Arabidopsis thaliana sequence analysis confirms the presence of cyt b-561 in plants: Evidence for a novel protein family

Han Asard*, Javier Terol-Aleaydeb, Valeria Pregerc, Jurgen Del Faveroa, Wim Verelsta, Francesca Sarlabc, Manuel Pérez-Alonsoa, Paolo Trostc

a Laboratory of Plant Physiology, University of Antwerp (RUCA), Groenenborgerlaan 171, B-2020 Antwerp, Belgium
b Departamento de Genética, University of Valencia, Dr. Moliner 50, Burjasot 46100, Spain
c Laboratory of Plant Physiology, Department of Biology, University of Bologna, Via Irnerio 42, 40126 Bologna, Italy
d Department of Molecular Genetics, Flanders Interuniversity Institute for Biotechnology (VIB), University of Antwerp (UIA), Universiteitsplein 1, 2610 Wilrijk, Belgium

Received 22 May 2000; accepted 16 August 2000

Abstract – Recent advances in the Arabidopsis sequencing project has elucidated the presence of two genes Atb561-A and Atb561-B that show limited homology to the DNA sequence encoding for the mammalian chromaffin granule cytochrome b-561 (cyt b-561). Detailed analysis of the structural features and conserved residues reveals, however, that the structural homology between the presumptive Arabidopsis proteins and the animal proteins is very high. All proteins are hydrophobic and show highly conserved transmembrane helices. The presumably heme-binding histidine residues in the plant and animal sequences as well as the suggested binding site for the electron acceptor, monodehydroascorbate, are strictly conserved. In contrast, the suggested electron donor (ascorbate) binding site is not very well conserved between the plant and animal sequences questioning the function of this motif. Sequence analysis of the Atb561-B gene demonstrates a different splicing than that initially predicted in silico resulting in a protein with nine extra amino acids and a significantly higher homology to the other cyt b-561 sequences. The homology between the plant and animal sequences is further supported by the strong similarity between a number of biochemical properties of the chromaffin cyt b-561 and the cyt b-561 isolated from bean hook plasma membranes. Since the mammalian cyt b-561 is considered specific to neuroadrenergic tissues, the identification of a closely related homologue in an aneural organism demonstrates that these proteins constitute a new class of widely occurring membrane proteins. Both the plant and animal cyt b-561 are involved in transmembrane electron transport using ascorbate as an electron donor. The similarity between these proteins therefore suggests, for the first time, that this transport supports a number of different cell physiological processes. An evolutionary relationship between the plant and animal proteins is presented. © 2000 Éditions scientifiques et médicales Elsevier SAS

ascorbic acid / chromaffin granules / cytochrome b-561 / electron transport / genome analysis / higher plants / plasma membrane

aa, amino acid / cyt b-561, cytochrome b-561 / E°, standard redox potential at pH 7 / EST, expressed sequence tag / MDHA, monodehydroascorbate / ORF, open reading frame

1. INTRODUCTION

Characterization of higher plant plasma membrane fractions has revealed the presence of a high potential (E°: +110 to +165 mV) ascorbate-reducible b-type cytochrome with an α-band wavelength maximum at 561 nm (cytochrome b-561, cyt b-561) [6, 7]. This cytochrome constitutes a common plant plasma membrane protein since it was found in a variety of tissues and in a large number of species [4, 5]. Recent experiments have demonstrated that the cyt b-561 is involved in the electrogenic transport of electrons across the plasma membrane [3, 14] using cytoplasmic ascorbate as an in vivo electron donor and monodehydroascorbate (MDHA) as an extracellular (apoplastic) electron acceptor. Although the physiological function

* Correspondence and reprints: fax +32 3 218 04 17.
E-mail address: hanasard@ruca.ua.ac.be (H. Asard).
of this protein remains to be elucidated, the transmembrane transport of electrons to MDHA suggests a role in the regeneration of ascorbate. Ascorbate in the plant apoplast plays an important role in antioxidative defence reactions, e.g. those induced under biotic and/or abiotic stress conditions [21]. Attempts to solubilize and purify the protein from plasma membrane fractions have recently been presented [8, 25].

