Plant Molecular Biology **18**: 467–476, 1992.
© 1992 *Kluwer Academic Publishers. Printed in Belgium.*

Heat shock Hsp70 protein is chloroplast-encoded in the chromophytic alga *Pavlova lutherii*

Carol D. Scaramuzzi, Harold W. Stokes and Roger G. Hiller
School of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

Received 29 May 1991; accepted in revised form 2 October 1991

Key words: chloroplast DNA, chromophytic alga, heat shock protein, Hsp70, *Pavlova lutherii*

Abstract

Heat shock proteins are ubiquitous and highly conserved. Recently they have become implicated in the import of proteins into organelles. All the heat shock genes characterized to date, however, are known or assumed to be encoded in the nuclear genome even if the corresponding protein can be localised in the mitochondrion or chloroplast. In contrast, we identify here an *hsp70* gene in the unicellular chromophytic alga *Pavlova lutherii* which is located on the chloroplast genome. Localisation of this gene to the chloroplast chromosome is confirmed by Southern blot analysis and pulse-field gel electrophoresis which also reveals that the length of the *P. lutherii* chloroplast chromosome is 115 kb. We compare the predicted protein of this *hsp70* gene with that of maize and of the analogous proteins in the prokaryotic organisms *Escherichia coli* and *Synechocystis* PCC6803. The greatest identity is found with the cyanobacterium *Synechocystis* PCC6803.

Introduction

The heat shock protein Hsp70 belongs to the major family of highly conserved proteins that are synthesized as a response to stress [16]. The heat shock response is probably ubiquitous and *hsp70* genes or the prokaryotic equivalent, *dnaK*, have been found in organisms from bacteria [1] to man [14]. Recently heat shock proteins, including Hsp70, have been implicated in the import of polypeptides to the endoplasmic reticulum, mitochondrion and chloroplast [5, 15, 19]. Import of heat shock proteins into organelles is presumed to be associated with their recognised role as molecular chaperones, which are proteins involved in the transport of proteins across organelle membranes [12, 24]. As part of this function, they act to ensure the correct folding of certain polypeptides and their assembly into functional oligomers without themselves forming part of the assembled product [12]. Several molecular chaperones are known to be associated with the heat shock response or constitutively expressed homologues of heat shock proteins [12, 15, 24].

Although Hsp70 homologues of eukaryotes are present in organelles, their genes are thought to be

The nucleotide sequence reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X59555.

located in the nucleus and indeed all previously identified eukaryotic *hsp70* genes are nuclear-encoded [8, 22]. For example, three Hsp70 homologues have recently been identified in the chloroplast of *Pisum sativum* [19], although the genes coding for these proteins are presumed to be located in the nucleus. The evidence for this is indirect. Firstly, the chloroplast genomes of both *Marchantia polymorpha* [21] and *Nicotiana tabacum* [28] have been completely sequenced and do not contain any region encoding a protein related to Hsp70. Secondly, a yeast mitochondrial Hsp70 homologue, Ssc1p, has been shown to be nuclear-encoded [8], as have the *hsp70* genes in maize [26] and petunia [31].

Apart from Hsp70, many chloroplast proteins are nuclear-encoded and are subsequently transported into the chloroplast [29]. *Pavlova lutherii* is a chromophytic alga belonging to the phylum Prymnesiophyta. The chloroplasts of these algae are surrounded by four membranes, the outermost of which is continuous with the nuclear envelope. Furthermore, vesicles have been observed between these membranes [10] which have been suggested [11] to have a role in the import of nuclear-encoded light-harvesting proteins. During sequencing of clones derived from *P. lutherii* chloroplast DNA, an open reading frame coding for an Hsp70 homologue was found. Here we present the sequence of this *hsp70* gene together with unequivocal evidence for it being located on the chloroplast chromosome.

Materials and methods

Strains, bacteriophages and plasmids

The host strain HB101 [18] was used for all manipulations. Chloroplast recombinants were generated using the vector pBR322 [3]. pZmc37 which contains a 2500 bp chloroplast fragment coding for a maize ribulose-1,5-bisphosphate carboxylase (Rubisco) large subunit (LSU) has been described [2]. pMAQ801 (this study) contains an 8.5 kb *P. lutherii* chloroplast DNA fragment cloned in the *Bam* HI site of pBR322 and includes the *hsp70* gene. pMAQ807 (this study) contains a 9.5 kb *Bgl* II chloroplast DNA fragment cloned into the *Bam* HI site of pBR322. This fragment encodes the *P. lutherii rpoB* and *rpoC1* genes. For DNA sequencing, fragments were subcloned into the M13 bacteriophage vectors mp18 or mp19 [32].

DNA cloning and fragment purification

Manipulations of DNA were performed according to standard protocols [18] or to the manufacturers' instructions when using DNA modifying enzymes. DNA fragments for subcloning into M13 or for use in hybridisation analysis were gel-purified using GeneClean II (BIO 101, La Jolla, CA, USA).

DNA sequencing

Sequencing was performed by the chain termination method [27] with the use of a Sequenase kit (United States Biochemical Corporation). Sequencing was done using both single- and double-stranded templates according to the manufacturers' protocols. Primers used were either the universal primer supplied or synthetic oligonucleotides synthesised on a Pharmacia Gene Assembler MarkII. Oligonucleotides were designed to provide extensive overlap between gels and, where necessary, compressions were resolved by the substitution of dITP for dGTP.

