The EMBO Journal vol.7 no.3 pp.575-581, 1988

A heat shock protein localized to chloroplasts is a member of a eukaryotic superfamily of heat shock proteins

Elizabeth Vierling[1,2], Ronald T.Nagao[3], Amy E.DeRocher[2] and Lisa M.Harris[1]

Departments of [1]Biochemistry and [2]Molecular and Cellular Biology, University of Arizona, Biological Sciences West, Tucson, AZ 85721 and [3]Department of Botany, University of Georgia, Athens, GA 30602, USA

Communicated by A.Tissieres

We have isolated cDNA clones from soybean and pea that specify nuclear-encoded heat shock proteins (HSPs) which localize to chloroplasts. The mRNAs for these HSPs are undetectable at control temperatures, but increase ~150-fold during a 2-h heat shock. Hybridization-selection followed by *in vitro* translation demonstrates that these HSPs are synthesized as precursor proteins which are processed by the removal of 5-6.5 kd during import into isolated chloroplasts. The nucleotide sequence of the cDNAs shows the derived amino acid sequences of the mature pea and soybean proteins are 79% identical. While the predicted transit peptide encoded by the pea cDNA has some characteristics typical of transit sequences, including high Ser content, multiple basic residues and no acidic residues, it lacks two domains proposed to be important for import and maturation of other chloroplast proteins. The carboxy-terminal region of the chloroplast HSP has significant homology to cytoplasmic HSPs from soybean and other eukaryotes. We hypothesize that the chloroplast HSP shares a common structural and functional domain with low mol. wt HSPs which localize to other parts of the cell, and may have evolved from a nuclear gene.

Key words: chloroplast/heat shock/protein import/transit peptide

Introduction

High temperature stress, or heat shock, induces dramatic changes in gene expression in a wide variety of prokaryotic and eukaryotic organisms (Lindquist, 1986; Nagao *et al.*, 1986). Typically, raising an organism's temperature five or more degrees leads to vigorous transcription of a discrete set of genes and predominant synthesis of the corresponding 'heat shock proteins' (HSPs). The evolutionary conservation of the heat shock response and the HSPs suggests that these proteins are involved in fundamental cell processes. Physiological and genetic data indicate that HSPs protect cells from stress damage, but the mechanism of protection is unknown.

The response of plants to heat shock exhibits many of the characteristics of the heat shock response in other organisms (Kimpel and Key, 1985). However, in contrast to other eukaryotes in which hsp70 is the predominant HSP, the most abundant HSPs in plants are a complex array of 10-30 low mol. wt polypeptides. Although the majority of these small HSPs are localized in the cytoplasm, we and others have shown that one of the chloroplast HSPs is localized to chloroplasts (Kloppstech *et al.*, 1985; Suss and Yordanov, 1986; Vierling *et al.*, 1986). This chloroplast-localized HSP ranges in size from 21 to 28 kd depending on the plant species. These HSPs are the products of nuclear genes; they are synthesized from poly(A) RNA *in vitro* and their biosynthesis *in vivo* is inhibited by cycloheximide. The localization of these HSPs within the chloroplast is unclear. In our studies (Vierling *et al.*, 1986) and the work of Suss and Yordanov (1986) the majority of the protein is recovered with the soluble fraction of the chloroplast. Other workers observe a significant proportion of the HSPs associated with thylakoid membranes (Kloppstech *et al.*, 1985; Schuster *et al.*, 1987). The function of this HSP within the chloroplast is unknown. Schuster *et al.* (1987) have proposed that the HSPs protect specifically photosystem II from stress damage, while we have suggested that the protein has a more general protective role (Vierling, 1987; Vierling *et al.*, 1987).

In order to investigate further the possible functions of chloroplast-localized HSPs, we have begun to examine their structure and expression. The present study reports the isolation and characterization of cDNA clones for a chloroplast HSP from both pea and soybean. We demonstrate that these proteins are synthesized as precursor polypeptides which are processed to their mature mol. wt upon import into chloroplasts. DNA sequence analysis indicates that they are homologous to a ubiquitous class of small HSPs found in the cytoplasm of plants and other eukaryotes.

Results

Identification of cDNA clones

To isolate a cDNA encoding the soybean chloroplast HSP, several cDNAs were identified as heat shock positive clones by differential colony hybridization. Plasmid DNA from each clone was used to hybrid-select mRNA from poly(A) RNA isolated from heat shocked soybean cells. The hybrid-selected mRNAs were translated *in vitro* in wheat germ extracts in the presence of [^{35}S]methionine and the polypeptides used for *in vitro* import into isolated pea chloroplasts. A single cDNA was obtained that hybrid-selected an mRNA for a 28-kd polypeptide which was processed to 22 kd during import into chloroplasts (Figure 1, panel A). The 22-kd polypeptide co-migrates with a major soybean chloroplast HSP (hsp22) (Figure 1 and Vierling *et al.*, 1986). More than 90% of hsp22 is recovered in the soluble protein fraction of the chloroplast. The hsp22 cDNA was estimated to be no more than 60% of the length of the corresponding mRNA (see below). Further screening of the cDNA library did not yield full-length hsp22 clones or clones corresponding to the putative 27-kd soybean chloroplast HSP (Figure 1, lane 4).

Low-stringency hybridization with the soybean hsp22 cDNA was used to isolate several cDNAs from a pea heat

