Isolation and characterization of the ATP-binding cassette (ABC) transporter system genes from loofah witches’ broom phytoplasma

CHUN-LIN HUANG† & KUO-CHIEH HO

Department of Life Science, Institute of Plant Biology, National Taiwan University, Taipei, Taiwan

(Received 26 July 2006)

Abstract
A clone containing a 3903 bp EcoRI-restriction fragment was obtained from a λZAP genomic library of loofah witches’ broom (LfWB) phytoplasma by plaque hybridization using a PCR fragment as a probe. Sequence analysis revealed that this fragment contained three open reading frames (ORFs). The deduced amino acid sequences of ORF 1 and ORF 2 showed a high homology with the ATP-binding proteins of the ABC transporter system genes of prokaryotes and eukaryotes, and encoded proteins with a molecular mass of 36 and 30 kDa, respectively. Based on amino acid sequence similarity, secondary structure, hydrophilicity and a signal peptide sequence at the N-terminus, we predicted that ORF 3 might encode a specific solute-binding prolipoprotein of the ABC transporter system with a molecular mass of 62 kDa. The cleavage site of this prolipoprotein signal peptide was similar to those of gram-positive bacteria. In addition to nutrient uptake, ABC transporter systems of bacteria also play a role in signal transduction, drug-resistance and perhaps virulence. The possible implications of the system to the survival and the pathogenesis of phytoplasma were discussed.

Keywords: Loofah witches’ broom (LfWB), phytoplasma, ATP-binding cassette (ABC) transporter system, open reading frames

Database accession number: AF086618 (P36), AF086619 (P30), AF086620 (P62)

Introduction
Phytoplasmas, a member of mollicutes, are the causal agents of more than 200 diseases of higher plants (Kirkpatrick 1989; McCoy et al. 1989). They have remained uncultured in vitro and their inability to grow in culture medium has severely hindered their studies (Lee and Davis 1986). As a result, phytoplasmas are the most poorly characterized groups of plant pathogens. Although phytoplasmas’ host ranges (plant and insect) and symptomatology (phenotype differences) are similar, their genotype and genome sizes (600–1200 kDa) are diverse (Razin 1992).

Phytoplasmas were known previously as mycoplasmalike organisms (MLOs). However, phylogenetic relationships based on the sequences of the full-length 16S rRNA, ribosomal protein genes and the 16S/23S spacer regions revealed that phytoplasmas form a large discrete monophyletic clade, paraphyletic to the Acholeplasma species, within the Anaeroplasma clade rather than Mycoplasma (Tully 1993; Gundersen et al. 1994; Sear and Kirpatrick 1994; Ho et al. 2001). Recently, phytoplasma can be detected, identified and classified clearly by molecular methods. While these phylogenetic advances continue, research has focus on the pathogenicity genes. These result in the novel approaches to achieve effective control of phytoplasmal diseases (Kirkpatrick and Smart 1995; Lee et al. 2000; Wagner et al. 2001) and the completion of two phytoplasma genome sequences (Oshima et al. 2004; Bai et al. 2006).

The ATP-binding cassette (ABC) superfamily is one of the largest protein families found in living systems, with several hundred different genes identified to date in organisms ranging from bacteria to man. Typically, ABC transporters utilize the energy of ATP to pump substrates across the membrane against
a concentration gradient. The common protein components of the ABC-type uptake systems include one or two transmembrane proteins which usually span the membrane five to six times each, one or two peripheral membrane ATP-binding proteins in the cytoplasmic side and a high-affinity extra-membrane solute-binding protein (Higgins 1992; Locher 2004; Khwaja et al. 2005). In gram-negative bacteria, the ligand-specific binding protein is soluble and periplasmic. In gram-positive bacteria, the soluble-binding protein is also extracellular and anchors to the membrane via an N-terminal hydrophobic lipid extension. The transmembrane protein components are believed to form solute-specific channels; the peripheral membrane ATP-binding proteins energize the systems and sometimes serve regulatory roles, and the ligand-binding proteins confer specificity and high affinity for the substrates to the transporter systems (Ames 1986; Shuman 1987; Tam and Saier 1993; Monnet 2003). In eukaryotic cells, ABC transporters have been predominantly found in the plasma membrane where they catalyze the efflux of various compounds out of the cell. Other ABC proteins are located in intracellular organelles such as peroxisomes, mitochondria, the endoplasmic reticulum and vacuoles; some of these proteins mediate the compartmentation of compounds into the organelles (Davies and Coleman 2000; Young et al. 2001).

