The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing α -crystallin domains (Acd proteins)

Klaus-Dieter Scharf,¹ Masood Siddique,² and Elizabeth Vierling²

¹Biocenter of the Goethe University, Marie-Curie-Strasse 9, D-60439 Frankfurt/Main, Germany

²Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721, USA

Abstract Comprehensive analysis of the *Arabidopsis* genome revealed a total of 13 sHsps belonging to 6 classes defined on the basis of their intracellular localization and sequence relatedness plus 6 ORFs encoding proteins distantly related to the cytosolic class CI or the plastidial class of sHsps. The complexity of the *Arabidopsis* sHsp family far exceeds that in any other organism investigated to date. Furthermore, we have identified a new family of ORFs encoding multidomain proteins that contain one or more regions with homology to the ACD (Acd proteins). The functions of the Acd proteins and the role of their ACDs remain to be investigated.

INTRODUCTION

The small Hsp (sHsp), or Hsp20, family of heat stress proteins is a nearly ubiquitous family of stress proteins that range in size from approximately 16–42 kDa. This diverse family of proteins, which includes the α -crystallin proteins (Acd proteins) of the vertebrate eye lens, is defined by a conserved C-terminal domain of approximately 90 amino acids referred to as the α -crystallin domain (ACD) (de Jong et al 1998). The ACD is flanked by an N-terminal domain and a short C-terminal extension (Nover 1991; Caspers et al 1995; Waters et al 1996; MacRae 2000). The sHsp N-terminal domains show no sequence conservation, except between evolutionarily closely related proteins. The sHsp family is, therefore, different from the Hsp70 and Hsp90 chaperone families, in which sequence identity, even between prokaryotic and eukaryotic members, is significant throughout the protein. Another defining feature of the sHsps is that, with few exceptions, all members of the family form homo-oligomeric complexes of 200–750 kDa (9 to >24 subunits), with the size dependent on the specific protein (Chen et al 1994; Lee et al

1995; Ehrnsperger et al 1997, 1999; Helm et al 1997; Kirschner et al 2000).

Plants are characterized by an unusual complexity of heat stress-inducible proteins belonging to the sHsp family. Plant sHsps appear to have evolved independently after the divergence of plants and animals (Waters 1995; de Jong et al 1998). Although in other organisms sHsps are found in the cytoplasm under most conditions, and at times in the nucleus, plants express both cytosolic sHsps and specific isoforms targeted to intracellular organelles. There are at least 2 forms of sHsps in the cytosol, referred to as class I and class II proteins (Vierling 1991), which share only approximately 50% identity in the ACD and are estimated to have diverged more than 400 million years ago (Waters and Vierling 1999). Three separate gene families have been described that encode mitochondrial, plastidial, and endoplasmic reticulum-localized sHsps, each with appropriate organelle-targeting signals. The organelle sHsp forms appear to be unique to plants, with the exception of Hsp22 in *Drosophila*, which was recently reported to be located in mitochondria (Morrow et al 2000).

Recent solution of the structure of Hsp16.5 from *Methanococcus jannaschii* at 2.9-Å resolution has provided gen-

Correspondence to: Klaus-Dieter Scharf, Tel: 49-69-798-29285; Fax: 49-69-798-29286; E-mail: scharf@cellbiology.uni-frankfurt.de.

Fig 1. Sequence comparison of sHsps. *Arabidopsis* (At) and selected representatives of the sHsp family from tomato (Lp), pea (Ps), and wheat (Ta) are included. A color code marks sequence motifs conserved among the members of a given class or for all sHsps. Putative cleavage sites for removal of the targeting peptides of the organelle proteins (classes P, ER, and M) are indicated by arrows. The Met bristle was defined as a unique, methionine-rich sequence motif characteristic for all plastidial sHsp (Chen and Vierling 1991). Closed arrowheads point to the positions of introns in the ORF coding sequences. The ACD (yellow color code) comprises 2 homologous regions, consensus region II and I, separated by a linker, which is more variable in sequence and size. In the corresponding vertebrate genes, homologous regions II and I are encoded by 2 separate exons. Potential positions of secondary structural elements are indicated above the alignment based on positions of these elements in *M jannaschii* Hsp16.5. The conserved sequence motif, basic-x-I/V-x-I/V, which is partially involved in formation of the C-terminal β -strand (β 10) is underlined for each class of sHsps.

eral insight into the organization of sHsp monomers and oligomers (Kim et al 1998a). The ACD consists of a β -sandwich of 2 antiparallel β -sheets similar to the fold of the unrelated Fc fragment of immunoglobulin, although the exact folding topology is different (Deisenhofer 1981). The dominant β -structure of the sHsps was previously predicted from circular dichroism (C.D.) spectroscopy (Merck et al 1993; Dudich et al 1995; Sun et al 1997), secondary structure analysis of sHsp sequences (Caspers et al 1995; de Jong et al 1998), and site-directed spin-labeling studies (Berengian et al 1999; McHaourab et al 2000). This same β -sandwich fold is predicted to be present in all sHsps, and the putative location of β -strands in *Arabidopsis* and other plant sHsps are indicated in the alignment in Figure 1.

In *M jannaschii* the Hsp16.5 monomers form a 24-subunit, hollow spherical oligomer with an outer diameter of 120 Å and an inner diameter of 65 Å. Virtually the entire N-terminal domain (32 residues) of the proteins was disordered and not resolved. These 32 N-terminal residues

(approximately 20% of the protein mass) could fill a significant portion of the interior space, and therefore the size of this space is debated. The basic building block of the *M jannaschii* oligomer is a dimer, another structural feature that is likely to be conserved in the plant sHsps. Within the resolved structure, the dimer interface has the strongest and most extensive contacts within the oligomer, being stabilized by both intermolecular hydrogen bonds and extensive hydrophobic contacts. The interface contacts comprise primarily an extended loop that forms the β 6-strand, which binds antiparallel to β 2 to extend one of the β -sheets in the β -sandwich fold. This loop region shows limited sequence conservation, but is primarily hydrophilic and capable of forming the necessary β -strand in all plant sHsps.

Interactions of the C-terminal extensions are critical in stabilizing the *M jannaschii* Hsp16.5 oligomer. A 3-residue β -strand (β 10) in the C-terminus extends the β -sheet of another dimer by binding to β 4. The β 10-strand is part of a conserved sequence motif recognized by de Jong et

Fig 1. Continued.

al (1998), basic-x-I/V-x-I/V, which is present in all 6 defined plant sHsp classes (Fig 1), suggesting this same contact is involved in formation of the plant sHsp oligomers. In addition to intermolecular hydrogen bonds in this contact, the side chains of the hydrophobic residues in this motif make extensive hydrophobic contacts with a hydrophobic groove formed by strands $\beta 4$ and $\beta 8$ on one monomer of the interacting dimer. Recent solution of the structure of wheat Hsp16.9 (van Montfort et al, in preparation) will add considerably to our detailed understanding of plant sHsp structure and test current structural predictions.

Almost all sHsps function as chaperones in vitro to prevent irreversible protein aggregation and insolubilization and to interact otherwise with nonnative proteins in an adenosine triphosphate (ATP)-independent fashion (Horwitz 1992; Jakob et al 1993; Knauf et al 1993; Merck et al 1993; Lee et al 1995; Plater et al 1996; Kim et al 1998b). Early studies with crude cell protein fractions from plants suggested that sHsps could protect other proteins from heat-induced insolubilization (Jinn et al 1989). Very likely, sHsps function in the chaperone network to maintain denatured proteins in a folding competent state (Ehrensperger et al 1997; Lee et al 1997). Substrate renaturation requires interaction with the Hsp70 chaperone system and ATP (Forreiter et al 1997; Leroux et al 1997; Veinger et al 1998;

Lee and Vierling 2000). Interaction with substrate appears to involve heat-induced exposure of hydrophobic-binding sites on the sHsp subunits (Lee et al 1997), which occurs because of heat-induced oligomer dissociation or increased subunit exchange rates (Bova et al 1997, 2000; Ehrensperger et al 1999; Haslbeck et al 1999). The sHsp complexes with denatured protein then reassembles into a larger complex, and reassembly may be critical for proper activity. This striking reorganization during stress may also initiate the assembly of cytosolic, multichaperone complexes, comprising 40-nm heat stress granules, which are found in heat-stressed plant cells (Nover et al 1983, 1989; Kirschner et al 2000). Full understanding of the importance of sHsp chaperone activity to plant stress tolerance will require identification of the critical substrates protected by these proteins.