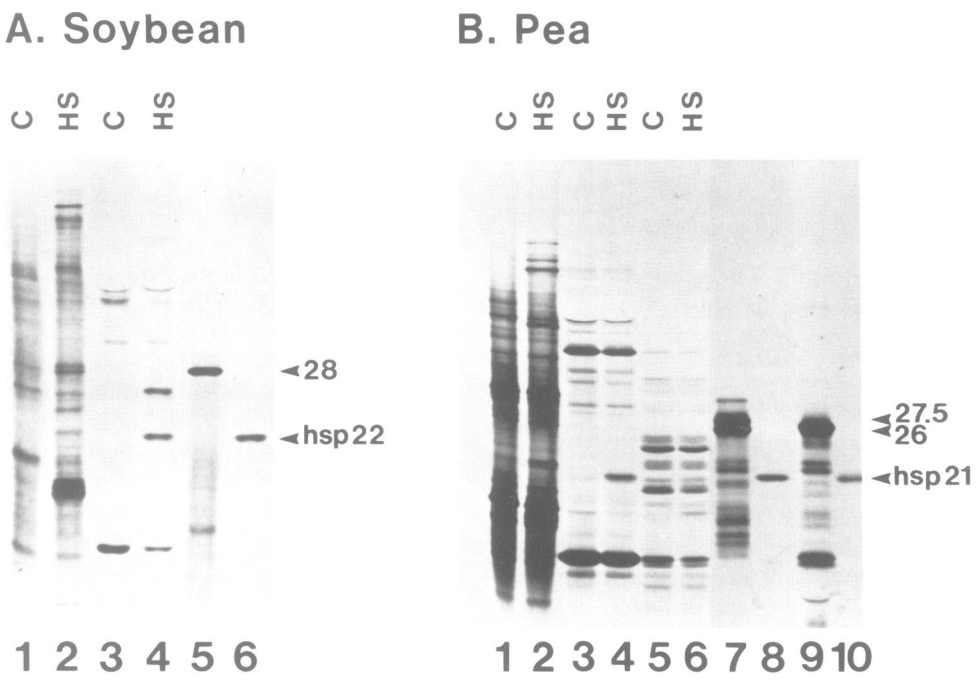

Fig. 1. Identification of cDNAs encoding chloroplast HSPs from soybean and pea. **(A)** Soybean samples. **Lane 1**, *in vitro* translation products of poly(A) RNA from control plants; **lane 2**, translation products of poly(A) RNA from heat-shocked plants; **lane 3**, polypeptides recovered in the chloroplast-soluble protein fraction following import of translation products shown in lane 1; **lane 4**, as for lane 3 only using translation products shown in lane 2; **lane 5**, *in vitro* translation product(s) of mRNA hybrid-selected using the hsp22 cDNA; **lane 6**, soluble chloroplast protein fraction following import of product(s) in lane 5. **(B)** Pea samples. **Lanes 1–4** as for soybean samples. **Lane 5**, chloroplast membrane fraction following import of translation products in lane 1; **lane 6**, chloroplast membrane fraction following import of translation products in lane 2; **lane 7**, *in vitro* translation products of mRNA hybrid-selected using the hsp21 cDNA; **lane 8**, soluble chloroplast protein fraction following import of products in lane 7; **lane 9**, translation products of RNA synthesized by SP6 transcription of an hsp21 cDNA; **lane 10**, soluble chloroplast protein fraction following import of products in lane 9. Sizes (in kd) of the precursor and mature chloroplast HSPs are indicated.

shock cDNA library. *In vitro* translation of mRNAs hybrid-selected by the pea cDNAs produced two polypeptides of 27.5 and 26 kd (Figure 1, lane 7). When these polypeptides were incubated with isolated pea chloroplasts a 21-kd polypeptide was recovered in the soluble protein fraction of the chloroplast. One cDNA clone which appeared to contain the AUG start codon was subcloned into an SP6 vector and transcribed *in vitro*. *In vitro* translation of the transcripts produced only the 26-kd precursor polypeptide (lane 9). The 26-kd polypeptide was imported and processed to the 21-kd form by isolated chloroplasts. These data confirm that the cDNA contains the entire coding region for the 26-kd polypeptide and that this polypeptide is an authentic precursor of the 21-kd mature form (hsp21). Minor low mol. wt polypeptides in lanes 7 and 9 are attributable to premature termination of transcription and/or translation in these samples. Significantly, none of these polypeptides co-migrate with the mature hsp21. A cDNA encoding the 27.5-kd polypeptide has not been identified.

Characterization of the precursor and mature proteins

The precursor and mature forms of the chloroplast HSPs were further characterized by two-dimensional gel electrophoresis (Figure 2). The precursor polypeptides from both pea and soybean are significantly more basic than the mature proteins. The pea precursor polypeptides of 27.5 and 26 kd have pIs of ~7.0 and ~6.7 respectively (panels C and E). The presence of additional isoforms in the translation products of the SP6 transcripts are attributed to artefactual *in vitro* modifications in this experiment. *In vitro* import into chloroplasts of either the products of hybrid-selected mRNAs or the SP6 transcripts produces the same mature forms of the protein, a 21-kd doublet at ~pH 5.5 (panels D and F). This doublet co-migrates with the major pea chloroplast HSP (panel B; see also Vierling *et al.*, 1986). These results suggest that both the 27.5- and 26-kd peptides are processed to the mature size of 21 kd. Therefore the difference in mol. wt of the 27.5- and 26-kd peptides is predicted to result from differences in the length of their transit peptides.

The 28-kd soybean precursor protein produced by hybrid-selection and *in vitro* translation focuses as two distinct polypeptides with pIs of ~7.4 and ~7.0 (panel G). Mature hsp22 also focuses as two polypeptides with pIs of 6.2 and 6.0 (panel H). HSPs of identical size and pI have been observed in whole cell extracts of soybean cells labeled *in vivo* (Vierling and Key, 1985), indicating that both pI forms of the mature protein represent correctly processed precursors. Whether these forms arise from one or both precursor polypeptides has not been determined.

Level of induction

Soybean hsp22 and pea hsp21 mRNAs are not detectable in control tissues but accumulate significantly during a 2-h heat shock treatment (Figure 3, panel A). The hsp22 mRNA is ~1100 bp and the hsp21 mRNA is ~1000 bp. The level and fold-induction of pea hsp21 mRNA were estimated by dot-blot hybridization of total poly(A) RNA using *in vitro* transcribed mRNA as a standard (panel B). Following a 2-h heat shock at 39°C, hsp21 mRNA represents ~0.75% of the total poly(A) RNA. Estimating the lower limit of mRNA detection as 10 pg, we conclude that the level of hsp21 in

