ELSEVIER

Biochimica et Biophysica Acta 1369 (1998) 7–13

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Short sequence-paper

Cloning of *AtMRP1*, an *Arabidopsis thaliana* cDNA encoding a homologue of the mammalian multidrug resistance-associated protein [1]

Elena Marin *, Nathalie Leonhardt, Alain Vavasseur, Cyrille Forestier

CEA / Cadarache, DSV, DEVM, Laboratoire de Bioenergétique Cellulaire F-13108, St. Paul-lez-Durance Cedex, France

Received 4 August 1997; accepted 1 September 1997

Abstract

A cDNA encoding a putative ATP-binding cassette (ABC) transporter from *Arabidopsis* was cloned and sequenced based on an EST clone homologous to ABC sequences in other species. The cDNA is 5.5 kb long and contains an ORF encoding a 1623 amino acids protein. This sequence is the first MRP-like protein found in plants. © 1998 Elsevier Science B.V.

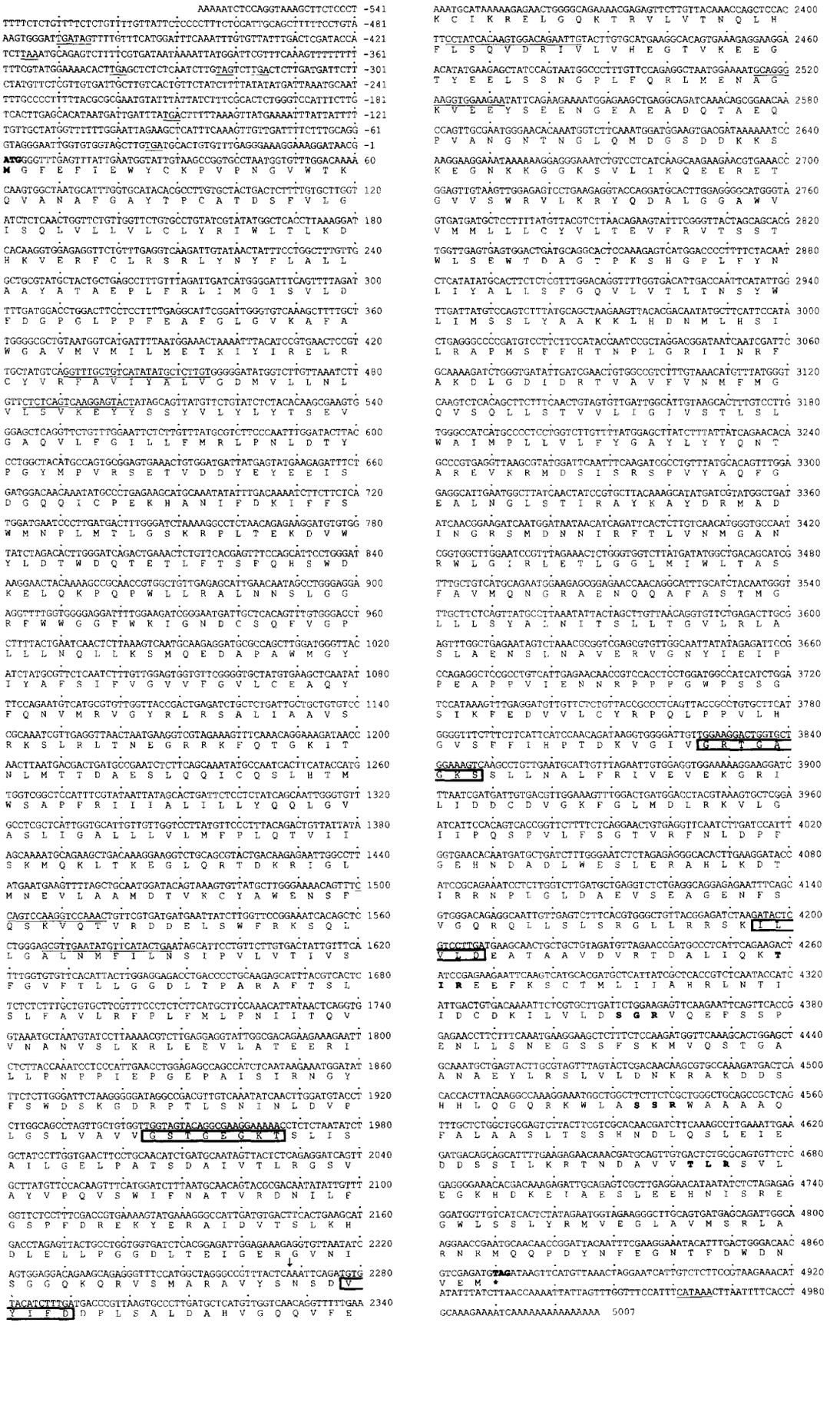
Keywords: ABC transporter; ATP-binding cassette; Mutidrug resistance-associated protein; cDNA sequence; (*Arabidopsis thaliana*)

