Volume 323, number 1,2, 51–54 FEBS 12444 May 1993

Occurrence and sequence of a DnaJ protein in plant (*Allium porrum*) epidermal cells

J.-J. Bessoule

Plant Research Laboratory, Michigan State University, MI. USA

Received 1 December 1992; revised version received 3 February 1993

Antibodies raised against a purified fraction from microsomal membranes of leek epidermal cells were used to screen a λzap expression library from epidermal cells of leek plants. A near full-length clone was isolated. This cDNA contains an open reading-frame of 1,191 bp coding for a DnaJ protein (leek DNAJ 1 or LDJ1). Leek DnaJ1 represents the second protein of this type described in a pluricellular organism, the first being that sequenced from human cells.

Plant; *Allium porrum*; DnaJ protein; Chaperone; Heat-shock protein; Prenylation; Zinc finger

1. INTRODUCTION

The first DnaJ protein to be sequenced [1] was first identified by the isolation of *E. coli* mutants unable to propagate bacteriophage λ. This protein is also involved in the replication of the plasmid P1 [2] and some recent studies have shown how this protein could be implicated in DNA replication [2–5]: for example, DnaJ protein interacts with DnaK (another heat-shock protein) to induce the active form of the P1 RepA initiator protein [3].

DnaJ protein also acts with DnaK protein and chaperonins (GroEL and GroES) in the folding of proteins in the cell: for example, DnaK, DnaJ and GroEL can interact with rhodanese and prevent its aggregation (induced by a denaturation) [6].

DnaJ proteins also seem to be implicated in the targeting of proteins in the cell: in yeast, two DnaJ proteins (NPL1 and SCJ1) have been isolated by analyzing mutants defective in the import of proteins to the nucleus [7,8]. One of them (NPL1) is encoded by a gene allelic to *sec63*, a gene that affects the transit of nascent secretory proteins across the endoplasmic reticulum membrane [7]. Another DnaJ protein from yeast (YDJ1P) was isolated by complementation of a mutant defective in mitochondrial protein import [9].

Seven DnaJ proteins have been sequenced from human cells [10], yeast [8,9,11,12] and prokaryotes [1,13]. No report has as yet described the presence of such proteins in plants.

Correspondence address: IBC-CNRS, 1 rue C. Saint-Saëns, 33077 Bordeaux Cedex, France.

Abbreviations. LB, Luria-Bertani medium (Bactotryptone 10 g, Bacto yeast extract 5 g, NaCl 10 g, H_2O 1 l; IPTG, isopropylthio-β-D-galactoside; PBS, phosphate-buffered saline; BCIP, 5-bromo-4-chloro-3-indolyl phosphate, disodium salt; NBT, Nitroblue tetrazolium chloride.

2. MATERIALS AND METHODS

2.1. *Preparation of antiserum*

The antiserum used was raised against the acyl-CoA elongating system from leek epidermal cell microsomes as described previously [14]. Before screening the library, the serum was diluted 10-fold and successively incubated for 30 min with 4 nitrocellulose filters (preincubated with an *E. coli* lysate) at room temperature. The serum was then diluted a further 10-fold and incubated with eight other filters. Just before use, it was further diluted 5-fold. This serum cross-reacted with a DnaJ protein during the experiments described below.

2.2. *Cloning of the LDJ1 gene*

An expression library of leek epidermal cells in λzap phage was purchased from Stratagene (La Jolla, CA, USA). An aliquot of a phage stock solution (titrated for confluence) was incubated at 37°C for 15 min with 600 μl of an overnight culture of *E. coli* (strain Sure) grown in LB supplemented with 1 mM $MgSO_4$. 3 ml of top agar (pre-warmed to 50°C) was then added and the mixture was plated out on LB plates and incubated at 42°C for 4 h. The plates were then overlayed with nitrocellulose filters pre-soaked with 10 mM IPTG and incubated overnight at 37°C. After washing in PBS, filters were incubated with BSA (1%) in PBS, 0.1% Triton, 0.2% sodium azide, and then with the antiserum prepared as described above (1/500). After 90 min at room temperature, the filters were washed in PBS, 0.1% Triton, 0.2% sodium azide, and incubated with anti-IgG alkaline phosphatase conjugate (1/5,000) for 1 h at room temperature. The protein–antibody complexes were visualized using BCIP (0.15 mg/ml) and NBT (0.3 mg/ml).