MEMBERS OF THE ARABIDOPSIS sHsp FAMILY

Searching the *Arabidopsis* genome revealed 19 open reading frames (ORFs) encoding proteins related to the previously characterized 5 classes of sHsps in plants (Vierling 1991; Waters et al 1996), of which 8 had already been cloned and characterized (see references given in Table 1). Details of the size, properties, and references to accession numbers and publications are compiled in Table 1

Table 1 Survey of the small Hsp family

Num- ber	Name (former name)	Chromo- somal localization	ORF (aa)	MW (kDa)	pI	Intracellular localization	Accession number (gene)	Accession number (protein)	Reference
1	Class Cl (cytoplasmic/nuclear)								
2	AtHsp17.8-CI	I	157	17.8	6.0	C	At1g07400	AAF79569	DS
3	AtHsp17.6A-CI	I	155	17.6	7.8	C	At1g59860	AAD39328	DS
4	AtHsp17.6B-CI	II	153	17.6	6.4	C	At2g29500	AAC95188	DS
5	AtHsp17.4-CI	III	156	17.4	5.0	C	At3g46230	CAB90950	DS, Takahashi and Komeda 1989
6	AtHsp17.6C-CI	I	157	17.6	5.2	C	X17293	CAA35182	DS, Helm and Vierling 1989
7	AtHsp17.6 (Hsp17.6)						At1g53540	AAG21467	
8	AtHsp18.1-CI	V	161	18.1	7.8	C	X16076	CAA34208	DS, Takahashi and Komeda 1989
9	LpHsp17.7-CI		154	17.7	5.2	C	At5g59720	BAB09509	Forreiter and Löw 1998a
10	PsHsp18.1-CI		158	18.1	5.7	C	X17295	CAA35183	
11	TaHsp16.9A-CI		151	16.9	5.7	C	AJ225047	CAA12388	
12	Class CII (cytoplasmic/nuclear)								
13	AtHsp17.6-CII	V	155	17.6	6.4	C	M33899	AAA33672	Lauzon et al 1990
14	AtHSP17.7-CII	V	156	17.7	5.4	C	X13431	CAA31785	McElwain and Spiker 1989
15	LpHsp17.3-CII		155	17.3	6.4	C	At5g12020	CAB87675	DS, Bartling et al 1992
16	PsHsp17.1-CII		152	17.1	6.4	C	X63443	CAA45039	DS, Prändl et al 1998
17	TaHsp17.4-CII		157	17.4	5.4	C	At5g12030	CAB87676	Forreiter and Löw 1998b
18	Class CIII (cytoplasmic/nuclear)						Y14070	CAA74399	
19	AtHsp17.4-CIII	I	155	17.4	9.0	(Nu)	AJ225049	CAA12390	Lauzon et al 1990
20	LpHsp16.1-CIII		144	16.1	8.5	(Nu)	M33901	AAA33670	Weng et al 1991
21	Class P (plastids)								
22	AtHsp25.3-P	IV	227	25.3	8.8	P	X58279	CAA41218	DS
23	LeHsp26.1-P		234	26.1	8.6	P	At1g54050	AAD25777	DS, Chen and Vierling 1991, Osteryoung et al 1993
24	PsHsp26.2-P		232	26.2	6.4	P	AF399821	CAB38279	Lawrence et al 1997
25	TaHsp26.6-P		237	26.6	9.9	P	X4g27670	AAA32818	Vierling et al 1988
26	Class ER (endoplasmic reticulum)						X54102	CAA38036	Nguyen et al 1993
27	AtHsp22.0-ER	IV	195	22.0	5.4	ER	M94455	AAB49626	
28	GmHsp22.0-ER		192	22.0	6.3	ER	U59917	CAA30167	DS
29	LeHsp21.5-ER		190	21.5	5.8	ER	X07187	CAA47745	DS, Helm et al 1995
30	PsHsp22.7-ER		197	22.7	6.8	ER	X63198	AAA19931	Helm et al 1993
31	Class M (mitochondria)								
32	AtHsp23.6-M	IV	210	23.6	7.3	M	AB026983	BAA97658	Lauzon et al 1990
33	Class M (mitochondria)						M33898	AAA33673	DS, Visioli et al 1997, Willett et al 1996
34	AtHsp23.6-mito						At4g25200	CAA23061	
35							X98375	CAA67022	
36							U72958	AAB38795	

Table 1 Continued

Num- ber	Name (former name)	Chromo- somal localization	ORF (aa)	MW (kDa)	pl	Intracellular localization	Accession number (gene)	Accession number (protein)	Reference
26	AtHsp23.5-M	V	210	23.5	9.6	M	At5g51440	BAB09755	DS
27	LeHsp23.8-M (mtHsp)		210	23.8	7.3	M	AB017134	BAA32547	DS
28	PsHsp22.9-M (Hsp22)		202	22.9	9.0	M	X86222	CA60120	Lenne et al 1995
29	TaHsp23.5-M sHsp-like proteins	IV	214	23.5	6.2	M	AF104107	AAD03604	Basha et al 1999
30	Class CI related								
31	AtHsp15.4-CI(r)		134	15.4	5.0	(C)	At4g21870	CAA17154	DS
32	AtHsp15.7-CI(r)		137	15.7	9.1	(Po)	At5g37670	BAB08313	DS
33	AtHsp18.5-CI(r)		162	18.5	4.2	(P)	At2g19310	AAC16461	DS
34	AtHsp21.7-CI(r)		192	21.7	5.3		At5g54660	BAB09345	DS
35	Class P related	I							
36	AtHsp14.2-P(r)		127	14.2	8.2		At5g47600	BAB09086	DS
37	AtHsp26.5-P(r)		232	26.5	7.3	(M)	At1g52560	AAD55608	DS

Hsps from *Arabidopsis thaliana* (At), *Lycopersicon peruvianum* (Lp), *Pisum sativum* (Ps), and *Triticum aestivum* (Ta) are listed with their names based on their structure and assignment to the 6 classes (see Figs 1 and 2). The molecular weight (MW) and isoelectric points (pI) were calculated on the basis of the amino acid (aa) sequences of the open reading frames (ORFs) using Clone Manager 5 software. The chromosomal localization of the genes and intracellular localization of the corresponding proteins are indicated: C, cytoplasm; P, plastids; ER, endoplasmic reticulum; M, mitochondria; Nu, nucleus. The target sites given in parentheses refer to predictions in the database. For identification, the accession numbers for the genes (http://mips.gsf.de/proj/thal/db/search/search_frame.html) and proteins (<http://www.ncbi.nlm.nih.gov/>) as well as references are given. DS, direct submission to the database as part of the *Arabidopsis* Genomic Sequencing Project.

and Figure 1. The classification is based on the intracellular localization of the sHsps in the cytoplasmic-nuclear compartment (classes CI and CII), plastids (class P), endoplasmic reticulum (class ER), and mitochondria (class M). From the sequence comparison in Figure 1, which also includes selected representatives from other plants, it is evident that this classification is reflected also by characteristic sequence motifs found primarily in the N-terminal domain and C-terminal extensions. In addition, class-specific features of the ACD are also notable. There are 6 proteins of cytosolic class CI and 2 proteins of class CII. One representative each of the sHsp family is targeted to the chloroplasts and endoplasmic reticulum, and 2 are targeted to mitochondria. In addition, there is one protein (number 15 in Table 1) distantly related to class CII proteins, which we designated class CIII. A closely related member of this class was found in tomato by yeast 2-hybrid screening using a class CI sHsp as bait. It has unique properties, because it interacts with both class CI and class CII sHsps in the yeast 2-hybrid system, and it is preferentially localized in the nucleus (K. D. Scharf et al, in preparation). These 13 core members of the sHsp family are closely related to similar proteins from other plants (Figs 1 and 2). Six additional sHsps included in Table 1 (numbers 30 to 35) and Figure 2 are more distantly related (Figs 2 and 4), but also appear to be bona fide members of this protein family.