Nucleic acid hybridisation

DNA fractionated on agarose gels was transferred onto Zeta-probe nylon membrane (Bio-Rad) by alkaline diffusion according to the manufacturers' instructions. DNA for dot blotting was prepared with a BioRad dot blot apparatus by direct application to the membrane after denaturation. Probes for hybridisation were made radioactive by oligo-labelling using a kit and protocols provide by Bresatec (SA, Australia). Hybridisations were routinely carried out at 65 °C in

the presence of Dextran sulphate (10% w/v), 4× SSPE (20× SSPE = 3.6 M NaCl, 0.2 M Na_2HPO_4, 0.02 M EDTA, pH 7.7), 1.0% SDS and 1.0% Blotto (not-fat powdered milk) for 16 h.

Cell culture

Stock cultures of *P. lutherii* were obtained from the CSIRO Fisheries Collection, Hobart, No. CS-23, and cultivated axenically in Provasoli's sterile enrichment seawater [23] at 18 °C under constant white light illumination of 20 μE m^{-2} s^{-1} with aeration.

Isolation of genomic DNA

The method of isolation of genomic DNA and separation of chloroplast DNA is modified from Rochaix [25]. 5×10^{10} cells were harvested by centrifugation at $3000 \times g$ and resuspended in 20 ml of A buffer. Pronase and SDS treatment was as described for the phenol extraction. The aqueous phase was collected and DNA was spooled onto a glass rod, after adding 0.1 volumes of 3 M sodium actetate (pH 5.1) and two volumes of ice-cold absolute ethanol. The DNA was washed in 80% and absolute ethanol (10 min each), air-dried and dissolved in 10 ml of TE. The DNA was then purified by CsCl density gradient centrifugation [18]. The DNA was recovered from the gradient, diluted four-fold with sterile water and then recovered by spooling as described above and finally resuspended in 200 μl of TE.

Separation of nuclear and chloroplast DNA

350 μg of genomic DNA was dissolved in 3.35 ml of TE (10 mM Tris pH 8.0, 1 mM EDTA) and sheared by a single passage through a 19 gauge syringe needle. CsCl (4.27 g) was added and dissolved and the refractive index adjusted to 1.3990. This solution was then centrifuged at 40000 rpm for 65 h (Ti80 rotor, Beckman Spinco). Following centrifugation, 150 μl fractions were collected from the bottom of the gradient. Absorbance readings at 260 nm and the refractive index of each fraction were recorded. Standards comprising DNA from *E. coli* and *Micrococcus lysodeikticus* were treated in an identical fashion for the purpose of determining the G/C content of *P. lutherii* DNA fractions.

Pulse-field gel electrophoresis

Lysis of *P. lutherii* cells and pulse-field gel electrophoresis has been previously described [7]. Approximately 10^9 cells were lysed.