The ABC transporters can transport a remarkable variety of substrates, including ions, carbohydrates, lipids, antibiotics, anti-cancer drugs, pigment molecules and even large peptides. They participate in many biological functions, e.g. phosphate regulation in *Escherichia coli* (Saier 1993), chemical signaling between *Agrobacterium tumefaciens* and its plant host (Ankenbauer and Nester 1990; Cangelosi et al. 1990), control of *Bacillus subtilis* sporulation (Perego et al. 1991; Koide and Hoch 1994), transport of sex pheromones in yeasts (McGrath and Varshavsky 1989), and chloride via the cystic fibrosis transport regulator protein in mammals (Riordan et al. 1989; Tata et al. 1991). The vast majority of studies on ABC transporters have been driven by their diverse importance.

In this communication, we reported the cloning and characterization of a DNA fragment containing the ABC transporter system genes from loofah witches' broom (LfWB) phytoplasma. The possible implications of the system to the survival and the pathogenesis of phytoplasma were also discussed.

**Materials and methods**

**Bacterium and plant**

The LfWB phytoplasmas were maintained by graft inoculation in periwinkles (Chen and Ho 1997). The original diseased plant was provided by Dr H.-J. Su, Professor of Department of Plant Pathology and Microbiology, National Taiwan University, Taiwan.

**Phytoplasma genomic library construction and screening**

The following methods, including healthy or diseased plant DNA extraction, phytoplasma DNA purification and phytoplasma genomic library construction were done as previously described (Ho et al. 2001).

The library was screened by plaque hybridization using a $^{32}$P-labelled PCR fragment amplified on the LfWB phytoplasma DNA with primers RN1 (5'-CGCTTAGAGTCTAGTGTA-3') and RN2 (5'-CCAACCAAGCTTC-3'), synthesized according to the DNA sequence of ribonuclease III of *Mycoplasma genitalium* (Fraser et al. 1995). The positive plaque areas were selected and rescreened until a single, isolated plaque could be picked up.

**Southern blot analysis**

Phytoplasma DNA (3 μg) or plant DNA (10 μg) were digested with different restriction enzymes and subjected to Southern blot analysis performed at 42°C in the presence of 50% formamide and 0.1% SDS using the $^{32}$P-labelled DNA probe. The filter was washed with 0.1 x SSC (1 x SSC: 150mM NaCl, 15 mM sodium citrate) containing 0.1% SDS at 50°C. The hybridized bands were detected by exposing the filter to a PhosphoImager screen (PhosphoImager 425; Molecular Dynamics).

**Sequence determination and analysis**

For DNA sequencing, the recombinant phage was converted into a phagemid by *in vivo* excision according to manufacturer's instructions (Stratagene). The sequence of DNA insert in the recombinant phagemid was determined by a DNA automated sequencer (ABI Prism Model 377, v.30; Applied Biosystems) using a step-by-step procedure in which synthetic primers for forward sequencing were designed from previously released sequences. DNA sequence analysis was performed using DNAStar software (DNASTAR). The deduced amino acid sequences were analyzed through the BLAST of National Center for Biotechnology Information sever (http://www.ncbi.nlm.nih.gov/BLAST), and submitted to the 3D-PSSM fold recognition server (http://www.sbg.bio.ic.ac.uk/~3dpssm/) for prediction of secondary structure if it is necessary (Kelley et al. 2000). The alignment of multiple sequences was performed by the program Clustalx ver 1.8 (Thompson et al. 1997).
Results and discussion

Nucleotide sequence analysis of the cloned DNA fragment

A 170-bp PCR product, named RN-P was amplified on the phytoplasma DNA. Three clones were obtained using RN-P as a probe to screen an EcoRI-genomic library of LfWB phytoplasma constructed on λZAP vector. The DNA sequences indicated that these clones contained the same DNA insert of 3903 bp.

The nucleotide sequence of the DNA insert was composed of 76.53% A + T which corresponded with the low G + C content of phytoplasma genome. No strong homology was found by comparing the nucleotide sequence with those deposited in GenBank. However, the deduced amino acid sequence revealed that this 3903-bp DNA insert contained three open reading frames (ORFs) (Figure 1). The ORF 1, from nucleotide 290 to 1237, encoded a protein of 36 kDa (designed as protein P36). The ORF 2, from nucleotide 1237 to 2031, had one base overlapping with ORF 1 and encoded a protein of 30 kDa (designed as protein P30). The ORF 3, from nucleotide 2090 to 3688 and separated by 60 nucleotides from ORF 2, encoded a protein of 62 kDa (designed as protein P62). All three ORFs had an ATG as a start codon, which was preceded by a ribosome binding site. There was a putative -10 sequence and a putative -35 sequence upstream from the ORF 1, however, no obvious sequence for a rho-dependent or -independent termination site was found downstream from the ORF 3 (Figure 2). Transcription termination at indiscrete sites has been reported for several protein-encoding genes and ribosomal genes of archaeobacteria, where transcription stops within or near a T-rich sequence (Ho et al. 2001 and references therein). It was possible that the transcription of this ABC transporter operon terminated in one of several pyrimidine-rich regions downstream from the ORF 3, as found in ribosomal gene of the same organism (Ho et al. 2001).