After purification of the 'positive' phages and in vivo excision (protocol from Stratagene), the 'packaged' pBluescript DNA was mixed with *E. coli* cells (strain XL1Blue) and spread on LB ampicillin + tetracycline plates to produce colonies. From these colonies, pBluescript plasmids containing the cDNA were further purified on a CsCl gradient.

2.3. *DNA sequencing*

Using the universal M13 and reverse M13 sequencing primers, as well as various other specifically synthesized oligonucleotide primers,

```
   1  AAA AAC GCG TCA CCA GAC GAT TTG AAG AAG GCA TAT CGA AAG GCT GCG    48
   1  Lys Asn Ala Ser Pro Asp Asp Leu Lys Lys Ala Tyr Arg Lys Ala Ala    16

  49  ATT AAG AAT CAT CCT GAT AAA GGT GGC GAT CCC GAG AAG TTT AAG GAG    96
  17  Ile Lys Asn His Pro Asp Lys Gly Gly Asp Pro Glu Lys Phe Lys Glu    32

  97  TTG GCT CAA GCT TAT GAT GTT CTA AGT GAC CCT GAA AAG CGT GAG ATA   144
  33  Leu Ala Gln Ala Tyr Asp Val Leu Ser Asp Pro Glu Lys Arg Glu Ile    48

 145  TAT GAT CAG TAT GGC GAG GAT GCT CTT AAG GAA GGA ATG GGT GGA GGC   192
  49  Tyr Asp Gln Tyr Gly Glu Asp Ala Leu Lys Glu Gly Met Gly Gly Gly    64

 193  GGT GGA GAC CAT GAT CCT TTC GAC ATC TTT CAG TCG TTC TTT GGT GGT   240
  65  Gly Gly Asp His Asp Pro Phe Asp Ile Phe Gln Ser Phe Phe Gly Gly    80

 241  GGA GGT TTT GGA GGT GGT GGT AGC AGC AGG GGC CGC AGA CAA AGG AGA   288
  81  Gly Gly Phe Gly Gly Gly Gly Ser Ser Arg Gly Arg Arg Gln Arg Arg    96

 289  GGT GAA GAT GTG GTT CAT CCT CTC AAA GTA TCT CTC GAG GAA CTT TAC   336
  97  Gly Glu Asp Val Val His Pro Leu Lys Val Ser Leu Glu Glu Leu Tyr   112

 337  AAT GGA ACT TCT AAG AAA CTC TCT TTG TCT AGA AAT GTT ATC TGC TCA   384
 113  Asn Gly Thr Ser Lys Lys Leu Ser Leu Ser Arg Asn Val Ile Cys Ser   128

 385  AAG TGC AAT GGC AAA GGA TCA AAA TCA GGT GCT TCA ATG AGA TGT GCA   432
 129  Lys Cys Asn Gly Lys Gly Ser Lys Ser Gly Ala Ser Met Arg Cys Ala   144

 433  TCT TGC CAA GGT TCT GGT ATG AAA GTT TCT ATT CGC CAG TTG GGT CCT   480
 145  Ser Cys Gln Gly Ser Gly Met Lys Val Ser Ile Arg Gln Leu Gly Pro   160

 481  GGA ATG ATT CAG CAG ATG CAG CAT CCT TGC AAT GAC TGT AAA GGC ACA   528
 161  Gly Met Ile Gln Gln Met Gln His Pro Cys Asn Asp Cys Lys Gly Thr   176

 529  GGA GAA ATG ATA AAT GAT AAG GAT AGG TGT CCA TTG TGC AAA GGT GAA   576
 177  Gly Glu Met Ile Asn Asp Lys Asp Arg Cys Pro Leu Cys Lys Gly Glu   192

 577  AAG GTC GTG CAA GAG AAG AAG GTT TTA GAA GTG CAT GTT GAG AAA GGG   624
 193  Lys Val Val Gln Glu Lys Lys Val Leu Glu Val His Val Glu Lys Gly   208

 625  ATG CAG AAT GGG CAG AGA ATT ACA TTC CCT GGC GAA GCT GAT GAA GCG   672
 209  Met Gln Asn Gly Gln Arg Ile Thr Phe Pro Gly Glu Ala Asp Glu Ala   224

 673  CCA GAT ACA GTT ACT GGG GAC ATT GTC TTT GTT CTA CAG CAG AAA GAA   720
 225  Pro Asp Thr Val Thr Gly Asp Ile Val Phe Val Leu Gln Gln Lys Glu   240

 721  CAT CCC AAA TTC CAA AGG AAA GGG GAT GAT TTA TTT TAT AAA CAT ACC   768
 241  His Pro Lys Phe Gln Arg Lys Gly Asp Asp Leu Phe Tyr Lys His Thr   256

 769  CTT TCT CTC ACT GAG GCC CTT TGC GGT TTC CAG TTT GTG TTA ACT CAC   816
 257  Leu Ser Leu Thr Glu Ala Leu Cys Gly Phe Gln Phe Val Leu Thr His   272

 817  TTG GAT GGC AGG CAA CTC CTT ATC AAG TCT AAC CCT GGA GAG GTG GTT   864
 273  Leu Asp Gly Arg Gln Leu Leu Ile Lys Ser Asn Pro Gly Glu Val Val   288

 865  AAG CCA GAT CAA TTC AAG GCG ATC AAT GAC GAA GGA ATG CCA ATG TAT   912
 289  Lys Pro Asp Gln Phe Lys Ala Ile Asn Asp Glu Gly Met Pro Met Tyr   304

 913  CAA AGG CCA TTC ATG AGG GGG AAG TTG TAC ATC CAG TTC TTG GTT GAT   960
 305  Gln Arg Pro Phe Met Arg Gly Lys Leu Tyr Ile Gln Phe Leu Val Asp   320

 961  TTC CCC GAT TCG CTT ACC CCA GAC CAG TGC AAA GTG ATT GAA AGC GTG  1008
 321  Phe Pro Asp Ser Leu Thr Pro Asp Gln Cys Lys Val Ile Glu Ser Val   336

1009  CTT CCT AGA AGT GCC TCT TCT CAG CTA ACA GAC ATG GAG ATC GAT GAA  1056
 337  Leu Pro Arg Ser Ala Ser Ser Gln Leu Thr Asp Met Glu Ile Asp Glu   352

1057  TGT GAA GAA ACG ACC ATG CAT GAT GTG AAC ATA GAA GAG GAG ATG AGG  1104
 353  Cys Glu Glu Thr Thr Met His Asp Val Asn Ile Glu Glu Glu Met Arg   368

1105  AGG AAG CAA CAT CAG CAC GCA CAG GAG GCT TAC GAT GAG GAT GAT GAA  1152
 369  Arg Lys Gln His Gln His Ala Gln Glu Ala Tyr Asp Glu Asp Asp Glu   384

1153  GGT CAT GGC GGT GGT CAG AGG GTG CAA TGT GCT CAG CAG TGA GAG CAT  1200
 385  Gly His Gly Gly Gly Gln Arg Val Gln Cys Ala Gln Gln ***

1201  TGT CTT TGC T
```