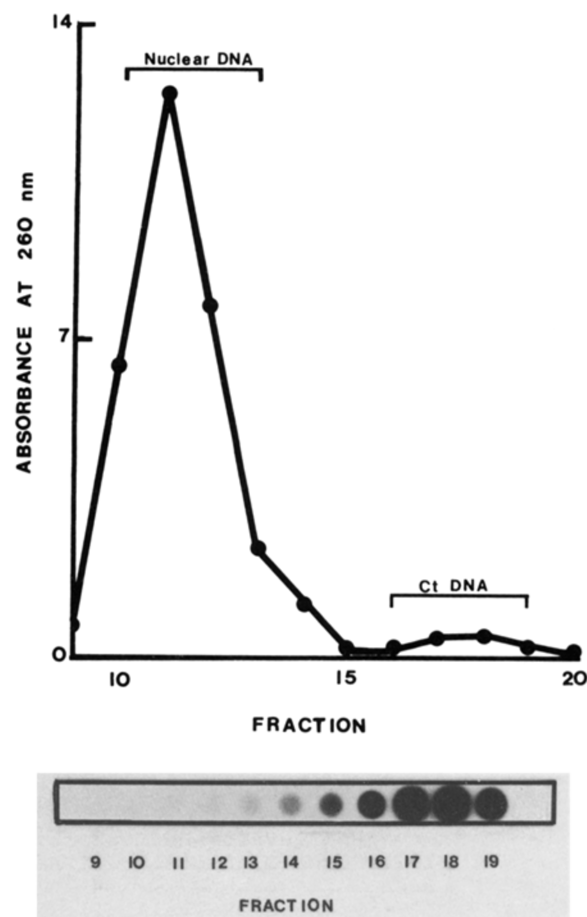

Fig. 1. A (top). Elution profile for *P. lutherii* genomic DNA subjected to caesium chloride density gradient centrifugation and fractionated. B (bottom). Dot blot of eluted fractions with pZmc37 used as a probe. The photo is the result of a 16 h exposure to X-ray film.

```
 -180 GGATTTTCAATCCACTGCTCTACCGACTGAGCTATCCCGGCTTTTTATGATGAAACCATT

 -120 ATACTAAAAAGAGTAAAAAAAATCAACCAAATTAAGGTTCGGTAAATACACCTTTAAGAA

  -60 ATACCGATTTAAAGTTATATTATCTAATATAGATAATATAATTTAGGAAGATAGAAAAA

      MetAlaLysValValGlyIleAspLeuGlyThrThrAsnSerValValAlaValMetGlu
   +1 ATGGCAAAAGTCGTTGGTATTGATTTAGGTACAACAAACTCGGTGGTCGCTGTTATGGAA

      GlyGlyLysProThrValIleThrAsnSerGluGlyGlyThrThrThrProSerValVal
  +61 GGAGGAAAACCTACTGTAATAACAAACTCAGAAGGTGGAACAACAACACCCTCTGTTGTG

      AlaTyrAlaLysAsnGlyAspLeuLeuValGlyGlnIleAlaLysArgGlnAlaValIle
 +121 GCTTATGCAAAAAATGGTGATCTTCTTGTAGGACAGATAGCAAAAAGGCAGGCCGTAATA

      AsnSerGluAsnThrPheTyrSerValLysArgPheIleGlyArgProSerLysGluVal
 +181 AATTCTGAAAATACATTTTATTCGGTAAAGAGGTTTATTGGAAGGCCATCAAAAGAGGTT

      SerAspGluLeuArgGlnThrProTyrLysIleGluAspSerGluGlyLysIleArgLeu
 +241 TCAGATGAACTGAGACAGACTCCTTACAAGATAGAAGATAGTGAAGGTAAGATAAGATTA

      LysCysProAsnLeuAsnLysAsnPheAlaAlaGluGluIleSerAlaGlnValLeuArg
 +301 AAATGCCCTAACTTAAATAAGAATTTTGCAGCAGAAGAAATTTCAGCACAAGTACTAAGA

      LysLeuValAsnAspAlaAsnLysTyrLeuGlyGluLysValGluLysAlaValIleThr
 +361 AAGCTTGTGAACGATGCAAATAAATATTTAGGAGAAAAAGTAGAAAAAGCTGTAATTACA
      HindIII
      ValProAlaTyrPheAsnAspSerGlnArgGlnAlaThrLysAspAlaGlyLysIleAla
 +421 GTACCTGCATATTTTAACGATTCTCAAAGACAAGCAACAAAAGATGCAGGTAAAATCGCA

      GlyLeuGluValLeuArgIleIleAsnGluProThrAlaAlaSerLeuAlaTyrGlyLeu
 +481 GGCTTAGAGGTTTTACGTATCATAAATGAACCTACAGCAGCTTCATTGGCTTATGGTTTA

      AspLysLysAspAsnGluThrIleLeuValPheAspLeuGlyGlyGlyThrPheAspVal
 +541 GATAAAAAAGATAACGAAACTATTTTAGTTTTTGATTTGGGTGGAGGTACGTTTGATGTA

      SerIleLeuGluValGlyAspGlyValPheGluValLeuSerThrSerGlyAspThrArg
 +601 TCTATACTCGAAGTAGGAGATGGTGTTTTTGAGGTACTTTCAACTTCAGGGGATACAAGA

      LeuGlyGlyAspAspPheAspGluLysIleValLysTrpLeuLeuAsnGluPheGluLys
 +661 CTGGGTGGAGATGATTTTGACGAGAAAATTGTTAAGTGGTTGTTAAACGAATTCGAAAAA

      GluGluLysPheSerLeuLysGlyAspSerGlnAlaLeuGlnArgLeuThrGluAlaAla
 +721 GAAGAGAAATTCAGTTTAAAAGGCGATAGCCAAGCTTTACAAAGGCTTACAGAAGCTGCA
                                     HindIII
      GluLysAlaLysIleGluLeuSerSerLeuSerGlnThrGluIleAsnLeuProPheIle
 +781 GAAAAAGCAAAAATCGAGCTCTCTAGTTTGAGTCAAACAGAAATAAATTTGCCATTTATA

      ThrAlaAsnGluAsnGlyAlaLysHisIleGluLysThrLeuThrGlyGluLysPheGlu
 +841 ACAGCCAACGAAAACGGGGCAAAACATATCGAAAAGACATTAACTGGTGAAAAATTTGAA

      SerLeuCysSerAspLeuPheAspArgCysArgIleProValGluAsnAlaLeuLysAsp
 +901 AGCTTATGTAGCGACTTGTTTGACAGATGTCGTATACCTGTAGAAAACGCTTTAAAAGAT
     HindIII
      AlaLysLeuLysProAsnGlnIleAspGluValValLeuValGlyGlySerThrArgIle
 +961 GCAAAACTAAAACCAAATCAAATCGATGAAGTTGTTCTTGTAGGTGGATCAACGCGAATT

      ProAlaValLysLysLeuValLysAspIleLeuGlyLysGluProAsnGluThrValAsn
+1021 CCCGCTGTAAAAAAACTTGTTAAAGATATACTAGGAAAGGAACCGAACGAAACAGTAAAT

      ProAspGluValValAlaIleGlyAlaAlaIleGlnAlaGlyValLeuSerGlyGluVal
+1081 CCTGATGAAGTAGTAGCTATAGGTGCTGCAATACAAGCGGGTGTACTTTCTGGAGAAGTA

      LysAspIleLeuLeuLeuAspValThrProLeuSerLeuGlyValGluThrLeuGlyGly
+1141 AAAGATATTTTACTATTAGATGTAACTCCTCTATCTTTAGGGGTAGAAACATTAGGGGGA

      ValThrThrLysIleIleProArgAsnThrThrValProThrLysLysSerGluIlePhe
+1201 GTTACAACAAAGATTATCCCTAGAAACACAACTGTACCTACTAAAAAGTCAGAAATTTTT
```

```
      SerThrAlaValAspAsnGlnProAsnValGluIleHisValLeuGlnGlyGluArgGlu
+1261 TCAACAGCGGTAGATAATCAACCAAACGTAGAAATACATGTCTTACAAGGAGAGCGTGAA

      PheAlaArgAspAsnLysSerLeuGlyThrPheArgLeuAspGlyIleLeuProAlaPro
+1321 TTTGCTAGAGACAATAAAAGTCTAGGTACTTTTAGACTAGATGGTATACTACCAGCCCCT

      ArgGlyIleProGlnIleGluValThrPheAspIleAspAlaAsnGlyIleLeuSerVal
+1381 AGAGGAATACCTCAAATTGAAGTAACTTTTGATATTGATGCAAATGGTATCCTTTCAGTA

      ThrAlaGlnAspLysGlyThrSerLysGlnGlnSerIleThrIleSerGlyAlaSerThr
+1441 ACTGCTCAAGATAAAGGAACATCAAAGCAGCAAAGTATAACTATATCTGGAGCATCTACT

      LeuProLysGluGluValGluLysMetValLysGluAlaGluGlnAsnAlaAlaAlaAsp
+1501 TTACCTAAAGAGGAAGTAGAAAAGATGGTTAAAGAAGCAGAGCAAAACGCTGCAGCAGAT

      LysGluLysGlyGluAsnIleArgValLysAsnGluAlaAspLeuTyrCysTyrGlnAla
+1561 AAAGAGAAAGGCGAAAATATACGCGTTAAGAATGAAGCAGACTTATACTGCTATCAAGCA

      GluLysGlnIleSerGluLeuProGluAlaLeuValAsnGluAsnGlnSerLeuIleLys
+1621 GAGAAACAAATAAGTGAGCTACCAGAAGCATTAGTAAATGAAAACCAATCTCTAATCAAA

      GluSerLysGluThrValGluMetLeuLysGluAsnIleLysLysGluAspTyrAspLys
+1681 GAAAGTAAAGAAACGGTCGAGATGTTAAAAGAGAACATCAAAAAAGAAGATTATGACAAA

      IleLysGluAsnLeuLysLysLeuGlnGluLysLeuMetGluIleGlyGlnLysAlaTyr
+1741 ATAAAAGAAAATTTAAAAAAATTACAGGAAAAGCTAATGGAGATTGGCCAAAAAGCTTAC
                                                           HindIII
      AlaLysLysGluProLeuLysAspGluAspSerAsnLysAlaGlySerGlnAspAspPhe
+1801 GCTAAAAAAGAGCCTTTGAAAGACGAAGATTCAAATAAAGCTGGATCTCAGGATGATTTT

      IleAspAlaAspPheThrGluSerLys***
+1861 ATTGACGCGGATTTTACCGAAAGTAAATAATTCTTAAAGTGTCTTGCTAATATAGAAACC

+1921 TCAAAATCATAACTAGAAATCCCAAAAATATAAGATTTTTCCGATTTCTAGTTATGATTT

+1981 TCAGGTTTAATTGAAAAAGTTAACACTTTTACGCATCAAATTAATGTACCATAAGTTGT
```