The primers RN1 and RN2 were originally designed from mycoplasma sequence to screen the library for the gene encoding ribonuclease III of LfWB phytoplasma. Instead, the DNA insert containing ABC transporter genes was obtained due to the existence of partial sequence homologies.

P36 and P30 have homology with ATP-binding proteins

The deduced amino acid sequences were used to search for protein homologies by BLAST from protein sequence database. Proteins P36 and P30 displayed a significant homology with several ATP-binding proteins. The features of ATP-binding proteins including (1) Walker A motif, a glycine-rich loop involved in ATP binding; (2) the ABC signature motif or C motif; (3) the Walker B motif that is associated with many nucleotide-binding proteins; and (4) a switch region characterized by an invariant histidine residue or H motif, which was suggested to be involved in ATP hydrolysis (Higgins 1992; Zhao et al. 2004) were found in P36 and P30. P30 has shorter sequence length truncating at C-terminus, as it’s homolog, DppD of onion yellow (OY) and aster yellows witches’-broom (AYWB) phytoplasmas (Figure 3). Many truncated proteins have been found in the OY and AYWB phytoplasma genomes (Bai et al. 2006). This might be the unique feature of phytoplasma associated with their small genome size.

Hydropathic analysis by the method of Kyte and Doolittle (1982) revealed that P36 and P30 were fairly hydrophilic peripheral membrane proteins as the ATP binding proteins of complex bacterial transporter system (Figure 4A, B). P36 and P30 could form a common engine, which binds and hydrolyzes ATP, and energizes unidirectional substrate transport as the common architecture of ABC transporters. This engine is attached to a specialized translocon composed of two transmembrane proteins.