Fig. 1. Nucleotide and deduced amino acid sequence of LDJ1. The number 1 in the nucleotide sequence indicates the beginning of the clone, and the stop codon is marked by ***.

the clone was sequenced by the Sanger dideoxy chain termination method [15], using [^{35}S]dATP as the label. The *Hin*cII- and *Bst*XI-digested fragments from the purified cDNA were subcloned into pBluescript and were used to determine the complete sequence. Dideoxynucleotide triphosphates were purchased from Pharmacia and the Sequenase 2.0 DNA Sequencing Kit from United States Biochemical Corp.

3. RESULTS AND DISCUSSION

Using antibodies against a purified fraction from microsomal membranes of leek epidermal cells and screening a λzap expression library from these cells, a clone coding for a polypeptide of 397 amino acids (44 kDa) was isolated (see Fig. 1). The sequence of the first 52 amino acids (N-terminal) of the protein shares homology with the consensus sequence of the previously sequenced DnaJ proteins (Fig. 2). For example, a 'HPD (-K in eukaryotes)' box is present in the eight DnaJ proteins, and sequences such as K-K/S-A-Y-R-K, EKFKE are found in both yeast DnaJ and LDJ1 proteins. Between amino acid number 77 and 87 there is a Gly(-Phe)-rich region that has also been described in several DnaJ proteins (see Table I).