Fig. 2. Nucleotide sequence of the *P. lutherii hsp70* gene and surrounding sequence. *Hin*d III restriction sites and the predicted Hsp70 protein sequence are indicated.

Results

Identification of chloroplast DNA

A total of 20 fractions were collected off the CsCl gradient designed to separate nuclear DNA from chloroplast DNA. The absorbance of each of these fractions was determined and the elution profile is shown in Fig. 1A. Two peaks were present, the first comprising fractions 9 to 14 and the second comprising fractions 16 to 19. The bouyant densities in CsCl of the two peaks were, respectively, 1.720 g/cm^3 and 1.690 g/cm^3. By comparison with the two standards (see Materials and methods) the respective G/C contents of the DNA in each peak are 67% and 37%.

To confirm that the A/T-rich DNA peak consisting of less than 10% of the total DNA was from the chloroplast, 10 μl of each fraction was dot-blotted onto nylon membrane and probed with pZmc37 which contains a fragment of the gene coding for the large subunit of Rubisco from maize [2]. This ubiquitous plant gene is highly conserved and is chloroplast-encoded in all eukaryotic plants and algae. The results of the hybridisation experiment, shown in Fig. 1B, confirms that the small peak containing less than 10% of the total DNA is highly enriched in chloroplast DNA.

Identification of a hsp70 *gene*

To investigate the molecular organisation of the chloroplast genome of *Pavlova lutherii*, DNA from

fractions 16–19 were pooled and used in subsequent cloning experiments. Chloroplast DNA was digested with *Bgl* II and cloned into *Bam* HI-digested, dephosphorylated pBR322. We identified, by DNA sequencing, a clone designated pMAQ801, possessing an open reading frame (ORF) which, when used in data base searches, showed remarkable similarity with previously identified Hsp70 proteins. Consequently, the sequence of a 2.0 kb region completely contained within this *Bgl* II fragment was determined and is shown in Fig. 2 along with the predicted amino acid sequence of the ORF. The ORF predicts a polypeptide of 629 amino acids with a molecular mass of 68714 Da which is obviously related to the Hsp70 family of proteins.