The ATP-binding cassette (ABC) superfamily is probably one of the largest and most diverse family of proteins that mediates the selective ATP-dependent movement of solutes across biological membranes [1]. Each protein is specific for a substrate or a group of related substrates that can range from inorganic ions to sugars, amino acids, complex polysaccharides and proteins. Besides possessing their own intrinsic transporter/channel activities, ABC proteins are also implicated in the regulation of heterologous ion channels. Many of the mammalian ABC proteins identified so far are of considerable clinical importance. The human P-glycoprotein (P-gp) confers resistance to chemotherapeutic drugs on tumor cells, a phenomena called "Multiple Drug Resistance" [2]. The receptor of sulphonylureas in pancreatic β cells (SUR1) is mutated in patients with persistent hyperinsulinemic hypoglycemia of infancy, and is the target of diabetes treatments [3]. Mutations in the CFTR gene, which codes for the cystic fibrosis transmembrane conductance regulator, provoke cystic fibrosis (for a review on CFTR see [4]). Moreover, proteins of this family play important roles in many different physiological processes. For example, they allow the secretion of α-factor and M-factor mating pheromones [5,6], the macrophages engulfment of corpses generated by apoptotic cell death [7] or heavy metal tolerance in yeast [8,9]. Animal, fungi and plant cells have the capacity to eliminate a broad range of lipophilic toxins from the cytosol after their conjugation with glutathione, glucoronide or glucose. The efflux of conjugated xenobiotics is directly driven by ATP hydrolysis [10,11]. The transporters involved are likely all members of the ABC family, at least for glutathione S-conjugated transport because these

* Corresponding author. Fax: 33 4 4225 4656; E-mail: marin@dsvcad.cea.fr

[1] The nucleotide sequence data reported in this paper will appear in the GenBank nucleotide sequence database with the following accession number AF014960.

PII S0005-2736(97)00213-7



pumps have been identified at the molecular level in both animal cells (MRP1, [12]) and yeast (YCF1, [8]).

All ABC transporters share a common structural organisation, with four "core" domains. Two transmembrane domains form a pore in the membrane and determine the substrate specificity of the transporter. Two cytosolic ATP-binding domains (also called nucleotide binding domains or NBD) couple ATP hydrolysis to solute movement. The NBD regions share extensive sequence homology. These domains present the highly conserved "Walker A" and "Walker B" sequences, most probably responsible for the interaction with ATP [13]. These Walker sequences are separated by a stretch of about 120–170 amino acids, including a short (12–13 amino acids) peptide motif, called the ABC transporter "signature" region [14]. The four core domains may be expressed as separate polypeptides, like TAP1 and TAP2, which form heterodimers in the membrane environment, or can be alternatively fused together into larger, multidomain proteins like CFTR or YCF1 [1]. In some bacterial and in all eukaryotic ABC transporters at least one transmembrane region and one NBD are encoded by a single gene [2]. Together with this basic four core domains, an additional N-terminal hydrophobic region can be present, establishing a subfamily called "MRP-like" group [15]. The conservation between ABC proteins in evolutionary distant organisms allows to conclude a common ancestral origin [16]. More than 150 ABC proteins have been identified in organisms ranging from bacteria to man, including 28 sequences from *Saccharomyces cerevisiae* [17]. Nevertheless, only one *Arabidopsis* sequence has been reported to date, AtPGP1, which encodes a P-gp homologue [18] with an unknown function. In an attempt to identify the *Arabidopsis* vacuolar glutathione-conjugated pump, we searched for EST sequences homologous to the ATP-binding domains of known eukaryotic ABC proteins.