Moreover, as in the case of DnaJ proteins from bacteria and in two of the four DnaJ proteins from yeast (Table I), LDJ1 contains four Cys-X-X-Cys-X-Gly-X-Gly units (underlined in Fig. 1) which can be organized into two series of direct repeats of the sequence Cys-X-X-Cys-X-Gly-X-Gly-$(X)_8$-Cys-X-X-Cys-X-Gly-X-Gly. These two series are the same as those described in YDJ1P and, as in the latter protein, there is also a missing final Gly, which is replaced by Lys in LDJ1. As noted by Caplan and Douglas [11], these units are similar to those described in the family of zinc finger proteins [16]. To our knowledge this is the first description of such units in plants.

Among all the known DnaJ proteins, the homology of LDJ1 is highest with YDJ1P (see [11,13]). In addition to the Gly(-Phe)-rich regions and the Cys-X-X-Cys-X-Gly-X-Gly units, there are (between amino acids 23 and 75 of YDJ1P) 34 identical and 14 isologous amino acids. After the zinc finger domain, the percent homology remains high: between amino acids 210 and 409 of YDJ1P, 80 were also found in LDJ1.

Furthermore, the C-terminal sequence of LDJ1 is VQCAQQ, which is very similar to the C-terminal sequence of YDJ1P: VQCASQ. As first suggested [11], and recently shown [17] by Caplan and co-authors, such a sequence corresponds to a farnesylation site in YDJ1P. The prenylation, which occurs on the cysteine residue is required for the protein to function at elevated temperature and to promote the attachment of proteins to the membranes [17,18]. Since the C-terminal sequence of LDJ1 is very similar to the C-terminal sequence of YDJ1, and since the C-terminal amino acid is a glutamine residue, it is possible that LDJ1 is prenylated and that the latter prenyl group is a farnesyl isoprenoid, rather than a geranylgeranyl group (see [19]). In any case, LDJ1 could be a membrane-bound protein (as seems to be the case for some other DNAJ proteins [1,11,18,20]) despite the absence of large hydrophobic regions (data not shown).

Some prenylated proteins (ras or GTP-binding proteins) are localized in the plasma membrane, while others (rab/YPT1) are located in the intracellular membranes (see [19] for review). Unfortunately, since the reported cDNA is not full length, it is unclear whether or not the protein is synthesized with an N-terminal

	n=	n … x
LeekDnaJ	?	K N A S P D D L K K A Y R K A A I K N H P D K - -
E coliDnaJ	1	M A K Q - D Y Y E I L G V S K T A E E R E I R K A Y K R L A M K Y H P D R N Q
MtDnaJ	5	E W V E K D F Y Q E L G V S S D A S P E E I K R A Y R K L A R D L H P D A N P
YDJ	1	M V K E T K F Y D I L G V P V T A T D V E I K K A Y R K C A L K Y H P D K N P
SIS	1	M V K E T K L Y H L L G V S P S A N E Q E L K K G Y R K A A L K Y H P D K - P
Sec 63	120	A T K L F O P Y E I L G I S T S A S D R D I K S A Y R K L S V K F H P D K L A
SCJ1	48	I L A D Q - Y Y A I L E I D K D A T E K E I K S A Y R Q L S K K Y H P D K N A
HDJ1	1	M G K - - D Y Y Q T L G - Q A A A L G R G D Q A G L P P P G L R Y H P D K - P

	x+1 …
LeekDnaJ	G G D P - - - E K F K - E - - - - - - L A Q A Y D V L S D P E K R E I Y D Q Y.
E coliDnaJ	- G D K E A E A K F K - E - - - - - - I K E A Y E V L T D S Q K R A A Y D Q Y.
MtDnaJ	- G N P A A G E R F K - A - - - - - - V S E A H N V L S D P A K R K E Y D E T
YDJ	- - S E E A A E K F K - E - - - - - - A S A A Y E I L S D P E K R D I Y D Q F
SIS	T G D T - - - E K F K - E - - - - - - I S E A F E I L N D P Q K R E I Y D Q Y
Sec 63	K G L T P D - E K S V M E E T Y V Q I T K A Y E S L T D E L V R Q N Y L K Y
SCJ1	- G S E E A H Q K F I - E - - - - - - V G E A Y D V L S D P E K K K I Y D Q F
HDJ1	T G D T - - - E K F K - E - - - - - - I S E A F E I L N D P Q K R E I Y D Q Y

Fig. 2. Regions of known DnaJ proteins with high homologies Bold letters represent amino acids present in at least 4 of the 8 reported sequences; n is the rank in the protein sequence (starting from the N-terminal) of the first amino acids of the presented region. Dashes were introduced to optimize the homologies between the proteins; A---B signifies that there are no amino acids present between amino acids A and B in the sequence of the protein.

