTATCGAAAGAAGCAGTAACTGTTCAGCAATGCT 967
             |||||||| | ||||||| |||||||||| || || ||||| |||||||
ATB1     968 CTTCCAGAGCCCAAAGCTCACGAAAGAAGCTGTGACAGTTCAACAATGCT 1017

ATA1     968 ACATTGAGCTCTCAAAGATGGTCAAGGAGAAGTTGAGCTCCTTAGACCCG 1017
             |||| ||||| || ||| |||| ||||||||| | ||| || | || ||
ATB1    1018 ACATGGAGCTGTCCAAGTTGGTGAAGGAGAAGCTAAGCCCCATTGATCCT 1067

ATA1    1018 TACTTTGACAAGCTTGCAGATGCCATGGTTACATGGATTGAAGCTTGGGA 1067
             |||||||||||||||||||||| ||||| || ||||||||||||||||
ATB1    1068 TACTTTGACAAGCTTGCAGATGCTATGGTCACTTGGATTGAAGCTTGGGA 1117

ATA1    1068 TGAGCTTAACCCACCAGCAGCCAGTGGCAAAAGCTTGAGAGCAGTATGAG 1117
             |||||||||||||||           | ||||||          ||||
ATB1    1118 TGAGCTTAACCCACC..........CACTAAAGCT..........TGAG 1146

ATA1    1118 CCAAAAAGAAAAAGCCACCAAAGTTTTGGTTATTTTTAGCTCAAATTATC 1167
             |  :||| |||   |||||  |||||||||||   |||||||||  ||||
ATB1    1147 CAGCAAAAAAACCACCACCGCAGTTTTGGTTA...TTAGCTCAACATATC 1193

ATA1    1168 GT.TACTTTTAAATTTCTGATTTTACGAACCTTTCTTGCTTTTTTTACAC 1216
              | ||  ||     || ||:||||
ATB1    1194 ATCTATCTTCTCCGTTTTGTTTTT..... 1217

ATA1    1217 ATTTGAGTAGTTTTCATCATCAGTACTTTCTC..ATTGTCCGGTTATGGT 1264
                     |||||        ||  ||||||  ||| |||||     |
ATB1    1218 .........GTTTT........GTCTTTTCTCAAATTTTCCGG....CGA 1246

ATA1    1265 TTTTGCATTTGGTTTAAATATCACCGGTTTATTTATAAACAGTGGTGGAT 1314
             || | || || ||||  |  |  ||||  | | | |||| ||||   | |
ATB1    1247 TTCTTCAGTTCGTTTTTAAGTTGCCGGAGTTTATTTAAATAGTGAATGGT 1296

ATA1    1315 TAGTAGTACTATT.....................TTCTGAGTTTTTT..TC 1342
               ||||| |||||                     || |||| |||||  |
ATB1    1297 ..GTAGTTCTATTTATGACCGAGACAATCTCTCTTTTGAGGTTTTTCGTT 1344

ATA1    1343 TTTGTTTCATTAATAAAAAGGCCTTTTCATAGGTGTTTGCAATTAAAAAA 1392
             ||||||||| |||||||||  | | | | |           |  ||||||
ATB1    1345 TTTGTTTCAATAATAAAAAATCATATCCCT..........TAAAAAAAAA 1384

ATA1    1393 AAAAAAAAAGGGC 1405
             |||||||||||||
ATB1    1385 AAAAAAAAAGGGC 1397
```

Figure 3-4. Aligment of ATA1 Nucleotide Sequence with That of ATB1

Legend to Figure 3-5. Aligment of the deduced ATA1 amino acid sequence with that of ATB1. The ATA1 encodes a 364 amino acid long protein and ATB1 encodes a 357 amino acid long protein. Their sequences similarity is 93%. The GCG GAP program was used to generate this figure.

```
ATA1    MVEPANTVGLPVNPTPLLKDELDIVIPTIRNLDFLEMWRPFLQPYHLIIV 50
        |||||||||:|||..|||||||||||||||||||||||:||||||||
ATB1    MVEPANTVGIPVNHIPLLKDELDIVIPTIRNLDFLEMWRPFFQPYHLIIV

ATA1    QDGDPSKKIHVPEGYDYELYNRNDINRILGPKASCISFKDSACRCFGYMV 100
        |||||||.| ||||:||||||||||||||||||||||||||||||||||
ATB1    QDGDPSKTIAVPEGFDYELYNRNDINRILGPKASCISFKDSACRCFGYMV

ATA1    SKKKYIFTIDDDCFVAKDPSGKAVNALEQHIKNLLCPSSPFFFNTLYDPY 150
        |||||||||||||||||||||||||||||||||||.||||||||||||
ATB1    SKKKYIFTIDDDCFVAKDPSGKAVNALEQHIKNLLCPSTPFFFNTLYDPY

ATA1    REGADFVRGYPFSLREGVSTAVSHGLWLNIPDYDAPTQLVKPKERNTRYV 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
ATB1    REGADFVRGYPFSLREGVSTAVSHGLWLNIPDYDAPTQLVKPKERNTRYV

ATA1    DAVMTNPKGTLFPMCGMNLAFDRDLIGPAMYFVLMGDGQPIGRYDDMWAG 250
        ||||| |||||||||||||:||:||||||||.||||||||||||||||||
ATB1    DAVMTIPKGTLFPMCGMNLAYDRELIGPAMYFGLMGDGQPIGRYDDMWAG

ATA1    WCIKVICDHLSLGVKTGLPYIYHSKASNPFVNLKKEYKGIFWQEEIIPFF 300
        ||||||||||:||||||||||||||||||||||||||||||:||||||
ATB1    WCIKVICDHLGLGVKTGLPYIYHSKASNPFVNLKKEYKGIFWQEDIIPFF

ATA1    QNAKLSKEAVTVQQCYIELSKMVKEKLSSLDPYFDKLADAMVTWIEAWDE 350
        |.:||.||||||||||:||||:||||||.:|||||||||||||||||||||
ATB1    QSPKLTKEAVTVQQCYMELSKLVKEKLSPIDPYFDKLADAMVTWIEAWDE

