Cell Stress & Chaperones (2001) **6** (3), 209–218
© Cell Stress Society International 2001
Article no. csac. 2001.315

The J-domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones

Jan A. Miernyk

Plant Genetics Research Unit, US Department of Agriculture, Agricultural Research Service, Curtis Hall, University of Missouri, Columbia, MO 65211, USA

Abstract A total of 89 J-domain proteins were identified in the genome of the model flowering plant *Arabidopsis thaliana.* The deduced amino acid sequences of the J-domain proteins were analyzed for an assortment of structural features and motifs. Based on the results of sequence comparisons and structure and function predictions, 51 distinct families were identified. The families ranged in size from 1 to 6 members. Subcellular localizations of the *A thaliana* J-domain proteins were predicted; species were found in both the soluble and membrane compartments of all cellular organelles. Based on digital Northern analysis, the J-domain proteins could be separated into groups of low, medium, and moderate expression levels. This genomics-based analysis of the *A thaliana* J-domain proteins establishes a framework for detailed studies of biological function and specificity. It additionally provides a comprehensive basis for evolutionary comparisons.

INTRODUCTION

DnaJ was originally characterized from *Escherichia coli* as a 41-kDa heat shock protein (Georgopoulos et al 1980). It was subsequently demonstrated in both biochemical and genetic experiments that DnaJ interacts directly with DnaK and GrpE (Liberek et al 1991; Scidmore et al 1993; Goffin and Georgopoulos 1998), constituting a molecular chaperone machine (Bukau and Horwich 1998; Miernyk 1997, 1999). Additionally, there is evidence that DnaJ can act independently as a chaperone (Laufen et al 1999). DnaJ binds to the adenosine triphosphate (ATP)–ligated form of DnaK and stimulates hydrolysis to adenosine $5'$-diphosphate plus inorganic phosphate (Pi) (Liberek et al 1991).

DnaJ can be envisioned as having a linear, modular sequence consisting of the J-domain, a proximal G/F-domain, and a distal zinc finger (CxxCxGxG)$_4$ domain, followed by less conserved C-terminal sequences (Caplan et al 1993; Silver and Way 1993). The J-domain consists of approximately 75 conserved amino acid residues that comprise 4 α-helices. The invariant tripeptide, HPD, which is both characteristic of and absolutely essential for the biological function of J-domains, is located between helices II and III. The G/F-domain, a sequence rich in Gly and Phe residues, comprises a flexible linker region that helps to convey specificity of interactions among DnaK, DnaJ, and target polypeptides (Wall et al 1995; Yan and Craig 1999). The zinc finger domain is believed to mediate protein-protein interactions among DnaK, DnaJ, and target polypeptides (Banecki et al 1996; Szabo et al 1996).

In recent years, a large number of DnaJ-related proteins have been nonsystematically characterized from a variety of different organisms. The relatively small genome size (The *Arabidopsis* Genome Initiative 2000), coupled with an abundance of well-defined mutants (Koncz et al 1992) and the ease of genetic manipulation (Redei 1975), has resulted in extensive study of *Arabidopsis thaliana* as a model flowering plant (Meinke et al 1998). Herein, a genomics approach to analysis of all the J-domain proteins from the simple flowering plant *A thaliana* is presented.

MATERIALS AND METHODS

Database analysis

The publication version of the *A thaliana* genome sequence was accessed via the San Diego Supercomputer Center (http://www.sdsc.edu/index.html). Using the National Center for Biotechnology Information (NCBI) Web site (http://www.ncbi.nlm.nih.gov/BLAST/), an it-

Accepted 25 April 2001.

Correspondence to: Jan A. Miernyk, Tel: 573-882-8167; Fax: 573-884-7850; E-mail: miernykj@missouri.edu.

erative BLAST search was conducted using the previously cloned J-domain sequences (Zhou et al 1995; Kroczyńska et al 1996, 2000; Lin and Lin 1997; Zhou and Miernyk 1999; Miernyk and Coop 2000) as the search terms.

Nomenclature

Eukaryotic proteins related to DnaJ were initially referred to as DnaJ homologues. As the sequences of an increasing number of divergent proteins accumulated, a more systematic nomenclature became necessary. The initial attempt separated protein sequences into groups I, II, and III (Cheetham and Caplan 1998). Type I sequences would contain the J-, G/F-, and zinc finger domains, type II sequences would have the J- plus either a G/F- or zinc finger domain, and type III sequences have only the J-domain. It was subsequently noted that Roman numerals are not allowed in gene names, and it was proposed that A, B, and C be substituted for I, II, and III (Ohtsuka and Hata 2000). Furthermore, a complete nomenclature was proposed: 2 lowercase letters for the genus and species, Dj, A, B, or C, plus an Arabic numeral, indicating chronology (Ohtsuka and Hata 2000). In this system the first of the *A thaliana* J-domain proteins is designated atDjB1. This nomenclature has been used with one minor modification. The original proposals required the (CxxCxGxG)$_4$ motif, which mediates protein-protein interactions, for type III/C sequences. Herein, any of the well-defined sequence motifs that mediate protein-protein interactions can be substituted for the DnaJ zinc-binding domain (other types of zinc finger sequences, coiled-coil motifs, tetratricopeptide repeat sequences).

J-domain sequence comparisons

Because the primary sequence lengths of the *A thaliana* J-domain proteins varied from as small as 112 to as large as 2535 amino acids in length, it seemed unlikely that overall group comparisons would yield meaningful information. Thus, only the J-domains were compared. The sequences were manually aligned using the absolutely conserved tripeptide HPD as the center point. The alignments were analyzed using SEQBOOT, and a tree constructed using PHYLIP (Felsenstein 1989).