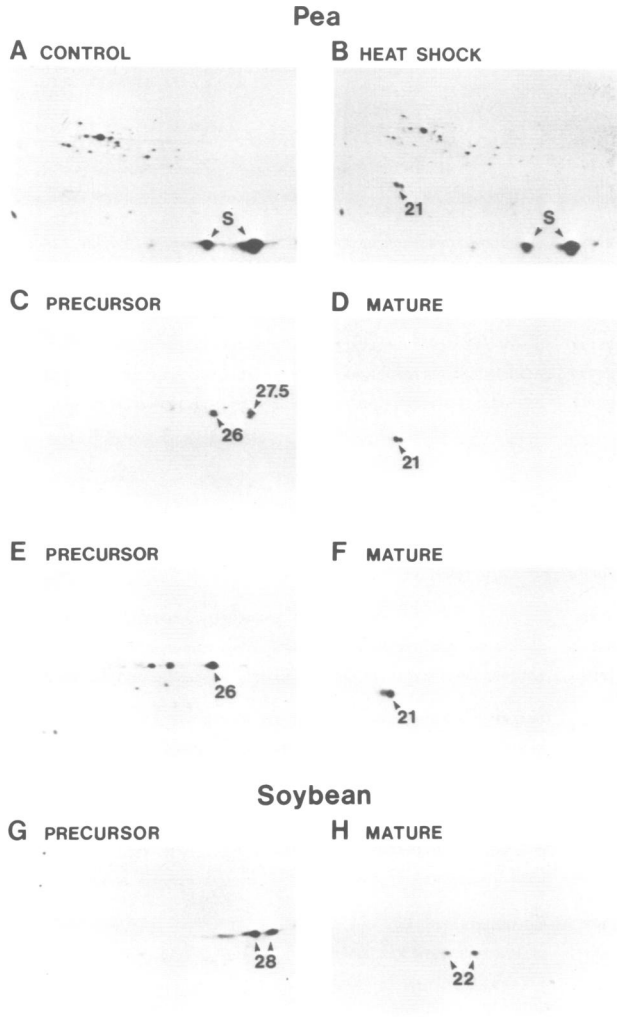

Fig. 2. Two-dimensional gel analysis of precursor and mature forms of the chloroplast HSPs. Panels A, B, D, F and H show proteins recovered in the soluble fraction of the chloroplast following import of the specified *in vitro* translation products. **Panel A**: import of translation products of poly(A) RNA isolated from control pea plants. **Panel B**: import of translation products of poly(A) RNA isolated from heat-shocked pea plants. **Panel C**: translation products of RNA hybrid-selected using pea hsp21. **Panel D**: import of proteins shown in panel C. **Panel E**: translation products of RNA transcribed from hsp21. **Panel F**: import of proteins shown in panel E. **Panel G**: translation products of RNA hybrid-selected using soybean hsp22. **Panel H**: import of proteins shown in panel G. The pI range of the gels is from pH 4.5 (left) to 7.5 (right).

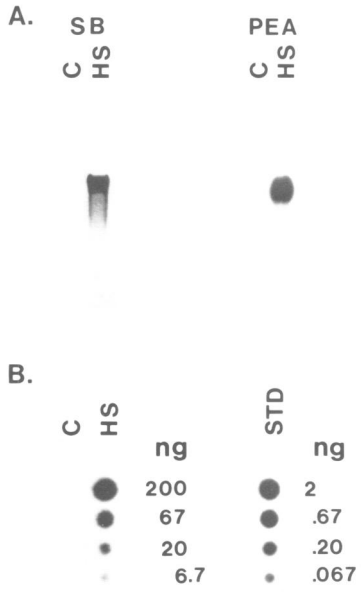

Fig. 3. Determination of the level of mRNA induction by heat. (**A**) Northern analysis of control (C) and heat shock (HS) mRNA isolated from soybean (SB) or pea. Blots were probed with ^{32}P-labeled hsp22 (soybean) or hsp21 (pea). (**B**) Dot-blot hybridization to quantify pea hsp21 mRNA. A dilution series of control mRNA, heat shock mRNA or SP6 transcripts from hsp21 (STD) was applied to nitrocellulose. The filter was hybridized with ^{32}P-labeled hsp21.

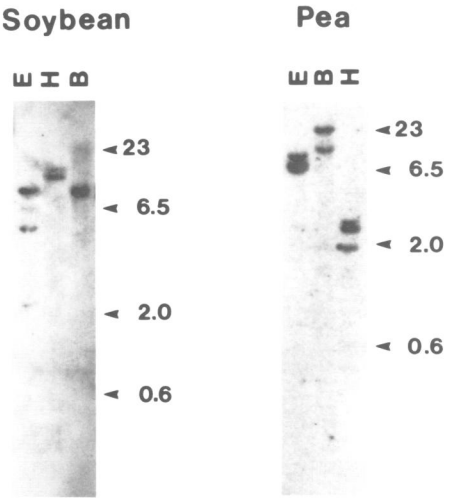

Fig. 4. Southern analysis of genomic DNA from soybean and pea probed with hsp22 or hsp21 respectively. Genomic DNA was cut with *Eco*RI (E), *Bam*HI (B) or *Hin*dIII (H).

control tissues is $<0.005\%$ of total poly(A) RNA. Therefore hsp21 is induced a minimum of 150-fold relative to the level in unstressed tissues. All blots were rehybridized with a cDNA encoding the small subunit of ribulose bisphosphate carboxylase to confirm that mRNA in the control samples was intact (not shown).

Southern analysis

Southern analysis of soybean genomic DNA suggests that chloroplast hsp22 is encoded either by a single gene or two closely linked genes (Figure 4). In pea, the Southern data suggest that there are at least two genes, most likely corresponding to a single gene for each of the observed hsp21 precursor polypeptides (27.5 and 26.0 kd). This small number of genes is in contrast to the low mol. wt cytoplasmic HSPs in soybean which are members of complex multigene families (Nagao *et al.*, 1985). Thus, the chloroplast HSP gene(s) must be very vigorously transcribed to result in the observed increase in mRNA.