ATA1    LNPPAASGKSLRAV*
        ||||.      :|
ATB1    LNPPT......KA*
```

Figure 3-5. Aligment of the Deduced ATA1 Amino Acid Sequence with That of ATB1

with Dhugga's 40 kD pea protein, it suggested that ATA1 and ATB1 may encode a protein similar to this 40 kD protein.

The hydropathy analysis (Figure 3-6) of ATA1 and ATB1 in Figure 3-6 indicated that both ATA1 and ATB1 protein are relatively hydrophilic with no obvious membrane-spanning domains.

Highly conserved in different species

The nucleotide sequences and the deduced amino acid sequences of ATA1 and pea clones (Dhugga personal communication) were aligned in Figure 3-7 and Figure 3-8. Comparision at the nucleotide level, the ATA1 and pea sequences are very similar with, about 71% identical. The amino acid sequence identity is 84%.

The complete nucleotide sequence of ATA1 was used to search dBEST. Fifteen Arabidopsis clones and seven rice clones were found highly similar to ATA1 as of June 15,1995. Some of the Arabidopsis clones(ATA1, ATA2, ATB1, ATB2, ATB3) and four rice clones (OS1, OS2, OS3, OS4) were used to do computer alignment analysis.

The nucleotide sequences alignment (Figure 3-9) shows that within the Arabidopsis group the identity is 83%, and they have 75% identity with the

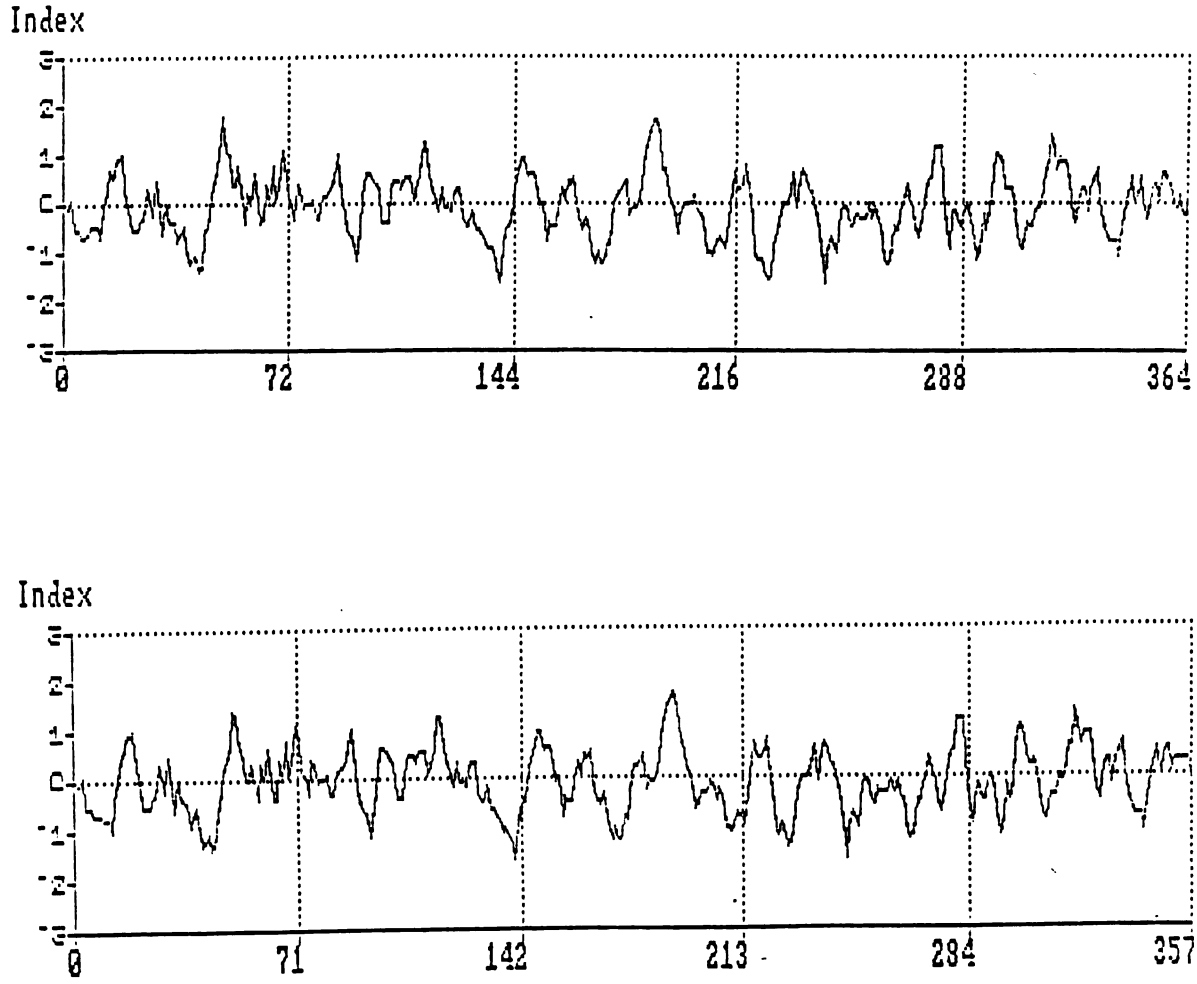
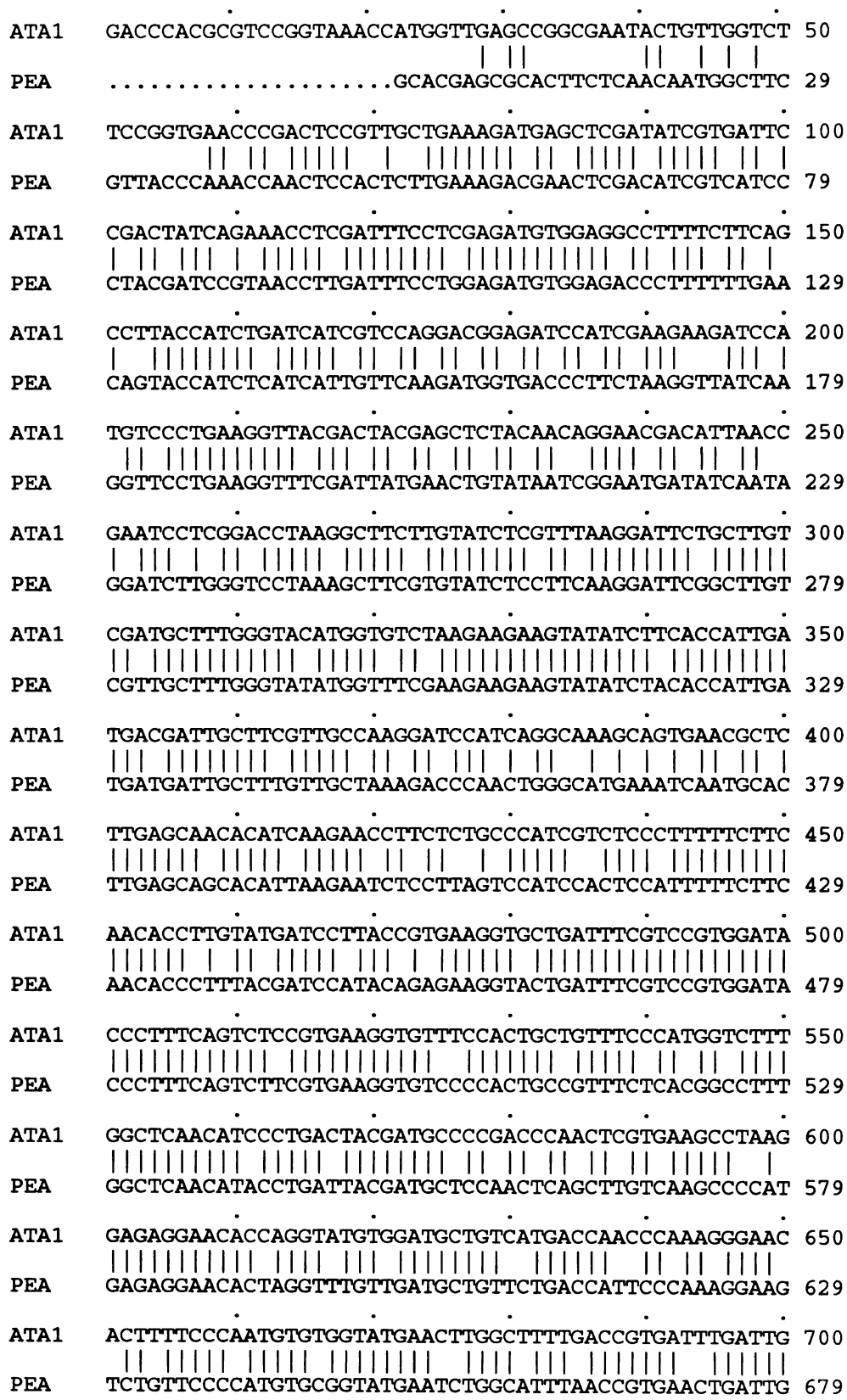


Figure 3-6. Hydropathy plots of the deduced amino acid sequences of ATA1 and ATB1. Both ATA1 and ATB1 are relatively hydrophilic. PROSIS program was used to generate this figure.

Legend to Figure 3-7. The nucleotide sequence of ATA1 with that of pea. Comparision at the nucleotide level, the ATA1 and pea clone sequences are very similar, about 71% identical. The pea cDNA clone sequence is from personal communication with Dhugga. The GCG GAP program was used to generate this figure.


ATA1 GACCCACGCGTCCGGTAAACCATGGTTGAGCCGGCGAATACTGTTGGTCT 50
PEA .....GCACGAGCGCACTTCTCAACAATGGCTTC 29
ATA1 TCCGGTGAACCCGACTCCGTTGCTGAAAGATGAGCTCGATATCGTGATTC 100
PEA GTTACCCAAACCAACTCCACTCTTGAAAGACGAACTCGACATCGTCATCC 79
ATA1 CGACTATCAGAAACCTCGATTTCCTCGAGATGTGGAGGCCTTTTCTTCAG 150
PEA CTACGATCCGTAACCTTGATTTCCTGGAGATGTGGAGACCCTTTTTTGAA 129