In silico protein localization

Predictions of subcellular location of the *A thaliana* J-domain proteins used the following algorithms to search the J-domain protein amino acid sequences: PSORT (http://psort.nibb.ac.jp/form.html) (Nakai and Horton 1999); TargetP (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al 2000); Predotar (http://www.inra.fr/Internet/Produits/Predotar/index.html); and MITOPROT (http://www.mips.biochem.mpg.de/cgi-bin/proj/medgen/mitofilter) (Claros and Vincens 1996) plus direct visual examination.

Table 1 Relationships between genome size and the compliment of Hsp70 and J-domain chaperone proteins

Organism	Reading frames	Hsp70 proteins	J-domain proteins
Escherichia coli	4288	3	5
Saccharomyces cerevisiae	5885	14	20
Drosophila melanogaster	13 601	14	22
Caenorhabditis elegans	19 099	15	32
Arabidopsis thaliana	25 498	17	89
Mus musculus	~35 000	16[a]	23[a]

[a] Sequencing not yet complete.

Digital analysis of gene expression

The advanced BLAST algorithm was used to search the NCBI dbEST for *A thaliana* J-domain protein sequences (31 014 entries). The number of expressed sequence tag (EST) clones for each of the J-domain proteins was used for digital Northern analysis (Audic and Claverie 1997; Ewing et al 1999). Only EST clones isolated from non-normalized libraries and sequenced from the 5′-end were scored (Audic and Claverie 1997; Mekhedov et al 2000; Ohlrogge and Benning 2000).

RESULTS AND DISCUSSION

The *A thaliana* J-domain proteins

An iterative BLAST search of the publication version of the *A thaliana* genome sequence (The *Arabidopsis* Genome Initiative 2000), using the deduced amino acid sequences of the previously analyzed J-domains as the search terms, unveiled a total of 89 proteins. Analyses aimed at reducing this final number, by exposing potential duplicity arising from errors in sequencing and gene identification, were not productive. The complete annotation of all 89 deduced amino acid sequences can be downloaded from http://www.agron.missouri.edu/ars_columbia/miernyk.html.

The relatively large number of *A thaliana* J-domain proteins was unexpected. Results from analyses of the yeast (Goffeau et al 1996), worm (The *C. elegans* Sequencing Consortium 1998), and fly (Adams et al 2000) genomes suggested a relatively constant relationship between the total number of 70-kDa stress proteins and J-domain proteins (Table 1). Taking into account the presence of the plastid class of organelles in plant cells, the number of 70-kDa stress proteins is similar to those of the other organisms (Lin et al 2001). Even considering what might be anticipated as additional plastidial J-domain proteins, the total number is unexpectedly large.

Table 2 The *Arabidopsis thaliana* J-domain proteins

Protein	Nucleotide accession number	Protein accession number	Chromo-some	Molecular weight (kDa)	J-domain position[a]	Localization[b]	Protein class	Phylo-genetic clade
B1	U16246	S71190	I	47 841	1	mm	1	2
A2	L36113	U64912	I	46 328	1	cyto/mom/pm	2	1
A3	U22340	S71199	III	46 444	1	cyto/mom/pm	2	1
C4	AC016661	AAF23300	I	12 109	1	sp	3	3
C5	AL163002	CAB86070	V	12 064	1	sp	3	3
C6	AF037168	AAB91418	V	32 574	1	nuc	4	2
C7	AC006841	AAD23695	II	38 838	1	cyto	5	2
C8	AF099906	AAF14680	I	18 325	1	ps	6	4
B9	AL161746	CAB81922	V	36 971	1	cyto	7	5
C10	Y11969	CAA72705	I	48 210	1	cyto	8	2
C11	AJ292973	T05496	IV	17 891	1	ps	6	4
B12	AC010871	AAF07844	III	35 797	1	cyto	9	5
B13	Z49238	S58287	II	59 361	1	sp/imp	10	3
B14	AL161572	T04618	IV	38 191	1	cyto	7	5
B15	AF089810	AAD13758	I	45 484	1	nuc	11	9
B16	AC002396	T00641	I	44 024	1	nuc	11	9
C17	AB007648	BAB11179	V	51 763	1	nuc/ps/mm	12	4
C18	AC006931	AAD21732	II	36 121	1	ps/nuc/mm	13	2
B19	AL162651	CAB83110	III	36 816	1	sp	14	1
C20	AF214107	T05252	IV	23 361	1	ps	6	4
C21	AC011717	AAF09056	I	73 935	1	sp/imp	15	4
B22	AC007109	AAD25656	II	37 109	1	cyto	7	5
B23	AC005489	AAD32885	I	38 757	1	cyto	7	5
A24	AL161596	T06102	IV	42 404	1	ps	16	8
C25	AL117386	CAB55698	IV	28 496	2	mim	17	4
A26	AC006592	AAD22362	II	42 163	1	ps/cyto	16	8
C27	AP000607	BAB10965	V	32 549	2	nuc	18	6
C28	AB018107	BAB08321	V	36 198	1	nuc	18	6
C29	AL021960	T04949	IV	73 515	1	sp/imp	15	4
A30	AB017064	BAB11067	V	50 265	1	mm	19	2
C31	AC007018	AAD29061	II	79 023	1	nuc	20	6
B32	AC007258	AAD39315	I	36 612	1	cyto	7	5
C33	AL049658	CAB41145	III	39 027	1	cyto	9	5
C34	AC006053	AAF18620	II	73 097	1	nuc	20	6
C35	AC012190	AAF80653	I	43 924	1	cyto	5	2
A36	AL163002	CAB86083	V	53 690	3	sp	21	5
B37	AP002063	BAB01967	III	30 152	1	cyto	4	2
C38	AB024034	BAB02800	III	12 320	1	ps	6	4
C39	AC005966	AAD14474	I	43 307	1	nuc	11	9
B40	AC020579	AAF31731	I	72 475	1	nuc	22	3
C41	AC003952	T00836	II	17 748	1	sp/ps	6	4
B42	AB006705	BAB09499	V	30 264	2	cyto/ps	23	9
B43	AL137080	CAB68140	III	41 320	1	sp/imp	24	6
B44	AL161551	CAB78959	IV	60 803	1	nuc	25	6
B45	AC002291	AAC00633	I	39 464	1	cyto	5	2
C46	AC010718	AAF04450	I	44 644	1	cyto	8	2
B47	AF007271	T01797	V	182 708	1	nuc	26	6
B48	AB005237	BAB09669	V	32 338	2	sp/imp	24	6
C49	AB023028	BAB10088	V	40 762	2	sp/imp	27	6
C50	AC005882	AAD21421	I	32 105	1	sp/imp	24	8
B51	AB017069	BAB09110	V	48 253	1	nuc	20	6
A52	AC009322	AAD55483	I	53 830	1	sp/ps	28	8
C53	AL161516	T01980	IV	19 685	1	cyto	29	4
A54	AB019230	BAB02706	III	48 133	1	mm/ps	28	8
C55	AC009243	AAF17683	I	31 907	1	mm/ps	29	4
C56	AB012246	BAB09472	V	37 618	2	nuc/ps	30	7
B57	AB023033	BAB10771	V	77 974	2	sp/imp	31	7
C58	AC002986	T01052	I	65 710	2	sp/imp	31	7
C59	AP000600	BAB02987	III	25 641	1	cyto	32	3
B60	AL021749	T04618	IV	38 191	1	pl	7	5
C61	AB009053	BAB10847	V	23 022	1	nuc	33	6
C62	AL161551	T06152	IV	39 165	1	nuc	18	6
A63	AC010871	AAF07843	III	62 494	1	sp	34	9
C64	AL035601	T04742	IV	57 134	1	mm	35	4
B65	AC006069	AAD12700	II	34 557	1	nuc	36	6
B66	AJ007450	CAA07520	I	50 264	3	nuc	37	3
B67	AC002510	T00808	II	105 493	3	nuc	21	3