cDNA sequence analysis

The complete DNA sequences and deduced amino acid sequences of the soybean and pea cDNAs are shown in Figure 5. The soybean cDNA consists of 546 bp of coding sequence and 150 bp 3′ non-coding sequence, but is truncated before the amino terminus of the protein. The coding region corresponds to 181 amino acids with a mol. wt of 20 512 which represents $>90\%$ of the mature 22-kd protein. The pea

```
          -40                        -20
ATTC TAACA AGCAA TTTTT GTCTC TCATT TCACA TTTTC CATAA CAACC ATG GCT CAG
                                                      Met Ala Gln

                          30                                  60
TCA GTA TCA TTG TCA ACA ATT GCT TCA CCT ATT CTA TCC CAA AAA CCA GGC AGC
Ser Val Ser Leu Ser Thr Ile Ala Ser Pro Ile Leu Ser Gln Lys Pro Gly Ser

                          90
TCT GTC AAA AGT ACA CCA CCT TGC ATG GCC TCA TTT CCA TTA AGG AGG CAA TTG
Ser Val Lys Ser Thr Pro Pro Cys Met Ala Ser Phe Pro Leu Arg Arg Gln Leu

120                       150
CCT AGG CTA GGA CTG AGA AAT GTG AGA GCT CAG GCT GGT GGT GAT GGT GAT AAC
                                                GGT GGT GAT AAC
Pro Arg Leu Gly Leu Arg Asn Val Arg Ala Gln Ala Gly Gly Asp Gly Asp Asn
                                                      Gly  *   *   *

        180                                 210
AAA GAT AAC TCT GTG GAA GTT CAT CGT GTT AAC AAA GAC GAT CAA GGA ACT GCT
AAA GAC AAC TCT GTG GAA GTA CAA CAT GTT AGC AAG GGT GAC CAA GGA ACC GCG
Lys Asp Asn Ser Val Glu Val His Arg Val Asn Lys Asp Asp Gln Gly Thr Ala
 *   *   *   *   *   *   Gln His  *  Ser  *  Gly  *   *   *   *   *

                  240                                 270
GTA GAA AGA AAG CCT CGC AGA AGT TCC ATC GAT ATT TCA CCA TTT GGT TTA TTG
GTT GAG AAG AAG CCA CGT AGA ACT GCC ATG GAC ATT TCA CCA TTT GGT ATC TTG
Val Glu Arg Lys Pro Arg Arg Ser Ser Ile Asp Ile Ser Pro Phe Gly Leu Leu
 *   *  Lys  *   *   *   *  Thr Ala Met  *   *   *   *   *   *  Ile  *

                          300                                 330
GAC CCA TGG TCA CCT ATG AGA AGC ATG CGT CAG ATG CTA GAC ACA ATG GAC AGA
GAC CCG TGG TCA CCC ATG AGG AGC ATG CGC CAG ATA TTG GAC ACC ATG GAC CGA
Asp Pro Trp Ser Pro Met Arg Ser Met Arg Gln Met Leu Asp Thr Met Asp Arg
 *   *   *   *   *   *   *   *   *  Ile  *   *   *   *   *   *   *

                                  360
ATT TTC GAA GAT GCA ATC ACA ATT CCT GGA AGA AAC ATT GGA GGA GGT GAG ATT
GTT TTC GAA GAT ACC ATG ACA TTC CCT GGA AGA AAC ATA GGA GGA GGG GAG ATC
Ile Phe Glu Asp Ala Ile Thr Ile Pro Gly Arg Asn Ile Gly Gly Gly Glu Ile
Val  *   *   *  Thr Met  *  Phe  *   *   *   *   *   *   *   *   *   *

390                               420
CGT GTG CCT TGG GAA ATC AAA GAT GAA GAA CAT GAA ATC AGA ATG CGT TTT GAT
CGT GCC CCT TGG GAC ATC AAA GAT GAA GAA CAT GAA ATC AGA ATG AGG TTT GAC
Arg Val Pro Trp Glu Ile Lys Asp Glu Glu His Glu Ile Arg Met Arg Phe Asp
 *  Ala  *   *  Asp  *   *   *   *   *   *   *   *   *   *   *   *   *

        450                                 480
ATG CCT GGT GTT TCT AAG GAA GAC GTT AAA GTA TCA GTT GAA GAT GAT GTT CTT
ATG CCG GGT CTT GCC AAG GAA GAT GTT AAG GTA TCA GTG GAG GAT GAT ATG CTT
Met Pro Gly Val Ser Lys Glu Asp Val Lys Val Ser Val Glu Asp Asp Val Leu
 *   *   *  Leu Ala  *   *   *   *   *   *   *   *   *   *  Met  *

                  510                                 540
GTC ATA AAA AGT GAT CAT AGA  -   -  GAG GAA AAT GGT GGA GAA GAT TGT TGG
GTT ATA AAA GGT GGC CAC AAA AGT GAA CAA GAA CAT GGT GGA GAT GAT TCT TGG
Val Ile Lys Ser Asp His Arg  -   -  Glu Glu Asn Gly Gly Glu Asp Cys Trp
 *   *   *  Gly Gly  *  Lys Ser Glu  *   *  His  *   *  Asp  *  Ser  *

                                  570
TCA AGA AAG AGT TAT AGT TGT TAT GAT ACT CGT CTC AAG CTT CCT GAT AAC TGT
TCA AGC AGG ACC TAT AGT TCC TAT GAT ACC CGT TTA AAG CTG CCG GAT AAT TGC
Ser Arg Lys Ser Tyr Ser Cys Tyr Asp Thr Arg Leu Lys Leu Pro Asp Asn Cys
 *  Ser Arg Thr  *   *  Ser  *   *   *   *   *   *   *   *   *   *   *

600                               630
GAG AAA GAG AAG GTT AAG GCT GAG TTG AAA GAT GGT GTG CTT TAT ATT ACT ATT
GAG AAG GAC AAG GTT AAG GCA GAG TTG AAA AAT GGT GTG CTT TAT ATC ACC ATT
Glu Lys Glu Lys Val Lys Ala Glu Leu Lys Asp Gly Val Leu Tyr Ile Thr Ile
 *   *  Asp  *   *   *   *   *   *  Asn  *   *   *   *   *   *   *

        660                                 690
CCT AAG ACT AAA ATT GAA CGC ACG GTT ATT GAT GTT CAA ATT CAG TGA TAGGAT
CCT AAG ACC AAG GTT GAA CGT AAG GTT ATT GAT GTC CAA GTT CAG TGA GATATG
Pro Lys Thr Lys Ile Glu Arg Thr Val Ile Asp Val Gln Ile Gln
 *   *   *  Val  *   *  Lys  *   *   *   *  Val  *

                  720                                 750
TAAAC TTGAA AAATG CCATG CTTTT TAATA GTATC CTCTA TGTGG AGGAG CTAAT GTGTT
CTTGG GCTTT GTATT ATGCA ACACG GTTTT TTACC ATGTT TTTAC GTTTT TAACA TATAT

                  780                                 810
ATTTT GTCTC TTTTT TTACC TATAA AAATG TTTTG TCTAA ATGTA GAATA AATAA GATCC
ACACC CATTA TAAGG GTTAA ATATG TTGTA TTAAG CGTAG AGAAG TTTAA ATAAA ACTGC

                  840                                 870
TTTAC TATTT TATAA ATGTT TTCAA ATGTC TAATT TATAT ATTTT CTATG TATTT GGTTT
TGTTG TATAA TTTTC TAATG TCAA

                  900                                 930
GCTGG AATAA CAATT AGTTT AGTAA CACAC TCGAA TGTGT CCCAT CGATGG
```

Fig. 5. Complete DNA sequence and derived amino acid sequence for the pea and soybean chloroplast hsp cDNAs. Pea cDNA and amino acid sequences are above the corresponding soybean sequences. Asterisks indicate identical amino acids. The nucleotide sequence is numbered using the start site of hsp21 translation as +1.

cDNA includes the complete coding region of 696 bp, 240 bp of 3′ non-coding region and 49 bp of 5′ non-coding region. We have not determined if this represents the entire 3′ or 5′ non-coding sequences of the pea cDNA. The predicted mol. wt of the pea protein, 26 149, is in good agreement with the size of the precursor protein estimated by SDS−PAGE.