Comparison of the conserved NBD amino acid sequences of mammalian ABC proteins with the *Arabidopsis* EST sequence database led to the identification of a homologous EST (accession number N37563). This clone was kindly provided by the *Arabidopsis* DNA Stock Center and fully sequenced by the dideoxy chain terminator method [19]. Two oligonucleotide primers were designed on the 5′ extremity of the EST cDNA. The sense primer sequence was 5′-GGATTGTTGGAAGGACTAGTGC-3′ and the antisens primer sequence was 5′-GTCC-CACGCTGGAATTCTCTCC-3′. They introduced a *Spe*I and an *Eco*RI site, respectively and amplified a 330 bp fragment on cDNA, comprising the nucleotides coding for the Walker A motif and part of the ABC signature motif within a NBD. A polymerase chain reaction (PCR) was performed on *Arabidopsis* Landsberg *erecta* (Ler) genomic DNA in the following conditions: 94°C 1 min, 30 cycles 94°C 30 s/60°C 30 s/72°C 1 min, and a final extension cycle of 3 min at 72°C, in a Perkin Elmer Cetus DNA thermal cycler. The amplified fragment was purified from an agarose gel with QIAEXII (Qiagen), digested with *Spe*I and *Eco*RI and then inserted into the plasmid Bluescript II KS(−) (Stratagene). The sequence of 8 independent clones revealed an identical 400 bp long fragment which contains a 75 bp intron (data not shown). The PCR amplified fragment was used as a ^{32}P-labelled probe to screen a λZAPII cDNA library kindly provided by Dr M. Thomas (Institut de Biotechnologie des Plantes, Orsay). The library was made from two-week old *Arabidopsis* seedlings of the Columbia (Col) ecotype [20]. From one million plaques screened, seven positive plaques were isolated. Subsequent sequence and restriction analysis showed that all the clones were cDNA copies of the same mRNA species, although they differed in length. The biggest of these clones (2740 bp insert) was completely sequenced in both strands. This partial cDNA has a 2601 bp open reading frame (ORF)

Fig. 1. Nucleotide sequence of the complete *AtMRP1* cDNA with the deduced amino acid sequence for the largest ORF starting with a methionine. The first initiation codon and the stop codon are indicated in bold, as well as the C-terminal potential protein kinase C phosphorylation sites. The stop codons in frame with the initial ATG in the 5′ untranslated sequence are underlined, as well as the putative polyadenylation site. The arrow indicates the 5′ extremity of the cDNA clone obtained in the expression library screening. The sequences underlined correspond to the positions of the oligonucleotides used to perform RACE PCR. The Walker A and Walker B motifs of both nucleotide binding domains are boxed.

with no initial methionine, followed by a stop codon and ending with a poly(A)$^+$ tail. The sequence of this clone was fully identical to the EST sequence and its 5′ extremity is pointed by an arrow in Fig. 1.

Based on this partial cDNA sequence, RACE PCR was performed using the 5′ RACE amplification kit (Life Technologies), on 1 μg total RNA from leaves of 4–5 weeks old *Arabidopsis* plants. RNA extraction was made according to the SNAP procedure (Invitrogen) with a DNAse step to avoid genomic DNA contamination of the sample. After reverse transcription with a specific primer (Race#a), a homopolymeric C$^+$ tail was added. The PCR amplification reaction was made with the second specific primer (Race#b) and a poly(G)$^+$ oligonucleotide, in the conditions recommended by the manufacturer. The product of this amplification reaction was cloned in the pCR2.1 vector (Invitrogen) and fully sequenced in both strands. Three sequential RACE PCR reactions were necessary to achieve the cloning of a complete cDNA, presented in Fig. 1. The primers used in these reactions are underlined in the figure and correspond to the following sequences:

Race1a: TTCTTCCACCTTCCCTGC
Race1b: CAATTCTGTCCACTTGTGATAGG
Race2a: TCAGTATGAACATATTCAACG
Race2b: CAGTTTGGACCTTGGACTGG
Race3a: AGTACTCCTTGACTGAGAG
Race3b: GAGCATATATGACAGCAAACC