Confirmation that the hsp70 *gene is on the chloroplast chromosome*

To verify that the *hsp70* gene isolated here is chloroplast-encoded, an 899 bp *Hind* III fragment internal to the gene (position +900 to +1798 in Fig. 2) was used as a probe in Southern blotting experiments against *Hind* III-digested nuclear, chloroplast and total genomic DNA fractions. The results shown in Fig. 3 reveal that only those fractions containing chloroplast DNA possess sequences homologous to this *P. lutherii hsp70* gene probe. Figure 3 is the result of a three-day autoradiograph exposure. Subsequent exposures of up to 14 days failed to reveal any evidence of sequences homologous to this *hsp70* probe in the nuclear fraction.

While the fraction from which pMAQ801 is derived is highly enriched for chloroplast DNA, it is nonetheless possible that it is contaminated with small amounts of either mitochondrial DNA or A/T-rich nuclear DNA and that one of these is the source of our *hsp70* gene. To confirm the chloroplast origin of this clone, *P. lutherii* cells were lysed and subjected to pulsed-field gel elec-

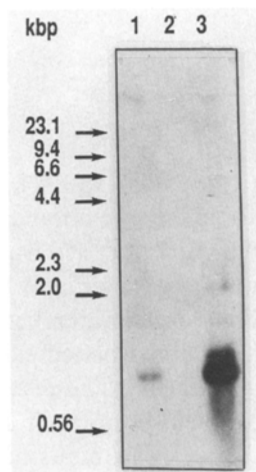

Fig. 3. Southern blot analysis of *P. lutherii* DNA. Two micrograms of total DNA (lane 1), two micrograms of nuclear DNA (lane 2) and half a microgram of chloroplast DNA (lane 3) were digested to completion with *Hind* III and separated by agarose gel electrophoresis. The major band seen in total DNA (lane 1) and chloroplast DNA (lane 3) are due to hybridisation to the 899 bp *Hind* III DNA fragment. The nature of the two minor bands in lane 3 at approximately 1.5 kb and 2.1 kb is not known but may imply the presence of a gene family.

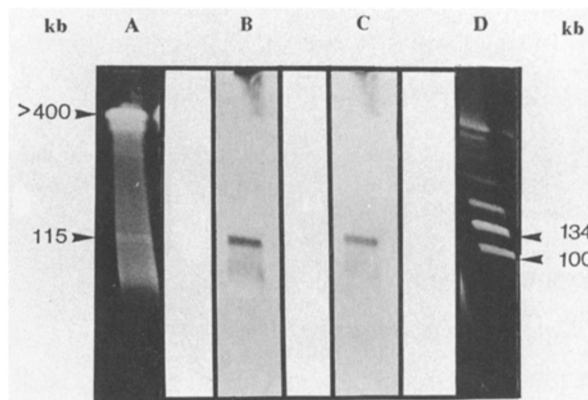

Fig. 4. Pulse-field gel electrophoresis of lysed *P. lutherii* cells. Lane A. Agarose gel of electrophoresed *P. lutherii* DNA. Lane B. Southern blot of A using a fragment of pMAQ807 (*rpoB* and *rpoC1*) as a probe. The filter was exposed to X-ray film for 3 h. Lane C. Southern blot of A using a fragment of pMAQ801 (*hsp70*) as a probe. The filter was exposed to X-ray film for 3 h. Prior to hybridisation with this probe the filter was stripped of pMAQ801 probe DNA by standard protocols [18], and exposed to X-ray film for 16 h to confirm complete removal of probe. Lane D. Agarose gel of molecular weight markers comprising concatamers of non-transcribed spacer DNA from *Dictyostelium discoideum* extrachromosomal ribosomal DNA (R. Cole and K. Williams, personal communication). Bands begin at 100 kb and increase in length in 34 kb increments.

trophoresis. Result shown in Fig. 4 reveal several bands, one corresponding to a length of approximately 115 kb, with the rest larger than 400 kb. DNA from this gel was transferred by Southern blotting to nylon membrane. In the first instance this filter was probed with pMAQ807, another recombinant clone derived from our enriched chloroplast fractions. We have found that this clone contains two open reading frames which predict proteins very closely related to RpoB and RpoC1 chloroplast RNA polymerase subunit proteins (data not shown). We therefore conclude that pMAQ807 is derived from the *P. lutherii* chloroplast genome. When the cloned fragment comprising pMAQ807 was used to probe the Southern blot of the pulsed-field gel, hybridisation occurs specifically to the 115 kb band identifying this band as chloroplast DNA. Cattolico [4] has stated that the chromophytic algae possess chloroplast genomes in the range of 100 to 160 kb. We have also estimated by restriction mapping that the *P. lutherii* chloroplast genome is in the range of 110 to 130 kb (data not shown). Thus a length of 115 kb is consistent with both of these sets of values.