Table 2 Continued

Protein	Nucleotide accession number	Protein accession number	Chromosome	Molecular weight (kDa)	J-domain position[a]	Localization[b]	Protein class	Phylogenetic clade
C68	AC008148	AAD55512	I	17 340	1	pl	32	3
C69	AC011808	AAG10814	I	44 274	2	sp	31	7
B70	AB025663	BAA97241	V	39 908	1	nuc	38	1
B71	AB025622	BAB08418	V	85 237	1	cyto	39	6
C72	AC004261	T02119	II	32 285	2	sp	40	3
B73	AC009465	AAD57012	III	129 903	1	nuc	41	6
C74	AB008270	BAB10192	V	15 017	1	cyto	42	2
C75	AC005314	AAC36167	II	65 132	1	nuc	43	6
C76	AL021768	T06151	IV	34 401	1	nuc	44	6
C77	AC011623	AAF08586	III	75 919	1	nuc	45	6
B78	AC004512	T02350	I	68 553	2	nuc	46	6
B79	AC008153	AAF00659	III	75 807	1	nuc	47	3
B80	AL049640	T06635	IV	100 307	3	nuc	37	7
C81	AC007727	AAD41419	I	56 910	3	cyto	37	7
C82	AF096371	T01950	IV	65 097	3	nuc	48	3
B83	AC013258	AAF21067	I	71 601	3	nuc	37	3
C84	AL050351	T08563	IV	38 462	1	cyto	5	2
C85	AC073506	AAG30962	I	50 706	3	cyto	48	3
B86	AC005936	AAC97214	II	68 333	3	sp/imp	48	7
B87	AC008262	AAF27053	I	76 377	3	nuc	49	3
B88	AC005168	T02646	II	277 102	2	sp/imp	50	1
B89	Z99708	CAB16841	IV	163 974	3	nuc	51	7

[a] J-domain position: 1, the N-terminal third; 2, the middle third; and 3, the C-terminal third of the deduced amino acid sequence.

[b] Localization: mm, mitochondrial matrix; mom, mitochondrial outer membrane; pm, peroxisomal membrane; pl, peroxisomal lumen; sp, secretory pathway; nuc, nucleus; cyto, cytoplasm; ps, plastidial stroma; and imp, integral membrane protein.

The 89 genes were distributed among the *A thaliana* chromosomes; I, 27; II, 14; III, 13; IV, 16; and V, 19 (Table 2). Of the 89 reading frames, 9 encode type A J-domain proteins, 35 type B, and 45 type C. There is no obvious position or location relationship among the genes, nor are there any structural or functional grouping of the encoded proteins.

Protein families

The deduced amino acid sequences of the *A thaliana* J-domain proteins were subjected to an extensive set of analyses to reveal the occurrence of structural and functional domains and motifs. The results from these analyses were a major consideration for division of the total into a series of families. The results from BLAST searches of the genomes of other organisms also played a role in defining the protein families. In total, 51 families were distinguished (http://www.agron.missouri.edu/ars_columbia/miernyk.html) (Table 2). The families vary in size from 1 to 6 members. Both median and mode values for the population are 1.

