Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 340 (2004) 111–121

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Ankyrin repeat-containing proteins in *Arabidopsis*: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins

Cristian Becerra[a], Torben Jahrmann[a,b], Pere Puigdomènech[a], Carlos M. Vicient[a,b,*]

[a] *Departament de Genètica Molecular, IBMB-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain*
[b] *Departament de Genètica Vegetal, Centre de Cabrils, IRTA, Ctra. de Cabrils s/n, 08348 Cabrils, Barcelona, Spain*

Received 3 November 2003; received in revised form 15 May 2004; accepted 1 June 2004
Available online 25 July 2004
Received by W. Martin

Abstract

Ankyrin repeats are present in a great variety of proteins of eukaryotes, prokaryotes and some viruses and they function as protein–protein interaction domains. We have search for all the ankyrin repeats present in *Arabidopsis* proteins and determined their consensus sequence. We identified a total of 509 ankyrin repeats present in 105 proteins. Ankyrin repeat containing proteins can be classified in 16 groups of structurally similar proteins. The most abundant group contains proteins with ankyrin repeats and transmembrane domains (AtANKTM). Sequence similarity analysis indicates that these proteins are divided in six families. Some of the *AtAnkTm* genes are organized in tandem arrays and others are present in duplicated parts of the *Arabidopsis* genome. The expression of several *AtAnkTm* genes was analyzed resulting in a wide variety of expression patterns even within the same family. The likely functions of these proteins are discussed in comparison with the known functions of proteins with similar organization in other species.

Keywords: Ankyrin; *Arabidopsis*; Protein domain; Transmembrane

1. Introduction

Ankyrin repeats (ANK repeats) is a commonly occurring protein repeat present in prokaryotes, eukaryotes and some viruses (Sedgwick and Smerdon, 1999). The primary structure of ANK repeats consists in 33 residues repeated in tandem that built a specific secondary and tertiary structure. Only few of the amino acids are invariant and correspond to hydrophobic positions which are necessary to maintain the secondary structure (Rhode and Bork, 1993; Bork, 1993; Mosavi et al., 2002). ANK repeat tandem arrays consists of two or more repeats separated by less than 20 amino acids (Sedgwick and Smerdon, 1999). ANK repeats are present in proteins involved in very different functions including cell cycle regulation, mitochondrial enzymes, cytoskeleton interactions, signal transduction or toxins (Sedgwick and Smerdon, 1999).

ANK repeats mediate protein–protein interactions. This function has been experimentally demonstrated in both binding heterologous proteins and mediating homodimerization (Bork, 1993; Lin et al., 1999). For example, ANK repeats are involved in binding together subunits of the GABA-binding protein β (GABPβ); the protein I-κBα is almost entirely comprised of ANK repeats and is able to bind the 65 kDa subunit of NF-κB; the α-latrotoxin from black widow spider venom, which contains 19 ANK repeats, associates with an extracellular protein target; and Su(H) and Deltex proteins bind to Notch ANK repeats.

The folding of the ANK repeats plays an important role for its function (Mosavi et al., 2002). The arrays of ANK repeats consist of pairs on antiparallel α-helices stacked side by side and linked by a series of intervening β-hairpin motifs. The structure is stabilized by extended antiparallel β-sheets formed between the repeats and by hydrophobic

Abbreviations: ANK, ankyrin.

* Corresponding author. Departament de Genètica Molecular, IBMB-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain. Tel.: +34-93-4006100; fax: +34-93-2045904.
E-mail address: cvsgmp@cid.csic.es (C.M. Vicient).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.06.006

bonds within the repeat and between the neighbouring repeats. The extended β-sheet projects away from the helical pairs almost at right angles, resulting in a L-shaped cross-section. The ability of ANK repeats to bind target proteins involves contact through the tips of the β-hairpins, which are exposed to the solvent, and the surface of the helical bundle facing the ankyrin groove. In general, the residues in the tips of the β-hairpins are not conserved highly in the ANK consensus. As such, they are not structurally constrained and are ideally located to perform binding roles and determine protein interaction specificities.

In plants, few proteins containing ANK repeats have been characterized and the molecular functions of the repeats have not been demonstrated. Here, a comprehensive analysis of the *Arabidopsis* genome detected several genes coding for proteins with ANK repeats that can be classified on 16 groups. We paid special attention to a group of genes coding proteins with ANK repeats and transmembrane domains.

2. Materials and methods

2.1. Plant material

Arabidopsis Col-0 plants were grown in soil, in 22 °C growth chambers, with 18 h light days. Plants used for RNA extractions from roots were grown on 0.8% (w/v) Murashige and Skoog basal salt mixture agar plates in 22 °C growth chambers under 18 h light days.

2.2. Sequence acquisition and analysis

The initial set of ANK repeats containing proteins was compiled from SMART v3.5 (http://smart.embl-heidelberg.de/) and TAIR Protein Search http://www.arabidopsis.org). The phasing of the repeats was that proposed by Michaely and Bennett (1992). Additional repeats were obtained using REP v1.1 (http://www.embl-heidelberg.de/~andrade/papers/rep/search.html). A multiple alignment of ANK repeats was obtained using CLUSTALW in order to construct representative sequences. With these sequences we screened protein databases using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and identified unique hits, removing duplications from our data set caused by the multiple identification numbers frequently assigned to the same DNA or protein sequence in the databases. The search was conducted in three steps, in each one searching and adding new repeats and recalculating consensus patterns. ANK repeats do not have high sequence conservation (Bork, 1993) and using this system we ensured isolation of all putative ANK repeats. However, we also risk including false hits. In order to remove noisy hits we took into account that ANK repeats are always found in arrays of at least two repeats separated by not more than twenty amino acids. We removed all the "isolated" ANK repeats obtained. We also removed all partial repeats and those putative ANK repeats that do not conserve at least two of the six most conserved amino acids defined by Bork (1993). In a final round, several candidates with weak signals were examined in detail. The less conserved ANK repeats were added only when they occur