In the coding regions the nucleic acid homology between the pea and soybean cDNAs is 81.2% with 50% of the differences representing synonomous changes. The predicted amino acid sequences of the two proteins are 79% identical. Of the non-identical amino acids, all changes represent evolutionarily conservative replacements with the exception of

A

```
                          20                  40                  60
HSP21   MAQSVSLSTIASPILSQKPGSSVKSTPPCMASFPLRRQLPRLGLRNVRAQAGGDGDNKDN

HSP22                                                       GGDNKDN

                          80                  100                 120
HSP21   SVEVHRVNKDDQGTAVERKPRRSSIDISPFGLLDPWSPMRSMRQMLDTMDRIFEDAITIP
                                                    .: :... .   . ...
HSP17                        MSLIPGFFGGRRSNVFDPFSLDMWDPFKDFHVPTSSVS
                                                    .  :...     . . .
HSP22   SVEVQHVSKGDQGTAVEKKPRRTAMDISPFGILDPWSPMRSMRQILDTMDRVFEDTMTFP

                        140                 160
HSP21   GRNIGGGEIRVPWEIKDEEHEIRMRFDMPGVSKEDVKVSVEDD-VLVIKSDHR-EENGGE
        . : .  ..:: :. . :.: ..   :.::..::.::: .::: :: :..... :... .
HSP17   AENSAFVSTRVDWKETPEAHVFKA--DIPGLKKEEVKVQIEDDRVLQISGERNVEKEDKN
        . : .  ..:. :. . :.: ..   :.::: ::.::: .:::... .:... .... .
HSP22   GRNIGGGEIRAPWDIKDEEHEIRMRFDMPGLAKEDVKVSVEDDMLVIKGGHKSEQEHGGD

      180                 200                 220
HSP21   DCWSR--KSYSCYDTRLKLPDNCEKEKVKAELKDGVLYITIPKTKIERTVIDVQIQ
        : : :  .: . .. :..::.: . ..:::....::: :.::..... .
HSP17   DTWHRVERSSGKFTRRFRLPENAKVNEVKASMENGVLTATVPKEEVKKPDVKAIEISG
        :.:  .:. .... :..::.: . ..:::...:::: :.::.:.. . .
HSP22   DSW--SSRTYSSYDTRLKLPDNCEKDKVKAELKNGVLYITIPKTKVERKVIDVQVQ
```

B

```
                          20                  40                  60
HSP21   MAQSVSLSTIASPILSQKPGSSVKSTPPCMASFPLRRQLPRLGLRNVRAQAGGDGDNKDN

26FF                                                        MSLSTS

                          80                  100                 120
22FF                      MRSLPMFWRMAEEMARMPRLSSPFHAFFHEPPVWSVALPRNW

HSP21   SVEVHRVNKDDQGTAVERKPRRSSIDISPFGLLDPWSPMRSMRQMLDTMDRIFEDAITIP

26FF    LSLVDELQDPRSPIYQLGLGLHPHSRYVLLPLGTQQRRSINGCPCASPICPSSPAGQVLA

                        140                 160
22FF    QHIARWQEQELPPPATVN--KDGYKLTLDVKDYSELKVKV-LDESVVLVEAKSEQQEAEQ
                .  ... .. .:. . :. .::: ....:..... .....
HSP21   GRNIGGGEIRVPWEIKDE--EHEIRMRFDMPGVSKEDVKVSVEDDVLVIKSDHREENGGE
              ..:  ... ..... .:..  . ....:.: :. .....: ::. ..
26FF    LRREMANRNDIHWPATAHVGKDGFQVCMDVAQFKPSELNVKVVDASILVEGKH-EERQDD

      180                 200                 220
22FF    GGYSSRHF---LRRFVLPEGYEADKVTSTLSSDGVLTISVPNPPGVQETLKEREVTIEQT
        . .:.. .    :. ::...: .::...: .:::: :..:..  .. :. . ..
HSP21   DCWSRKSYSCYDTRLKLPDNCEKEKVKAEL-KDGVLYITIPKTK-IERTVIDVQIQ
        .  .    .  : :.::... :.: ..: .:::: ..:::
26FF    HG---HIMGHFVRRYKVPDGYKAEQVVSQLSSDGVLTVSIPKPEAVEDKSKERIIQIQQV

22FF    GEPAKKSAEEPKDKTASQ

26FF    GPAHLNVKANDSEVKGKENGAPNGKDK
```

Fig. 6. Homology relationships of chloroplast HSPs to cytoplasmic HSPs from soybean and *Drosophila*. Double dots indicate identical amino acids, single dots indicate conservative replacements. Numbers refer to amino acid positions in the pea hsp21 sequence. (**A**) Pea hsp21 (HSP21) and soybean hsp22 (HSP22) compared with soybean hsp17.5-E (HSP17). (**B**) Pea hsp21 (HSP21) compared with *Drosophila* hsp23 (23FF) or hsp26 (26FF).

a two amino acid insertion in soybean hsp22 at residues 120 and 121. The proteins are primarily hydrophilic and have no hydrophobic segments longer than 10 amino acids, indicating that they are not integral membrane proteins.

The degree of identity between the soybean and pea HSPs is very similar to that observed in comparisons between the small subunit of ribulose bisphosphate carboxylase from the two species. Sequences of the pea and soybean small subunit proteins are 75 and 76% identical at the amino acid and nucleotide levels respectively (Berry-Lowe *et al.*, 1982).