Several of the *A thaliana* J-domain protein families are distinguished by the relationship of the members with proteins previously characterized from other organisms. Family 2 (proteins A2 and A3), family 5 (C7, C35, B45, and C84), and family 15 (C21 and C29) are orthologous with the *Saccharomyces cerevisiae* YDJ1, DJP1/CAJ1, and SEC63 families, respectively. Families 7 (B9, B14, B22, B23, B32, and B60) and 9 (C12 and C33) contain paralogous members of the Hsp40 superfamily. Family 14 includes a single protein, atDjB19, related to members of the *Homo sapiens* hDj9/ERj3/HEDJ family. The family 21 (A36 and B37), 24 (B43, B48, and C50), and 32 (C59 and C68) proteins are orthologous with members of the mouse mmDj11, mmDj10, and mmDj6 families, respectively (Ohtsuka and Hata 2000). Families 37 (B66, C80, C81, C83) and 48 (C82, C85, C86) comprise 2 variants of the auxilin family of J-domain proteins (Quesada et al 1997; Umeda et al 2000). The atDjB88 protein, the only member of family 50, is very large, at 2535 amino acid residues. Initially, it seemed likely that a stop codon had been missed during gene annotation. Recently, however, an almost equally large homolog, RME-8, was identified from *Caenorhabditis elegans* (Zhang et al 2000).

A few of the J-domain protein families have been defined during genetic studies of *A thaliana* or other plants. The atDjB13 protein, the sole member of family 10, was identified in a study of plant genes that allowed yeast to survive exposure to the thiol-oxidizing agent diamide (Kushnir et al 1995). The family 11 proteins (B15, B16, and C39) were initially characterized during studies of the altered response to gravity *A thaliana* mutants (Sedbrook et al 1999). The members of family 31 (B57, C58, C69) are very closely related to the *Brassica napus* S-locus protein 5 family. The atDjB65 protein, family 36 is orthologous with the AHM1 protein from wheat. The wheat protein is a nuclear matrix–localized, MAR-binding pro-

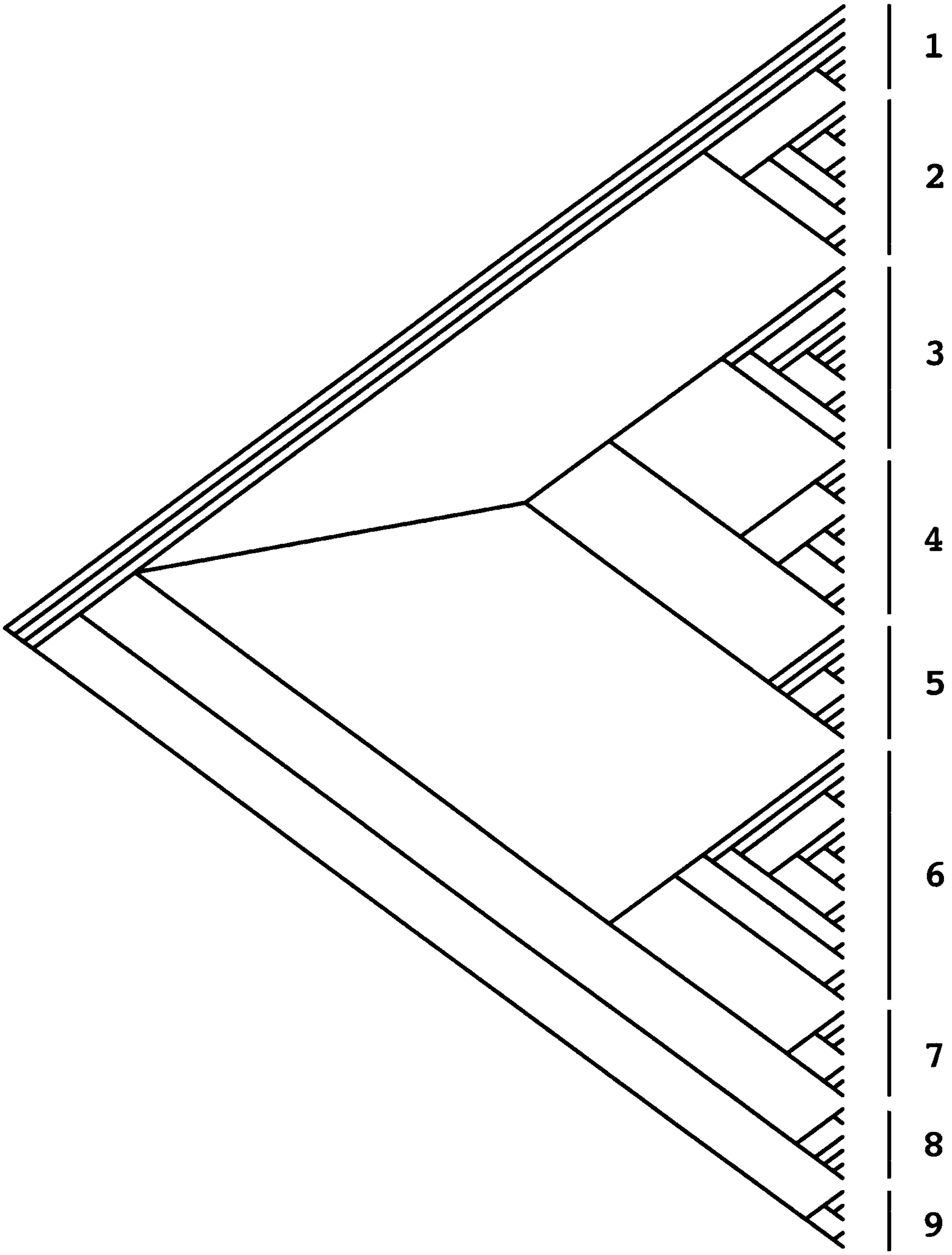
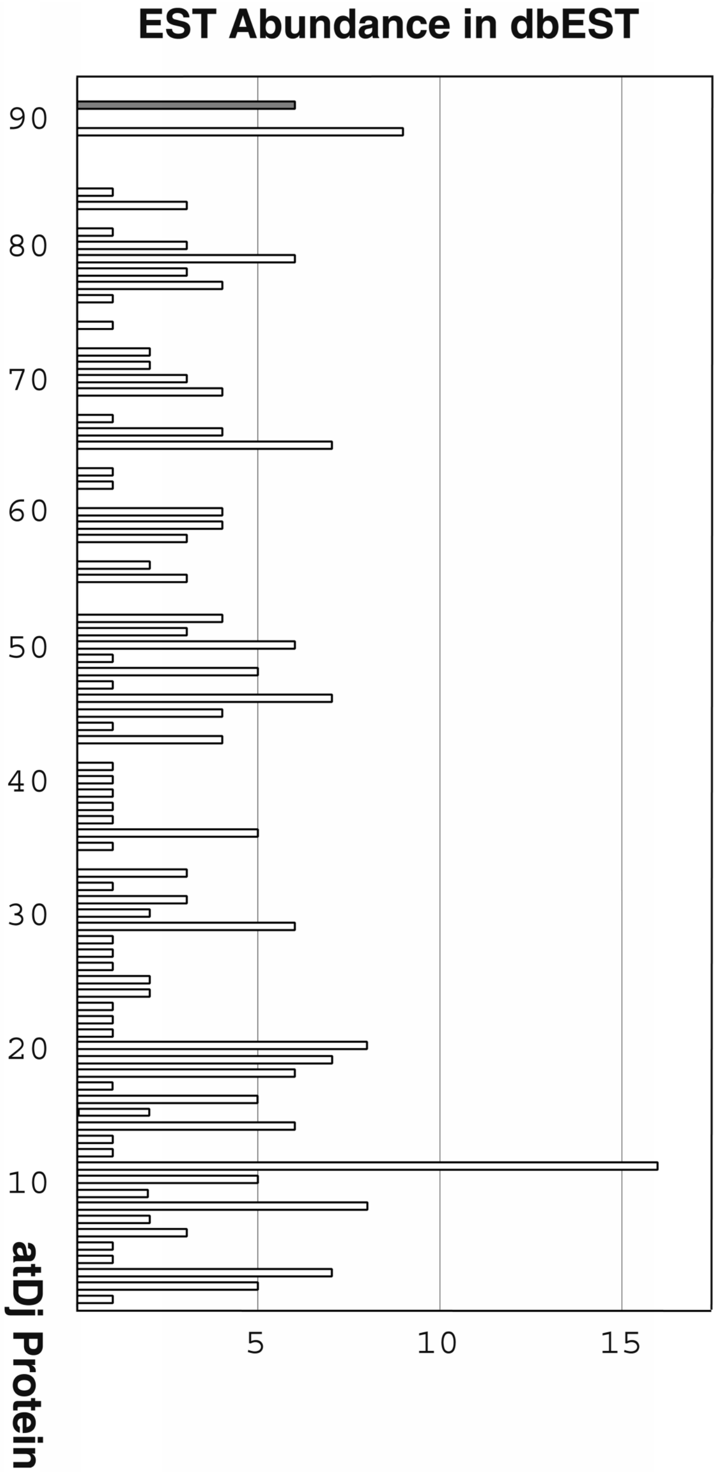