Table 1
PCR primers used for RT-PCRs

Atg Number	Primer sequence 5′	Primer sequence 3′
At4g03440	CTTGGATTTGCTACGTCGTAGCC	GAGATACTGCTCTCCACTCAGCC
At4g03450	CTTGTACACGCGGCTCTAAAGGC	ACCCTCTTGGCGAACAAGTGCAC
At4g03480	CAGATTCCGCTTCATGTGGCCGC	CAGGCAGCTGTCTCCATTTGGCG
At4g05040	GCAGGTAACAATGACCTTGAAGGG	CCTGCAGCAAACGTCATTGTGGC
At4g14400	GACAACGTGGACCGTGAAGTGAGG	AGAGCCGCTACCACGAGAAGAGC
At1g14480	CGGGATGGATCCAGAGAATGAGCC	GGAGGGCCATCTGATAAGTGGCTG
At1g14500	GCCAAAAAGATTCTGCTTCCACCG	GCTGGAGTGCAGTTTGATAAGTGG
At4g10720	GCGAGAAAACTTAACACATACGGG	GGTTATGTAAGATATCTAGGGCGG
At4g11000	GCTAAACGTGTCAGGTTTCAGCCC	GGAAGGTGAGCACTGAGAGATAGC
At5g15500	AACGCCGACGGACTTACAC	ATTCCCAAAACCAAACTACC
At5g51160	GGGGGTTGAAAAGAAGCTTTGCCG	GCTTCGCTCGGAAACATAACCAGC
At5g54610	GGGTGGATGCAGAAAATGCGCG	GGGCTGCAGTCTGAAAAGTGGC
At5g54620	GCGAGACTGCTCTACATATTGCGG	GCGATCTCAAGAAGACAGATGCGG
At2g24600	CTTGAGCTTGTCGAGGGAGAAGG	CTCTCTTCCCGTAACGCGTACGG
At5g54710	GGCCCAGAGTGCAAACATACGCC	CACCACCAGGAGGGTTTATCCCG
At1g07710	GGGGAAACAGAACCAGTCAGGCG	CCATGTGAAGGGCTGTTTGGCCC
At2g01680	CGCTTTTCATGTCGCTGCCAAGCG	CCGTAAAGGTCAGCAATAAGCTCG
At3g09550	GTTGCTCCTTCGAGCTGATCCGG	AGCTCCTTGGCGATGCCATCGACG
At3g12360	GCTGAAGTTGCGGAGATTCGAGC	GTCAAGCGCTGTCTTGTGATCGC
At5g02620	CCGGAACCAAAGCCAAGAACGGC	AGCTGAGTATGCACCTCATGGCC
At5g60070	AGGGTCAGACGCCACTTCACATG	CGCTTCTCCGAGAGATTGTCCCG
At5g04690	AACACGCAGATGGAGATGGCTCG	AACGTTCCTTCAAGCCCAAGGCC
At5g09810 (Actin)	GGCCGATGGTGAGGATATTC	CTGACTCATCGTACTCACTC
At2g40170 (*AtEm6*)	GGCGTCTCAACAAGAGAAGAAGC	GGGGAAGTTTGATTTAGGTCTTG

between clearly identified repeats or were located in the extreme of a tandem array.

The analysis of the presence of additional domains other than ANK repeats was performed using SMART v3.5 (http://smart.embl-heidelberg.de/). The alignment of the protein sequences was done with CLUSTALW and phylogenetic analysis were performed using the neighbor-joining method. The analysis of chromosomal duplications in the *Arabidopsis* genome was done using the on-line facility provided by Ken Wolfe (http://wolfe.gen.tcd.ie/athal/dup).

The position of the introns, start and stop codons of the predicted AtANKTM proteins was checked by comparing the genomic sequences with the cDNA sequences deposited into the GenBank, comparing the proteins of the same family in the alignments and by direct sequencing of RT-PCR fragments. We found small differences from the prediction in sixteen of the forty genes. All the analyses shown here were done with the reviewed sequences.

2.3. RNA extraction and RT-PCR

Total RNAs were extracted from frozen organs of *Arabidopsis* as described (Vicient and Delseny, 1999) and treated with DNAse I (RNAse-free DNAseI, Promega). Total pretreated RNA (2 μg) was reverse transcribed with Omniscript reverse transcriptase kit (Qiagen) using an oligo-dT primer. cDNAs were amplified with specific primers designed flanking introns (Table 1). Reaction controls with nonreverse transcribed RNA were also used to detect gDNA contamination. The actin gene was used as a control of RNA loading. *AtEm6* gene was used as a control of expression in mature seeds (Vicient et al., 2000). PCR reactions were performed using 0.2 mM each dNTP, 360 μg/ml BSA and 1 pmol μl^{-1} each primer in a final volume of 50 μl. The reaction mixtures were heated to 95 °C for 5 min, followed by 28 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 90 s. Reactions were completed with one incubation at 72 °C for 10 min. Reactions were performed in a Minicycler (MJ Research, Waltham, MA) thermal cycler. Reaction products were cloned in pGEM-T (Promega) for sequencing.

3. Results

3.1. Generation of a nonredundant ankyrin containing protein set

The target of this work is to identify all the ankyrin repeats in *Arabidopsis* proteins using database searches. Our searching scheme (see Section 2) allowed us to identify a total of 509 ANK repeats coded by 105 genes. The number of ANK repeats in the same array ranks between 2 and 10 and the average is 4.5. Some of the proteins contain two separate arrays of ANK repeats. A summary, including Atg number for each of these proteins is provided in Table 2. For

Table 2
Arabidopsis ankyrin repeat containing proteins

EN	Atg no.	Fa.[a]	Name	Reference
Cluster A: Ankyrin-transmembrane proteins				
1	At1g03670	1	n.a.	n.a.
2	At4g03440	1	n.a.	n.a.
3	At4g03450	1	n.a.	n.a.
4	At4g03460	1	n.a.	n.a.
5	At4g03470	1	n.a.	n.a.
6	At4g03480	1	n.a.	n.a.
7	At4g03490	1	n.a.	n.a.
8	At4g03500	1	n.a.	n.a.
9	At4g05040	1	n.a.	n.a.
10	At4g14390	1	n.a.	n.a.
11	At4g14400	1	ACD6	Lu et al., 2003
12	At1g14480	2	n.a.	n.a.
13	At1g14500	2	n.a.	n.a.
14	At4g10720	2	n.a.	n.a.
15	At4g11000	2	n.a.	n.a.
16	At5g15500	2	n.a.	n.a.
17	At5g51160	2	n.a.	n.a.
18	At5g54610	2	n.a.	n.a.
19	At5g54620	2	n.a.	n.a.
20	At1g10340	3	n.a.	n.a.
21	At1g34050	3	n.a.	n.a.
22	At2g24600	3	n.a.	n.a.
23	At5g50140	3	n.a.	n.a.
24	At5g54700	3	n.a.	n.a.
25	At5g54710	3	n.a.	n.a.
26	At5g54720	3[b]	n.a.	n.a.
27	At1g05640	4	n.a.	n.a.
28	At1g07710	4	n.a.	n.a.
29	At2g01680	4	n.a.	n.a.
30	At2g31820	4	n.a.	n.a.
31	At3g09550	4	n.a.	n.a.
32	At3g12360	4	n.a.	n.a.
33	At5g02620	4	n.a.	n.a.
34	At5g60070	4	n.a.	n.a.
35	At3g18670	5	n.a.	n.a.
36	At3g54070	5	n.a.	n.a.
37	At5g04690	5	n.a.	n.a.
38	At5g35830	5[b]	n.a.	n.a.
39	At2g14250	6[b]	n.a.	n.a.
40	At5g20350	6	n.a.	n.a.
Cluster B: Proteins with only ankyrin repeats				
41	At1g04780	1	n.a.	n.a.
42	At1g11740	1	n.a.	n.a.
43	At1g62050	1	n.a.	n.a.
44	At3g04470	1	n.a.	n.a.
45	At3g24210	1	n.a.	n.a.
46	At2g17390	2	ART2	Peck et al., 2001
47	At4g35450	2	ARP2	Yan et al., 2002
48	At5g40160	3	EMB506	Albert et al., 1999
49	At5g66055	3	AKR	Zhang et al., 1992
50	At5g07840	4	n.a.	n.a.
51	At5g61230	4	n.a.	n.a.
52	At3g01750	5	n.a.	n.a.
53	At3g04140	5	n.a.	n.a.
54	At5g65860	6	n.a.	n.a.
55	At4g19150	7	n.a.	n.a.
56	At2g03430	8	n.a.	n.a.
57	At5g12320	9	n.a.	n.a.
58	At3g09890	10	n.a.	n.a.