The number of sHsps found in different organisms varies significantly. Prokaryotes contain 1 or 2 sHsps, except species of *Rhizobium*, some of which have at least 10 sHsps (Munchbach et al 1999). *Saccharomyces cerevisiae* has 2 sHsps encoding Hsp26 (Petko and Lindquist 1985) and Hsp42 (Wotton et al 1996), one of the largest family members. In humans, only 5 sHsps, including α A and α B-crystallin (de Jong et al 1998), were described, until recently when 4 additional transcribed sequences were identified (Boelens et al 1998; Suzuki et al 1998; Krief et al 1999; W. de Jong, personal communication). *Caenorhabditis elegans* has 16 proteins classified as sHsps (Ding and Candido 2000) and *Drosophila melanogaster* has 12 (Sébas-tien et al 2001).

A NEW FAMILY OF PROTEINS WITH ACDs (Acd PROTEINS)

An interesting new group of proteins is represented by 25 ORFs in the *Arabidopsis* genome encoding proteins with one or more ACDs. They are compiled in Table 2 as Acd proteins specified with numbers derived from their molecular weights. Although they are unrelated to the 6 classes of sHsps defined herein, it is an intriguing observation that a structural motif related to the ACD may be functional also in other proteins. Unfortunately, our knowledge about this new class of proteins is very lim-

Fig 2. Phylogenetic relationship of proteins of the sHsp family. The figure was created by Clustal alignment on the basis of the full-length amino acid sequences of all proteins compiled in Table 1. Proteins belonging to the 6 classes of sHsp are marked by shaded areas. The 6 proteins (numbers 30 to 35 in Table 1) with a more distant relationship to class CI and class P are clearly separated from the others.

ited. However, 3 examples indicate that this family deserves special attention.

One Acd protein (RTM2 equals AtAcd41.3 in Table 2) was recently characterized as an essential component of a specific resistance mechanism against systemic spreading of tobacco etch virus in *Arabidopsis* (Whitham et al 2000). In addition to the N-terminal ACD, the RTM2 protein contains a central region with highly charged repeat motifs and a long C-terminal domain with a putative membrane-spanning domain. The RTM2 defect in the *Arabidopsis* mutant used for cloning of the gene is due to a point mutation introducing a stop codon in the ACD. The RTM2 protein is not heat shock inducible, and the *rtm2* mutant plants are not impaired in their thermotolerance. The authors hypothesize that RTM2 has no typical chap-

erone function but rather functions as part of a localized high-molecular-weight complex restricting virus movement.

Another intriguing group in the Acd family is represented by Acd55.2 and Acd44.3 (numbers 18 and 19 in Table 2). They belong to the growing number of ARID proteins (ARID indicates AT-rich interaction domain, also called BRIGHT domain), which are DNA-binding proteins with a helix-turn-helix motif (Tucker et al 2000). Some of them were characterized as developmental regulators in *D melanogaster* and mammals (Tucker et al 1995; Treisman et al 1997). In addition, the p270 subunit of the SWI-SNF chromatin remodeling complex is a member of the ARID protein family (Kobayashi et al 2000). Two additional ARID proteins in the *Arabidopsis* genome (acces-

Table 2 Survey of *Arabidopsis* Acd proteins, proteins with domain(s) homologous to the α -crystallin domain (ACD)

Num- ber	Name (former name)	Chromosomal localization	ORF (aa)	MW (kDa)	pI	Target predicted ^a	ACD (aa) ^b	Gene number ^c	Accession number (protein)	Reference
Proteins with 1 ACD										
1	AtAcd31.2	I	285	31.2	9.7	M	188–284	At1g06460	AAF24817 ^d	DS
2	AtAcd39.0	I	345	39.0	6.2		239–338	At1g54840	AAC64893	DS
3	AtAcd51.9	I	463	51.9	7.1		360–459	At1g20870	AAG51105	DS
4	AtAcd21.4	I	197	21.4	4.6		93–192	At1g54850	AAD30605	DS
5	AtAcd15.5	I	143	15.5	7.7		37–135	At1g76440	AAC64894	DS
6	AtAcd55.5	V	498	55.5	6.4		373–477	At5g02480	AAG51117	DS
7	AtAcd55.8 (Slt1)	II	494	55.8	6.0		364–468	At2g37570	AAF16657	DS
8	AtAcd55.3	III	489	55.3	6.0		365–469	At3g12570	CAB85985	DS
9	AtAcd16.9	IV	152	16.9	5.4		51–139	At5g14830	AAAK25847 ^d	Lin et al 1999
10	AtAcd22.1	III	198	22.1	5.5	SP, TMD	82–171	At3g22530	AAC23628	Lin et al 1999
11	AtAcd28.7	I	244	28.7	4.9		33–127	At1g76770	AAG51034	Kaneko et al 2000
12	AtAcd216.9	I	1871	216.9	4.6		1–50	At1g76780	CAB10262	DS
13	AtAcd41.3 (Rtm2)	V	366	41.3	9.3	TMD	23–103	At5g04890	CAB78525	Kaneko et al 2000
14	AtAcd25.1 (Atuk)	II	224	25.1	9.7	TMD	21–100	At2g27140	BAB01468 ^d	DS
15	AtAcd28.1	V	249	28.1	10.0	TMD	17–96	At5g20970	AAF04451	DS
16	AtAcd16.6	I	153	16.6	10.4	TMD	8–70	At1g54400	AAF04452	DS
17	AtAcd54.2	III	490	54.2	9.4	P, TMD	30–110	At3g10680	BAB08993	Kaneko et al 1999,
18	AtAcd55.2	I	494	55.2	5.8	Nu	410–492	At1g76510	AAF61902	Whitham et al 2000
19	AtAcd44.3	I	398	44.3	6.4	Nu	314–396	At1g20910	AAD22497	Lin et al 1999,
20	AtAcd32.4	II	287	32.4	6.6	SP, TMD	201–283	At2g17410	AAB86518	Whitham et al 2000
Proteins with more than 1 ACD										
21	AtAcd25.4	IV	232	25.4	8.8		33–117, 149–231	At4g16540	CAB10430	DS
22	AtAcd26.0	II	239	26.0	9.4		36–120, 156–238	At2g03020	AAC32920	Lin et al 1999
23	AtAcd28.9	V	264	28.9	6.0		39–121, 166–249	At5g47590	BAB09085	DS
24	AtAcd81.4	IV	743	81.4	6.6	TMD	33–127, 188–272, 405–488, 514–597, 660–742	At4g16550	CAB78697 CAB10431	DS
25	AtAcd114.3	IV	1046	114.3	7.7	SP, TMD	548–632, 765–847, 964–1045	At4g16560	CAB78698 CAB10432	DS

Following the procedure for sHsps (Table 1), Acd proteins are identified by numbers indicating their molecular weight (MW) in kilodaltons as derived from the amino acid (aa) sequence of the predicted open reading frames (ORFs), pI, isoelectric points; DS, direct submission to the database of the *Arabidopsis* Genomic Sequencing Project.

^a No entry, no prediction in the database; M, mitochondria; Nu, nucleus; P, plastids; SP, secretory pathway; TMD, transmembrane domain.

^b The position of the ACDs are indicated by the aa residues. For sequence details, see Fig 5 online.

^c Gene numbers refer to the MIPS database (see also legend to Table 1).

^d Protein accession number refers to an ORF that is different to that predicted by the assignment of the corresponding chromosomal DNA. In case of AtAcd44.3 (number 19), the ORF was corrected according to a corresponding expressed sequence tag (accession number AV441277).

^e Acd proteins with ARID domain (see text for explanation). The ARID domain in Acd55.2 comprises aa residues 138–263, and in Acd44.3, aa residues 103–190.

Fig 3. Promoter architecture of *Arabidopsis* genes encoding sHsps with respect to the positioning of HSE clusters. For orientation and positioning of the putative TATA box, we used the well-defined start site of the ORFs and the data for genomic clones with experimentally defined transcription start sites. These are Hsp17.4-CI and Hsp18.1-CI (Takahashi and Komeda 1989) and Hsp25.3-P (Osteryoung et al 1993). Symbols are explained at the bottom of the figure. Numbers on the left refer to the corresponding numbers in Table 1.