ATA1 CCTTACCATCTGATCATCGTCCAGGACGGAGATCCATCGAAGAAGATCCA 200
PEA CAGTACCATCTCATCATTGTTCAAGATGGTGACCCTTCTAAGGTTATCAA 179
ATA1 TGTCCCTGAAGGTTACGACTACGAGCTCTACAACAGGAACGACATTAACC 250
PEA GGTTCCTGAAGGTTTCGATTATGAACTGTATAATCGGAATGATATCAATA 229
ATA1 GAATCCTCGGACCTAAGGCTTCTTGTATCTCGTTTAAGGATTCTGCTTGT 300
PEA GGATCTTGGGTCCTAAAGCTTCGTGTATCTCCTTCAAGGATTCGGCTTGT 279
ATA1 CGATGCTTTGGGTACATGGTGTCTAAGAAGAAGTATATCTTCACCATTGA 350
PEA CGTTGCTTTGGGTATATGGTTTCGAAGAAGAAGTATATCTACACCATTGA 329
ATA1 TGACGATTGCTTCGTTGCCAAGGATCCATCAGGCAAAGCAGTGAACGCTC 400
PEA TGATGATTGCTTTGTTGCTAAAGACCCAACTGGGCATGAAATCAATGCAC 379
ATA1 TTGAGCAACACATCAAGAACCTTCTCTGCCCATCGTCTCCCTTTTTCTTC 450
PEA TTGAGCAGCACATTAAGAATCTCCTTAGTCCATCCACTCCATTTTTCTTC 429
ATA1 AACACCTTGTATGATCCTTACCGTGAAGGTGCTGATTTCGTCCGTGGATA 500
PEA AACACCCTTTACGATCCATACAGAGAAGGTACTGATTTCGTCCGTGGATA 479
ATA1 CCCTTTCAGTCTCCGTGAAGGTGTTTCCACTGCTGTTTCCCATGGTCTTT 550
PEA CCCTTTCAGTCTTCGTGAAGGTGTCCCCACTGCCGTTTCTCACGGCCTTT 529
ATA1 GGCTCAACATCCCTGACTACGATGCCCCGACCCAACTCGTGAAGCCTAAG 600
PEA GGCTCAACATACCTGATTACGATGCTCCAACTCAGCTTGTCAAGCCCCAT 579
ATA1 GAGAGGAACACCAGGTATGTGGATGCTGTCATGACCAACCCAAAGGGAAC 650
PEA GAGAGGAACACTAGGTTTGTTGATGCTGTTCTGACCATTCCCAAAGGAAG 629
ATA1 ACTTTTCCCAATGTGTGGTATGAACTTGGCTTTTGACCGTGATTTGATTG 700
PEA TCTGTTCCCCATGTGCGGTATGAATCTGGCATTTAACCGTGAACTGATTG 679

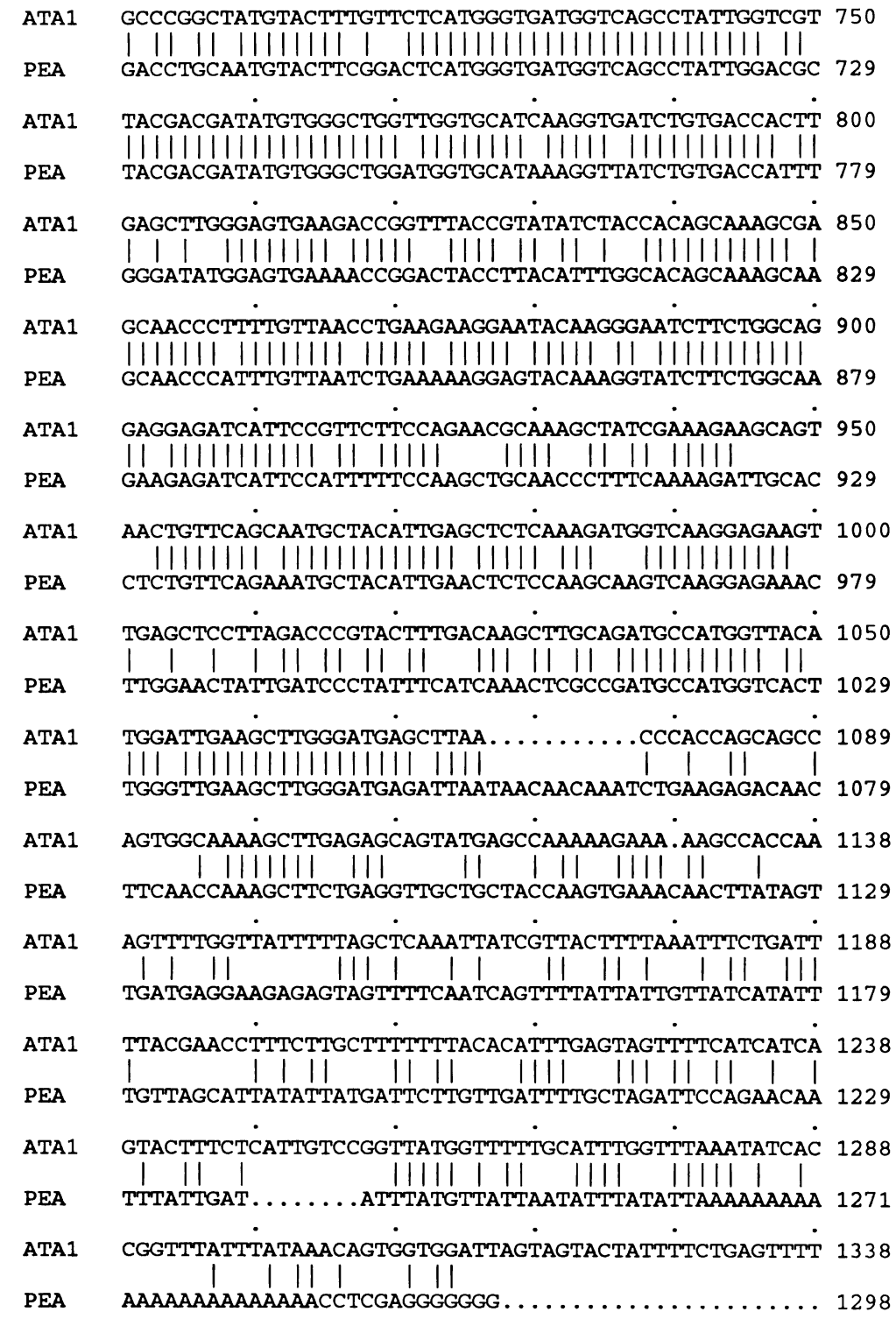


Figure 3-7 The Nucleotide Sequence of ATA1 with that of Pea

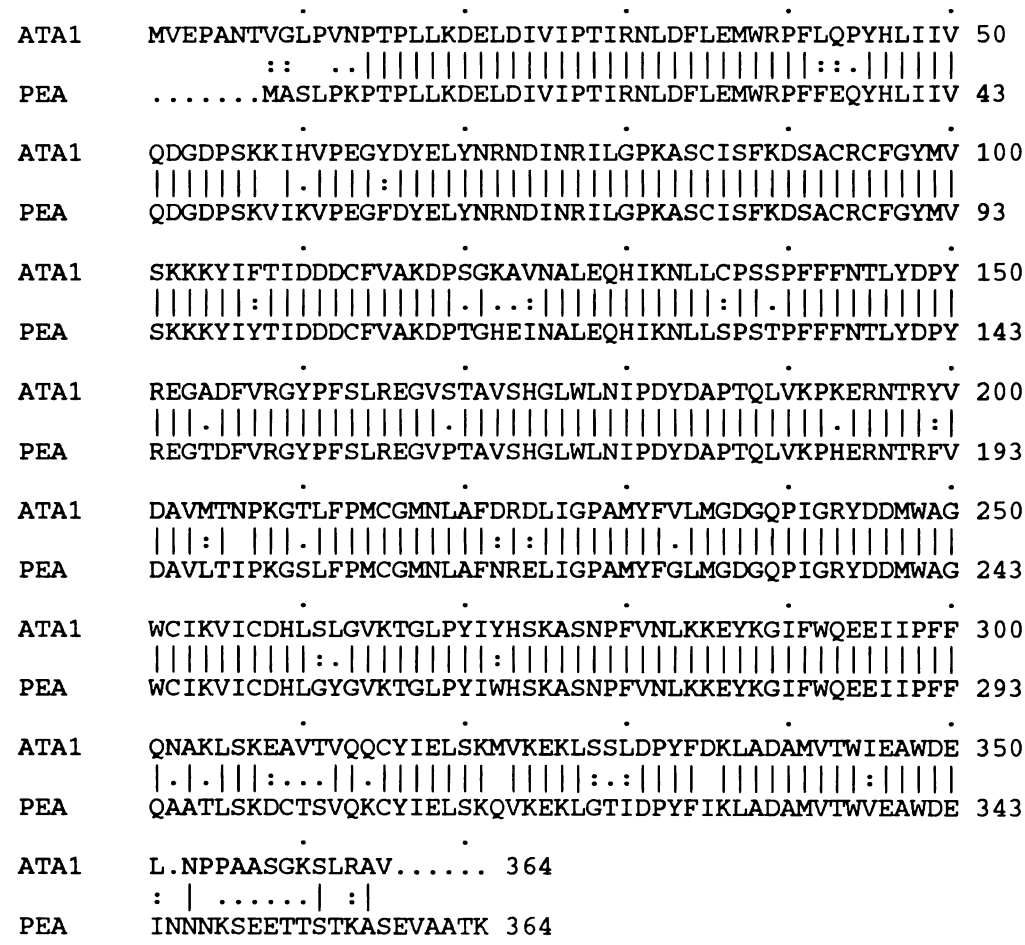


Figure 3-8 Aligment of ATA1 Deduced Amino Acid Sequence with that of Pea. Both ATA1 and pea amino acid sequences are 364 a.a. long. The identity is 84%. The GCG GAP program was used to generate this figure.

Legend to Figure 3-9. Aligment of Arabidopsis nucleotide sequences with that of rice group. The OS stands for rice group. The complete nucleotide sequence of ATA1 was used to search dBEST. Fifteen Arabidopsis clones and seven rice clones were found highly similar to ATA1 as of June 15, 1995. Both the Arabidopsis and rice clone showed here are longer than ATA1 at the 5′ end.. The identity within the Arabidopsis group is 83%, and they have 75% identity with the rice group. The GCG PILEUP program was used to generate this figure.