(continued on next page)

Table 2 (*continued*)

EN	Atg no.	Fa.[a]	Name	Reference
Cluster C: Proteins with BTB domain				
59	At1g64280	1	NPR1	Cao et al., 1997
60	At2g41370	1	NPR1	Cao et al., 1997
61	At3g57130	1	NPR1	Cao et al., 1997
62	At4g19660	1	NPR1	Cao et al., 1997
63	At4g26120	1	NPR1	Cao et al., 1997
64	At5g45110	1	NPR1	Cao et al., 1997
65	At2g04740	2	n.a.	n.a.
Cluster D: Protein kinases				
66	At1g14000	1	APK-like	Chinchilla et al., 2003
67	At2g31800	1	Atapk2	Chinchilla et al., 2003
68	At3g58760	1	APK-like	Chinchilla et al., 2003
69	At3g59830	1	Atapk3	Chinchilla et al., 2003
70	At4g18950	1	APK-like	Chinchilla et al., 2003
71	At2g43850	2	Atapk1	Chinchilla et al., 2003
72	At5g13530	3	n.a.	n.a.
Cluster E: Zinc-finger proteins				
73	At2g40140	1	n.a.	n.a.
74	At2g41900	1	n.a.	n.a.
75	At3g55980	1	n.a.	n.a.
76	At5g12850	1	n.a.	n.a.
77	At5g58620	1	n.a.	n.a.
78	At3g28880	2	n.a.	n.a.
Cluster F: Potassium channels				
79	At2g25600	1	AKT	Pilot et al., 2003
80	At2g26650	1	AKT	Pilot et al., 2003
81	At3g02850	1	AKT	Pilot et al., 2003
82	At4g22200	1	AKT	Pilot et al., 2003
83	At4g32500	1	AKT	Pilot et al., 2003
84	At5g37500	1	AKT	Pilot et al., 2003
Cluster G: Ring Finger proteins				
85	At3g23280	1	n.a.	n.a.
86	At4g14365	1	n.a.	n.a.
87	At5g07270	2	n.a.	n.a.
88	At5g57740	2	n.a.	n.a.
89	At2g28840	3	n.a.	n.a.
Cluster H: ARF GTPase-activating domain-containing protein				
90	At1g10870	1	n.a.	n.a.
91	At1g60860	1	n.a.	n.a.
92	At5g13300	1	n.a.	n.a.
93	At5g61980	1	n.a.	n.a.
Cluster I: Calmodulin binding motif-containing protein				
94	At1g67310	1	CAMTA	Bouché et al., 2002
95	At2g22300	1	CAMTA	Bouché et al., 2002
96	At5g09410	1	CAMTA	Bouché et al., 2002
97	At5g64220	1	CAMTA	Bouché et al., 2002
Cluster J: Acyl-CoA binding protein				
98	At4g27780	1	ACBT	Chye et al., 2000
99	At5g53470	1	ACBT	Chye et al., 2000
Cluster K: Chromodomain protein				
100	At2g47450	1	CAO	Klimyuk et al., 1999
Cluster L: Helicase				
101	At1g06670	1	DEAH	Isono et al., 1999

Table 2 (*continued*)

EN	Atg no.	Fa.[a]	Name	Reference
Cluster M: Protein with RCC-1 domains				
102	At3g03790	1	n.a.	n.a.
Cluster N: Protein with tetratricopeptide repeats				
103	At3g04710	1	n.a.	n.a.
Cluster O: Protein with pH domain				
104	At5g14230	1	n.a.	n.a.
Cluster P: Protein with ATPase associated motif				
105	At3g24530	1	n.a.	n.a.

[a] Families inside each cluster based on sequence similarity.
[b] Proteins that only have ankyrin repeats but by sequence similarity they are grouped with proteins containing additional domains.

convenience, we assigned an ‘‘entry number’’ (EN) for each protein.

3.2. Determination of a consensus sequence of arabidopsis ankyrin repeats

A consensus sequence was determined based on the alignment of the sequences of the 509 *Arabidopsis* ANK repeats (Table 3). None of the positions was completely conserved but we identified 7 residues with at least 50% conservation across all repeats. Five of them are located between positions 2 and 9, which is the region of overall higher conservation. All these residues have been previously reported as highly conserved based largely on animal proteins and the presence of strictly conserved hydrophobic and hydrophilic positions has also been observed (Michaely and Bennett, 1992; Bork, 1993).

3.3. Phylogenetic and structural analysis of ankyrin-containing proteins in Arabidopsis

The 105 ANK repeat containing proteins were analyzed for the presence of additional recognizable protein motifs and 16 clusters of structurally related proteins were identified (Table 4). Protein sequences of each cluster were aligned and a similarity tree was generated (not shown), giving rise to a classification in families of sequence related proteins. Thirty seven *Arabidopsis* genes code for proteins containing ANK repeats and transmembrane domains (cluster A) whereas 21 code proteins with only ANK repeats as recognizable motif (cluster B). Genes EN26, EN38 and EN39 code for proteins structurally belonging to cluster B, however, their sequences are more related to some proteins in cluster A. They probably correspond to truncated forms of proteins of the cluster A and we include them in this group.

Among the 105 proteins isolated only 31 have been previously characterized to any extent. Clusters with higher number of genes are those with a lower proportion of studied genes. Only one of the 40 genes in cluster A has

Table 3
Consensus sequence of *Arabidopsis* Ankyrin repeats

aa[a]	*Arabidopsis*		Consensus sequence by	
	Consensus	Most abundant amino acids	Bork, 1993	Michaely and Bennett, 1992
1	Hyphyl.	D 21%	–	–
2	G	G 63%	Turn-like or polar	G
3	Hyphyl.	–	–	–
4	T	T 56%	T or S	T
5	P, A	P 37%; A 29%	P	A, P
6	L	L 77%	L	L
7	H	H 64%	H	H
8	Hyphob.	L 23%	Hydrophobic	L, I, V
9	A	A 84%	A	A
10	A	A 48%; V 20%	Hydrophobic	A, S
11	Hyphyl.	–	–	R, Q, K
12	Hyphyl.	–	–	–
13	G	G 62%	Turn-like or polar	G, N
14	Hyphyl.	H 24%	Turn-like or polar	H, N
15	–	–	–	V, L, T
16	Hyphyl.	E 29%	Turn-like or polar	E, D
17	Hyphob.	V 23%; I 20%	hydrophobic	V, I, M
18	V	V 45%	hydrophobic	V, A
19	Hyphyl.	K 27%	Turn-like or polar	K, E, R
20	–	L 18%	–	L, V
21	L	L 63%	L	L
22	L	L 46%	L	L
23	Hyphyl.	E 19%	Turn-like or polar	D, K, Q, E
24	Hyphyl.	–	–	–
25	–	G 30%	Turn-like or polar	G
26	–	A 24%; P 18%	–	A
27	Hyphyl.	D 20%	–	D, N, S
28	–	L 19%	–	V, P, I
29	–	–	–	N, D
30	–	–	–	A
31	–	–	–	–
32	Hyphyl.	D 19%	–	T, D, N
33	Hyphyl.	N 19%	–	K

The consensus contains a single amino acid when it represents alone more than 40% and amino acid class when it represents more than 60%. Hyphen means no special amino acid or amino acid type in that position. Most abundant amino acids are only indicated if they represent 18% or more of the total. Percentages refer to the 509 ankyrin repeats identified in this study.

[a] Amino acid position.

been recently characterized (Lu et al., 2003). For this reason we decided to study this family in a greater depth in Section 3.4.