Fig 3. Continued.

Table 3 Expression of the genes encoding small Hsps (sHsps) in *Arabidopsis*: survey of expressed sequence tag (EST) clones

Num- ber ^a	Name of sHsp	EST accession number	Length (nt) ^b	Source of EST ^c					
				DS	GS	FB	IF	L	M
1	AtHsp17.8-CI	AI998089	455 (C)					+	
2	AtHsp17.6A-CI								
3	AtHsp17.6B-CI	AI998064	532 (C)					+	
		AA042551	502						+
		AA597558	410						+
4	AtHsp17.4-CI	AI995439	509				+		
		Z34874	242			+			
		AV555356	249		+				
		AA597555	218						+
5	AtHsp17.6C-CI	AI998904	616 (C)					+	
		AW004417	497						+
6	AtHsp18.1-CI	AV441076	618 (C)					+	
		AI995531	580 (C)				+		
		AV442469	346					+	
		AA597611	417						+
		AA395897	400						+
10	AtHsp17.6-CII	Z46682	307	+					
		AA042544	636						+
11	AtHsp17.7-CII	AI998408	541 (C)					+	
		Z27013	382	+					
		AA650758	477						+
		W43073	467						+
		W43152	404						+
		Z29845	308	+					
15	AtHsp17.4-CIII	W43132	460						+
17	AtHsp25.3-P								
21	AtHsp22.0-ER	AA597970	262						+
25	AtHsp23.6-M								
26	AtHsp23.5-M								
	sHsp-like protein								
	Class CI related								
30	AtHsp15.4-CI(r)	AV522597	394 (C)					+	
31	AtHsp15.7-CI(r)								
32	AtHsp18.5-CI(r)	AA404830	578						+
		T46311	536						+
		AA728682	425			+			
		T21216	278						+
33	AtHsp21.7-CI(r)	N97114	492						+
	Class P related								
34	AtHsp14.2-P(r)	AI999186	554 (C)					+	
		AA728700	380 (C)			+			
35	AtHsp26.5-P(r)	AW004197	458 (C)				+		

^a Numbers refer to the corresponding numbers in Table 1.

^b (C), coding sequence complementary to the given database entrance.

^c Sources of RNA were exclusively from non-heat shock material: DS, dry seeds; GS, green siliques (Asamizu et al 2000); FB, flower buds; IF, inflorescence tissue; L, leaves; M, mixed pool of RNA (Newman et al 1994).

sion numbers AAF17649 with 338 amino acid residues and AAF40449 with 448 amino acid residues) have no ACD but an HMG box in their C-terminal parts.

The first examples of proteins with multiple ACDs were 2 major egg protein antigens of *Schistosoma mansoni* with 40 kDa and 2 ACDs (Nene et al 1986). Later, they were characterized as soluble Ca²⁺-binding proteins (Moser et al 1992). Similar to some members of the *Arabidopsis* Acd proteins, their duplicate ACDs are separated by a short proline-rich linker. Particularly interesting in this group of *Arabidopsis* Acd proteins is Acd114.3 (number 25 in Table 2) with 3 ACDs linked to a N-terminal domain exhibiting high-sequence homology to the tobacco protein CND41. This protein was described before as

a chloroplast nucleoid DNA-binding protein (Nakano et al 1997) with aspartic protease activity (Murakami et al 2000). CND41 might be involved in down-regulation of transcription in chloroplasts.

A detailed sequence comparison and phylogenetic analysis of all proteins contained in Tables 1 and 2 are presented online in the attached material (Figs 4 and 5).

PROMOTER STRUCTURE AND EXPRESSION OF sHsps

Except for some well-defined developmental stages, eg, during seed maturation (Hightower and Nover 1991; Waters et al 1996; Wehmeyer et al 1996), sHsps of plants are

exclusively found after heat shock induction. We therefore analyzed the putative promoter structures of the genes listed in Table 1. Because in most cases experimental data about the transcription start site are lacking, we used the ATG start codon of translation as the only reliable point to search upstream for presumptive TATA box and heat shock element (HSE) motifs. The results are summarized in the block diagrams in Figure 3.

In practically all cases, putative promoter regions are characterized by clusters of HSEs frequently found in plant heat stress genes. These clusters are formed of palindromic repeats of the head (aGAAn) and tail (nTTcT) modules. Frequently, they are imperfect, because defective 5-nucleotide modules are interspersed (shaded blocks in Fig 3). In agreement with mutational analysis in our and other laboratories (Amin et al 1987; Bharti et al, unpublished data), modules are considered defective if the invariant G or C (letters underlined) and/or the 2 highly conserved A or T residues (uppercase letters) are lacking (see explanations given in Fig 3). The importance of such imperfect clusters for the regulatory fingerprint of the heat shock genes was recently demonstrated in our group by a strong synergistic activation of reporter gene expression with a mixture of tomato HsfA1 and HsfB1 (Bharti et al, unpublished data).

Searching in the *Arabidopsis* expressed sequence tag (EST) libraries, which were exclusively prepared from non-heat shock tissues, revealed ESTs for practically all cytoplasmic sHsps (Table 3). The only exception is Hsp17.6A-CI. Interestingly, no ESTs were found for the plastidial and mitochondrial representatives either. As expected, ESTs encoding class CI and class CII sHsps are detectable in libraries derived from RNAs from flower buds, green siliques, and dry seeds. However, surprisingly, some related ESTs were also found in rosette leaves. It is, of course, an open question, whether the presence of messenger RNAs also correlates with the accumulation of the corresponding proteins, which are usually not detectable in nonstressed leaves.

ACKNOWLEDGMENTS

We thank Lutz Nover (Frankfurt/Main, Germany) and Wilfried de Jong (Nijmegen, Netherlands) for many helpful discussions and comments during the preparation of the manuscript. The work was supported by grants from the Deutsche Forschungsgemeinschaft to K.-D.S. (SCHA 577/6-1) and by the Fonds der Chemischen Industrie. E.V. acknowledges research support from the National Institutes of Health (GM-42762) and the US Department of Agriculture (NRICGP) and support during preparation of this review from the National Science Foundation (PO-WRE Award), the Guggenheim Foundation, and the Dutch National Science Foundation.