The AtMRP1 cDNA is 5574 bp long and comprises a 5′ non-translated sequence of 567 bp, an ORF of 4869 bp and a 3′ non-translated region of 110 bp followed by a poly(A)$^+$ tail. The starting ATG codon is indicated in bold and preceded by several in frame stop codons (underlined in Fig. 1). The translation initiating sequence (AACG**ATG**G) match the plant consensus translation initiation codon [21]. The TAG ending codon is followed by a putative polyadenylation signal underlined in Fig. 1 [22]. These features supported our assignment of the 4869 bp ORF to the AtMRP1 cDNA. According to this assignment, the biggest ORF starting with a methionine encodes a protein of 1623 amino acids with an estimated molecular mass of 182 kDa. The deduced polypeptide sequence encompasses all the features of ABC proteins. AtMRP1 presents two transmembrane regions of 6α helices each and two hydrophilic NBD. Each NBD consists in classical Walker A and Walker B consensus sequences, boxed in Fig. 1. Between them, the sequence ISGGQKQRVSMARA in the first NBD and FSVGQRQLLSLSRGL in the second NBD correspond to the ABC signatures.

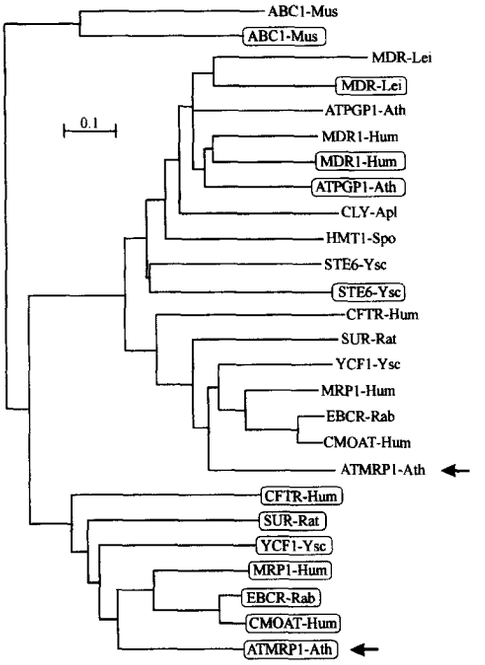


Fig. 2. Phylogenetic analysis of the amino acid sequences of the extended NBD of a representative group of ABC proteins from various organisms. The cladogram was constructed with the N-terminal and the C-terminal NBD regions from 13 sequences found in the Gene Bank database plus the *A. thaliana* MRP-like sequence reported here, which is pointed by arrows (accession number AF014960). The alignment and the tree were calculated by the neighbour joining method with the program Clustal W [30]. Distances along the horizontal axis are proportional to the difference between sequences. The sequences corresponding to the N-terminal NBD motifs are boxed, whereas those of C-terminal NBD motifs are not. Species legend and accession numbers (in brackets): ABC1-Mus, *Mus musculus* (X75926); MDR-Lei, *Leishmania enrietti* (Q06034); ATPGP1-Ath, *Arabidopsis thaliana* (419760); MDR1-Hum, *Homo sapiens* (P08183); CLY-Apl, *Actinobacillus pleuropneumoniae* (1170305); HMT1-Spo, *Schizosaccharomyces pombe* (Q02592); STE6-Ysc, *Saccharomyces cerevisiae* (134967); CFTR-Hum, *H. sapiens* (P13569); SUR-Rat, *Rattus norvegicus* (Q09429); YCF1-Ysc, *S. cerevisiae* (P39109); MRP1-Hum, H. sapiens (P33527); EBCR-Rab, *Oryctolagus cuniculus* (Z49144); CMOAT-Hum, *H. sapiens* (1574998).

The N-terminal part of AtMRP1 possesses an hydrophobic extension of 200 residues which likely forms a series of 4 transmembrane α helices. This extension allows to conclude that AtMRP1 is a member of the MRP-like subfamily of ABC proteins [15]. The phylogeny tree presented in Fig. 2 was made with the extended NBD amino acid sequences from 14 different ABC proteins. The *Arabidopsis* sequence we described here is more closely related to vertebrate MRP-like proteins than to the already described AtPGP1 sequence, an *Arabidopsis* homologue to MDR-like sequences. This observation means that, from an evolutionary point of view, the separation between the two subfamilies of ABC transporters is older than the separation of different kingdom. The classification of AtMRP1 in the MRP-like group is also supported by a stronger degree of homology with the sequences of human MRP1 [12], yeast YCF1 [8], mammals EBCR [23] and CMOAT [24], which belong all to the MRP-like subgroup.