3.3.1. Proteins with only ankyrin repeats

Twenty one *Arabidopsis* genes encode for proteins with ANK repeats as the only recognizable motif (Table 2): families B1 to B10 and genes EN26, EN38 and EN39. The size of the encoded proteins range from 144 to 664 amino acids and the number of ANK repeats from 2 to 10. In the most extreme case, ANK repeats comprise 87% of the protein (EN57), in other cases ANK repeat arrays are concentrated in the N- or C-terminal regions. Four of these genes have been previously characterized. Two of them (EN46 and EN47) code proteins similar to tobacco ANK1 and are involved in pathogen defense (Peck et al., 2001; Yan et al., 2002) and the roles of EN48 and EN49 are related to embryogenesis and development (Zhang et al., 1992; Albert et al., 1999).

3.3.2. BTB domain cluster

Analysis of the *Arabidopsis* genome revealed seven genes that encode proteins with a BTB domain and ANK repeats. The BTB domain, also known as POZ (poxvirus and zinc finger), is known to be a protein–protein interaction motif found at the N-terminus of several C2H2-type transcription factors as well as Shaw-type potassium channels (Bardwell and Treisman, 1994). They are divided into two families (C1 and C2) with a different order in the domains. The C1 family contains six proteins in which the BTB domain is located in the N-terminus and the ANK repeats in the C-terminus. These genes encode proteins similar to NPR1, a protein involved in the control of the onset of systemic acquired resistance to a broad spectrum of pathogens (Cao et al., 1997). Proteins similar to NPR1 have been observed in other plant species such as rice, *Brassica* and tobacco, but any protein with a similar domain structure has been found in animals or fungi. The C2 family contains one gene coding for a protein with the domains in the reverse order to the C1 proteins. Proteins with similar domain organization have been observed in animals and fungi, but their functions are unknown except for a human elongation factor 1A binding protein (Unoki and Nakamura, 2001).

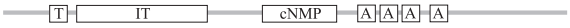
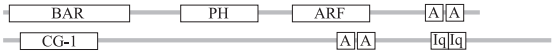
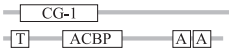
3.3.3. Protein kinases

Seven genes code proteins with ankyrin repeats and protein kinase domains, divided into three families. In D1 (five genes) and D2 (one gene) families, the ANK repeats are located in the N-terminus and the kinase domain in the C-terminus. D1 proteins are similar to the *Medicago* APK ankyrin protein kinase (Chinchilla et al., 2003). Structural homologues of D1 and D2 proteins also exist in animals, for example, the human cardiac ankyrin repeat kinase (Acc. NM_015978). The D3 family contains a single gene (EN72) that codes a protein with an N-terminal RING finger domain, a kinase domain in the middle and C-terminal ANK repeats. The gene EN66 seems to encode a tyrosine kinase, but the kinase specificity of the other proteins is not known.

3.3.4. Zinc-finger proteins

Six genes code proteins with ANK repeats and zinc-finger domains. They are divided into two families (E1 and E2). Family E1 contains five genes coding proteins with short arrays of two or three ANK repeats at the N-terminus and one or two zinc-fingers in the central part of the protein.

Table 4
Clusters of ANK containing Proteins in *Arabidopsis*

Class	Description	Number of genes	Scheme of a representative protein
A	Proteins with transmembrane domains	40	
B	Proteins only with ankyrin repeats	18	
C	Proteins with ankyrin and BTB domain	7	
D	Kinases	7	
E	Proteins with Zinc finger	6	
F	Potassium channel proteins	6	
G	Proteins with Ring Finger	5	
H	Proteins with ARF GTPase-activating domain	4	
I	Proteins with calmodulin binding motif	4	
J	Proteins with Acyl CoA binding domain	2	
K	Proteins with chromodomain	1	
L	Helicase	1	
M	Regulator of chromosome condensation	1	
N	Protein containing Tetratricopeptide repeats	1	
O	Proteins with PH domain	1	
P	Proteins with ATPase associated domain	1	

A ankyrin repeat; T transmembrane; BTB, Broad-Complex, Tramtrack and Bric a brac; KIN, protein kinase domain; IT, ion transport protein domain; Z, zinc-finger domain; cNMP, cyclic nucleotide-monophosphate binding domain; R, RING finger; BAR, BAR domain; PH, Pleckstrin homology domain; ARF, Putative GTP-ase activating proteins for the small GTPase; CG-1, CG-1 domain; Iq, Short calmodulin-binding motif containing conserved Ile and Gln residues; ACBP, Acyl CoA binding protein domain; C, chromodomain; R, putative single-stranded nucleic acids-binding domain; DEXD, DEAD-like helicases superfamily; HE, helicase superfamily c-terminal domain; HA, helicase associated domain; R, regulator of chromosome condensation; Tt, tetratricopeptide repeats; AAA, ATPase associated domain.

Similar proteins are present in rice. The E2 family contains one gene that codes a protein with similar domains but having an array of six ANK repeats. The functions of none of these proteins have been determined.

3.3.5. Potassium channels

The first plant protein described containing an ANK repeat was AKT1 (EN80) which codes a protein similar to a shaker-like K^+ channel located in the plasma membrane (Sentenac et al., 1992). Shaker potassium channels play an important role in the uptake of K^+ from the soil. The *Arabidopsis* genome contains nine genes coding Shaker potassium channels and six of them contain ANK repeats (Pilot et al., 2003). Plant Shaker channels share a common structure: a hydrophobic core composed of six transmembrane segments, a long cytoplasmic C terminal region containing a putative cyclic nucleotide binding domain, and a KHA domain. Many channels, but not all, also contain ANK repeats between the putative cyclic binding domain and the KHA domain. Similar proteins have been described in many other plant species including dicots and monocots (Pilot et al., 2003).

3.3.6. Ring finger proteins

There are six *Arabidopsis* genes encoding proteins with RING domains and ANK repeats. One of them also has a kinase domain and for this reason has been included in the family D3 (EN72). The remaining five contain four to six ANK repeats at the N-terminus and a RING finger at the C-terminus. They are divided into three families (G1–G3). Similar proteins have been found in rice and in animals, but the functions are unknown.

3.3.7. ARF GTPase-activating domain-containing protein

Four *Arabidopsis* genes with a similar complex organization were identified (Family H1), from N to C terminus: a BAR domain, a PH domain, a GTPase activating domain and two to three ANK repeats. None of these genes have been studied in any plant species.

3.3.8. Calmodulin binding motif-containing protein

Four *Arabidopsis* genes (family I1) code proteins with a similar organization: an N-terminal CG-1 domain, two or three ANK repeats in the central region, and two calmodulin binding motifs in the C-terminus. They were named CAMTAs (Calmodulin-binding transcription activators) and have been described in *Arabidopsis* and other plant species (Bouché et al., 2002). CG-1 domains are highly conserved with about 130 amino acid residues containing a predicted bipartite NLS and named after a partial cDNA clone isolated from parsley encoding a sequence-specific DNA-binding protein (da Costa e Silva, 1994). CG-1 domains are associated with CAMTA proteins.

3.3.9. Acyl-CoA binding protein

Two different genes encoding cytosolic acyl-CoA-binding proteins were identified (family J1). These proteins also contain ANK repeats in the C-terminal region (Chye et al., 2000) and a transmembrane motif at the N-terminus.

3.3.10. Chromodomain protein

The *Arabidopsis* CAO gene (chlorophyll a/b binding protein harvesting-organelle specific protein) codes a protein with ANK repeats and chromodomains (Klimyuk et al., 1999). This is a nuclear gene encoding a chloroplast signal

recognition particle, which is part of a protein complex. The ANK repeats are necessary for the formation of the complex.