Characteristics of the pea transit peptide

Estimating the size of the mature pea chloroplast HSP to be 21 kd (as determined by SDS−PAGE), the transit peptide has a predicted size of ~5 kd which would include the first 46−47 amino acids. This segment of hsp21 has several features common to transit peptides of chloroplast proteins (Karlin-Neuman and Tobin, 1986; Schmidt and Mishkind, 1986). The initiator Met is followed directly by Ala and multiple Ser residues appear in the next six positions. The peptide is rich in Ser, 19%, compared with 6% in the mature protein. Six basic residues (Arg + Lys) and no acidic resi-

Table I. Summary of homology relationships between pea HSP21 and other proteins

	Soybean HSP17.5 (22–147)[a]	*Drosophila* HSP22 (60–151)	*Drosophila* HSP23 (15–145)	*Drosophila* HSP26 (77–164)	Bovine α-crystallin B-chain (67–161)
Amino acid overlap[b]	104–227	138–231	83–219	131–219	134–231
No. (%) identical amino acids[c]	40 (33)	21 (23)	30 (22)	22 (25)	21 (22)
No. (%) conservative replacements	54 (44)	51 (55)	64 (47)	44 (50)	50 (52)
z value[d]	26.4	11.8	12.4	9.5	9.9

[a]Numbers in parentheses indicate the amino acid segment of the indicated protein that has homology to pea hsp21.
[b]Pea hsp21 segment with homology to the indicated protein.
[c]The percentage is calculated for the homologous segment, not the entire protein.
[d]Calculated according to Lipman and Pearson (1985): $z < 6$ is not significant; $6 < z < 10$ is probably significant; $z > 10$ is significant.

dues are located in this protein segment. In contrast, acidic and basic residues are almost equally distributed throughout the mature protein. The removal of six basic residues from the amino terminus is consistent with the pI difference between the precursor and mature proteins.

It is notable that two sequence domains, proposed as important components of chloroplast transit peptides, are not present in the proposed pea hsp21 transit peptide. A domain surrounding the sequence GLK identified by Schmidt and Mishkind (1986) and included in the framework structure proposed by Karlin-Neuman and Tobin (1986) (block II, P-F-G-K) is not found in the transit peptide region of the HSP. Second, a site proposed to be required for recognition by the transit peptidase, which includes the sequence GRV preceding the cleavage site, is not present within 20 amino acids to either side of the predicted hsp21 cleavage site.

Proteins homologous to chloroplast HSPs

The PIR protein data base from the National Biomedical Research Foundation (March 1987 revision) was searched for amino acid sequences similar to pea hsp21 or soybean hsp22 using the FASTP program (Lipman and Pearson, 1985). Three groups of proteins with significant sequence similarities were identified; low mol. wt cytoplasmic HSPs from soybean, low mol. wt HSPs from *Drosophila*, and α-crystallins from several eukaryotes (Figure 6 and Table I). The strongest homology of the chloroplast HSPs with all of these proteins is found in the carboxy-terminal half of the protein. The NH_2-terminal 50 amino acids of the mature chloroplast HSPs showed no significant relationship to any protein in the data bank.

Comparison of the chloroplast HSPs to soybean hsp17.5-E (Czarnecka *et al.*, 1985), one member of a major family of soybean cytoplasmic HSPs, shows two regions with > 40% identity corresponding to amino acids 138–163 and 192–220 of pea hsp21 (Figure 6, panel A). These segments are also the regions of greatest homology between the chloroplast HSPs and low mol. wt HSPs from *Drosophila* (panel B). Between these segments is a region of sequence heterogeneity, both in comparisons between the two chloroplast HSPs or between the different low mol. wt HSPs. Despite this sequence flexibility the segment is markedly hydrophilic in all the HSPs, suggesting these segments are structurally similar.

The weakest homology relationship is to the α-crystallins (Table I). Amino acid sequence similarity between low mol. wt HSPs and α-crystallins has been previously noted (Ingolia and Craig, 1982), and it has been proposed that these two groups of proteins evolved from a common ancestor (van den Heuvel *et al.*, 1985).

Discussion

We have isolated cDNAs for nuclear-encoded genes from soybean and pea which specify chloroplast-localized HSPs. The cDNAs were demonstrated to encode chloroplast proteins by the criterion that they hybrid-select mRNAs which translate to precursor polypeptides that can be imported and processed by chloroplasts *in vitro*. Two additional criteria support the identification of the pea cDNA: (i) translation of transcripts synthesized *in vitro* produces a 26-kd precursor of a 21-kd chloroplast protein and (ii) the mature chloroplast-localized hsp21 observed in the *in vitro* experiments is identical in size and pI to the major HSP previously shown to be present in chloroplasts isolated from heat-stressed plants (Vierling *et al.*, 1986).

As is typical of HSP gene regulation in higher plants and other organisms, the chloroplast HSP mRNAs are undetectable at normal temperatures. The 150-fold increase in chloroplast HSP mRNA, to a level representing 0.75% of total poly(A) RNA after 2 h of stress, is the most dramatic example of environmental regulation of a chloroplast protein known to date. The abundance of the mRNA is at least one-third that of the mRNA for the small subunit of ribulose bisphosphate carboxylase [~2% of the total poly(A) RNA in pea leaves; Bedbrook *et al.*, 1980]. For comparison, we have also estimated mRNA abundance of a 20-kd cytoplasmic HSP from pea to be 3% of the poly(A) RNA after a 2-h heat shock, which is significantly higher than that of the chloroplast HSP (A.DeRocher, unpublished observations). Whether the difference in mRNA levels of the chloroplast and cytoplasmic HSPs is a result of differences in gene number, promoter strength or mRNA stability remains to be determined.

Although the proposed transit peptide of the pea chloroplast HSP has several characteristics in common with other transit peptides (rich in Ser, basic, not hydrophobic), it lacks sequences similar to proposed recognition sites for intermediate and maturation processing (Karlin-Neumann and Tobin, 1986; Schmidt and Mishkind, 1986). The absence of these sequences suggests either that import and processing of chloroplast proteins does not require strict sequence specificity, or that the chloroplast HSPs are processed by different receptors/proteases than those which recognize the small subunit of ribulose bisphosphate carboxylase. The lack of a sequence framework in mitochondrial transit peptides

(von Heijne *et al.*, 1986), along with recent experimental data indicating the intermediate processing site is not required for import (Reiss *et al.*, 1987) would support the former interpretation. If the chloroplast HSPs are processed by different receptors/proteases, such proteins are present in the chloroplast envelope prior to heat stress, as shown by import and processing of HSPs by chloroplasts isolated from control plants.