Fig 1. Phylogenetic relationships among the J-domain sequences of the *A thaliana* J-domain proteins. The sequences were manually aligned, then analyzed using SEQBOOT. The resulting cladogram was constructed using the PHYLIP program. Members of the clades, from top to bottom, are as follows: 1, A3, A2, B19, C29-2, B70, C21-2, B88; 2, C18, C35, C10, C46, B45, C7, C84, A30, B1, C74, B37, C6; 3, B13, B40, C72, B79, B66, C85, B87, B83, C82, B67, C4, C5, C59, C68; 4, C25, C20, C11, C41, C38, C8, C17, C21-1, C29-1, C64, C53, C55; 5, A36, B9, B60, B14, B22, C12, B23, B32, C33; 6, C61, C75, C31, B51, C34, B71, C77, B47, B73, B44, C76, C62, C27, C28, B78, B65, C49, B43, B48; 7, B86, B89, B80, C81, B57, C69, C58; 8, C50, C56, A52, A54, A26, A24; 9, B15, B16, C39, A63, B42.



Confidence Interval

ESTx	ESTy
	Ymin-Ymax
0	* - 5
1	* - 7
2	* - 9
3	* - 11
4	* - 12
5	0 - 14
6	0 - 16
7	1 - 17
8	1 - 19
9	2 - 20
10	2 - 22
11	3 - 23
12	4 - 24
13	4 - 26
14	5 - 27
15	6 - 28
16	6 - 30
17	7 - 31
18	7 - 32
19	8 - 34
20	8 - 35
20	55% - 75%
25	52% - 64%
30	47% - 60%
40	40% - 50%

Fig 2. Digital Northern analysis of the *A thaliana* J-domain proteins. The EST abundance in dbEST, as of January 2001, is presented. The histograms are, from bottom to top, atDjB1 to atDjC89. The shaded histogram, included for the sake of comparison, is for the plastidial enzyme dihydrolipoyl acetyltransferase. The differences between numbers, at the 95% confidence interval, are also presented. For a given EST number, indicated in the first column, the number of ESTs immediately beyond the confidence interval is indicated in the second column (Audic and Claverie 1997; Mekhedov et al 2000).

tein that, in addition to the J-domain, contains an AT-hook motif (Morisawa et al 2000).

Additionally, it can be assumed that the type A J-domain proteins that are localized within mitochondria or plastids (families 19 and 28) act as general chaperones in these organelles. This leaves 31 families (44 proteins) that cannot be assigned a function based on analogy. The only known role for J-domain proteins is in association with the 70-kDa stress proteins as molecular chaperones (Greene et al 1998). The very large number of *A thaliana* J-domain proteins with unassigned functions suggests possible roles in plant-specific cellular processes and signal transduction pathways, possible association with partners other than the AtHsp70 proteins, or both.