3.3.11. Helicase

Gene EN101 codes a protein similar to the DEAH family of RNA/DNA helicases (Isono et al., 1999). The protein contains two ANK repeats.

3.3.12. Other proteins

Four more genes coding proteins with ANK repeats have been found. They also contain some other recognizable protein motifs such as four RCC-1 domains (EN102), three tetratricopeptide repeats (EN103), a PH domain (EN104) and an ATPase associated motif (EN105). None of these, or similar proteins, have been characterized in plants and their functions are unknown.

3.4. Ankyrin-transmembrane proteins in Arabidopsis. Phylogenetic analysis and genome distribution

Thirty-seven proteins in cluster A contain four to eleven N-terminal ANK repeats and two to five C-terminal transmembrane domains (Table 2). These proteins were named AtANKTM proteins (*Arabidopsis thaliana* ankyrin transmembrane). Three additional genes (EN26, EN38 and EN39) code for proteins without transmembrane domains but with sequence similarity to the ANK repeat region of the AtANKTM proteins.

The predicted amino acid sequences of the ANKTM proteins were aligned and a similarity tree constructed. Sequence comparison demonstrates that they are divided into six distinct families named 1–6. Similar results were obtained when the analysis was performed using only the N-terminal ankyrin containing part of the proteins (Fig. 1A) or only the C-terminal transmembrane containing part. Based on the alignments, a consensus protein model for each family is shown in Fig. 1B. Some of the ANK repeats are present in most or all the proteins of the family but some are lacking in some proteins due to insertions, deletions or single mutations. The position and number of transmembrane domains is conserved in all proteins of the same family except for the three truncated proteins.

We analyzed the intron distribution of all the AtANKTM genes reported here (data not shown). Only one gene has no introns. The great majority of introns (89 of 96) are located in the ankyrin coding regions of the genes. The position of the introns is not correlated with the position of the ANK repeats as has been observed for some mammalian and plant genes (Albert et al., 1999). Forty-one of the introns interrupt the region coding for an ANK repeat. We analyzed whether the intron/exon number and distribution patterns are related to the phylogenetic distribution in families. Although some introns are present in more than one gene of the same family, in general, their position and number are not conserved.

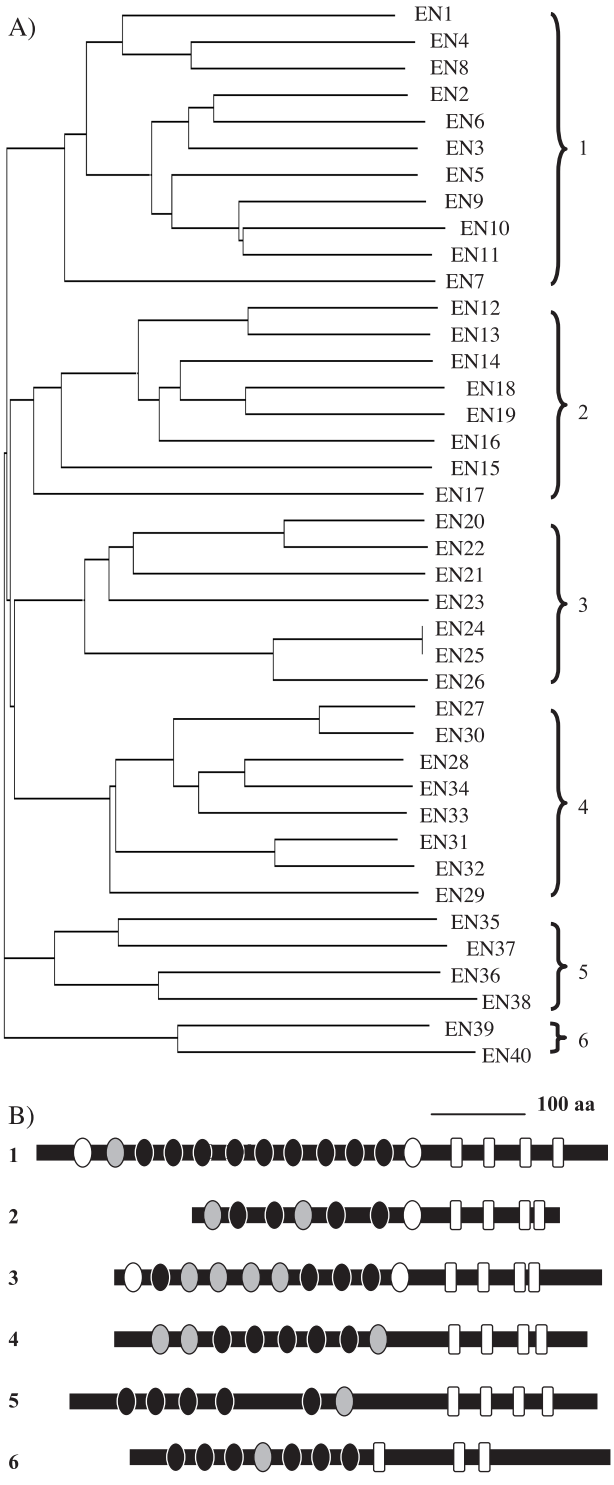


Fig. 1. Families of *Arabidopsis* proteins containing Ankyrin repeats and transmembrane domains (ANKTM). (A) Neighbor-joining tree of the N-terminal region of the ANKTM proteins containing the ankyrin repeats. Numbers in the right indicate the different families. (B) Schematic representation of the hypothetic consensus proteins of each of the six ANKTM protein families. Circles represent ankyrin repeats. Black circles represent ankyrin repeats present in >90% of the proteins of the family, grey circles present in 50–89% of the proteins and white circles present in <50%. Empty rectangles represent transmembrane domains.

AtAnkTm genes are distributed in all chromosomes although not uniformly (Fig. 2). Whereas chromosome II and III contain four genes, chromosome V contains 13. There are some areas with a high density of genes such as at the bottom of chromosome V and the top of chromosome I. Conversely, there are large regions that are devoid of *AtAnkTm* genes, including the bottom of chromosomes I and IV. There are five cases of two ore more genes arranged in tandem (Fig. 2). Three of them correspond to couples of genes (EN10 and EN11, EN12 and EN13, and EN18 and EN19), one include three genes (EN24, EN25 and EN26) and finally, there are seven genes arranged in tandem in chromosome IV (EN2 to EN8). All these genes are closely related in the phylogenetic analysis (Fig. 1A) with the exception of the group of seven genes which are not so closely related although they belong to the same family.

Analysis of the inter- and intrachromosome duplicated areas of the *Arabidopsis* genome indicates that there are three cases of correlation between gene localization and genome duplications. The gene EN27 in chromosome I seems to be duplicated in chromosome II (EN30). Accordingly, genes EN27 and EN30 are closely related in the phylogenetic analysis (Fig. 1A). Genes EN28, EN33 and EN34 are located in an area repeated three times in the *Arabidopsis* genome, twice in chromosome V and one in chromosome I. These genes are also closely related according to the phylogenetic analysis. Finally, gene EN1 is located in a region of chromosome I duplicated in the area of chromosome IV containing the group of seven *AtAnkTm* genes in tandem. In the phylogenetic analysis EN1 is closely related to EN4 and EN8.