The homology of the chloroplast HSPs to low mol. wt HSPs found in the cytoplasm of all eukaryotes suggests that these proteins have a common evolutionary ancestor. HSPs with a similar domain structure have not been identified among the HSPs in *Escherichia coli*. Using heterologous hybridization, we have also been unable to detect similar genes in cyanobacteria (E.Vierling and D.Bryant, unpublished observations). Failure to identify prokaryotic homologs of the eukaryotic low mol. wt HSPs may indicate that these proteins are specific to eukaryotes. Assuming this to be the case, we propose that the gene for the chloroplast HSP has evolved from a nuclear gene encoding a cytoplasmically localized HSP. At some point the gene could have acquired sequences that created an amino-terminal protein segment sufficient to direct import into the chloroplast. These changes would have been stabilized by subsequent selection if this new protein had a beneficial function within the chloroplast. A chloroplast glutamine synthetase with stronger homology to eukaryotic than to prokaryotic forms of the enzyme (Tingey *et al.*, 1987) may have evolved in a similar fashion. As genes for more chloroplast proteins are identified, many may be genes which were not encoded by the chloroplast genome before the original proposed establishment of its endosymbiotic relationship with eukaryotes.

Identification of the chloroplast HSPs as members of a superfamily of low mol. wt HSPs may indicate that these proteins have a similar function to their counterparts in the cytoplasm. However, the functions of small HSPs (as well as other HSPs) have yet to be elucidated. Experiments with yeast, which appears to have only one small HSP, have failed to identify a phenotype in mutants in which this gene was deleted or disrupted (Petko and Lindquist, 1986). Loomis and Wheeler (1982) reported that *Dictyostelium* mutants incapable of synthesizing small HSPs had a thermosensitive phenotype. Mutants in other organisms have not been identified. In many organisms, small HSPs can be isolated as large protein aggregates (Lindquist, 1986). This is also true of the crystallin proteins and has been suggested to result from similar properties of the two classes of proteins. However, no biological activity has been associated with HSP particles. Whether the chloroplast HSP forms similar aggregates has not been tested.

Ohad and colleagues have proposed that a small chloroplast HSP in *Chlamydomonas* functions to protect photosystem II activity during high temperature stress (Schuster *et al.*, 1987). This conclusion is based on their observations that (i) cycloheximide inhibition of HSP synthesis prevents the development of photosystem II thermotolerance and (ii) high temperatures *in vivo* result in the association of the chloroplast HSP with thylakoid membranes. As reported previously, we have not found a similar thylakoid association in pea *in vivo* (Vierling *et al.*, 1986). It is possible that formation of aggregates of the chloroplast HSP, as mentioned above, produces particles with a similar density to thylakoid membranes. Further studies of HSP localization within the chloroplast are required to resolve this discrepancy.

We have recently observed that approximately equal levels of chloroplast HSP mRNA are present in both roots and leaves of pea during heat stress (A.DeRocher, unpublished data). Based on this result and the observed homology of the chloroplast and cytoplasmic HSPs, we hypothesize that these proteins are present in all types of plastids and perform a function similar to their cytoplasmic counterparts. If they also affect photosynthesis, it may be as an indirect consequence of a more general function, for example membrane stabilization or protection from proteolysis. Until an *in vitro* assay for the function of these proteins is developed, or specific mutants can be isolated, their protective function remains speculative.

Although many of the characteristics of the heat shock response and HSP structure have been conserved during evolution, higher plants are unique in the complexity and abundance of their low mol. wt HSPs. Five distinct gene families of small HSPs have been identified in soybean (Nagao *et al.*, 1985, 1986) including the family encoding the chloroplast HSP. We have also identified three of these families in pea and two in petunia (Vierling, 1987). The common domain shared by the chloroplast and 17.5-kd cytoplasmic HSP is present in the other three families of soybean cytoplasmic HSPs as well (R.T.Nagao, in preparation). The amino-terminal domains, which have diverged between families, may specify each protein's interactions with different cell structures. An understanding of the significance of such a complex and diversified group of small HSPs in higher plants will require that the functions of these ubiquitous proteins be determined. The discovery of small HSPs that localize to the chloroplast extends the potential significance this particular group of proteins may have to eukaryotic cell function during stress.

Materials and methods

cDNA cloning

A cDNA library in pUC9 was made from soybean (*Glycine max*, cv. Corsoy) poly(A) RNA enriched for the chloroplast HSP mRNA by size fractionation (Vierling, 1987). The cDNA synthesis was according to Gubler and Hoffman (1983). The library was screened by differential hybridization with single-stranded ^{32}P-labeled control or heat shock cDNA to identify heat shock positive clones (Maniatis *et al.*, 1982). The heat shock positive clones were rescreened by hybridization to previously cloned cDNAs encoding 15- to 18-kd soybean HSPs (Nagao *et al.*, 1985) to eliminate these from further consideration.

A cDNA library of total pea (*Pisum sativum*, cv. 'Little Marvel') poly(A) RNA from heat shock tissue was constructed in λgt10 as described by Huynh *et al.* (1985) and screened with the soybean hsp21 clone at reduced stringency in 6 × SSC, 5 × Denhardt's, 50 μg/ml salmon sperm DNA and 0.2% SDS, at 60°C. The cDNA inserts were subcloned from the phage into pUC or SP6 plasmids.

RNA isolation

Soybean poly(A) RNA for library construction and all experiments was obtained from a photosynthetic line of tissue culture cells as described previously (Vierling and Key, 1985). Pea poly(A) RNA was isolated from 8- to 10-day-old plants grown in vermiculite under a 16-h light, 23°C day/17°C night regime. For both control and heat shock samples, whole seedlings or tissue culture cells were incubated in H_2O in the light in shaking water baths. Heat treatments were for 2 h at 39°C for both pea and soybean. Control plants were treated under identical conditions except at 23°C.

In vitro transcription

A pea cDNA containing the AUG start codon for translation was cloned into SP65 (Promega). Transcription of the linearized plasmid was performed using SP6 polymerase as recommended by the supplier (Promega). The transcripts were translated without prior capping in a reticulocyte lysate system (Bethesda Research Laboratories).

Hybridization-selection of mRNA

Hybridization-selection reactions were performed as described previously by Gantt and Key (1985). Eluted mRNAs were translated in either wheat germ extracts (Gantt and Key, 1985) or reticulocyte lysates.