P62 is a hypothetical extracellular solute-binding lipoprotein

The BLAST database search result showed that the P62 was most similar to the solute-binding protein DppA of the OY and AYWB phytoplasma (Bai et al. 2006), but with only 54% similarity and 48% similarity, respectively. The DppA of ABC dipeptide transporter system belongs to the extracellular solute-binding lipoprotein cluster 5. The proteins of cluster 5
-35
-10
50
P36
60
D0
80
P80
100
120
140
160
180
Figure 2. Nucleotide sequence of the 3903 bp cloned-DNA fragment and deduced amino acid sequences of P36, P30 and P62. The putative -10 and -35 regions of the promoter, and the pyrimidine-rich sequences where the transcription possibly terminates are boldfaced. The Shine-Dalgarno (SD) sequences for each gene and the possible lipoprotein signal peptidase cleavage site are underlined. The vertical arrow indicates the processing site of the P62 precursor protein. The inverted sequences of the 3'-untranslated region are numbered and indicated by arrows.
**Figure 3.** Comparison of P36, P30 and ATP-binding proteins of other microorganisms’ ABC transporter systems. The sequences of Walker A, C motif, Walker B and H motif are underlined and indicated.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Organism</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>DppD_Tpe</td>
<td><em>Thermotoga petrophila</em> (ZP_01652599)</td>
<td></td>
</tr>
<tr>
<td>DppD_Tte</td>
<td><em>Thermoanaerobacter tengcongensis</em> (NP_624050)</td>
<td></td>
</tr>
<tr>
<td>OppD_Bcl</td>
<td><em>Bacillus clausii</em> (YP_177152)</td>
<td></td>
</tr>
<tr>
<td>OppF_Bau</td>
<td><em>Bacillus licheniformis</em> (YP_078436)</td>
<td></td>
</tr>
<tr>
<td>OppD_Gka</td>
<td><em>Geobacillus kaustophilus</em> (YP_146668)</td>
<td></td>
</tr>
<tr>
<td>OppF_AYWB527</td>
<td>Aster yellows witches'-broom phytoplasma (YP_456723)</td>
<td></td>
</tr>
<tr>
<td>DppD_ONY192</td>
<td>Onion yellows phytoplasma (NP_950444)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C motif</th>
<th>Walker B</th>
</tr>
</thead>
<tbody>
<tr>
<td>DppD_Tpe</td>
<td>EARKMAIDLL RKGKPEEPK RVDYEHQPSL GMMGRQAMIA MALSCRPSLL IADEPTTALD VTIQAQ1IL EMLKQKEYGM</td>
</tr>
<tr>
<td>DppD_Tte</td>
<td>EAAMKIAEML KQVIPRAEQ IMSYTHPELS GMMQRKMIA MAISCFQRLD IADEPTTALD VTIQAQ1IL LRLKEEKKM</td>
</tr>
<tr>
<td>OppD_Bcl</td>
<td>EAAMKIAEML KQVIPRAEQ IMSYTHPELS GMMQRKMIA MAISCFQRLD IADEPTTALD VTIQAQ1IL LRLKEEKKM</td>
</tr>
<tr>
<td>OppF_Bau</td>
<td>ERWQHRVEHL ETGVLM--KE HRYHDPEH QGQRQUIGIA RALVDVEPEF IADEPTIALD VSIQIQVNL MELKQKEYGM</td>
</tr>
<tr>
<td>OppD_Gka</td>
<td>ERWQHRVEHL ETGVLM--KE HRYHDPEH QGQRQUIGIA RALVDVEPEF IADEPTIALD VSIQIQVNL MELKQKEYGM</td>
</tr>
<tr>
<td>DppD_AWYB527</td>
<td>DKPKYKMLD IMKKGIDMY LCDYEHQPSL QGQRQUIGIA RALIKPKFY VCDEIVSALD VSIQIQVNL MLKDQKDY31</td>
</tr>
<tr>
<td>P30</td>
<td>DPPKLKPQIE LVNGLQSL NLXGQRQIA NLXLQKFLP IEKQMLNLK LQEELKDY31</td>
</tr>
<tr>
<td>DppD_AWYB528</td>
<td>DPPKLKPQIE LVNGLQSL NLXGQRQIA NLXLQKFLP IEKQMLNLK LQEELKDY31</td>
</tr>
<tr>
<td>OppF_AWYB528</td>
<td>DPPKLKPQIE LVNGLQSL NLXGQRQIA NLXLQKFLP IEKQMLNLK LQEELKDY31</td>
</tr>
<tr>
<td>P30</td>
<td>DPPKLKPQIE LVNGLQSL NLXGQRQIA NLXLQKFLP IEKQMLNLK LQEELKDY31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H motif</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DppD_Tpe</td>
<td>KHPFRCPEFE KGKQDVTPPEE LEDLDGVRK MCQPRWQLDE MRHAK3-234</td>
</tr>
<tr>
<td>DppD_Tte</td>
<td>YPFRPCEVYD-VCKRMMP LVAQEMKVG AKNLKG--227</td>
</tr>
<tr>
<td>OppD_Bcl</td>
<td>YPFRPCEVYD-VCKRMMP LVAQEMKVG AKNLKG--227</td>
</tr>
<tr>
<td>OppF_Bau</td>
<td>YPFRPCEVYD-VCKRMMP LVAQEMKVG AKNLKG--227</td>
</tr>
<tr>
<td>OppD_Gka</td>
<td>YPFRPCEVYD-VCKRMMP LVAQEMKVG AKNLKG--227</td>
</tr>
<tr>
<td>DppF_AWYB527</td>
<td>YPFRPCEVYD-VCKRMMP LVAQEMKVG AKNLKG--227</td>
</tr>
<tr>
<td>P30</td>
<td>YPFRPCEVYD-VCKRMMP LVAQEMKVG AKNLKG--227</td>
</tr>
</tbody>
</table>

For personal use only.
Figure 4. Hydropathy plots of the polypeptides P36 (A), P30 (B) and P62 (C). The hydropathy profiles are obtained according to the method of Kyte and Doolittle (1982), using a window length of 19. Hydrophilic regions are above the line and hydrophobic regions are below.
are peptide- and nickel-binding proteins that the average sequence similarity score is 0.52 and the identity is low (Tam and Saier 1993). The lipoprotein is the key to understanding the function of the ABC transporter system. However lipoproteins seldom show homology between each other. It is hard to determine function from comparison of the protein sequences (Tam and Saier 1993). The extracellular solute-binding receptors of bacteria were grouped into eight clusters based on their sequence similarity, and these groupings were generally found to correlate with the molecular sizes and solute-binding specificities of the proteins. The proteins of cluster 5 are the largest of the solute-binding proteins with a size range of 493–543 residues. They are more than 100 residues larger than any of the other binding proteins (Tam and Saier 1993). The molecular size of P62 was in the range of the cluster 5 binding proteins.