3.5. Expression analysis of the AtAnkTm genes

Total RNA was extracted from different *Arabidopsis* organs (roots, leaves, flowers, stems, caulinar leaves and siliques at three different stages of development) and analyzed by semiquantitative RT-PCR using pairs of primers specific to 22 of the *AtAnkTm* genes (Table 1) (Fig. 3). Controls for DNA contamination, RNA integrity and equalization of the quantities of cDNA are described in Section 2. The actin gene was expressed at similar levels in all organs and the PCR amplification using *AtEm6* primers was according to the previous data (a gene specifically expressed in seed maturation) (Vicient et al., 2000). On the other hand, databases were screened for *Arabidopsis* ESTs corresponding to the *AtAnkTm* genes (Table 5). We can not observe any clear correlation between gene family and organ or stress-response transcription specificity.

RT-PCR amplification gave bands for 13 of the genes in at least one of the samples. Primers for genes EN25 and EN32 produced bands in all the organs tested except for mature siliques. EN25 is highly expressed in leaves and stem and an EST was found corresponding to an ‘‘above-ground tissues’’ library. Gene EN32 is also expressed in all organs but is highly expressed in leaves. Primers for gene EN29 produced amplification in all tissues except roots and in a low level in stems. Several ESTs from different tissues were found in databases corresponding to this gene. Other genes from which we detected amplification in several organs are EN12, EN16, EN18, EN19, EN22 and EN33.

Some of the genes seem to have more specific patterns of expression. For example, primers for genes EN14 and EN17 amplified only in roots and primers for gene EN11 only from leaves. The pattern of expression of EN11 is consistent with the expression pattern reported by Lu et al. (2003) for the ACD6 gene.

RT-PCR did not amplify bands using eight pairs of primers (genes EN2, EN3, EN6, EN9, EN13, EN15, EN28, EN31 and EN34). Accordingly, there are not corresponding ESTs for genes EN6, EN13, EN15 and EN34. ESTs corresponding to genes EN2, EN3 and EN9 came from stressed tissues, conditions that we did not tested. Gene EN3 has a corresponding EST from root and

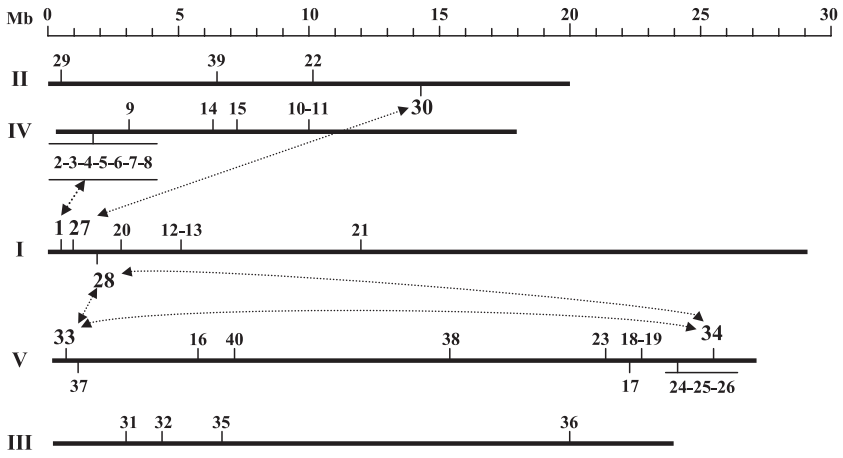


Fig. 2. Chromosomal distribution and duplication events for *Arabidopsis AnkTm* genes. Deduced chromosomal positions of the *AtAnkTm* genes are indicated by EN. The scale is in Megabases (Mb). Numbers separated by hyphens represent genes arranged in tandem. Connecting pointed lines mark a correlation between duplicated genomic region and the presence of *AnkTm* genes.

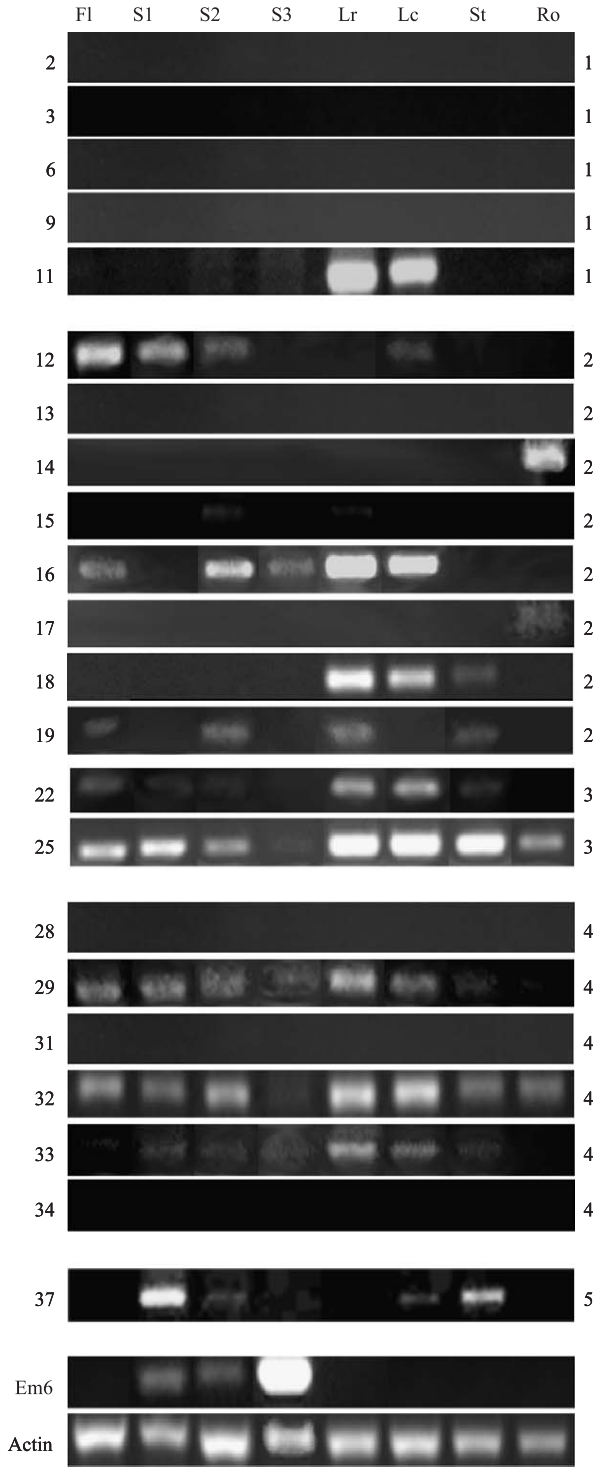


Fig. 3. RT-PCR analysis of the expression profiles of 22 *Arabidopsis* genes encoding ANKTM proteins. Ethidium bromide-stained 1.5% agarose gels showing RT-PCR products. The EN number of the corresponding genes is shown in the left. The corresponding gene family is shown on the right. RNAs used were extracted from flowers (Fl), immature siliques 1–7 dap (S1), intermediate siliques 8–14 dap (S2), mature siliques 14–21 dap (S3), rosette leaves (Lr), caulinar leaves (Lc), stems (St) and roots (Ro). Total RNA (2 μg) was used to synthesize cDNA. A fraction (1/20) of the synthesized cDNA was used to amplify gene transcripts by PCR. The bands shown corresponded to the expected size in each case. *Actin* and *AtEm6* genes were used as controls.