```
                                                                    50
OS1        GCGAGATCCA gGGAGaGgGA GAgGGAGAgA TCATGGcgGg GaCGGtGAcG
OS2        GCGAGATCCA gGGAGaGgGA GAgGGAGAgA TCATGGcgGg GaCGGtGAcG
OS3        GCGAGATCCA gGGAGaGgGA GAgGGAGAgA TCATGGcgGg GaCGGtGAcG
OS4        GCGAGATCCA gGGAGaGgGA GAgGGAGAgA TCATGGcgGg GaCGGtGAcG
ATB2       cCaAatTNTN ttctNtNtGA atcaatccAA TCATGGTTGA GCCGGCGAAc
ATB1       cCaAatTCTC ttctCtCtGA atcaatccAA TCATGGTTGA GCCGGCGAAc
ATB3       cCaAatTCTC ttctCtCtGA atcaatccAA TCATGGTTGA GCCGGCGAAc
ATA2       .....cgANC tcagCtCtct cAtttcGAAA cCATGGTTGA GCCGGCGAAt
ATA1       .......... .GacCcAcGc GtccGgtAAA cCATGGTTGA GCCGGCGAAt
Consensus  GCGAGATCCA -GGA--G-GA GA-GGAGAAA TCATGGTTGA GCCGGCGAAG