Chloroplast import, gel electrophoresis

For all chloroplast import experiments, *in vitro* translation in wheat germ extracts was performed as described by Vierling *et al.* (1986). Reticulocyte lysate was used according to the supplier. All translations were performed in the presence of [^{35}S]methionine (Du Pont/NEN, >1000 Ci/mmol) at 1 μCi/μl. Isolation of intact pea chloroplasts, import reactions, protease treatment and reisolation of chloroplasts were carried out as described previously (Vierling *et al.*, 1986).

One-dimensional SDS−PAGE was performed on 12.5% acrylamide or 10−16% gradient acrylamide gels. Two-dimensional electrophoresis was according to O'Farrell (1975), using 12.5% SDS−acrylamide gels for the second dimension.

Northern and dot-blot hybridizations

Northern and dot-blot hybridizations were performed according to Vierling and Key (1985). Poly(A) RNA amounts were quantified using a poly(U) hybridization assay (Bishop *et al.*, 1974). Probes for hybridization were prepared using isolated cDNA inserts. Fragments were labeled with [^{32}P]-dATP by the random-primer labeling method (Feinberg and Vogelstein, 1983).

DNA isolation and Southern analysis

Total genomic DNA was isolated according to Shure *et al.* (1983). DNA was digested with restriction enzymes according to the supplier (Boehringer Mannheim Biochemicals), electrophoresed on 0.8% agarose gels and blotted to nitrocellulose (Maniatis *et al.*, 1982). Blots were hybridized with probes prepared as for Northern analysis, using 5−10 × 10^6 c.p.m. per filter. Hybridization and washing was performed as for Northern analysis.

DNA sequencing and sequence analysis

The complete DNA sequence of both strands of the cDNAs was obtained by either the Maxam and Gilbert chemical sequencing technique (soybean) (Maniatis *et al.*, 1982) or the M13 dideoxy method (pea) (Messing, 1983). Sequences were analyzed on a microcomputer using the programs of Schwindinger and Warner (1984). Analysis of protein homologies was performed using the FASTP program described by Lipman and Pearson (1985).

Acknowledgements

We thank Dr J.L.Key in whose laboratory this work was initiated, Dr R.Hallick for help with DNA sequence analysis and Drs C.Wasmann and J.Kimpel for critical advice concerning the manuscript. Special regards to J.Laswick for photographic work. This research was supported by NIH BRSG 828657 and NSF DCB-8517576 to E.V.

References

Bedbrook,J.R., Smith,S.M. and Ellis,R.J. (1980) *Nature*, **287**, 692−697.
Berry-Lowe,S.L., McKnight,T.D., Shah,D.M. and Meagher,R.B. (1982) *J. Mol. Appl. Genet.*, **1**, 483−498.
Bishop,J.O., Rosbach,M. and Evans,D. (1974) *J. Mol. Biol.*, **85**, 75−80.
Czarnecka,E., Gurley,W.B., Nagao,R.T., Mosquera,L.A. and Key,J.L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3726−3730.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 36−38.
Gantt,J.S. and Key,J.L. (1985) *J. Biol. Chem.*, **260**, 6175−6181.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263−268.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning*. IRL Press, Oxford, Vol. I, pp. 49−78.
Ingolia,T.D. and Craig,E. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2360−2364.
Karlin-Neumann,G.A. and Tobin,E.M. (1986) *EMBO J.*, **5**, 9−13.
Kimpel,J.A. and Key,J.L. (1985) *Trends Biochem. Sci.*, **10**, 353−357.
Kloppstech,K., Meyer,G., Schuster,G. and Ohad,I. (1985) *EMBO J.*, **4**, 1901−1909.
Lindquist,S.L. (1986) *Annu. Rev. Biochem.*, **55**, 1151−1191.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435−1441.
Loomis,W.F. and Wheeler,S. (1982) *Dev. Biol.*, **79**, 399−408.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20−78.
Nagao,R.T., Czarnecka,E., Gurley,W.B., Schoffl,F. and Key,J.L. (1985) *Mol. Cell. Biol.*, **5**, 3417−3428.
Nagao,R.T., Kimpel,J.A., Vierling,E. and Key,J.L. (1986) In Miflin,B.J. (ed.), *Oxford Surveys of Plant Molecular and Cell Biology*. Vol. 3, pp. 384−438.
O'Farrell,P.H. (1975) *J. Biol. Chem.*, **250**, 4007−4013.
Petko,L. and Lindquist,S. (1986) *Cell*, **45**, 885−894.
Reiss,B., Wasmann,C.C. and Bohnert,H.J. (1987) *Mol. Gen. Genet.*, **209**, 116−121.
Schmidt,G.W. and Mishkind,M.L. (1986) *Annu. Rev. Biochem.*, **55**, 879−912.
Schuster,G., Evan,D., Kloppstech,K. and Ohad,I. (1987) In Biggins,J. (ed.), *Progress in Photosynthesis Research*. Martinus Nijhoff, Dordrecht.
Schwindiger,W.F. and Warner,J.R. (1984) *Nucleic Acids Res.*, **12**, 601−604.
Shure,M., Wessler,S. and Federoff,N. (1983) *Cell*, **35**, 225−233.
Suss,K.-H. and Yordanov,I.T. (1986) *Plant Physiol.*, **81**, 192−200.
Tingey,S.V., Walker,E.L. and Corruzi,G.M. (1987) *EMBO J.*, **6**, 1−10.
van den Heuvel,R., Hendriks,W., Quax,W. and Bloemendal,H. (1985) *J. Mol. Biol.*, **185**, 273−284.
Vierling,E. (1987) In McIntosh,L. and Key,J.L. (eds), *UCLA Symposia Mol. Cell. Biol.* New Series, Vol. 62, pp. 99−108.
Vierling,E. and Key,J.L. (1985) *Plant Physiol.*, **78**, 155−162.
Vierling,E., Mishkind,M.L., Schmidt,G.W. and Key,J.L. (1986) *Proc. Natl. Acad. Sci. USA.*, **83**, 361−365.
Vierling,E., Roberts,J.K., Nagao,R.T. and Key,J.L. (1987) In Biggins,J. (ed.), *Progress in Photosynthesis Research*. Martinus Nijhoff, Dordrecht.
von Heijne,G. (1986) *EMBO J.*, **5**, 1335−1342.

Received on December 4, 1987