The abundance of very short J-domain proteins from *A thaliana* is particularly noteworthy. In many cases, these proteins are little more than a J-domain. For example, the atDjC4/5 proteins are only 112 amino acids long, including an presumptive N-terminal targeting peptide of approximately 20 amino acids. This type of J-domain protein has not previously been reported in other systems. Functional analysis of one of these proteins, atDjC8, is currently in progress. Although the recombinant protein stimulates ATPase activity of the cognate 70-kDa stress protein, it does not stimulate chaperone activity (JA Miernyk, in preparation).

J-domain sequence comparisons

The relationships among the *A thaliana* J-domain sequences that were apparent after 500 rounds of bootstrapping are presented as a cladogram (Fig 1). The sequences have been resolved into 9 clades containing 5 to 18 members. There is excellent agreement between the inclusion of protein species into a family defined by the deduced amino sequences exclusive of the J-domain and the within-clade grouping. Nearly all members of any given family are located within the same clade. The most notable exception is with members of the 2 classes of auxilin-related sequences, which are more widely dispersed throughout the cladogram.

The bases of the interrelationships among the clades are less apparent. In some instances, adjacent clades contain species predicted to have the same subcellular localization (Fig 1, clades 1 and 2). There seem, however, to be as many instances where this is not the case. It is likely that including still more distinguishing characteristics in the comparisons will help clarify the phylogenetic organization of this protein family.

Hennessy et al (2000) have presented an extensive analysis of the relationship between J-domain sequence and the class of the J-domain protein (I/A, II/B, III/C). They observed that the J-domain sequence of a type A protein is more likely to be closely related to other type A proteins, even from different species or kingdoms, than to type B or C sequences. Herein, the results of analysis of all of the members of a single complex organism are presented for the first time. In this study, no clear grouping of type A or B sequences is apparent (Fig 1). It should be noted, however, that this analysis includes a relatively small number of type A sequences.

Protein localization

Each of the 89 *A thaliana* J-domain protein sequences was subjected to analysis using multiple localization prediction algorithms. Examples of both soluble and integral membrane species were identified for virtually every subcellular compartment (Table 2). For the analysis presented herein, all proteins with a predicted localization in the secretory pathway downstream of the endoplasmic reticulum (Golgi apparatus, vacuole/lysosome, plasma membrane, extracellular) were combined under the heading ‘‘secretory pathway.’’ After some refinement through direct examination, a consensus for protein localization was reached for 80 of the 89. For the remaining 9 proteins, there are 2 to 3 contradictory predictions.

By far the most abundant predicted sites of J-domain protein localization are the cytoplasm and the nucleus (Table 2). Such a large proportion (27/89) of nuclear-localized chaperones is unprecedented. The most immediate question arising from the nuclear localization predictions is the identity of potential interacting proteins. Hsp70 is the only know partner for J-domain proteins (Greene et al 1998). Of the 17 AtHsp70 proteins (Lin et al 2001), 7 contain potential nuclear localization signal (NLS) sequences (Athsp70-1, -2, -3, -4, -5, -14, and -15) and are likely in vivo partners.

Six of the 9 instances of multiple location predictions involved potential N-terminal organelle-targeting peptides. The algorithms are apparently not yet sufficiently robust that it is always possible to make a high-probability distinction among a signal sequence, a transit peptide, and a mitochondrial-targeting peptide. The smallest group of predicted localization sites are those for peroxisomal proteins (Table 2), and the instances of peroxisome-cytoplasm predictions are suspect. These predictions are from PSORT, and this algorithm considers internal PTS1 sequences that are likely not functional in vivo (Subramani 1998). The only other instance of a predicted peroxisomal localization (atDjB66) is from visual examination. The N-terminus of this sequence contains a PTS2 motif (**RLX$_5$HL**): A_{11}E**RLLGIAEKLL**$_{22}$. In vivo, however, there are good data supporting a transient association of members of the YDJ1-type (class 2) J-domain proteins with the outer surface of the peroxisome boundary membrane, mediated by protein prenylation (Preisig-Muller et al 1994; Diefenbach and Kindl 2000).

In the few instances to date where localization of *A thal-*

iana J-domain proteins has been determined, there is agreement between the computer predictions and experimental data (cf atDjB1, Kroczyńska et al 1996). However, in no case is there a published study of a J-domain protein with a controversial location. The subcellular location of atDjC8, variously predicted as plastidial, as mitochondrial, or in the secretory pathway (Table 2), has now been experimentally determined to be plastidial (JA Miernyk, in preparation). Thorough analyses of the subcellular localization of other *A thaliana* J-domain proteins are in progress, using green fluorescent protein (GFP) chimera. An unequivocal knowledge of cellular localization will substantially enhance subsequent studies of protein function.

Digital Northern analysis

The concept of digital Northern analysis is based on the assumption that the number of EST clones will be a direct reflection of the abundance of messenger RNA (mRNA) in the population used to prepare the library (Audic and Claverie 1997; Ewing et al 1999; Mekhedov et al 2000). Just as with analog Northern data, it is equally accepted that there will be variance between ''signal strength'' and actual protein levels due to posttranslational regulation, mRNA, and protein turnover. The results from digital Northern analysis do, however, provide a first approximation of protein abundance.