REFERENCES

- Amin J, Mestrlil R, Schiller P, Dreano M, Voellmy R. 1987. Organization of the *Drosophila melanogaster* hsp70 heat shock regulation unit. *Mol Cell Biol* 7: 1055–1062.
- Asamizu E, Nakamura Y, Sato S, Tabata S. 2000. A large scale analysis of cDNA in *Arabidopsis*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Res* 7: 175–180.
- Bartling D, Bulter H, Liebeton K, Weiler EW. 1992. An *Arabidopsis thaliana* cDNA clone encoding a 17.6 kDa class II heat shock protein. *Plant Mol Biol* 18: 1007–1008.
- Basha EM, Waters ER, Vierling E. 1999. *Triticum aestivum* cDNAs homologous to nuclear-encoded mitochondrion-localized small heat shock proteins. *Plant Sci* 141: 93–103.
- Berengian AR, Parfenova M, McHaourab HS. 1999. Site-directed spin labeling study of subunit interactions in the α -crystallin domain of small heat-shock proteins—comparison of the oligomer symmetry in α A-crystallin, HSP 27, and HSP 16.3. *J Biol Chem* 274: 6305–6314.
- Boelens WC, Van Boekel MAM, de Jong WW. 1998. HspB3, the most deviating of the six known human small heat shock proteins. *Biochim Biophys Acta Protein Struct Mol Enzymol* 1388: 513–516.
- Bova MP, Ding LL, Horwitz J, Fung BKK. 1997. Subunit exchange of α A-crystallin. *J Biol Chem* 272: 29511–29517.
- Bova MP, McHaourab HS, Han Y, Fung BKK. 2000. Subunit exchange of small heat shock proteins: analysis of oligomer formation of α A-crystallin and Hsp27 by fluorescence resonance energy transfer and site-directed truncations. *J Biol Chem* 275: 1035–1042.
- Caspers GJ, Leunissen JAM, de Jong WW. 1995. The expanding small heat-shock protein family, and structure predictions of the conserved α -crystallin domain. *J Mol Evol* 40: 238–248.
- Chen Q, Osteryoung K, Vierling E. 1994. A 21-kDa chloroplast heat shock protein assembles into high molecular weight complexes in vivo and in organelle. *J Biol Chem* 269: 13216–13223.
- Chen Q, Vierling E. 1991. Analysis of conserved domains identifies a unique structural feature of a chloroplast heat shock protein. *Mol Gen Genet* 226: 425–431.
- Deisenhofer J. 1981. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry* 20: 2361–2370.
- de Jong WW, Caspers GJ, Leunissen JAM. 1998. Genealogy of the α -crystallin-small heat-stress superfamily. *Int J Biol Macromol* 22: 151–162.
- Ding L, Candido EPM. 2000. HSP25, a small heat shock protein associated with dense bodies and M-lines of body wall muscle in *Caenorhabditis elegans*. *J Biol Chem* 275: 9510–9517.
- Dudich IV, Zav'yalov VP, Pfeil W, Gaestel M, Zav'yalova GA, Denesyuk AI, Korpela T. 1995. Dimer structure as a minimum cooperative subunit of small heat-shock proteins. *Biochim Biophys Acta Protein Struct Mol Enzymol* 1253: 163–168.
- Ehrnsperger M, Graber S, Gaestel M, Buchner J. 1997. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO J* 16: 221–229.
- Ehrnsperger M, Lilie H, Gaestel M, Buchner J. 1999. The dynamics of Hsp25 quaternary structure. *J Biol Chem* 274: 14867–14874.
- Forreiter C, Kirschner M, Nover L. 1997. Stable transformation of an *Arabidopsis* cell suspension culture with firefly luciferase providing a cellular system for analysis of chaperone activity in vivo. *Plant Cell* 9: 2171–2181.
- Forreiter C, Löw D. 1998a. Cloning and characterization of two dif-

- ferent cDNAs coding for cytoplasmic small heat stress proteins in *Lycopersicon peruvianum*. *Plant Physiol* 117: 718.
- Forreiter C, Löw D. 1998b. Characterization of a cDNA coding for cytoplasmic class II small heat stress protein in *Lycopersicon peruvianum*. *Plant Physiol* 117: 1125.
- Haslbeck M, Walke S, Stromer T, Ehrnsperger M, White HE, Chen S, Saibil HR, Buchner J. 1999. Hsp26: a temperature-regulated chaperone. *EMBO J* 18: 6744–6751.
- Helm KW, LaFayette PR, Nagao RT, Key JL, Vierling E. 1993. Localization of small HSPs to the higher plant endomembrane system. *Mol Cell Biol* 13: 238–247.
- Helm KW, Lee GJ, Vierling E. 1997. Expression and native structure of cytosolic class II small heat-shock proteins. *Plant Physiol* 114: 1477–1485.
- Helm KW, Schmeits J, Vierling E. 1995. An endomembrane-localized small heat-shock protein from *Arabidopsis thaliana*. *Plant Physiol* 107: 287–288.
- Helm KW, Vierling E. 1989. An *Arabidopsis thaliana* cDNA clone encoding a low molecular weight heat shock protein. *Nucleic Acids Res* 17: 7995.
- Hightower L, Nover L. 1991. *Heat Shock and Development*. Springer, Berlin.
- Horwitz J. 1992. Alpha-crystallin can function as a molecular chaperone. *Proc Natl Acad Sci U S A* 89: 10449–10453.
- Jakob U, Gaestel M, Engel K, Buchner J. 1993. Small heat shock proteins are molecular chaperones. *J Biol Chem* 268: 1517–1520.
- Jinn TL, Yeh YC, Chen YM, Lin C-Y. 1989. Stabilization of soluble proteins in vitro by heat shock proteins-enriched ammonium sulfate fraction from soybean seedlings. *Plant Cell Physiol* 30: 463–469.
- Kaneko T, Katoh T, Sato S, Nakamura Y, Asamizu E, Kotani H, Miyajima N, Tabata S. 1999. Structural analysis of *Arabidopsis thaliana* chromosome 5, IX: sequence features of the region 1,011,550 bp covered by seventeen P1 and TAC clones. *DNA Res* 6: 183–195.
- Kaneko T, Katoh T, Sato S, Nakamura Y, Asamizu E, Tabata S. 2000. Structural analysis of *Arabidopsis thaliana* chromosome 3, II: sequence features of the 4,251,695 bp regions covered by 90 P1, TAC and BAC clones. *DNA Res* 7: 217–221.
- Kelley LA, MacCallum RM, Sternberg MJE. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* 299: 501–522.
- Kim KK, Kim R, Kim SH. 1998a. Crystal structure of a small heat-shock protein. *Nature* 394: 595–599.
- Kim R, Kim KK, Yokota H, Kim SH. 1998b. Small heat shock protein of *Methanococcus jannaschii*, a hyperthermophile. *Proc Natl Acad Sci U S A* 95: 9129–9133.
- Kirschner M, Winkelhaus S, Thierfelder J, Nover L. 2000. Transient expression and heat stress induced aggregation of endogenous and heterologous small heat stress proteins in tobacco protoplasts. *Plant J* 24: 397–412.
- Knauf U, Jakob U, Engel K, Buchner J, Gaestel M. 1993. Stress- and nitrogen-induced phosphorylation of the small heat shock protein Hsp25 by MAPKAP kinase 2 is not essential for chaperone properties and cellular thermoresistance. *EMBO J* 13: 54–60.
- Kobayashi R, Dallas PB, Pacchione S, Wilsker D, Bowrin V, Moran E. 2000. The human SWI-SNF complex protein p270 is an ARID family member with non-sequence-specific DNA binding activity. *Mol Cell Biol* 20: 3137–3146.
- Krief S, Faivre JE, Robert P, et al. 1999. Identification and characterization of cvHsp: a novel human small stress proteins selectively expressed in cardiovascular and insulin-sensitive tissues. *J Biol Chem* 274: 36592–36600.
- Lauzon LM, Helm K, Vierling E. 1990. A cDNA clone from *Pisum sativum* encoding a low molecular weight heat shock protein. *Nucleic Acids Res* 18: 4274.
- Lawrence SD, Cline K, Moore GA. 1997. Chromoplast development in ripening tomato fruit—identification of cDNAs for chromoplast-targeted proteins and characterization of a cDNA encoding a plastid-localized low-molecular-weight heat shock protein. *Plant Mol Biol* 33: 483–492.
- Lee GJ, Pokala N, Vierling E. 1995. Structure and in vitro molecular chaperone activity of cytosolic small heat shock proteins from pea. *J Biol Chem* 270: 10432–10438.
- Lee GJ, Roseman AM, Saibil HR, Vierling E. 1997. A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J* 16: 659–671.
- Lee GJ, Vierling E. 2000. A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiol* 122: 189–198.
- Lenne C, Block MA, Garin J, Douce R. 1995. Sequence and expression of the mRNA encoding Hsp22, the mitochondrial small heat-shock protein in pea leaves. *Biochem J* 311: 805–813.
- Leroux MR, Ma BJ, Batelier G, Melki R, Peter E, Candido M. 1997. Unique structural features of a novel class of small heat shock proteins. *J Biol Chem* 272: 12847–12853.
- Lin X, Kaul S, Rounsley SD, et al. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761–768.
- MacRae TH. 2000. Structure and function of small heat shock/alpha-crystallin proteins: established concepts and emerging ideas. *Cell Mol Life Sci* 57: 899–913.
- McElwain EF, Spiker S. 1989. A wheat cDNA clone which is homologous to the 17kd heat-shock protein gene family of soybean. *Nucleic Acids Res* 17: 1764.
- McHaourab HS, Berengian AR, Koteiche HA. 2000. Site-directed spin-labeling study of the structure and subunit interactions along a conserved sequence in the α -crystallin domain of heat-shock protein 27: evidence of a conserved subunit interface. *Biochemistry* 36: 14627–14634.
- Merck KB, Groenen PJTA, Voorter CEM, de Haard-Hoekman WA, Horwitz J, Bloemendal H, de Jong WW. 1993. Structural and functional similarities of bovine α -crystallin and mouse small heat-shock protein: a family of chaperones. *J Biol Chem* 268: 1046–1052.
- Morrow G, Inaguma Y, Kato K, Tanguay RM. 2000. The small heat shock protein Hsp22 of *Drosophila melanogaster* is a mitochondrial protein displaying oligomeric organization. *J Biol Chem* 275: 31204–31210.
- Moser D, Doenhoff MJ, Klinkert MQ. 1992. A stage-specific calcium-binding protein expressed in eggs of *Schistosoma mansoni*. *Mol Biochem Parasitol* 51: 229–238.
- Munchbach M, Nocker A, Narberhaus F. 1999. Multiple small heat shock proteins in Rhizobia. *J Bacteriol* 181: 83–90.
- Murakami S, Kondo Y, Nakano T, Sato F. 2000. Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. *FEBS Lett* 468: 15–18.
- Nakano T, Murakami S, Shoji T, Yoshida S, Yamada Y, Sato F. 1997. A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *Plant Cell* 9: 1673–1682.
- Nene V, Dunne DW, Johnson KS, Taylor DW, Cordingley JS. 1986. Sequence and expression of a major egg antigen from *Schistosoma mansoni*: homologies to heat shock proteins and alpha-crystallins. *Mol Biochem Parasitol* 21: 179–188.
- Newman T, deBruijn FJ, Green P, et al. 1994. Genes galore: a summary of methods for accessing results from large-scale partial