Fig. 3 represents a schematic view of several ABC proteins, as fusion of core domains. The percentage of identity at the amino acid level between AtMRP1 and the other sequences for each of these domains is also indicated. As expected, the regions where amino acid sequence homology is stronger are the ATP-binding domains, with a maximal percentage of identity for the second NBD of the EBCR protein. For instance, it is also the EBCR sequence which shows a stronger homology in the C-terminal part of the protein (20.9% identity over 85 residues). It is important to notice that the C-terminal part of AtMRP1 is longer than other protein sequences reported in the literature. This extension is 144 amino acids longer than the corresponding C-terminal part of MRP1, and 116 amino acids longer than the COOH region of EBCR. Such a large cytosolic extension has never been found in the sequences of other ABC proteins and shows little homology to other proteins in the database. The significance of this C-terminal part of AtMRP1 is at present unknown, but it is interesting to notice that it possesses several potential phosphorylation sites for protein kinase C (in bold in Fig. 1). Regulation of numerous ABC proteins is at least in part driven by protein kinase A and/or C phosphorylation [25]. The possible regulatory role of this part of AtMRP1 will be investigated by means of heterologous expression systems producing a whole protein or a truncated form.

To further characterise the isolated locus, genomic DNA extraction and Southern analysis were performed as described previously [26]. *Arabidopsis* plants of the Wassilewskija (WS), Col, C24 and Ler ecotypes were cultured as described in [27]. 2 μg of DNA were analysed by digestion with different enzymes, run on 0.7% agarose gels and blotted by standard procedures [28]. In all the cases, the hybridisation pattern was identical whatever the ecotype used to do the DNA extraction. Fig. 4 shows a typical Southern blot analysis obtained with 50% formamide

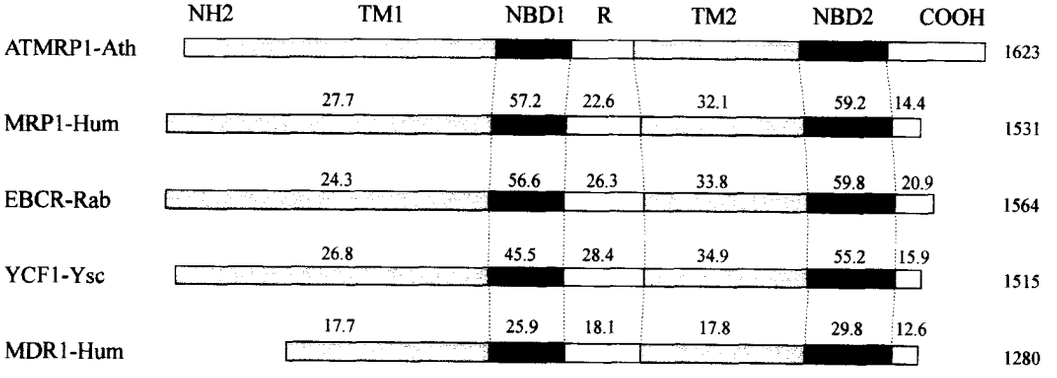


Fig. 3. Comparison between the different domains of AtMRP1 and a series of ABC proteins. Analysis of protein sequence identity was made with the Clustal W program [30] to compare potential functional domain sequence homology of AtMRP1 to MRP1, EBCR, YCF1 and MDR1. The core domains, indicated by white, grey or black boxes, correspond to: NH2, amino terminal extremity of the protein; TM, transmembrane domain; R, regulatory domain; COOH, C-terminal sequence behind the second NBD. The numbers in the right indicate the amino acid length of the sequences. The numbers on top of each box indicate the percent of identity between each domain and the corresponding AtMRP1 domain. The species and accession numbers are indicated in the legend of Fig. 2.