                                                                   100
OS1        gtGccgtTGG cgtCGGTGcc gtCGACGCCG cTGCTcAAGG AcGAGCTgGA
OS2        gtGccgtNGG cgtCGGTGcc gtCGACGCCG cTGCTcAAGG AcGAGCTgGA
OS3        gtGccgtCGG cgtCGGTGcc gtCGACGCCG cTGCTcAAGG AcGAGCTgGA
OS4        gtGccgtTGG cgtCGGTGcc gtCGACGCCG cTGCTcAAGG AcGAGCTgGA
ATB2       ACcGTTGGaa TTCCGGTGAA CCacAtcCNa TTGtTGAAGG ATGAGCTCGA
ATB1       ACcGTTGGaa TTCCGGTGAA CCacAtcCCa TTGtTGAAGG ATGAGCTCGA
ATB3       ACcGTTGGaa TTCCGGTGAA CCacAtcCCa TTGtTGAAGG ATGAGCTCGA
ATA2       ACtGTTGGtc TTCCGaNGAA CCCGACtCCG TTGCTGAAaG ATGAGCTCGA
ATA1       ACtGTTGGtc TTCCGGTGAA CCCGACtCCG TTGCTGAAaG ATGAGCTCGA
Consensus  ACGGTTGGGG TTCCGGTGAA CCCGACGCCG TTGCTGAAGG ATGAGCTCGA