From the early investigations on the plant plasma membrane cyt b-561, it was recognized that this protein shows remarkable biochemical similarities to the mammalian cyt b-561 characterized from chromaffin granules [12, 20]. The latter component (E°: -140 mV) transfers electrons from cytoplasmic ascorbate to MDHA present inside the chromaffin vesicles [19, 20, 32]. Intravesicular ascorbate provides reducing equivalents for the monoxygenase dopamine β-hydroxylase involved in dopamine biosynthesis. Aside from hematopoietic tumour cells [26], this particular cytochrome has so far only been demonstrated in catecholamine and neuropeptide secretory vesicles and its expression is coupled to neuronal cell differentiation [27, 29]. The primary structure of the chromaffin granule cyt b-561 has been elucidated from a number of mammalian species including humans [28], sheep, pig, bovine [23] and mouse [29]. The presence of similar genes was demonstrated in Xenopus laevis [27] and in Caenorhabditis elegans [33].

On the basis of the amino acid sequence of the mammalian cyt b-561, predictions on the secondary and tertiary structure of the protein have been formulated. These include the prediction of five [27, 28] or six [10, 22, 23] hydrophobic transmembrane helices, respectively. By different experimental approaches the mammalian protein has been shown to contain two distinct hemes [11, 16, 31] and suggestions have been presented for the binding sites of the electron donor (ascorbate) and electron acceptor (MDHA) [22].

Recent progress in the sequencing of the Arabidopsis genome has revealed the presence of two distinct sequences showing homology to the mammalian cyt b-561 genes. Although the overall homology at the DNA level is relatively low, a detailed comparison of the predicted protein structure, together with new sequence information presented in this paper, and the biochemical properties of the plant plasma membrane cyt b-561 strongly supports the idea that the plant and animal proteins are highly similar. This analysis demonstrates that the high potential cyt b-561 constitutes a novel protein family present in neural and aneural eukaryotic species.

2. RESULTS AND DISCUSSION

2.1. Identification of Atb561-A and Atb561-B

As a result of the participation of the Valencia group in the Arabidopsis Genome Sequencing Project, the sequence of the genomic clone M7J2 from Arabidopsis chromosome IV was obtained. An exhaustive search for similarities was performed on the 81.6 kb genomic sequence using the BLAST program. A small region showed a weak similarity to the cyt b-561 gene from different animals. That region of DNA was analysed to find the existing coding sequences, using different approaches (see Methods). An ORF, 1752 bp long and containing four exons, was predicted.

The conceptual translation of the ORF produced a 280-amino acid protein (GenBank accession number: A1022197), which was used to perform a new search for homologies using the BLAST program. This protein showed a clear similarity to cyt b-561 proteins derived from human, mouse, pig, cow, sheep, X. laevis and the nematode C. elegans. Nine Arabidopsis EST clones were found (GenBank: AA720180, AA728528, F19849, R65413, T46072, T04182, T04807, T08282 and Z37601) and the corresponding cDNAs were sequenced from both strands. The resulting sequences confirmed the in silico cDNA prediction and was named Atb561-A.

A putative protein showing homology to Atb561-A was found in the KAOS database [24] (GenBank accession number AB005231), in a genomic clone (BAC clone) MBB18 from chromosome V. However, the theoretical protein of 223 aa predicted from this sequence [24] was not correct. On the basis of sequencing the corresponding cDNA from the ESTs (GenBank accession numbers: Z26702, Z33925) we determined that the second exon of the ORF was 27 bp longer. The additional nine amino acids (Ala^79–Glu^87 in figure 2) showed a high homology to Atb561-A and the animal proteins (see below). The new protein was named Atb561-B (GenBank accession number: AF132115).

2.2. Intron-exon analysis

Combining the data from the genomic and EST sequences showed a transcription unit for Atb561-A comprised of four exons, of 267, 48, 201 and 324 bp, separated by introns of variable size (figure 1). The 5'-untranslated region is approximately 100 bp long and we could not identify any sequence resembling the TATA box. In this respect, Atb561-A is similar to the human cyt b-561 gene which has no TATA box [26]. Furthermore, two putative initiation sites are present in
the Ath561-A ORF, with the second one 123 bp downstream of the first (not shown). On the basis of the multiple alignment with the other cytochromes (see figure 2), the second methionine codon in the ORF is most likely the starting point for translation. In addition, two ESTs (GenBank accession numbers: R65413, T04182), containing the full length Ath561-A, begin downstream of the first ATG. Therefore, only the short version of the protein is included in figure 2, starting at the second methionine of the ORF. A similar problem with the bovine protein, which also contains two methionine codons in the beginning of the sequence, was solved by analysis at the protein level [27], suggesting that the protein was probably 21 aa shorter than that previously reported [23]. The polyadenylation signal for Ath561-A is 157 bp downstream of the termination codon. In contrast to Ath561-A the 5'-UTR for Ath561-B contains a putative TATA box 30 bp upstream of the first methionine codon.