Quantitation of the *A thaliana* J-domain protein ESTs in dbEST is presented in Figure 2. The EST clones can be separated into 3 classes: (1) those not significantly different (at the 95% confidence interval) from 0; (2) those that fall between 5 and 7; and (3) those with an abundance greater than 7. Seventy of the 89 proteins fall into class 1, 15 into class 2, and only 4 into class 3. It can be concluded that 19 of the *A thaliana* J-domain proteins are expressed at a relatively high level. Within this group, the number of ESTs for atDjC11 (16) is significantly larger than that for any other protein, suggesting a very high level of expression (Fig 2). The number of ESTs for *A thaliana* plastidial dihydrolipoyl acetyltransferase, a component of the pyruvate dehydrogenase complex that participates in fatty acid biosynthesis (Mooney et al 1999), is included for the sake of comparison (White et al 2000).

There is no obvious pattern of grouping or association (localization, type, size) among the J-domain proteins relative to gene expression. It will be necessary to validate the digital Northern results with those from analog Northern results. To date, the analog data are available for only a few proteins (Kroczynska et al 1996, 2000; Zhou and Miernyk 1999). The analog Northern data from atDj1, atDj3, and atDj6 are all consistent with constitutive expression. This is in marked contrast to *E coli* DnaJ, which is a classic heat shock protein (Georgopoulos et al 1980). In addition to making comparisons among the 89 species of *A thaliana* J-domain proteins, it will be both interesting and informative to determine which is up- or down-regulated in response to changes in environmental conditions. For example, experimental results suggest that the mRNA levels for atDj8, a small plastidial protein, are responsive to light (JA Miernyk, in preparation), whereas those of atDj11, another small plastidial protein, are not (W. Orme and J. C. Gray, personal communication). Expression of atDjC8 is also circadian and diurnally regulated (Schaffer et al 2001). Furthermore, it has been recently reported that expression of atDjB19 was increased nearly 3-fold by a transient water stress, but was largely unaffected by cold temperature treatment (Seki et al 2001).

The J-domain proteins from the model flowering plant *A thaliana* comprise a family of chaperones unprecedented in both size and complexity. It is likely that this has resulted from the necessity for plants to continuously respond to a myriad of environmental insults (Nover and Miernyk 2001). The J-domain protein catalog provided herein should provide a useful framework for future studies that will append ''functional'' to the extant genomics.

ACKNOWLEDGMENTS

F. Marsilio conducted the phylogenetic analyses presented in Figure 1, and B. P. Mooney assisted with preparation of Figure 2. Professor W. H. Miernyk provided a lucid explanation of the Poisson distribution and Bayesian statistics necessary for analysis of the digital Northern data. The efforts and patience of K. L. O'Donnell in explaining the bases of phylogenetics are greatly appreciated.

REFERENCES

Adams MD, Celniker SE, Holt RA, et al. 2000. The genome sequence of *Drosophila melanogaster. Science* 287: 2185–2195.

The *Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* 408: 796–815.

Audic S, Claverie JM. 1997. The significance of digital gene expression profiles. *Genome Res* 7: 986–995.

Banecki B, Liberek K, Wall D, Wawrzynow A, Georgopoulos C, Bertoli E, Tanfani F, Zylicz M. 1996. Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone. *J Biol Chem* 271: 14840–14848.

Bukau B, Horwich AL. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell* 92: 351–366.

The *C. elegans* Genome Initiative. 2000. Sequence and analysis of the genome of *C. elegans. Science* 282: 2012–2018.

Caplan AJ, Cyr DM, Douglas MG. 1993. Eukaryotic homologues of *Escherichia coli* DnaJ: a diverse protein family that functions with hsp70 stress proteins. *Mol Biol Cell* 4: 555–563.

Cheetham ME, Caplan AJ. 1998. Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress Chaperones* 3: 28–36.

Claros MG, Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 241: 779–786.

Diefenbach J, Kindl H. 2000. The membrane-bound DnaJ protein located at the cytosolic site of glyoxysomes specifically binds the cytosolic isoform 1 of Hsp70 but not other Hsp70 species. *Eur J Biochem* 267: 746–754.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300: 1005–1016.

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM. 1999. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res* 9: 950–959.

Felsenstein J. 1989. PHYLIP-Phylogeny Inference Package (version 3.2). *Cladistics* 5: 164–166.

Georgopoulos CP, Lundquist-Heil A, Yochem J, Feiss M. 1980. Identification of the *E. coli dna*J gene product. *Mol Gen Genet* 178: 583–588.

Goffeau A, Burrell BG, Bussey H, et al. 1996. Life with 6000 genes. *Science* 274: 563–567.

Goffin L, Georgopoulos C. 1998. Genetic and biochemical characterization of mutations affecting the carboxy-terminal domain of the *Escherichia coli* molecular chaperone DnaJ. *Mol Microbiol* 30: 329–340

Greene MK, Maskos K, Landry SJ. 1998. Role of the J-domain in the cooperation of Hsp40 with Hsp70. *Proc Natl Acad Sci U S A* 95: 6108–6113.

Hennessy F, Cheetham MSDE, Dirr HW, Black GL. 2000. Analysis of the levels of conservation of the J domain among various types of DnaJ-like proteins. *Cell Stress Chaperones* 5: 347–358.

Koncz C, Nemeth K, Redei GP, Schell J. 1992. T-DNA insertional mutagenesis in Arabidopsis. *Plant Mol Biol* 20: 963–976.

Kroczyńska B, Coop NE, Miernyk JA. 2000. AtJ6, a unique J-domain protein from *Arabidopsis thaliana. Plant Sci* 151: 19–27.

Kroczyńska B, Zhou R, Wood C, Miernyk JA. 1996. AtJ1, a mitochondrial homologue of the *Escherichia coli* DnaJ protein. *Plant Mol Biol* 31: 619–629.

Kushnir S, Babiychuk E, Kampfenkel K, Belles-Boix E, Van Montagu M, Inze D. 1995. Characterization of *Arabidopsis thaliana* cDNAs that render yeasts tolerant toward the thiol-oxidizing drug diamide. *Proc Natl Acad Sci U S A* 92: 10580–10584.