                                                                   150
OS1        cATCGTGATC CCGACGATCC GCAACCTgGA cTTCCTGGAG ATGTGGcGGC
OS2        cATCGTGATC CCGACGATCC GCAACCTgGA cTTCCTGGAG ATGTGGcGGC
OS3        cATCGTGATC CCGACGATCC GCAACCTgGA cTTCCTGGAG ATGTGGcGGC
OS4        cATCGTGATC CCGACGATCC GCAACCTgGA cTTCCTGGAG ATGTGGcGGC
ATB2       TATCGTNATC CCNACGATCC GtAACCTCGA TTTCNTGGAG ATGTGGAGGC
ATB1       TATCGTGATC CCCACGATCC GtAACCTCGA TTTCCTGGAG ATGTGGAGGC
ATB3       TATCGTGATC CCCACGATCC GtAACCTCGA TTTCCTGGNG ATGTGGAGGN
ATA2       TATCNaGATt CCGACtATCa GaAACCTCGA TTTCCTcGAG ATGTGGAGGC
ATA1       TATCGTGATt CCGACtATCa GaAACCTCGA TTTCCTcGAG ATGTGGAGGC
Consensus  TATCGTGATC CCGACGATCC GCAACCTCGA TTTCCTGGAG ATGTGGAGGC

                                                                   200
OS1        CcTTCTTcCA GCCcTACCAc CTCATCATCG TgCAGGACGG CGA.CCCGaC
OS2        CcTTCTTcCA GCCcTACCAc CTCATCATCG TgCAGGACGG CGA.CCCGaC
OS3        CcTTCTTcCA GCCcTACCAc CTCATCATCG TgCAGGACGG CGA.CCCGaC
OS4        CcTTCTTcCA GCCcTACCAc CTCATCATCG TgCAGGACGG CGA.CCCGaC
ATB2       CTTTNTTTCA GCCTTACCAT CTNATtATNG TCCAaGAtGG AGA.TCCATC
ATB1       CTTTTTTTCA GCCTTACCAT CTGATtATCG TCCAaGAtGG AGA.TCCATC
ATB3       CTTTTTTTtA GCCTTACCAT CTGATtATNG gCCAaGAtNG NGNtNCNNTC
ATA2       CTTTTcTTCA GCCTTACCAT CTGATCATCG TCCAGGACGG AGA.TCCATC
ATA1       CTTTTcTTCA GCCTTACCAT CTGATCATCG TCCAGGACGG AGA.TCCATC
Consensus  CTTT-TTTCA GCCTTACCAT CT-ATCATCG TCCAGGACGG -GA--CC-TC

                                                                   250
OS1        cAAGACCATC CGcGTCCC G AgGGCTTCGA CTACGAGCTC TACAACcGcA
OS2        cAAGACCATC CGcGTCCCCG AgGGCTTCGA CTACGAGCTC TAC.......
OS3        cAAGACCATC CGcGTCCCCG AgGGCTTCGA CTACNAGCTC TACAACcGcA
OS4        cAAGACCATC CGcGTCCCCG AgGGCTTCGA CTACGAGCTC TACAACcGcA
ATB2       GAAGACCATt gcTGTCCCTG AAGGgTTCGA tTACGgaCTC TACAACAGG.
ATB1       GAAGACCATt gcTGTCCCTG AAGGgTTCGA tTACGAaCTC TACAACAGGA
ATB3       GgAGACCATt gcTNTNCCNG AAGGgggtcg gTtd...... ..........
ATA2       GAAGAagATC CaTGTCCCTG AAGGtTaCGA CTACGAGCTC TACAACAGGA
ATA1       GAAGAagATC CaTGTCCCTG AAGGtTaCGA CTACGAGCTC TACAACAGGA
Consensus  GAAGACCATC CGTGTCCC-G AAGGCTTCGA CTACGAGCTC TACAACAGGA
```

Figure 3-9. Aligment of Arabidopsis Nucleotide Sequences with That of Rice group

rice group. The deduced amino acid sequences (Figure 3-10) shows that it is 93% identity within the Arabidopsis group and they have 69% identity with the rice group. Comparing the variation at nucleotide sequence level and deduced amino acid sequence level, it can be found that there are a lot of codon degeneration variation which will not effect sequence. It may indicate that this gene has some change during evolution and use codon degeneration to keep its basic function which might be important to the higher plants.

Although the deduced amino acid sequences of the rice and Arabidopsis clones are very similar, no homologous protein can be found in the data base. Thus we conclude that these proteins are both highly conserved and abundant in plants, yet unique to plants, as similar sequences have not been observed in yeast, animals, etc.

ATA1 fusion protein was overexpressed

We sought to prepare chemical quantities of the 40 kD protein so that it could be used as antigen for production of specific antiserum. It was decided to attempt this via expression of a fusion protein in *E. coli.* For this purpose, the ATA1 gene was subcloned into pET22b(+) vector(Figure 3-11). The chimeric construct was transformed into *E.coli* strain BL21(DE3), and was IPTG induced and ovexpressed. The overexpressed protein, which was in "inclusion bodies", was isolated by centrifugation, washed with buffers solubilized in 6M guanidine HCl, and purifed by nickel affinity column. The protein has been injected into rabbit and immune serum will be available soon.

```
          1                                                  50
ATB3      ........AQ ISQSFLSKFS SL*INPIMVE PANTVGIPVN HIPLLKDELD
ATB2      ........AQ X*QSFLSKXX SX*INPIMVE PANTVGIPVN HIXLLKDELD
OS4       .....HLLLL LLFLRAREIQ GEGEGEIMAG TVTVPLASVP STPLLKDELD
OS1       ..HHHHLLLL LLFLRAREIQ GEGEGEIMAG TVTVPLASVP STPLLKDELD
OS2       .....HLLLL LLFLRAREIQ GEGEGEIMAG TVTVPXASVP STPLLKDELD
OS3       TPHHHHLLLL LLFLRAREIQ GEGEGEIMAG TVTVPSASVP STPLLKDELD
ATA2      .......... ........RX QLSHFETMVE PANTVGLPXN PTPLLKDELD
ATA1      .......... .......... .......MVE PANTVGLPVN PTPLLKDELD
ATB1      .......... .......... .......MVE PANTVGIPVN HIPLLKDELD