Although the similarity between Ath561-A and -B genes is only 57.4 % at the DNA level, the organization of the two transcription units in A. thaliana is quite similar. Both ORFs have similar lengths, have four exons and code for proteins of 239 and 230 aa, respectively. As could be expected, the introns have not been conserved (figure 1). The intron-exon structure of the Arabidopsis genes was compared to that of the human cyt b-561 homologue, which is the only other protein of this family described at the genomic level so far [26]. The human protein contains five exons and two of the four splicing sites are perfectly conserved between the Arabidopsis and human sequences (figure 2).

According to our prediction of the Ath561-A and -B structure, the intron-exon organization does not reflect the structural domains of the protein. There are four exons and six predicted transmembrane domains and their position does not indicate that the exons reflect the structural domains of the protein. Transmembrane helix I (TR1) is present in exon 1 (E1), helix II expands the end of E1, E2 and the beginning of E3, helix III is located on E3, helix IV is divided between exons 3 and 4 and helices V and VI are present in E4 (figure 1). In contrast in the five transmembrane helix model suggested by Srivastava [27] for the human protein, each of the exons contains a single transmembrane domain.

2.3. Multiple alignment and structural analysis

Based on the Kyte and Doolittle [17] algorithm, hydropathy plots were constructed for the two Ath561 proteins (figure 3) and, as an example, compared to that of the human protein. The hydropathy plot shows six highly hydrophobic regions for each sequence separated by short hydrophilic loops. The hydrophobicity profiles of the Arabidopsis and human proteins closely overlap which is also reflected in the similar locations of the predicted transmembrane helices in the multiple alignment (figure 2). The very high hydrophobicity of the plant proteins is consistent with the absolute need for detergent in order to solubilize the protein from purified plasma membrane fractions [8, 25].

Recent predictions on the transmembrane structure of cyt b-561 from chromaffin granules are still contradictory. On one hand, the analysis of Srivastava et al. [27, 28] indicates the presence of five membrane spanning helices with the N-terminus facing the chromaffin vesicle interior. This prediction was partially based on the inaccessibility of the N-terminus to
helix model for the mammalian sequence, one of the two heme molecules was suggested to be co-ordinated by the pair His$^{64}$-His$^{140}$, whereas the other heme could be co-ordinated by the pair His$^{98}$-His$^{179}$ or, alternatively by the pair His$^{102}$-His$^{179}$ [22]. However, since the His$^{102}$ is not conserved in the Atb561-B sequence, we suggest that this residue is not involved in heme ligation.

Another remarkable site of strict conservation among the Arabidopsis and animal proteins is the polar five-amino acid stretch $^{138}$SLHSW$^{142}$. This sequence is located at the intravesicular end of $\alpha$-helix IV in the chromaffin vesicle membrane. The sequence is also strictly conserved in a partial EST-derived Drosophila melanogaster sequence (191 aa) (GenBank accession number: AI403697) that shows a high similarity (36 % identity) to the human cyt $b$-561 sequence. A tentative cyt $b$-561 sequence identified in Mesembryanthemum crystallinum (GenBank accession number: AF097661) contains a conservative replacement for the Ser$^{138}$ residue by Thr. By sequence comparison, Tsubaki’s group [22] suggested that this sequence could provide the binding site for MDHA. MDHA functions as an electron acceptor to the plant cyt $b$-561-mediated electron transport as demonstrated by using ascorbate-loaded plasma membrane vesicles [14]. The high conservation of the possible MDHA-binding site supports these results and points to the functional similarity between the plant and animal proteins.

Similarly, a stretch of nine amino acids ($^{79}$ALLVYRVFR$^{87}$) was indicated as the binding site for ascorbate and is strictly conserved in animal sequences [22]. However, comparison with plant sequences demonstrates that this sequence is only partially conserved in A. thaliana (figure 2) and M. crystallinum (not shown). In addition this stretch is also not conserved in the partial D. melanogaster sequence. These observations question the function of this sequence and are of importance with respect to the elucidation of the electron transfer mechanism from ascorbate to one of the heme groups of cyt $b$-561.

Finally, comparison of potential N-glycosylation motifs (at the PROSITE database), showed that Atb561-A contains two potential N-glycosylation sites (Asp$^{119}$, Asp$^{223}$, figure 2), whereas no potential glycosylation sites were found in the Atb561-B or human

antibodies and proteases [15]. On the other hand, Perin et al. [23], Degli Esposti et al. [10] and Okuyama et al. [22] suggested the presence of six transmembrane helices in the mammalian cyt $b$-561. The latter model accounts for the accommodation of two heme molecules per protein. Secondary structure predictions (TMpred software, [13]), for the Arabidopsis proteins also suggested six transmembrane helices (figure 2). A comparison of the predicted membrane spanning domains for all known cyt $b$-561 proteins indicates the high conservation of this structural feature (figure 2).