Table 5
Summary of information on ESTs of the *AtAnkTm* genes

EN	F[a]	ESTs. Accession Number and origin
1	1	None
2	1	AV798477; dehydration and cold
3	1	AI997466; root
		BE662952; ozone
4	1	None
5	1	None
6	1	None
7	1	None
8	1	None
9	1	AV793119, cold
10	1	None
11	1	AI998090; leaf
		AV440557; aboveground
		AI100255, H37081, T42092, T46084; Pooled
		BE662757; ozone
		AV822072, AV793013; cold
		AV794986, AV826452, AV830548; dehydration
		AV813132, AV815244, AV807835; dehydration and cold
		AU231325; various stresses
12	2	AV549133, AV537342; root
13	2	None
14	2	None
15	2	None
16	2	AU230590, AU239298; silique and flower
		BE038658; salt
17	2	AI996466; root
18	2	AU229652; AU238464; dehydration
19	2	None
20	3	AV798019, AV816532, AV827371; dehydration and cold
		BE662714; ozone. BU635941, CB074318; infected leaf
21	3	None
22	3	AI996591; root. AV825484, AV791380; cold
		AV795644; dehydration. BE662700, BE662788; ozone
23	3	None
24	3	AU237468, T45361; pooled
25	3	AV522917; aboveground. BE662949; ozone
26	3	AV528599; aboveground
		BF381542; ozone
27	4	None
28	4	AI996003; inflorescence
29	4	AA728512; inflorescence. AI996553; root
		AV557218, AV558991, AV561448; green siliques
		AV439689, AV441879; aboveground
		AI993199, R64963, T43424, T46195; pooled
		AV784945, AV787218, AV823873; dehydration and cold
30	4	CB185927; infected leaf
31	4	T04286, AA585945; pooled
32	4	H36055; pooled
33	4	AV534234; flower buds. AV782571, AV821976; cold
		AV803115, AV815435; dehydration and cold
34	4	None
35	5	None
36	5	None
37	5	AV564960; green siliques
38	5	None
39	6	N96603; pooled
40	6	R90543, AA605447, AI994371; pooled
		AV519957; aboveground
		BE521378, BE521377, AV808021, AV814010; mature seed
		AU227998, AU236986; silique and flower
		AV566955; green siliques. AV539138; roots

[a] Family.

EN28 from inflorescence. This apparent inconsistencies could be due to low levels of expression. For example, it was possible to amplify a very low intensity band corresponding to gene EN28 in flowers and young siliques using higher quantities of cDNA template and a high number of PCR cycles (data not shown).

4. Discussion

We identified a total of 509 ANK repeats in the *Arabidopsis* proteins coded by 105 genes, which represent 0.4% of the total *Arabidopsis* genes. This number is higher to a previous estimation of 0.25%, but is similar to the percentages estimated for humans, *Drosophila* and *C. elegans* (Jebanathirajah et al., 2002). Few of the amino acids in the ANK repeats are well conserved (Table 3). The use of more precise criteria to identify ANK repeats allow us to recognize many previously nonannotated repeats. Evaluating sequence conservation it becomes obvious that the terminal repeats in the arrays deviate more from the general consensus than those located centrally and the same was observed in animal proteins (Bork, 1993).

Although few of the 33 amino acids that compose the ANK repeats are conserved, several strictly conserved hydrophobic positions can be observed either in *Arabidopsis* and animals. This conservation is necessary to maintain the secondary structure that is essential for its function in protein–protein interactions (Bork, 1993; Rhode and Bork, 1993). Few plant proteins containing ANK repeats have been characterized and any experimental data demonstrates the role of the ANK repeats in plants. However, the conservation of the strictly hydrophobic positions in *Arabidopsis* suggests that ANK repeats may have similar functions in plants and animals. Many plant proteins containing ANK repeats are multidomain molecules in which ANK repeats are combined with other unrelated structural modules. The presence of ANK repeats on so diverse proteins makes a common function such as an enzymatic activity extremely unlikely and supports the idea that ANK repeats are involved in mediate protein–protein interactions also in plants.

The most abundant group of structurally similar ANK repeat-containing proteins counts for 40 elements: 37 proteins containing ANK repeats and transmembrane domains (Fig. 4), and three more that may represent truncated forms containing only the N-terminal ANK repeats. The 40 *AnkTm* genes are divided into six families. The evolution of these gene families has been complex. *AtAnkTm* genes are distributed all over the genome but not uniformly. This type of distribution seems to be common to other gene families and, for example, is similar to the distribution of the bHLH genes, although the regions of high and low density are not the same in both cases (Toledo-Ortiz et al., 2003). Inter- and intrachromosome and tandem array duplications have certainly plaid a role in amplify the number of *AtAnkTm* genes.

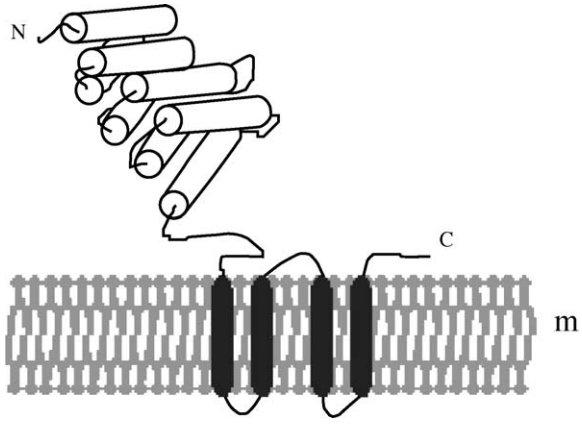


Fig. 4. Schematic representation of the *Arabidopsis* ANKTM proteins. Hypothetic representation of an ANKTM protein and the associated membrane (m). N and C ends of the protein are indicated by letters. Grey cylinders are the ankyrin repeats and black cylinders the transmembrane domains.

Sixteen of the *AtAnkTm* genes are located in different tandem arrays. Sets of genes organized in tandem arrays are very common in *Arabidopsis*.

The interaction specificity of the ANK repeats seems to be determined by its amino acid sequence (Bennett, 1992). When comparing AtANKTM families, the sequences of the ANK repeats are not well conserved between families but highly conserved within the same family. This conservation suggests that AtANKTM proteins of the same family may interact with the same or similar proteins. The variability in the expression patterns of genes of the same family indicates that although they could interact with similar proteins, their roles may not be redundant. Some of the *AtAnkTm* genes are widely expressed suggesting that they may have pleiotropic or general functions, whereas others have a more restricted expression and perhaps more specific functions. Only one of the *AtAnkTm* genes have been studied previously (*ACD6,* EN11) and codes a protein involved in salicylic acid signaling in defense responses (Lu et al., 2003). A mutant of this gene shows spontaneous cell-death and increased disease resistance. The function of this protein at the molecular level remains unknown. Proteins with similar domain organization have been found in other plant species (monocot and dicot) but their functions also remain unknown. Similar domain organization have been found also in some animal proteins whose functions are known and can suggest some roles for the *AtAnkTm* genes: (a) Membrane receptors, as ANKTM1, a menthol- and cold-activated channel (Story et al., 2003) and the human vanilloid receptor; (b) membrane channels, as the human CaT1 and CaT2 calcium entry channels (Peng et al., 2001), the OTRPC4 cation channel or the transient receptor potential channel present in taste receptor cells; or (c) membrane anchorage proteins that attach other proteins to the membrane in a similar role as ankyrin protein does (Bennett, 1992). Any of the three possibilities is compatible with the function of ACD6 (Lu et al., 2003).