The Southern blot of the pulse-field gel was also probed with the 899 bp *Hin*d III fragment from pMAQ801 which is internal to the *hsp70* gene. As shown in Fig. 4, this fragment also specifically hybridises to the 115 kb chloroplast band. This experiment therefore excludes a nuclear origin for the *hsp70* gene cloned here. While neither the CsCl density gradient nor the pulse-field gel revealed the presence of mitochondrial DNA, we believe it is extremely unlikely that this *hsp70* gene could be derived from DNA from the chromosome of this organelle. The mitochondrial DNA would need to be of both the same length and the same density as the chloroplast genome to produce the results described. Additionally, if our chloroplast fraction did contain mitochondrial DNA of the same length as the chloroplast DNA, the sum of the fragments generated by restriction digests would be approximately 230 kb, which is not the case. Consequently we conclude that the *hsp70* gene identified here is located on the chloroplast genome.

Comparison with other Hsp70 proteins

In Fig. 5 we compare the predicted *P. lutherii* Hsp70 protein sequence with that of *Zea mays* and the analogous prokaryotic protein, DnaK, from the cyanobacterium *Synechocystis* PCC6803 and *E. coli.* The *P. lutherii* sequence displays significant homology to all three, although the greatest similarity is seen with the prokaryotic proteins and the relatedness to DnaK from the photosynthetic prokaryote *Synechocystis* PCC6803 is striking. Specifically, the percentage of identity/similarity of *Synechocystis* PCC6803 DnaK [6], *E. coli* DnaK [1] and *Zea mays* Hsp70 [26] to *P. lutherii* Hsp70 is 70/91, 53/87 and 44/80 respectively.

Discussion

We report here the first case of an *hsp70* gene being encoded outside the nucleus. That the gene identified here is located on the chloroplast chromosome is supported by four observations.

1. The DNA from which the clone is derived is highly purified for chloroplast DNA.
2. Pulse-field gel electrophoresis and subsequent Southern blotting identifies the chloroplast chromosome to be 115 kb and that the *hsp70* gene hybridises specifically to it.
3. Restriction mapping of DNA derived from the fractions used for cloning is consistent with a genome size of 115 kb, excluding the possibility that these fractions were contaminated with mitochondrial DNA of the same size and same density as the chloroplast genome.
4. The A/T content of the *hsp70* gene is 65% which matches very closely the A/T content of the chloroplast as a whole which we estimate by caesium chloride density gradient centrifugation to be 63%.

When this chloroplast *hsp70* gene is used as a probe against nuclear, genomic and chloroplast DNA fractions, there appear to be no homologous sequences in the nucleus. These data might suggest that the only *hsp70* gene in *P. lutherii* is located in the chloroplast. It is probable however

```
1   MAK----VVGIDLGTTNSVVAVMEGGKPTVITNSEGGTTTPSVVAYAKNGDLLVGQIAKR
2   *G*----***********C**************A**FR******G******R********
3   *G*----II*********C**I*D*TT*R*LE*A**DR****II**TQD*ET****P***
4   ***SEGPAI*******Y*C*GLWQHDRVEI*A*DQ*NR****Y*GFT-DTER*I*DA**N

    QAVINSENTFYSVKRFIGRPSK--EVSDELRQTPYKI-E-DSEGKIRLKCPNLNKNFAAE
    ***M*PG************KFD--*ITN*ATEVA*SVVK-*GN*NVK*D**AQG*Q**P*
    ***T*PQ**LFAI**L***RFQDE**QRDVSIM*F**IA-ADN*DAWVE--VKGQKM*PP
    *VAM*PT**VFDA**L***RFSSPA*QSSMKLW*SRHLGLGDKPM*VFNYKGEE*Q****

    EISAQVLRKLVNDANKYLGEKVEKAVITVPAYFNDSQRQATKDAGKIAGLEVLRIINEPT
    **********D**S*****T*TQ**************************I**********
    Q***E**K*MKKT*ED****P*TE***********A*********R******K*******
    ***SM**I*MKEI*EA***STIKN**V******************V****N*M*******

    AASLAYGLDK----KDNETILVFDLGGGTFDVSILEV----GDGVFEVLSTSGDTRLGGD
    **********----***********************----*E******A*****H****
    **A*******---GTG*R**A*Y********I**I*IDEVD*EKT****A*N***H***E
    **AI******KATSSGEKNV*I***********L*TIEE----*I***KA*A***H***E

    DFDEKIVKWLLNEFEKEEKFSLKGDSQALQRLTEAAEKAKIELSSLSQTEINLPFITANE
    ***K***DF*AG**Q*A*GID*RK*K******************GV************TQ
    ***SRLINY*VE**K*DQGID*RN*PL*M***K************AQ**DV***Y***DA
    ***NRM*NHFVQ**KRKN*KDIS*NPR**R**RT*C*R**RT***TA**T*EIDSLFEGI

    NGAKHIEKTLTGEKFESLCSDLFDRCRIPVENALKDAKLKPNQIDEVVLVGGSTRIPAVK
    D*P**LDT**SRA***EI****I***G******IR***IDKSAL**I************Q
    TGP**MNIKV*RA*L***VE**VN*SIE*LKV**Q**G*SVSD**D*I****Q**M*M*Q
    D----FTPRSSRAR**E*NM***RK*ME***KC*R***MDKSSVHD***********K*Q

    KLVKDILGKEPNETVNPDEVVAIGAAIQAGVLSGEVK---DILLLDVTPLSLGVETLGGV
    EV**K****D**QG********V*****G********---*******S************
    *K*AEFF****RKD*****A******V*G***T*D**---*V************I**M***
    Q*QDFFN***LCKSI****A**Y***V**AI****GNERS*L***********L**A***

    TTKIIPRNTTVPTKKSEIFSTAVDNQPNVEIHVLQGEREFARDNKSLGTFRLDGILPAPR
    M*********I*************G*S************M*N*************P****
    M*TL*AK***I***H*QV****E***SA*T********KR*A******Q*N****N****
    M*VL******I****EQV***YS****G*L*Q*YE***ARTK**NL**K*E*S**P****

    GIPQIEVTFDIDANGILSVTAQDKGTSKQQSITISGAST-LPKEEVEKMVKEAEQNAAAD
    *V***************N***K*R**G*E***SIT****-**D***DR******S*****
    *M***********D***H*S*K**NSG*E*K***KAS*G-*NED*IQ***RD**A**E**
    *V***T******V*N**N*S*E**T*GQKNK***TNDKGR*S***I****Q***KYK*E*

    KEKGENIRVKNEADLYCYQAEKQISE--LPEALVNENQSLIKESKETVEMLKENIKKEDY
    **RR*K*DR**Q**SLV*******T*--*GDKVPAADK--**-AEGLIKD***AVAQ**D
    RKFE*LVQTR*QG*HLLHSTR**VE*--AGDK*PADDKTA*---ESALTA*ETAL*G**K
    E*VKKKVDA**ALEN*A*NMRNT*KDDKIASK*PA*DKKK*EDAVDGAISWLDSNQLAEV

    DKIKENLKKLQE---KLMEIGQKAYAKKEPLKD--------EDSNKAGSQDDFIDADFTE
    A**QTVMPEL*Q---V*YS**SNM*QQAGAEAGVGAPGAGP*AGTSS*GG**V***E*S*
    AA*EAKMQE*AQ---*****A*QQH*QQQT--------AGADA*ANNAKD**VV**E*E*
    EEFEDKM*E*EGICN---P*IA*M*?GEGA--GMGAAAGMD**APSG**GAG------PK

    --S-K
    --PE*
    VKDK*
    IEEVD
```