          51                                                 100
ATB3      IVIPTIRNLD FLXMWRXFF* PYHLIXGQDX XXXGDHCXXX RGSV......
ATB2      IVIPTIRNLD FXEMWRPXFQ PYHLIXVQD. ...GDPSKTI AVPEGFDYGL
OS4       IVIPTIRNLD FLEMWRPFFQ PYHLIIVQD. ...GDPTKTI RVPEGFDYEL
OS1       IVIPTIRNLD FLEMWRPFFQ PYHLIIVQD. ...GDPTKTI RVPEGFDYEL
OS2       IVIPTIRNLD FLEMWRPFFQ PYHLIIVQD. ...GDPTKTI RVPEGFDYEL
OS3       IVIPTIRNLD FLEMWRPFFQ PYHLIIVQD. ...GDPTKTI RVPEGFDYXL
ATA2      IXIPTIRNLD FLEMWRPFLQ PYHLIIVQD. ...GDPSKKI HVPEGYDYEL
ATA1      IVIPTIRNLD FLEMWRPFLQ PYHLIIVQD. ...GDPSKKI HVPEGYDYEL
ATB1      IVIPTIRNLD FLEMWRPFFQ PYHLIIVQD. ...GDPSKTI AVPEGFDYEL
```

Figure 3-10. Alignment of the Deduced Arabidopsis Amino Acid Sequences with That of Rice Group. The deduced amino acid sequences shows that it is 93% identity within the Arabidopsis group and they have 69% identity with the rice group. The GCG PILEUP program was used to generate this figure.

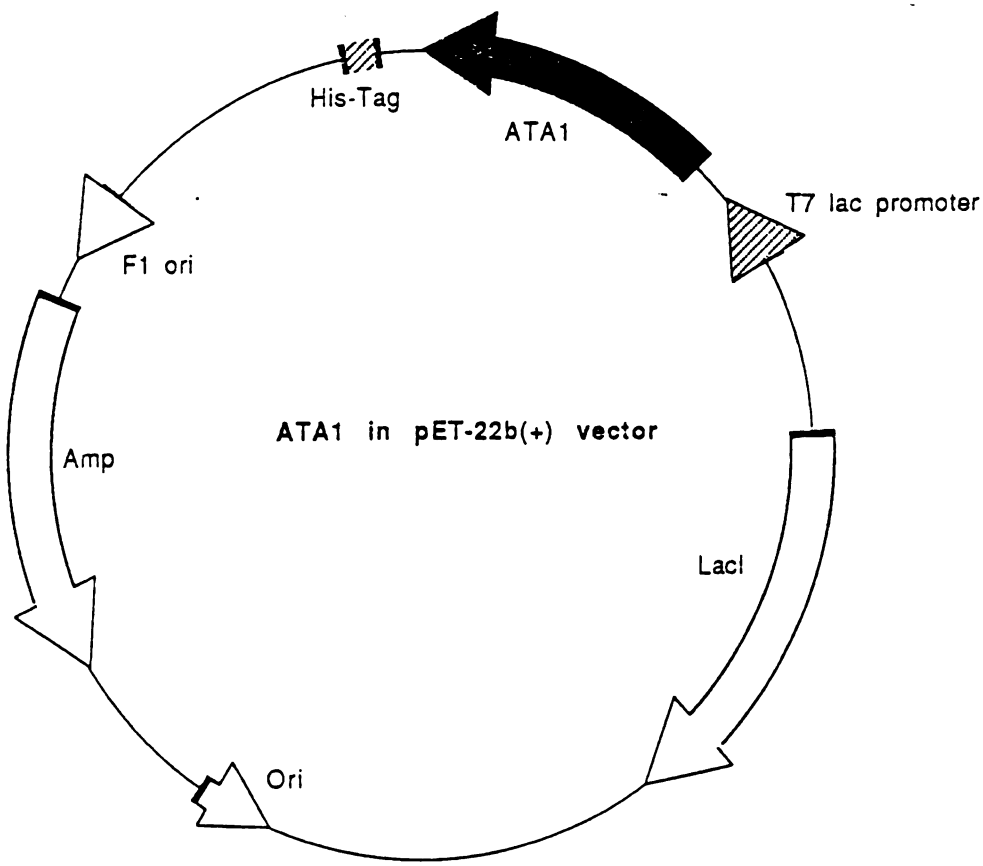


Figure 3-11. The construct of ATA1 in pET22b(+) vector.
The Bam HI/HindIII fragment of ATA1 was subcloned into corresponding sites in pET22b(+) expression vector.

CHAPTER 4

DISCUSSION AND CONCLUSION

Based on my initial work, several conclusions can be drawn. 1) From the restriction map, sequences and the aligment, there are at least two genes in Arabidopsis, ATA and ATB gene. 2) Similar genes that are highly conserved at amino acid sequence level are found in other plants. 3) They are moderately abundant in different species of plants, eg. Arabidopsis and rice, because fifteen Arabidopsis cDNA clones and seven rice clones have been found by using ATA1 to do dBEST search as of June 15, 1995. 4) No sequence similarity has been found in animals, yeast, etc. All these suggest that this protein probably plays an important role unique to plants. Based on Dhugga's work, one likely possibility is that it is involved in cell wall synthesis.

From Dhugga's observation, this 40 kD doublet protein can be reversibly labeled by UDP-[14C]Glc under the conditions of the GS-I assay (Dhugga et al., 1991). A possible function of GS-I is to form the (1,4)-β-glucan backbone of xyloglucan which is synthesized in the Golgi system. Both GS-I activity and labeling of the 40 kD protein are inhibited by UDP-xylose or UDP-Gal, but not by UDP-Man. Xylose and galactose, but not mannose, are components of xyloglucan. The above observation suggest that the 40 kD protein participates in GS-I activity.

Dhugga's experiment also shows that glucose was attached to the protein by a covalent bond (Dhugga et al., 1991). This evidence indicates that the 40 kD polypeptide is not a sugar nucleotide transmembrane translocase, instead it

could be a glycosyl transfer intermediate. It was observed that a saturating glycosylation of approximately one glucose/40 kD polypeptide. The 40 kD protein was found both associated with Golgi membranes and in the soluble fraction. The soluble form of the 40 kD protein appears to be auto-glycosylated. This is probably also true of the membrane bound form.

By comparing what is known regarding the synthesis of complex polysaccharides in other systems with the situation in plants, we may be able to gain new insight into the role of the 40kD protein in plants. The synthesis of bacteria lipopolysaccharide (O-Antigen) is one process that may serve as a model for investigating plant cell wall polysaccharide biosynthesis. Lipopolysaccharide is a combination of polysaccharide attached to a lipid core that forms a large part of the outer membrane of gram-negative enteric bacteria. During its synthesis, first the monosaccharides are transferred sequentially from their nucleotide carriers to the phosphomonoester of a specific lipid component of the membrane, known as antigen carrier lipid (ACL). (Figure 4-1, Robbins and Wright 1967), The trisaccharide repeat, eg. . mannosylrhamnosylgalactose in Fig 4-1, is formed, then polymerized, and finally it is transferred to lipopolysaccharide core acceptor. It is possible that the 40 kD protein is a substitute for ACL, playing the role during xyloglucan biosynthesis that ACL performs during lipopolysaccharide biosynthesis.

Another process that may provide insight into the possible role of PSI is glycogen biosynthesis. Glycogen is a storage form of sugar in animals. Its synthesis is primed via a protein called glycogenin. Glycogenin is the core protein of glycogen proteoglycan and is, at the same time, a self-glucosylating

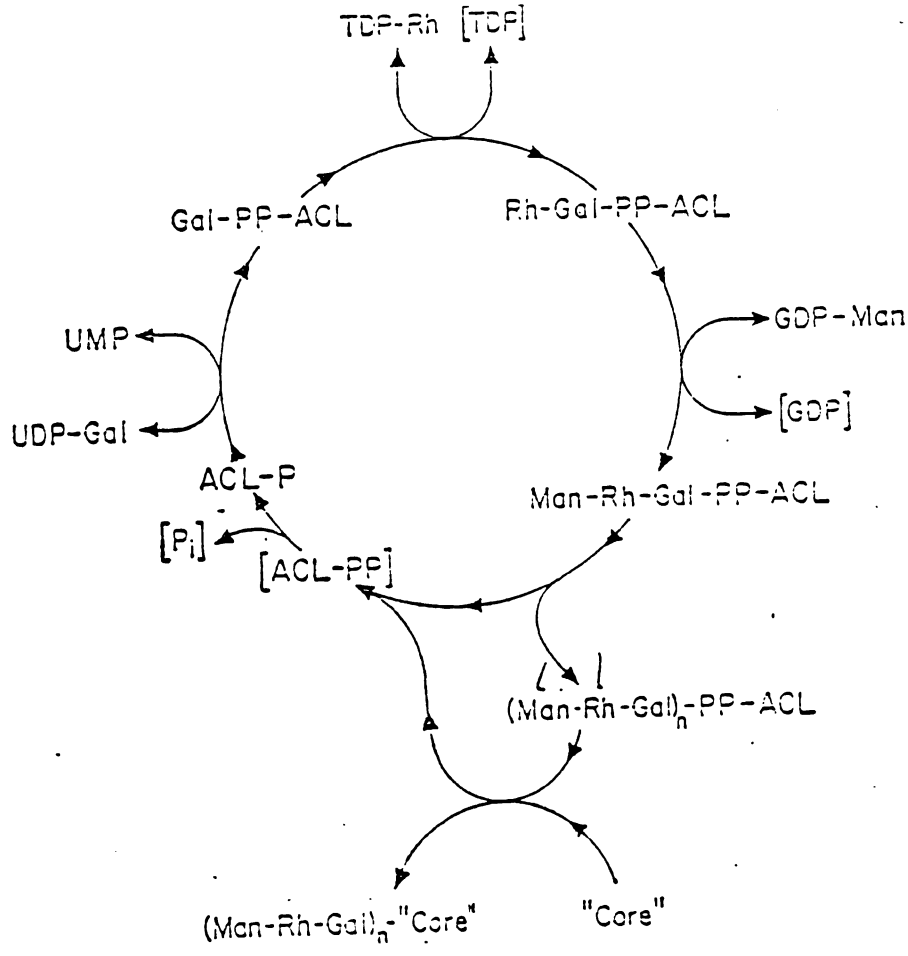


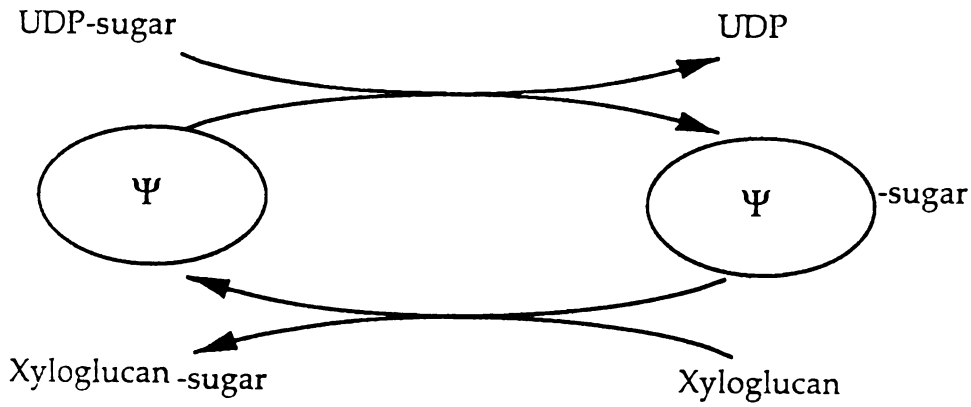
Figure 4-1. Scheme for the biosynthesis of the *Salmonella* O-antigen. (Robbins and Wright, 1967)

enzyme which catalyses early glucosyl transfer steps in the biosynthesis of glycogen (Manzella et al. 1995). It's not a single reaction, but with several consecutive glycosyl transfer steps, in which the product of one reaction becomes the acceptor and enzyme in the subsequent step. The end product of glycogenin action is an oligosaccharide composed of 8-11 glucose residues and the fully glucosylated form of glycogenin serves as the oligosaccharide primer for further chain elongation by proglycogen synthase and glycogen synthase (Pitcher et al. 1987; Lomako et al. 1988). Glycogenin can also be considered as a homologous protein of PSI ,though there are no obvious amino acid sequence similarity between glycogenin and PSI protein.

From Dhugga's work, the function of ACL in bacteria cell wall biosynthesis and the role of glycogenin in glycogen biosynthesis, we propose two possible functions of PSI protein. Model A) Monosaccharide intermediate. Monosaccharide are transferred from nucleotide sugar to PSI, then from PSI to the growing xyloglucan chain. Model B) Oligosaccharide Intermediate. Several sugar residues are transferred to PSI to form a oligosaccharide repeat unit, which is transferred to the growing xyloglucan chain (Figure 4-2). Where these reactions occurs within a cell is not clear at present. Dhugga claims that all the 40 kD protein is associated with Golgi (Dhugga et al., 1991;: Dhugga personal communication). However it remains to be determined on which side of the Golgi membrane this protein is located.

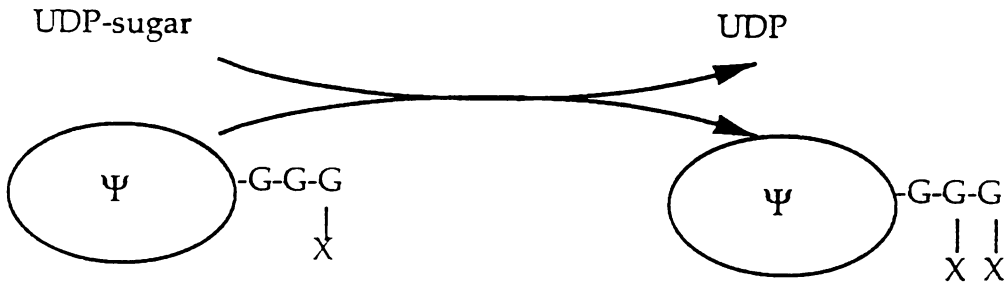