Multiple alignment revealed the strict conservation of five histidine residues (His$^{64}$, His$^{98}$, His$^{120}$, His$^{140}$, His$^{179}$, figure 2) located in the helices II to V. In the six

Figure 3. Hydrophathy plot analysis for Atb561-A and -B and the human cytochrome $b$-561 protein sequence.

Figure 2. Multiple alignment analysis (Clustal W, [30]) between the predicted Arabidopsis cytochrome $b$-561 homologues, and other cytochrome $b$-561 homologues identified in animal cells. The shaded areas indicate the predicted transmembrane helices using the TMpred software [13]. The boxed areas represent the predicted ascorbate and MDHA binding sequences. $^{\star}$, splicing site (first aa of new exon); N, Asp glycosylation sites; H, conserved His residues.
sequences. Both Atb561-A and -B also showed a high probability for a plasma membrane (64 % for Atb561-A and B) or ER (68.5 % for Atb561-B and 37 % for Atb561-A) localization as identified by comparison to known signal peptide sequences (PSORT version 6.4, [18]). Although these remain predictions, they are consistent with the experimental evidence that at least one of these proteins is probably located at the plasma membrane. In addition, some evidence indicates that an ascorbate-reducible cytochrome b is also found in other subcellular fractions, different from mitochondria and endoplasmic reticulum, in cauliflower [2] and bean hypocotyls [25].

On the basis of the conservation and similarities between the plant and animal sequences, a model is proposed for the Atb561-A protein using the six-helix model (figure 4). Based on the conservation of the MDHA binding site and the potential glycosylation sites, the C- and N-terminus are located on the cytoplasmic site of the membrane.

2.4. Evolutionary relationships

The evolutionary relationship between the proteins analysed in this study (figure 5) clearly reflects the systematic positions of the species, suggesting that the protein behaves as a molecular clock. Most of the sequences available so far originate from vertebrate species. However, an increasing number of ESTs from plants becoming available shows homology to the cyt b-561 protein family and will allow a further refinement of the evolutionary tree. Remarkably, no homologous protein has been found in Saccharomyces cerevisiae, whose full genome has been sequenced. The protein is present in insects, as there are several D. melanogaster ESTs with a high degree of similarity to the human cyt b-561.

3. CONCLUSION

This paper presents the first evidence for the existence of a new protein family, consisting of b-type cytochromes common to eukaryotic species. It is obvious that the transmembrane electron transport capacity of all of these proteins is probably very similar, and that this reaction may support a number of different physiological processes. Future work will be directed towards the further characterization of these functions in plants and of the regulation of the expression of the cyt b-561.

4. METHODS

The Arabidopsis thaliana P1 genomic clone M712, corresponding to the bottom arm of chromosome IV, was completely sequenced in the frame of the European Arabidopsis Genome Sequencing Project. A shotgun library approach was used to determine the DNA sequence of the insert. The DNA sequence was obtained by using the IR Taq DNA sequencing kit (Boehringer Mannheim) and a LI-COR Automated DNA Sequencer. The sequencing project was managed with the Staden Software Package, a total of 81.6 kb were analysed and the average redundancy was 8. The search for ORFs in the genomic clone was performed with the GENSCAN program [9], in collaboration with the Munich Information Centre for Protein Sequences.
(MIPS). The sequence databases were searched for sequence similarities using the BLAST program of the National Center for Biotechnology Information [1]. EST clones were sequenced with an ABI automatic DNA sequencer and compared to the genomic sequences using the GAP and BESTFIT programs from the GCG Software Package (Wisconsin University). The Arabidopsis genes will be referred to as Arb561-A (for Arabidopsis thaliana cyt b-561) and Ath561-B.

NOTE ADDED IN PROOF

In a recent paper (Asard et al. Protoplasma, in press), the suggested nomenclature for the plant Arabidopsis cyt b-561 genes is Arth561-n. A third putative Arabidopsis cyt b-561 has also been identified (GenBank accession number AC006917).

Acknowledgments. The authors gratefully acknowledge the financial support of the Fund for Scientific Research Flanders (FWO-Flanders), by CNR (Target project on Biotechnology) and by PRIN Bioenergetics and Membrane Transport, MURST, Italy. HA and WV are Research Associate and doctoral student respectively at the FW-Flanders.

REFERENCES