Fig. 5. Protein comparison of predicted peptide sequence of *P. lutherii* Hsp70 (line 1), *Synechocystis* PCC6803 DnaK (line 2), *E. coli* DnaK (line 3) and *Zea mays* Hsp70 (line 4). Protein alignments were generated using the CLUSTAL program [13]. The complete predicted peptide sequence for *P. lutherii* Hsp70 is shown, amino acid residues identical to *P. lutherii* are indicated by an asterisk.

that other *hsp70* genes are present in the nucleus but are not sufficiently similar to our chloroplast probe to be detected. Although Hsp70 proteins are very highly conserved, the dramatic difference between the G/C contents of the *P. lutherii* nuclear and chloroplast genomes (67% and 37% respectively) may have resulted in relatively diverged coding sequences through codon bias. It is noteworthy that the coding sequence of the gene reported here favours A/T-rich codons by 10/1 for some amino acids.

The exact map position of the *hsp70* gene on the chloroplast genome is, at present, unknown. To date we have identified, by DNA sequencing, four other *P. lutherii* chloroplast genes (unpublished observation) but we have not yet been able to link *hsp70* to any of them by restriction mapping. Additionally no other identifiable ORFs are present in the sequenced 250 bases immediately 3′ and 5′ to the *hsp70* gene.

The finding of a chloroplast-encoded Hsp70 protein and its close relatedness to DnaK of a cyanobacterium has several important implications. Firstly, it is widely accepted that the chloroplasts of plants and algae originated as endosybionts. In the case of the Chromophyta this was probably a two-step process; nonetheless the primary event is presumed to have involved a photosynthetic bacterium [9, 17]. *P. lutherii* Hsp70 is more closely related to prokaryotic DnaK proteins generally than to the Hsp70 of higher eukaryotes, strongly implying a prokaryotic origin for the *P. lutherii* gene. In particular, the similarity between *P. lutherii* Hsp70 and the corresponding protein from *Synechocystis* PCC6803 suggests that the *P. lutherii* chloroplast and the cyanobacteria have a very close common ancestor and supports the theory that chloroplasts originated as the result of an endosymbiotic event [30].

This is the first reported case of an *hsp70* gene to be chloroplast-encoded. Why would *P. lutherii* possess an *hsp70* gene in its chloroplast? The chloroplast of *P. lutherii* is surrounded by four membranes, a feature confined to chromophytic algae. The outermost membrane, the chloroplast endoplasmic reticulum (CER) is continuous with the nuclear envelope and vesicles have been found between the CER and the chloroplast [10]. Further, it has been suggested [11] that these vesicles have a role in the import of nuclear-encoded light-harvesting proteins into the chromophytic chloroplast. Given that current models for protein stabilisation and transport across organelles require the presence of Hsp70 on both sides of the membrane [12, 20] a chloroplast-encoded heat shock protein may facilitate chloroplast protein transport in this organism. It is possible that the organisation of protein transport genes in *P. lutherii* may be novel and that this organism will be important in understanding the biochemistry and evolution of chloroplast protein transport.