- sequencing of anonymous Arabidopsis cDNA clones. *Plant Physiol* 106: 1241–1255.
- Nguyen HT, Weng J, Joshi CP. 1993. A wheat (*Triticum aestivum*) cDNA clone encoding a plastid-localized heat-shock protein. *Plant Physiol* 103: 675–676.
- Nover L. 1991. *Heat Shock Response*. CRC Press, Boca Raton, FL.
- Nover L, Miernyk J. 2001. A genomic approach to the chaperone system of *Arabidopsis thaliana*. *Cell Stress Chaperones* 6: 173–174.
- Nover L, Scharf KD, Neumann D. 1983. Formation of cytoplasmic heat shock granules in tomato cell cultures and leaves. *Mol Cell Biol* 3: 1648–1655.
- Nover L, Scharf KD, Neumann D. 1989. Cytoplasmic heat shock granules are formed from precursor particles and are associated with a specific set of mRNAs. *Mol Cell Biol* 9: 1298–1308.
- Osteryoung KW, Sundberg H, Vierling E. 1993. Pola(A) tail length of heat shock protein RNA is increased by severe heat stress, but intron splicing is unaffected. *Mol Gen Genet* 239: 323–333.
- Petko L, Lindquist S. 1985. Hsp26 is not required for growth at high temperatures, nor for thermotolerance, spore development, or germination. *Cell* 45: 885–894.
- Plater ML, Goode D, Crabbe MJC. 1996. Effects of site-directed mutations on the chaperone-like activity of α B-crystallin. *J Biol Chem* 271: 28558–28566.
- Prändl R, Hinderhofer K, Eggers-Schumacher G, Schöffl F. 1998. Hsf3, a new heat shock factor from *Arabidopsis thaliana*, derepresses the heat shock response and confers thermotolerance when overexpressed in transgenic plants. *Mol Gen Genet* 258: 269–278.
- Sébastien M, Geneviève M, Julie M, Tanguay RM. 2001. *Drosophila* small heat shock proteins: cell and organelle-specific chaperones? In: *Small Stress Proteins*, ed Arrigo A-P and Müller WEG. Heidelberg, Springer-Verlag.
- Sun T-S, Das BK, Liang JJN. 1997. Conformational and functional differences between recombinant human lens α A- and α B-crystallin. *J Biol Chem* 272: 6220–6225.
- Suzuki A, Sugiyama Y, Hayashi Y, et al. 1998. MKBP, a novel member of the small heat shock protein family, binds and activates the myotonic dystrophy protein kinase. *J Cell Biol* 140: 1113–1124.
- Takahashi T, Komeda Y. 1989. Characterization of two genes encoding small heat-shock proteins in *Arabidopsis thaliana*. *Mol Gen Genet* 219: 365–372.
- Treisman JE, Luc A, Rubin GM, Heberlein U. 1997. Eyelid antagonizes wingless signaling during *Drosophila* development and has homology to the Bright family of DNA-binding proteins. *Genes Dev* 11: 1949–1962.
- Tucker PW, Herrscher RF, Kaplan MH, Lelsz DL, Das C, Scheuermann R. 1995. The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev* 9: 3067–3082.
- Tucker PW, Kortschak RD, Saint R. 2000. ARID proteins come in from the desert. *Trends Biochem Sci* 25: 294–299.
- Veinger L, Diamant S, Buchner J, Goloubinoff P. 1998. The small heat-shock protein IbpB from *Escherichia coli* stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. *J Biol Chem* 273: 11032–11037.
- Vierling E. 1991. The roles of heat shock proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol* 42: 579–620.
- Vierling E, Nagao RT, DeRocher AE, Harris LM. 1988. A heat shock protein localized to chloroplasts is a member of a eucaryotic superfamily of heat shock proteins. *EMBO J* 7: 575–581.
- Visioli G, Maestri E, Marmioli N. 1997. Differential display-mediated isolation of a genomic sequence for a putative mitochondrial LMW HSP specifically expressed in condition of induced thermotolerance in *Arabidopsis thaliana*(L.) Heynh. *Plant Mol Biol* 34: 517–527.
- Waters ER. 1995. The molecular evolution of small heat-shock proteins in plants. *Genetics* 141: 785–795.
- Waters ER, Lee GJ, Vierling E. 1996. Evolution structure and function of the small heat shock proteins in plants. *J Exp Bot* 47: 325–338.
- Waters ER, Vierling E. 1999. The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. *Mol Biol Evol* 16: 127–139.
- Wehmeyer N, Hernandez LD, Finkelstein RR, Vierling E. 1996. Synthesis of small heat shock proteins is part of the developmental program of late seed maturation. *Plant Physiol* 112: 747–757.
- Weng J, Wang ZF, Nguyen HT. 1991. A *Triticum aestivum* cDNA clone encoding a low-molecular weight heat shock protein. *Plant Mol Biol* 17: 273–275.
- Whitham SA, Anderberg RJ, Chrisholm ST, Carrington JC. 2000. Arabidopsis *RTM2* gene is necessary for specific restriction of Tobacco Etch Virus and encodes an unusual small heat shock-like protein. *Plant Cell* 12: 569–582.
- Willett DA, Basha EM, Vierling E. 1996. Nucleotide sequence of a cDNA encoding a mitochondrion-localized small Hsp from *Arabidopsis thaliana*: AtHsp23.6. *Plant Physiol* 112: 1399.
- Wotton D, Freeman K, Shore D. 1996. Multimerization of Hsp42p, a novel heat shock protein of *Saccharomyces cerevisiae*, is dependent on a conserved carboxyl-terminal sequence. *J Biol Chem* 271: 2717–2723.

Fig 4. Clustal analysis of the phylogenetic relationship based on the sequence comparison of the ACDs only. Members of the sHsp family are underlayed with gray boxes. Arrows point to the new members in *Arabidopsis*. For proteins with more than one ACD, the corresponding amino acid residues are given in brackets.