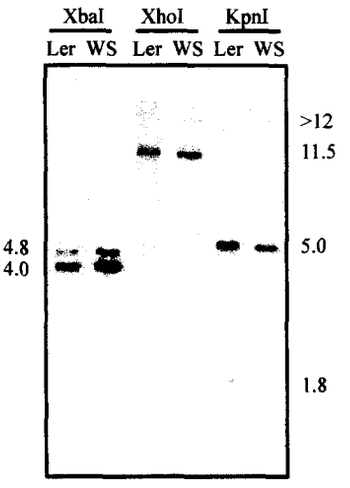


Fig. 4. Southern hybridisation analysis for the *Arabidopsis AtMRP1* gene. 2 μg of total genomic DNA of the Ler and WS ecotypes were digested with *Xba*I, *Xho*I or *Kpn*I, blotted and hybridised under stringent conditions with a probe corresponding to the *Xba*I insert (240 bp long) of the cloned PCR amplified fragment. This genomic probe does not present any restriction sites for the enzymes used. The numbers correspond to the size of the bands in kbp.

hybridisation buffer [28] at 42°C, followed by 2–3 washes 10 min each at 55°C on 2 × SSC, 0.2% sodium tetrapyrophosphate and 0.5% *N*-lauroylsarcosine (Sigma). A PhosphoImager (Molecular Dynamics) was used to reveal the bands. In this particular case, three restriction enzymes (*Xba*I, *Xho*I and *Kpn*I) were used to analyse DNA from WS and Ler leaves. The probe corresponds to the 240 bp long *Xba*I insert of the PCR amplified genomic fragment. This probe revealed two bands corresponding to two copies of the corresponding locus per genome. Using moderate stringency hybridisation conditions (hybridisation buffer without formamide, at 55°C) the same bands were revealed (data not shown). As the locus AtMRP1 appears in two copies in the *Arabidopsis* genome, we performed a PCR genomic library screening with the same oligonucleotides and conditions used to amplify the EST. The genomic library was a gift of Dr C. Lister (John Innes Institute, Norwich). Partial sequencing of the single genomic clone obtained by PCR screening allowed to confirm that the sequenced cDNA was an entire copy of a unique transcript (data not shown). Finally, to determine the organs that express the *Arabidopsis AtMRP1* gene, a Northern blot analysis was performed as described previously [29]. The PCR amplified probe revealed a 7 kb transcript on 10 μg total RNA in all the tissues tested including roots, stems, cotyledons, young and old leaves, flower buds and mature flowers (data not shown).

The conservation of all essential motifs of ABC proteins in the predicted amino acid sequence of AtMRP1 leaves little doubt that the cloned cDNA specifies an ABC protein. Importantly, alignments of the NBD regions of representative ABC proteins in the current protein databases clearly place AtMRP1 closest to the proteins belonging to the MRP-like subgroup, as human MRP1 [12], yeast YCF1 [8], mammalian CMOAT [24] and EBCR [23]. All these proteins were previously proposed to function as membrane ATP-driven pumps in detoxification by means of export of conjugated substrates. Cloning of AtMRP1 cDNA constitutes an important molecular tool for the study of glutathione S-conjugated export in plant cells, and will allow a better understanding of detoxification processes mediated by ABC proteins in plants.

We are very grateful to Dr M. Thomas (Institut de Biotechnologie des Plantes, Orsay) for the gift of an *Arabidopsis* cDNA library and to Dr C. Lister (John Innes Institut, Norwich) for the gift of a genomic library. Dr. V. Klimyuk (John Innes Institute, Norwich) is gratefully acknowledged for performing the PCR screening of the genomic library. The *Arabidopsis* DNA Stock Center and Dr K. Davis are gratefully acknowledged for providing the EST clone, as well as Dr Ph. Lessard for his precious help in the phylogenic tree construction.

Note added in proof

After the acceptance of our manuscript, Lu et al. (Proc. Natl. Acad. Sci. USA 94 (1997) 8243-8248) published the sequence of an *Arabidopsis* cDNA, also called *AtMRP1*, which is 85% identical to the sequence we describe here. So we propose to call our sequence *AtMRP2*.

References

[1] C.F. Higgins, Cell 82 (1995) 693–696.
[2] M.M. Gottesman, I. Pastan, Annu. Rev. Biochem. 62 (1993) 385–427.