Acknowledgements

We would like to thank Keith Williams for suggesting the pulse-field experiment and Rob Cole for running the pulse-field gel. We also thank S.W. Jeffrey for providing the *P. lutherii* strain. This research was supported, in part, by Macquarie University Research Grants to H.W.S and R.G.H.

References

1. Bardwell JCA, Craig EA: Major heat shock gene of *Drosophila* and the *Escherichia coli* heat-inducible *dnaK* gene are homologous. Proc Natl Acad Sci USA 81: 848–852 (1984).
2. Bedbrook JR, Coen DM, Beaton AR, Bogorad L, Rich A: Location of the single gene for the LSU of RUBISCO on the maize chloroplast chromosome. J Biol Chem 254: 905–910 (1979).
3. Bolivar F, Rodriguez RL, Greene PJ, Betlach MC, Huyneker HL, Boyer HW, Crosa JH, Falkow S: Construction and characterisation of new cloning vehicles. II. A multipurpose cloning system. Gene 2: 95–113 (1977).
4. Cattolico RA: Chloroplast evolution in algae and land plants. TREE 1: 64–67 (1986).
5. Chirico WJ, Waters MG, Blobel G: 70K heat shock related proteins stimulate protein translocation into microsomes. Nature 332: 805–810 (1988).
6. Chitnis PR, Nelson N: Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis* sp. PCC6803. J Biol Chem 266: 58–65 (1991).

7. Cole RA, Williams KL: Insertion of transformation vector DNA into different chromosomal sites of *Dictyostelium discoideum* as determined by pulse field gel electrophoresis. Nucl Acids Res 16: 4891–4902 (1988).
8. Craig EA, Kramer J, Shilling J, Werner-Washburne M, Holmes S, Kosic-Smithers J, Nicolet CM: *SSC1*, an essential member of the yeast *hsp70* multigene family, encodes a mitochondrial protein. Mol Cell Biol 9: 3000–3008 (1989).
9. Douglas SE, Murphy CA, Spencer DF, Gray MW: Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350: 148–151 (1991).
10. Gibbs SP: The chloroplast of *Euglena* has evolved from symbiotic green algae. Can J Bot 56: 2883–2889 (1978).
11. Grossman A, Manodori A, Snyder D: Light harvesting proteins of diatoms: Their relationship to the chlorophyll a/b binding proteins of higher plants and their mode of transport into plastids. Mol Gen Genet 224: 91–100 (1990).
12. Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, Georgopoulos CP, Hendrix RW, Ellis RJ: Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333: 330–334 (1988).
13. Higgins DG, Sharp PM: Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5: 151–153 (1989).
14. Hunt C, Morimoto RI: Conserved features of eukaryotic *hsp70* genes revealed by comparison with the nucleotide sequence of human *hsp70*. Proc Natl Acad Sci USA 82: 6455–6459 (1985).
15. Kang P-J, Ostermann J, Schilling J, Neupert W, Craig EA, Pfanner N: Requirement for *hsp70* in the mitochondrial matrix for translocation and folding of precursor proteins. Nature 348: 137–143 (1990).
16. Lindquist S, Craig EA: The heat-shock proteins. Annu Rev Genet 22: 631–677 (1988).
17. McFadden GI: Evidence that cryptomonad chloroplasts evolved from photosynthetic eukaryotic endosymbionts. J Cell Sci 95: 303–308 (1990).
18. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
19. Marshall JS, DeRocher AE, Keegstra K, Vierling E: Identification of heat shock protein *hsp70* homologues in chloroplasts. Proc Natl Acad Sci USA 87: 374–378 (1990).
20. Neupert W, Hartl F-U, Craig EA, Pfanner N: How do polypeptides cross the mitochondrial membranes? Cell 63: 447–450 (1990).
21. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S-I, Inokuchi H, Ozeki H: Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322: 572–574 (1986).
22. Powell MJ, Watts FZ: Isolation of a gene encoding a mitochondrial *hsp70* from *Schizosaccharomyces pombe*. Gene 95: 105–110 (1990).
23. Provasoli L: Growing marine seaweeds. In: DeVirville D, Feldman J (eds) Proc Int Seaweed Symp 4: 9–17. Pergamon, Oxford (1963).
24. Reading DS, Halberg RL, Myers AM: Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. Nature 337: 655–659 (1989).
25. Rochaix JD: Isolation of chloroplast DNA from *Chlamydomonas reinhardtii*. In: Edelman M, Hallick RB, Chua NH (eds) Methods in Chloroplast Molecular Biology, pp. 295–300. Elsevier Biomedical Press, Amsterdam (1982).
26. Rochester DE, Winer JA, Shah DM: The structure and expression of maize genes encoding the major heat shock protein, *hsp70*. EMBO J 5: 451–458 (1986).
27. Sanger F, Coulson AR, Barell BG, Smith AJH, Roe BA: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143: 161–178 (1980).
28. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chungwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusada J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049 (1986).
29. von Heijne G, Steppuhn J, Herrman RG: Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem 180: 535–545 (1989).
30. Whatley JM, Whatley FR: Chloroplast evolution. New Phytol 87: 233–247 (1981).
31. Winter J, Wright R, Duck N, Gasser C, Fraley E, Shah D: The inhibition of petunia *hsp70* mRNA processing during $CdCl_2$ stress. Mol Gen Genet 211: 315–319 (1988).
32. Yanisch-Perron C, Viera J, Messing J: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene 33: 103–119 (1985).