Class CI:											
AtHsp17.8-CI	50:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KQD-TW--HRVRS--SG--QFSRK--FLPENVM-DQKASM--ENG--VLINIVPK--	140							
AtHsp17.6A-CI	48:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KQD-TW--HRVRS--SG--QFSRK--FLPENVM-DQKASM--ENG--VLINIVPK--	138							
AtHsp17.6B-CI	48:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	138							
AtHsp17.4-CI	51:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	141							
AtHsp17.6C-CI	52:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	142							
AtHsp18.1-CI	54:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	144							
LpHsp17.7-CI	49:	IDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	139							
PaHsp18.1-CI	53:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	143							
TaHsp16.9A-CI	46:	VDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	136							
ZmHsp17.2-CI	47:	IDWKET-AAAHVFKADLP	GMKK--EE-VK--VE-ED-DSVLKIS--SRHVEKEE-----KND-TW--HRVRS--SG--QFTRR--FLPENVM-DQKASM--ENG--VLINIVPK--	137							
Class CI related:											
AtHsp15.4-CI(r)	28:	VRMSQS-PDSHTFVDLP	GLRK--EE-IX--VE-ED-SIYLIR-T--EATPMSP--DQPLK--TFKRE--FLPEISIDM-IGBAGY--ENG--VLINIVPK--	110							
AtHsp15.7-CI(r)	24:	IDWMS-NNSHIFKIN	VPQNK--ED-IX--VQ-EE-GNVLIR--GEIKE-KE--KEN-LVM--HVAEREAFSGGSEFLRR--FLPENVM-DQKASM--ENG--VLINIVPK--	119							
AtHsp18.5-CI(r)	62:	LNWTET-PTAHVFKAL	PGVDQ--DE-VI--AF-VDE-EGYLQCT--D--N--KFMRS--FLPNALTD-DQKASM--ENG--VLINIVPK--	134							
AtHsp21.7-CI(r)	89:	VDMSQT-DQAYVLKSD	IPVGR--NN-VQ--VY-VQINGRVMS--SNNSSKK--AATNSD--RSGRWSEH--GYVRR--FLPEISIDA-KYSFAFLNNNDYFLEIRPK--	185							
Class ER:											
AtHsp22.0-ER	73:	VDWKET-AEGHEIMLD	IPGLKK--DE-VK--IE-VSE-NGVLRS--SRKREEK--KGD-CW--HRVRS--YG--KFWRQ--FLPENVDL-DSKAKL--ENG--VLINIVPK--	163							
GmHsp22.0-ER	69:	VDWKET-PEGHVIMLD	IPGLKK--DE-VK--IE-VSE-NGVLRS--SRKREEK--KGD-HW--HRVRS--YG--KFWRQ--FLPENVDL-DSKAKL--ENG--VLINIVPK--	159							
LeHsp21.5-ER	69:	VDWKET-ARGHVISV	EPGLNK--DD-IX--IE-VSE-NGVLRS--SRKREEK--KGD-HW--HRVRS--YG--KFWRQ--FLPENADL-DQKASM--ENG--VLINIVPK--	161							
PaHsp22.7-ER	77:	VDWKET-PEGHVIMLD	IPGLKK--DD-IX--IE-VSE-NGVLRS--SRKREEK--KGD-HW--HRVRS--YG--KFWRQ--FLPENVDL-DSKAKL--ENG--VLINIVPK--	167							
Class CII:											
AtHsp17.6-CII	47:	ADVIEH-PNAYAFVDM	PGIKG--DE-IX--VQ-VEN-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	138							
AtHsp17.7-CII	48:	ADVIEH-PNAYAFVDM	PGIKG--DE-IX--VQ-VEN-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	139							
LpHsp17.3-CII	48:	ADVIEY-PNAYAFVDM	PGIKG--DE-IX--VQ-VEN-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	138							
PaHsp17.1-CII	45:	ADVIEH-PNAYAFVDM	PGIKG--DE-IX--VQ-VEN-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	135							
TaHsp17.4-CII	52:	ADVIEH-PNAYAFVDM	PGIKG--DE-IX--VQ-VEN-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	140							
ZmHsp17.0-CII	48:	ADVIEH-PNAYAFVDM	PGIKG--DE-IX--VQ-VEN-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	137							
Class CIII:											
AtHsp17.4-CIII	44:	IDILES-PKEYIFYL	IPGLSK--SD-LQ--VT-VSE-ERTLVKSN--KKKKDDDES--EBG-SKY--IRLERR--LAQ--NLVKK--FLPEISIDM-ASVTAKY--ENG--VLINIVPK--	139							
LpHsp16.1-CIII	35:	VDILT-PKEYIFYM	IPGLSK--SD-LQ--VS-VSD-ERTLVKSN--KKKKREEE--EBG-CKY--VRLERN--PPL--KLMRK--FLPEISIDM-ASVTAKY--ENG--VLINIVPK--	129							
Class P:											
AtHsp25.3-P	129:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	214							
LeHsp26.1-P	134:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	221							
PaHsp26.2-P	133:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	219							
TaHsp26.6-P	133:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	224							
ZmHsp26.4-P	133:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	227							
Class P related:											
AtHsp14.2-P(r)	28:	YEYET-KKSCVTRVDM	PGCP--SD-LT--YVVD-ANNVHF--ADEPAMPE--Y--ENAGR--KYGGSM--IPN-F-EAYDVKKTKVKL--ING--VLINIVPK--	110							
AtHsp26.5-P(r)	127:	GVVKEQDDCYK-LRYE	VGLTK--ED-VK--IT-VN-DGILITK--SOKHAKK--GSPEDD--YWS--SK-SYG--YYNTS--FLPENADL-DKSAVC--ENG--VLINIVPK--	218							
Class M:											
AtHsp23.6-M	109:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	195							
AtHsp23.5-M	111:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	195							
LeHsp23.8-M	113:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	195							
PaHsp22.9-M	103:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	187							
TaHsp23.5-M	116:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	199							
ZmHsp23.8-M	119:	WDIKKEEHIK-MRDM	PGLSK--ED-VK--IS-VS-DNLVMS--SRQRENK--NRG-VKY--VRMERR--MG--KFMRR--FLPENADL-DKSAVC--ENG--VLINIVPK--	203							

Consensus region II

Consensus region I

Fig 5. Sequence alignment of the ACD only. Highly conserved amino acid residues are marked with boxes. Closed arrowheads show the position of introns. Interesting is the high conservation between Acd members, forming more closely related subgroups. For further explanation, see Figure 1. Secondary structure predictions were made using the program 3D-PSSM (Kelley et al 2000). Amino acid residues fitting to the β -fold motifs in the ACD are given in boldface letters.