Laufen T, Mayer MP, Beisel C, Klostermeier D, Moor A, Reinstein J, Bakau B. 1999. Mechanism of regulation of hsp70 chaperones by DnaJ cochaperones. *Proc Natl Acad Sci U S A* 96: 5452–5457.

Liberek K, Marszalek J, Ang D, Georgopoulos C, Żylicz M. 1991. *Escherichia coli* DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. *Proc Natl Acad Sci U S A* 88: 2874–2878.

Lin W, Lin B-L. 1997. *AtJ10,* an *Arabidopsis dnaJ* homologue resembling calmodulin-binding *CAJ1* in yeast. *Plant Physiol* 115: 863.

Lin B-L, Wang J-S, Liu H-C, Chen R-W, Meyer Y, Barakat A, Delseny M. 2001. Genomic analysis of the HSP70 superfamily in *Arabidopsis thaliana. Cell Stress Chaperones* in press.

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. 1998. *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282: 679–682.

Mekhedov S, de Ilárduya OM, Ohlrogge J. 2000. Towards a functional catalog of the plant genome: a survey of genes for lipid biosynthesis. *Plant Physiol* 122: 389–401.

Miernyk JA. 1997. The 70 kDa stress-related proteins as molecular chaperones. *Trends Plant Sci* 2: 180–187.

Miernyk JA. 1999. Protein folding in the plant cell. *Plant Physiol* 121: 695–703.

Miernyk JA, Coop NE. 2000. AtJ20 (accession No. AF214107), a plastid-localized type III J-domain protein from *Arabidopsis thaliana* (PGR00–027). *Plant Physiol* 122: 619.

Miller SM, Kirk DL. 1999. *gls*A, a *Volvox* gene required for asymmetric division and germ cell specification, encodes a chaperone-like protein. *Development* 126: 649–658.

Mooney B, Miernyk JA, Randall DD. 1999. Cloning and characterization of the dihydrolipoamide S-acetyltransferase (E2) subunit of the plastid pyruvate dehydrogenase complex from *Arabidopsis thaliana. Plant Physiol* 120: 443–452.

Morisawa G, Han-yama A, Moda I, Tamai A, Iwabuchi M, Meshi T. 2000. AHM1, a novel type of nuclear matrix-localized, MAR binding protein with a single AT hook and a J domain-homologous region. *Plant Cell* 12: 1903–1916.

Nakai K, Horton P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* 24: 34–36.

Nover L, Miernyk JA. 2001. A genomics approach to the chaperone network of *Arabidopsis thaliana. Cell Stress Chaperones* 6: 175–176.

Ohlrogge J, Benning C. 2000. Unraveling plant metabolism by EST analysis. *Curr Opin Plant Biol* 3: 224–228

Ohtsuka K, Hata M. 2000. Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal for their classification and nomenclature. *Cell Stress Chaperones* 5: 98–109.

Preisig-Muller R, Muster G, Kindl H. 1994. Heat shock enhances the amount of prenylated Dnaj protein at membranes of glyoxysomes. *Eur J Biochem* 219: 57–63.

Quesada V, Ponce MR, Micol JL. 1999. OTC and AUL1, two convergent and overlapping genes in the nuclear genome of *Arabidopsis thaliana. FEBS Lett* 461: 101–106.

Redei GP. 1975. Arabidopsis as a genetic tool. *Annu Rev Genet* 9: 111–127.

Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E. 2001. Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. *Plant Cell* 13: 113–123.

Scidmore MA, Okamura HH, Rose MD. 1993. Genetic interactions between KAR2 and SEC63, encoding eukaryotic homologues of DnaK and DnaJ in the endoplasmic reticulum. *Mol Biol Cell* 4: 1145–1159.

Sedbrook JA, Chen R, Masson PH. 1999. *ARG1* (altered response to gravity) encodes a DnaJ-like protein that potentially interacts with the cytoskeleton. *Proc Natl Acad Sci U S A* 96: 1140–1145.

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y, Shinozaki K. 2001. Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* 13: 61–72.

Silver PA, Way JC. 1993. Eukaryotic DnaJ homologs and the specificity of Hsp70 activity. *Cell* 74: 5–6.

Subramani S. 1998. Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. *Physiol Rev* 78: 171–188.

Szabo A, Korszun R, Hartl FU, Flanagan J. 1996. A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO J* 15: 408–417.

Umeda A, Meyerholz A, Ungewickell E. 2000. Identification of the universal cofactor (auxilin 2) in clathrin coat dissociation. *Eur J Cell Biol* 79: 336–342.

Wall D, Zylicz M, Georgopoulos C. 1995. The conserved G/F motif of the DnaJ chaperone is necessary for the activation of the substrate binding properties of the DnaK chaperone. *J Biol Chem* 270: 2139–2144.

White JA, Todd J, Newman T, et al. 2000. A new set of Arabidopsis ESTs from developing seeds: the metabolic pathway from carbohydrates to seed oil. *Plant Physiol* 124: 1582–1594.

Yan W, Craig EA. 1999. The glycine-phenylalanine-rich region determines the specificity of the yeast Hsp40 Sis1. *Mol Cell Biol* 19: 7751–7758

Zhang Y, Grant B, Hirsh D. 2000. RME-8, a conserved DnaJ domain containing protein, functions in the late endosome in *C. elegans. Mol Biol Cell Suppl* P-139.

Zhou R, Kroczyńska B, Hayman GT, Miernyk JA. 1995. *At*J2, an *Arabidopsis* homologue of *Escherichia coli dna*J. *Plant Physiol* 108: 821–822.

Zhou R, Miernyk JA. 1999. Cloning and analysis of *AtJ*3 gene in *Arabidopsis thaliana. Acta Bot Sinica* 41: 597–602.