[3] P.M. Thomas, G.J. Cote, N. Wohlik, B. Haddad, P.M. Mathew, W. Rabl, L. Aguilar-Bryan, R.F. Gagel, J. Bryan, Science 268 (1995) 426–429.
[4] P.B. Davis, M. Drumm, M.W. Konstan, Am. J. Respir. Crit. Care Med. 154 (1996) 1229–1256.
[5] J.P. McGrath, A. Varshavsky, Nature 340 (1989) 400–404.
[6] P.U. Christensen, J. Davey, O. Nielsen, Mol. Gen. Genet. 255 (1997) 226–236.
[7] M.F. Luciani, G. Chimini, EMBO J. 15 (1996) 226–235.
[8] Z.S. Li, M. Szczypka, Y.P. Lu, D.J. Thiele, P.A. Rea, J. Biol. Chem. 271 (1996) 6509–6517.
[9] D.F. Ortiz, L. Kreppel, D.M. Speiser, G. Scheel, G. McDonald, D.W. Ow, EMBO J. 11 (1992) 3491–3499.
[10] E. Martinoia, E. Grill, R. Tommasini, K. Kreuz, N. Amrhein, Nature 364 (1993) 247–249.
[11] M. Klein, G. Weissenböck, A. Dufaud, C. Gaillard, K. Kreuz, E. Martinoia, J. Biol. Chem. 271 (1996) 29666–29671.
[12] S.P. Cole, G. Bhardwaj, J.H. Gerlach, J.E. Mackie, C.E. Grant, K.C. Almquist, A.J. Stewart, E.U. Kurz, A.M.V. Duncan, R.G. Deeley, Science 258 (1992) 1650–1654.
[13] J.E. Walker, M. Saraste, M.J. Runswick, N.J. Gay, EMBO J. 1 (1982) 945–951.
[14] V. Shyamala, V. Baichwal, E. Beall, G.F-L. Ames, J. Biol. Chem. 266 (1991) 18714–18719.
[15] G.E. Tusnady, E. Bakos, A. Varadi, B. Sarkadi, FEBS Lett. 402 (1997) 1–3.
[16] B. Sarkadi, M. Müller, Z. Hollo, Immunol. Lett. 54 (1996) 215–219.
[17] H.W. Mewes, K. Albermann, M. Bähr, D. Frishman, A. Gleissner, J. Hani, K. Heumann, K. Kleine, A. Maierl, S.G. Oliver, F. Pfeiffer, A. Zollner, Nature 387 (1997) 7–65.
[18] R. Dudler, C. Hertig, J. Biol. Chem. 267 (1992) 5882–5888.
[19] R. Sanger, S. Nicklen, A.R. Coulson, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 5463–5467.
[20] M.W. Bianchi, D. Guivarc'h, M. Thomas, J.R. Woodgett, M. Kreis, Mol. Gen. Genet. 242 (1994) 337–345.
[21] H.A. Lütcke, K.C. Chow, F.S. Mickel, K.A. Moss, H.F. Kern, G.A. Scheele, EMBO J. 6 (1987) 43–48.
[22] C.P. Joshi, Nucleic Acids Res. 15 (1987) 6643–6653.
[23] M.A. Van Kuijck, R.A.M.H. Van Aubel, A.E. Busch, F. Lang, F.G.M. Russel, R.J.M. Bindels, C.H. Van Os, P.M.Y. Deen, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 5401–5406.
[24] C.C. Paulusma, P.J. Bosma, G.J. Zaman, C.T.M. Bakker, M. Otter, G.L. Scheffer, R.J. Scheper, P. Borst, P.J. Oude Elferink, Science (1996) 1126–1128.
[25] Y.L. Jia, C.J. Mathews, J.W. Hanrahan, J. Biol. Chem. 272 (1997) 4978–4984.
[26] A. Marion-Poll, E. Marin, N. Bonnefoy, V. Pautot, Mol. Gen. Genet. 238 (1993) 209–217.
[27] G. Lascève, J. Leymarie, A. Vavasseur, Plant Cell Environ. 20 (1997) 350–358.
[28] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning, Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[29] E. Marin, L. Nussaume, A. Quesada, M. Gonneau, B. Sotta, P. Hugueney, A Frey, A. Marion-Poll, EMBO J. 15 (1996) 2331–2342.
[30] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.