Proteins with one ACD:												
AtAc31.2	188:	ENVAES-THSYVAIV	ELFGASI--ND-IR--	VE-VD-DNLVIT	--GRRTSICQKVDAG--TKASILGYHKQBIQ--	PFFVS--	WPLESNVIV	DNVSAEF--MDG--	TLNIVPK	:	283	
AtAc39.0	239:	MDIGEC-DDAYLFR	SLPGVKRDY--FS--	CE-VE-DNGKVLIR	--GVITTTGGKR--VKRYS--HVFEMQ--	TRSLCPGP--	NFSVS--	FLPGGVPHPE--FSGNFG--TDG--	ILEGVITR	:	338	
AtAc51.9	360:	VDIGVN-KVAYFP	QALPGVRDYE--FS--	CE-IE-DNGKVLIR	--GVITTTGGKR--VKRYS--HVFEMQ--	TRSLCPGP--	NFSVS--	FLPGGVPHPE--FSGNFG--TDG--	ILEGVITR	:	459	
AtAc21.4	93:	VDIAES-EDSYFPR	ALPGVRDYE--FS--	CE-IE-DNGKVLIR	--GVITTTGGKR--VKRYS--HVFEMQ--	TRSLCPGP--	NFSVS--	FLPGGVPHPE--FSGNFG--TDG--	ILEGVITR	:	192	
AtAc15.5	37:	VDIGVS-EADYFPR	ALPGVRDYE--FS--	CE-IE-DNGKVLIR	--GVITTTGGKR--VKRYS--HVFEMQ--	TRSLCPGP--	NFSVS--	FLPGGVPHPE--FSGNFG--TDG--	ILEGVITR	:	135	
AtAc55.5	373:	KTYVED-EALYVIT	LEPVDNLVTK--VS--	MR-NLITNGILKVS	--CLSTRASRF--VKRRD--RTFKLV	DQMAEHCPGP--	EPMRE--	IQLNRIPEANIEAYF--DGTGP--	VLEIVPK	:	477	
AtAc55.8	364:	KTYIED-EGEYLII	SLPVDNLVTK--VS--	MR-NLITNGILKVS	--CLSTRASRF--VKRRD--RTFKLV	DQMAEHCPGP--	EPMRE--	IQLNRIPEANIEAYF--DGTGP--	VLEIVPK	:	468	
AtAc55.3	365:	KTYIED-DRGLI	VLMSLPVDNLVTK--VS--	MR-NLITNGILKVS	--CLSTRASRF--VKRRD--RTFKLV	DQMAEHCPGP--	EPMRE--	IQLNRIPEANIEAYF--DGTGP--	VLEIVPK	:	469	
AtAc16.9	51:	VAVEES-HDCFR	FAVTDG--GGGG--	VRAIMVE-IPGVTIK	LIVRTNSSLGLSLDELEL--	DVNR--	FLPESTRPEL	VTVDG--DGDG--	ELIVIVPK	:	139	
AtAc22.1	82:	VAVEER-HDCFR	FAVTDG--GGGG--	VRAIMVE-IPGVTIK	LIVRTNSSLGLSLDELEL--	DVNR--	FLPESTRPEL	VTVDG--DGDG--	ELIVIVPK	:	171	
AtAc28.7	33:	SLSRET-NSVFF	LILHLKPKK--DG--	IDI-EINK--EG--NLIKIS	--GRK--QVEEMVLV--	KWVKKKETEIK--	EPKKV--	FRIPDIVL--DKIKARF	DDDA--TLTIVPK	:	127	
AtAc216.9	1:					MMVI--	RWMNRKEVEFS--	FRIPDIVL--DKIKARF	DDDA--TLTIVPK	:	50	
AtAc41.3	23:	SEWKQ-PEATL	NLIDLPK--EQ--	KK-VTVHS--SKMIVT	--GR--PL--ANRMSR--	FNEV--	FLIPONCIV	--DKIHGSP--KNN--	VLINIVPK	:	103	
AtAc25.1	21:	SNWKE-QPEAL	TLIVLPRK--EQ--	KK-VTVHS--SKMIVT	--GR--PL--ANRMSR--	FNEV--	FLIPONCIV	--DKIHGSP--KNN--	VLINIVPK	:	100	
AtAc28.1	7:	TRWTE-PDAE	VLADLPK--EQ--	KK-VTVHS--SKMIVT	--GR--PL--ANRMSR--	FNEV--	FLIPONCIV	--DKIHGSP--KNN--	VLINIVPK	:	96	
AtAc16.6	18:			GLKK--EH-LK--	IQ-LIN--SGVLIT--	SCC--PVQCTKIR--	FMIKE--	TKWAKDKR--NEAKRF--	SKG--VLYNIVPK	:	70	
AtAc54.2	30:	AQWMS-GSIF	LYVNLFGYR--DO--	IE-LK-KOERTIV	QIQ--GR--PLSAQT	KAR--	FSEA--	YRVDTCDM--TKLSTF--	BHG--LLTIRFPA	:	110	
AtAc55.2	410:	INVRET-KDCFE	IFALVGLLR--EE--	VR--VQ-SDP-AGRLVIA	--QPEQLDN--PWGIT--	PFFK--	VNFPARID	LPHTSVAVS--LHG--	RLPVRVFP	:	492	
AtAc44.3	314:	INVKEE-KDSFE	IFALVGLLR--EE--	VR--VQ-SDP-AGRLVIA	--QPEQLDN--PWGIT--	PFFK--	VNFPARID	LPHTSVAVS--LHG--	RLPVRVFP	:	396	
AtAc32.4	201:	INVQRT-QDCFE	YALVGLVLR--EE--	VR--VQ-SDP-AGRLVIA	--QPEQLDN--PWGIT--	PFFK--	VNFPARID	LPHTSVAVT--LNG--	QLFVRVPL	:	283	
Proteins with more than one ACD:												
AtAc25.4 (ACD1)	33:	LKILAN-GDMF	--VRMDLPGLVPD--EGGMS--	VY-HNRSEES	VVVY--AKAPKIH	THDSTER--	RYLTMG--	IGSCCAISSMTHM--	SDG--VFRLLTSK	:	117	
AtAc26.0 (ACD1)	36:	LKILAN-EDMF	--LRMDLPGLVPD--EGGMS--	VY-HNRSEES	VVVY--AKAPKIH	THDSTER--	RYLTMG--	IGSCCAISSMTHM--	SDG--VFRLLTSK	:	120	
AtAc25.4 (ACD2)	149:	SVILPN-GKLF	--VRADMPGLVPK--EN--FT--	VS-VT--NGRVVT	--GRAP--ALSHDSSGR--	FYSGDG--	AMLSTP	FDIPS--RRIKIA--KNG--	VIRLLIIP	:	231	
AtAc26.0 (ACD2)	156:	SKQLPN-GKLF	--VRADMPGLVPK--EN--FT--	VS-VT--NGRVVT	--GRAP--ALSHDSSGR--	FYSGDG--	AMLSTP	FDIPS--RRIKIA--KNG--	VIRLLIIP	:	238	
AtAc81.4 (ACD5)	660:	SKQLPN-GGLY	--VRADMPGLVPK--EN--FT--	VS-VT--NGRVVT	--GRAP--ALSHDSSGR--	FYSGDG--	AMLSTP	FDIPS--RRIKIA--KNG--	VIRLLIIP	:	742	
AtAc114.3 (ACD3)	964:	TKQLSN-GGLY	--MRVDMFGVPS--EK--FM--	VA-VDG--DGVVITM--	GRAP--VIMHDS	SGR--TVAK--	VANVPR	GDG--GRILVP--KHG--	VIRLLIIP	:	1045	
AtAc81.4 (ACD3)	405:	TKQLSN-GGLY	--LRIDMFGVPS--DG--FI--	VA-VDG--NGVVITM--	GRAP--ATMHD	SG--FYAMNFP	QANGKGFAP--	FKYLED	THL--FVRLDLP	:	488	
AtAc114.3 (ACD2)	765:	TKQLSN-GGLY	--LRIDMFGVPS--DG--FI--	VA-VDG--NGVVITM--	GRAP--ATMHD	SG--FYAMNFP	QANGKGFAP--	FKYLED	THL--FVRLDLP	:	847	
AtAc81.4 (ACD1)	33:	AKTLN-ESY	--IRPDLGPVPR--DO--VT--	VL-LDPNNT	VIL--ADAPKEHKYDSE--	RNYDGT--	KLRLPKIS	VGGINTIK--PKCCCE	FPFGTSHM--ADG--	VLRVLT	:	127
AtAc81.4 (ACD4)	514:	PKMLN-EDMY	--VRIDFPGVPR--DO--VR--	VF-LDQSKA	VCIP--ADAPKEHKYDSE--	RNYDGT--	KLRLPKIS	VGGINTIK--PKCCCE	FPFGTSHM--ADG--	VLRVLT	:	597
AtAc81.4 (ACD2)	188:	PKYLETHDLF	--VRIDFPGVPR--ES--VI--	IL-LPESK	AVIT--GRAPKEHKYDSE--	RNYDGT--	KLRLPKIS	VGGINTIK--PKCCCE	FPFGTSHM--ADG--	VLRVLT	:	272
AtAc114.3 (ACD1)	548:	PKYLETHDLF	--VRIDFPGVPR--ES--VI--	IL-LPESK	AVIT--GRAPKEHKYDSE--	RNYDGT--	KLRLPKIS	VGGINTIK--PKCCCE	FPFGTSHM--ADG--	VLRVLT	:	632
AtAc28.9 (ACD1)	39:	VKVLN-DNLV	--VRVDLPGLVPD--DA--VR--	YR-VDAVR	KLVPF--GETLANNK--EG--	VRQYSGT--	AGLQDCC	CEITGVDAK--ENG--	VLRMLTR	:	121	
AtAc28.9 (ACD2)	166:	IGVTRAGGGA	--FAVDMFGVPS--DD--VE--	VY-AD--EKEIR	PK--ATKIDVYHDESG--	RTYLSGVQ--	SFFPALISN	--NTIAMD--EFG--	VLRVLT	:	249	
Consensus region II												
Consensus region I												

Fig 5. Continued.