Since moderate amino acid sequence similarity to ABC-type dipeptide transporter protein, P62 was further submitted to the 3D-PSSM fold recognition server for prediction of secondary structure and then aligned to secondary structure elements of proteins with a solved crystal structure. The secondary structure of P62 showed the most similarity to the ABC oligopeptide-binding protein AppA of B. subtilis (Levdikov et al. 2005) with an E-value of 1.4e-41 and 100% estimated precision. AppA is a lipid-modified, membrane-anchored extracellular binding-protein that serves as the receptor for the transport system. It plays important roles in the signal pathway leading to the development of competence and sporulation (Levdikov et al. 2005).

The hydropathy analysis revealed that P62 was a hydrophilic protein with an N-terminal hydrophobic signal peptide and had no obvious membrane spanning domains (Figure 4C). Because the phytoplasma has no periplasmic space, there should be some mechanism to anchor hydrophilic solute-binding proteins to the membrane. There are three types of export sequence of membrane proteins: (1) non-cleavable, membrane-spanning sequence; (2) signal peptides with signal peptidase (SPase) I-like cleavage sites; and (3) signal peptides with SPase II-like lipoprotein-cleavage sites. There is no strict consensus sequence in the mycoplasma lipoprotein cleavage site besides a Cys in the +1 position. Leu is common at the −3 position in other eubacteria, but not in mycoplasmas (Cleavinger et al. 1995; Sutcliffe and Russell 1995). The region of the putative cleavage site for signal peptide (Figure 2) in P62 had a sequence L-D-F-C-S-N which was consistent with the consensus sequence (L-Y–Z-cleavage site-C-y–z) of the lipoprotein precursors (Gilson et al. 1988; Sutcliffe and Russell 1995). In these proteins the N-terminal cysteine is modified into a lipo amino acid that is thought to anchor them to the membrane

![Figure 5. Southern blot analysis of phytoplasma and healthy periwinkle DNA. Lanes 1–4 are the phytoplasma DNA (3 µg/lane) digested with BamHI, EcoRI, PstI and SalI, respectively. Lane 5 is the healthy periwinkle DNA (10 µg) digested with EcoRI. The probe used was a 32P-labeled RN-P fragment.](Image 492x15 to 602x29)

(Nielsen and Lampen 1982). These analyses suggested that P62 was a solute-binding lipoprotein specific for peptides and attached to the membrane via the modified N-terminal as in gram-positive bacteria.

**Southern blot analysis**

In an attempt to find out the copy number of the ABC transporter system genes, phytoplasma genomic DNA digested with each of the following restriction enzyme: BamHI, EcoRI, PstI and SalI, and analyzed with a Southern blot probed by a 32P-labeled RN-P fragment. There was single band in the individual lanes containing phytoplasma DNA but not in the lane of the healthy periwinkle DNA digested with EcoRI, suggesting that there was only a single copy of the genes in the phytoplasma genome and the genes were LfWB phytoplasma-, but not host- specific (Figure 5).

Phytoplasmas are obligatory parasites in the phloem of host plants, and have a small genome. The complete genome of Candidatus phytoplasma encodes even fewer metabolic functions than that of mycoplasma genomes. The loss of some biosynthetic pathways during the course of evolution results in their inability to be cultured in vitro. They must take up their basic nutrients from the environment through transporter systems to survive. The genomes of OY and AYWB phytoplasmas, and Spiroplasma kunkelii have many genes encoding transporter systems (Oshima et al. 2004; Zhao et al. 2004; Bai et al. 2006). The ABC transport systems of some bacteria are connected with virulence and pathogenesis in the host (Dudler et al. 2004).
In conclusion, we report in this communication an operon isolated from \( L_{fwb} \) phytoplasma. The operon contains three structural genes encoding three proteins constructing an ABC transporter which shares a common ABC structure and might have the same mechanism of translocation of specific substrates. Southern blot analysis indicated that this ABC transporter system was \( L_{fwb} \) phytoplasma-, but not host-specific. It could be a target for designing therapeutic agents against the pathogen.

Acknowledgements

We thank Dr Te-Chang Lee, Researcher of the Institute of Biomedical Sciences, Academia Sinica, Taiwan for providing aid in DNA autosequencing. Thanks also go to Dr Michael Conrad of University of North Carolina at Chapel Hill for his critical reading and suggestions. This work was financially supported by the National Science Council, Taiwan to K.-C. Ho (Grant No. NSC93-2621-B-002-011).

References