Legend to Figure 4-2. The proposed functions of polysaccharide synthesis intermediate (PSI) protein. Model A) Monosaccharides intermediate. Monosaccharide are transferred from nucleotide sugar to PSI, then from PSI to the growing xyloglucan chain. B) Oligosaccharide intermediate. Several sugar residues are transferred to PSI to form a oligosaccharide repeat unit, which is transferred to the growing xyloglucan chain.

A) Monosaccharide Intermediate



B) Oligosaccharide Intermediate

a)



b)

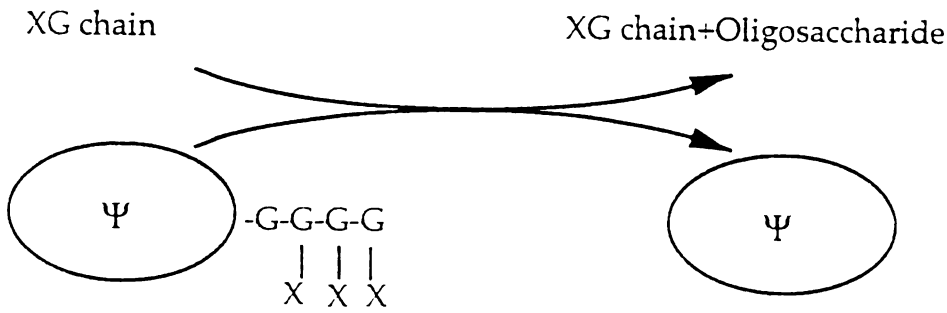


Figure 4-2 . Proposed functions of PSI protein

As noted above, futher work can be done to investigate the role of the PSI protein in plants. Considering its function, this protein should be Golgi-localized. Antibody against ATA1 can be used to do cellular immunolocalization to confirm this hypothesis. The antibody can also be used to against the crude protein extracts of Arabidopsis and other plants to detect the protein expression level in different species.

This cDNA clone can be constructed into an expression vector to over express this protein in E. coli. The overexpressed protein can be UDP[14C]Glc labelled to detect its function in an in vitro xyloglucan synthesis system.

The Agrobacterium transformation can be used to overexpress this protein in vivo in Arabidopsis to see how the transgenic plant will function, or to create a mutation that eliminate the function of this gene. The antisense technology or gene-disruption techniques can be used to creat a mutant plant to block the expression of this gene to see how this will affect the function of the plant.

LITERATURE CITED

Carpita NC and Delmer D (1981) Concentration and metabolic turnover of UDP-glucose in developing cotton fiber. J. Biol. Chem. 256: 308-315

Carpita NC and Gibeaut DM (1993) Structure 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

Cosgrove DJ (1993) How do plant cell walls extend. Plant Physiol. 102: 1-6

Delmer DP, Cooper G, Alexander D, Cooper J, Hayashi T, Nitsche C, Thelen M (1985) New approaches to the study of cellulose biosynthesis. J. Cell Sci. Suppl. 2: 33-50

Delmer DP and Stone BA (1988) Biosynthesis of plant cell wall. The biochemistry of plants. Vol. 14 pp. 373-416. Academic Press, Inc.

Devlin RM and Withman FH (1983) Plant cells: structure and function. Plant physiology. Fourth Edition. pp. 1-25. PWS Publishers.

Dhugga KS, Ray PM (1994) Purification of the reversibly glycosylated polypeptides from pea: purified polypeptides exhibit the same properties as the Golgi-bound form. Plant Physiol. Suppl. 105: 126

Dhugga KS, Ulvskov P, Gallagher SR, Ray PM (1991) Plant polypeptides reversibly glycosylated by UDP-glucose. Possible components of Golgi β-glucan synthase in pea cells. J. B. Chem. 266: 21977-21984

Driouich A, Faye L, Staehelin A (1993) The plant Golgi apparatus: a factory for complex polysaccharides and glycoproteins. TIBS 18: 210-214

Elbein AD (1979) The role of lipid-linked saccharides in the biosynthesis of complex carbohydrates. Annu. Rev. Plant Physiol. 30: 239-272

Fincher GB and Stone BA (1981) Encycl. Plant Physiol. New Ser. 13B: 68-132

Hayashi T (1989) Xyloglucans in the primary cell wall. Annu. Rev. Plant Physiol. Plant Mol. Biol. 40: 139-68

Hayashi T, Maclachlan G (1984) Biosynthesis of pentosyl lipids by pea membrane. Biochem. J. 217: 791-803

Hayashi T, Matsuda K (1981) Biosynthesis of xyloglucan in suspension-cultured soybean cells. Evidence that the enzyme system of xyloglucan synthesis does not contain β-1,4-glucan 4-β-D-glucosyltransferase activity. Plant Cell Physiol. 22: 1571-1584

Hayashi T, Nakajima T, Matsuda K (1984) Biosynthesis of xyloglucan in suspension-cultured soybean cells. Processing of the oligosaccharide building blocks. Agric. Biol. Chem. 48: 1023-1027

Hirschberg CB (1987)Topography of glycosylation in the rough endoplasmic reticulum and Golgi apparatus. Annu. Rev. Biochem. 56: 63-87

Keegstra K, Talmadge K, Bauer WD, Albersheim P (1973) The structure of plant cell walls. III. A model of the walls of suspension-culture sycamore cells based on the interconnection of the macromolecular components. Plant Physilo 51: 188-196

Kornfeld R and Kornfeld S (1985) Assembly of asparagine-linked oligosaccharide. Annu. Rev. Biochem. 54: 631-664

Lomaco J, Lamaco WM, Whelan WJ (1988) A self-glucosylating protein is the primer for rabbit muscle glycogen biosynthesis. FASEB J. 2: 3097-3103

Maclachlan G (1985) Are lipid-linked glycosides required for plant polysaccharide biosynthesis? In Biochemistry of plant cell wall. pp. 199-220. Cambridge Univ. Press. Cambridge

Manzella SM, Roden L, Meezan E (1995) Dodecyl-β-D-maltoside as a substrate for glucosyl and xylosyl transfer by glycogenin. Glycobiology 5: 263-271

McNeil M, Darvill AG, Fry SC, Albersheim P (1984) Structure and function of the primary cell walls of plants. Ann. Rev. Biochem. 53: 625-63

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L,Ohlrogge John, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clone. Plant Physiol. 106: 1241-1255

Northcote DH (1985) Cell organelles and their function in biosynthesis of cell wall components: control of cell wall assembly during differentiation. Biosynthesis and biodegradation of wood components. pp. 87-108. Academic press. Orlando, Florida

Pitcher J, Symthe C, Campbell DG, Cohen P (1987) Identification of the 38-kD subunit of rabbit skeletal muscle glycogen synthase as glycogenin. Eur. J. Biochem. 169: 497-502

Ray PM (1980) Cooperative action of beta-glucan synthetase and UDP-xylose xylosyl transferase of Golgi membranes in the synthesis of xyloglucan-like polysaccharide. Biochim. Biophys. Acta 629: 431-444

Robbins PW, Bray D, Dankert M, Wright A (1967) Direction of the chain growth in polysaccharide synthesis. Work on a bacterial polysaccharide suggests that elongation can occur at the "reducing" end of the growthing chain. Science 158: 1536-1542

Robbins PW and Wright A (1971) Biosynthesis of O-antigens. Microbial Toxins. pp. 351-367 Academic Press, Inc., New York and London

Talbott LD, Ray PM (1992) Molecular size and separability features of pea cell wall polysaccharides. Implication for models of primary wall structure. Plant Physiol. (1992) 98: 357-368

Talmadge K, Keegstra K, Bauer WD, Albersheim P (1973) The structure of plant cell walls. I. The macromolecular components of the walls of suspension-cultured sycamore cells with a detailed analysis of the pectic polysaccharides. Plant Physiol 51: 158-173

Stoddart RW (1984) Glycosylation in bacteria. The biosynthesis of polysaccharides. pp 58-87 Macmillan, New York

Varner JE, Lin LS (1989) Plant cell wall architecture. Cell 56: 231-239