Table I

Some characteristics of DNAJ proteins

Name	*E. coli* DnaJ	MtdnaJ	NPL1/ Sec63	SIS	YDJ1	SCJ1	HDJ1	LDJ1
Origin	*E coli*	*Myc tub.*	Yeast	Yeast	Yeast	Yeast	Human	Leek
Number of amino acids	375	356	663	352	409	404	339	> 397
Heat-shock protein	Yes	?	?	Yes	Yes	?	?	?
Glycine-rich region	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Number of zinc fingers	4	4	0	0	4	4	0	4
C-terminal					VQCASQ (farnesyl)	KDEL		VQCAQQ (farnesyl?)
Reference	1	13	7	12	9,11	8	10	This study

signal peptide and, therefore, from looking at the sequence, it is difficult to address the subcellular localization of the corresponding protein. Nevertheless, it can be noted that the high homology of LDJ1 with other DnaJ proteins that have less than 410 amino acids (except NPL1/Sec63), and especially with YDJ1P (409 amino acids) (Table I), strongly suggests that the sequenced cDNA (coding for 397 amino acids) is nearly full length.

Acknowledgements: This work was supported by CNRS and NSF and was carried out in the Plant Research Laboratory. The advice and facilities provided by Prof. C.R. Somerville (Plant Research Laboratory; Michigan State University) are greatly appreciated. The helpful reading by Dr. M.A. Heape is gratefully acknowledged.

REFERENCES

[1] Bardwell, J.C.A., Tilly, K., Craig, K., King, J., Zylicz, M. and Georgopolous, C. (1986) J. Biol. Chem. 261, 1782–1785.
[2] Wickner, S H. (1990) Proc Natl. Acad. Sci. USA 87, 2690–2694
[3] Wickner, S., Hoskins, J. and McKenney, K. (1991) Nature 350, 165–167.
[4] Zylicz, M., Ang, D., Liberek, K. and Georgopoulos, C. (1989) EMBO J. 8, 1601–1608.
[5] Liberek, K., Marzalek, J., Ang, D., Georgopoulos, C. and Zylicz, M. (1991) Proc. Natl. Acad. Sci. USA 88, 2874–2878.
[6] Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K. and Hartl, F.L. (1992) Nature 356, 683–689.
[7] Sadler, L., Chiang, A., Kurihara, T., Rothblatt, J., Way, J. and Silver, P. (1989) J. Cell Biol. 109, 2665–2675.
[8] Blumberg, H. and Silver, P A. (1991) Nature 349, 627–630.
[9] Atencio, D.P. and Yaffe, M.P. (1992) Mol Cell. Biol. 12, 283–291.
[10] Raabe, T. and Manley, J.L. (1991) Nucleic Acids Res. 19, 23.
[11] Caplan, A.V. and Douglas, M.G. (1991) J Cell Biol. 114, 609–621.
[12] Luke, M.M., Sutton, A and Arndt, K.T. (1991) J. Cell Biol 114, 623–638.
[13] Lathigra, R B., Young, D.B., Sweester, D. and Young, R.A. (1988) Nucleic Acids Res. 16, 1636.
[14] Bessoule, J.-J., Creach, A., Lessire, R. and Cassagne, C. (1992) Biochim. Biophys Acta 1117, 78–82.
[15] Sanger, S., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[16] Berg, J.M. (1990) J. Biol. Chem. 265, 6513–6516.
[17] Caplan, A.V., Tsai, J., Casey, P.J. and Douglas, M G. (1992) J. Biol. Chem. 267, 18890–18895.
[18] Glomset, J.A., Gelb, M.H. and Farnsworth, C.C. (1990) Trends Biochem. Sci. 139–142.
[19] Casey, P.J. (1992) J. Lipid Res. 33, 1731–1740.
[20] Zylicz, M., Yamamoto, T., McKittrick, N., Sell, S. and Georgopoulos, C. (1985) J. Biol Chem. 260, 7591–7598.