Proteins of *AtAnkTm* family 2 have certain sequence similarity with the maize TM20 protein in the region of the transmembrane domains. TM20 is a protein necessary for normal embryo development and contains twenty hydrophobic segments that can be grouped in five repeats formed by four segments (Stiefel et al., 1999). A possible function of TM20 is to act as an auxin membrane transporter (T Jahrmann, personal communication). In *Arabidopsis*, no gene coding for a protein with 20 transmembrane domain is present. A possibility could be that in *Arabidopsis* the function of the maize TM20 protein is carried out by a complex of AtANKTM proteins bound by the ANK repeats.

Mutant analysis, double-hybrid assays and cell localization experiments will give us the necessary clues to understand the functions and molecular interactions of AtANKTM proteins.

Acknowledgements

C.B. was the recipient of a fellowship of the Universitat Autonoma de Barcelona - Fundación Presidente Allende. C.M.V. is recipient of a ‘‘Ramon y Cajal’’ contract from the Spanish Ministry of Science. This work has been carried out thanks to the grant BIO2001-1721 from Plan Nacional de Investigación Científica y Técnica, to a grant from programme MAZE (European Union) and was done within the framework of Centre de Referència de Biotecnologia de la Generalitat de Catalunya.

References

Albert, S., Despres, B., Guilleminot, J., Bechtold, N., Pelletier, G., Delseny, M., Devic, M., 1999. The EMB 506 gene encodes a novel ankyrin repeat containing protein that is essential for the normal development of *Arabidopsis* embryos. Plant J. 17, 169–179.

Bardwell, V.J., Treisman, R., 1994. The POZ domain: a conserved protein–protein interaction motif. Genes Dev. 8, 1664–1677.

Bennett, V., 1992. Ankyrins: adaptors between diverse plasma membrane proteins and the cytoplasm. J. Biol. Chem. 267, 8703–8706.

Bork, P., 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? Proteins 17, 363–374.

Bouché, N., Scharlat, A., Snedden, W., Bouchez, D., Fromm, H., 2002. A Novel family of calmodulin-binding transcription activators in multicellular organisms. J. Biol. Chem. 277, 21851–21861.

Cao, H., Glazebrook, J., Clarke, J.D., Volko, S., Dong, X., 1997. The *Arabidopsis* NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. Cell 88, 57–63.

Chinchilla, D., Merchan, F., Megias, M., Kondorosi, A., Sousa, C., Crespi, M., 2003. Ankyrin protein kinases: a novel type of plant kinase gene whose expression is induced by osmotic stress in alfalfa. Plant Mol. Biol. 51, 555–566.

Chye, M.L., Li, H.Y., Yung, M.H., 2000. Single amino acid substitutions at the acyl-CoA-binding domain interrupt 14[C]palmitoyl-CoA binding of ACBP2, an *Arabidopsis* acyl-CoA-binding protein with ankyrin repeats. Plant Mol. Biol. 44, 711–721.

da Costa e Silva, O., 1994. CG-1, a parsley light-induced DNA-binding protein. Plant Mol. Biol. 25, 921–924.

Isono, K., Yamamoto, H., Satoh, K., Kobayashi, H., 1999. An *Arabidopsis* cDNA encoding a DNA-binding protein that is highly similar to the DEAH family of RNA/DNA helicase genes. Nucleic Acids Res. 27, 3728–3735.

Jebanathirajah, J.A, Peri, S., Pandley, A., 2002. Toll and interleukin-1 receptor (TIR) domain-containing proteins in plants: a genomic perspective. Trends Plant Sci. 7, 388–391.

Klimyuk, V.I., Persello-Cartieaux, F., Havaux, M., Contard, P., Schuenemann, D., Meierhoff, K., Gouet, P., Jones, J.D.G., Hoffman, N.E., Nussaume, L., 1999. A chromo protein encoded by the *Arabidopsis* gene CAO is a plant specific component of the chloroplast signal recognition particle. Plant Cell 11, 87–99.

Lin, J.H., Makris, A., McMahon, C., Bear, S.E., Patriotis, C., Prasad, V.R., Brent, R., Golemis, E.A., Tsichilis, P.N., 1999. The ankyrin repeat-containing adaptor protein Tvl-1 is a novel substrate and regulator of Raf-1. J. Biol. Chem. 274, 14706–14715.

Lu, H., Rate, D.N., Song, J.T., Greenberg, J.T., 2003. ACD6, a novel ankyrin protein, is a regulator and an effector of salicylic acid signaling in the *Arabidopsis* defense response. Plant Cell 15, 2408–2420.

Michaely, P., Bennett, V., 1992. The ANK repeat: a ubiquitous motif involved in macromolecular recognition. Trends Cell Biol. 2, 127–129.

Mosavi, L.K., Minor Jr., D.L., Peng, Z.Y., 2002. Consensus-derived structural determinants of the ankyrin repeat motif. Proc. Natl. Acad. Sci. 99, 16029–16034.

Peck, S.C., Nühse, T.S., Hess, D., Iglesias, A., Mein, F., Boller, T., 2001. Directed proteomics identifies a pnalt-scpecific protein rapidly phosphorilated in response to bacterial and fungal elicitors. Plant Cell 13, 1467–1475.

Peng, J.B., Brown, E.M., Hediger, M.A., 2001. Structural conservation of the genes encoding CaT1, CaT2, and related cation channels. Genomics 76, 99–109.

Pilot, G., Pratelli, R., Gaymard, F., Meyer, Y., Sentenac, H., 2003. Five-Group distribution of the shaker-like K^+ channel family in high plants. J. Mol. Evol. 56, 418–434.

Rhode, K., Bork, P., 1993. A fast, sensitive pattern-matching approach for protein sequences. CABIOS 9, 183–189.

Sedgwick, S.G., Smerdon, S.J., 1999. The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem. Sci. 24, 311–316.

Sentenac, H., Bonneaud, N., Minet, M., Lacroute, F., Salmon, J.M., Gaymard, F., Grignon, C., 1992. Cloning and expression in yeast of a plant potassium ion transport system. Science 256, 663–665.

Stiefel, V., Becerra, E.L., Roca, R., Bastida, M., Jahrmann, T., Graziano, E., Puigdomenech, P., 1999. TM20, a gene coding for a new class of transmembrane proteins expressed in the meristematic tissues of maize. J. Biol. Chem. 274, 27734–27739.

Story, G.M., Peier, A.M., Reeve, A.J., Eid, S.R., Mosbacher, J., Hricik, T.R., Earley, T.J., Hergarden, A.C., Andersson, D.A., Hwang, S.W., McIntyre, P., Jegla, T., Bevan, S., Patapoutian, A., 2003. ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. Cell 112, 819–829.

Toledo-Ortiz, G., Huq, E., Quail, P.H., 2003. The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell 15, 1749–1770.

Unoki, M., Nakamura, Y., 2001. Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway. Oncogene 20, 4457–4465.

Vicient, C.M., Delseny, M., 1999. Isolation of total RNA from *Arabidopsis thaliana* seeds. Anal. Biochem. 268, 412–413.

Vicient, C.M., Hull, G., Guilleminot, J., Devic, M., Delseny, M., 2000. Differential expression of the *Arabidopsis* genes coding for Em-like proteins. J. Exp. Bot. 51, 1211–1220.

Yan, J., Wang, J., Zhang, H., 2002. An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. Plant J. 29, 193–202.

Zhang, H., Scheirer, D.C., Fowle, W., Goodman, H.M., 1992. Expression of antisense or sense RNA of an ankyrin repeat-containing gene blocks chloroplast differentiation in *Arabidopsis*. Plant Cell 4, 1575–1588.