The Organization and Evolution of the Spinach Stress 70 Molecular Chaperone Gene Family

Charles L. Guy and Qin-Bao Li

Plant Molecular and Cellular Biology Program, Department of Environmental Horticulture, 1545 W.M. Fifield Hall, University of Florida, Gainesville, Florida 32611-0670

The stress 70 molecular chaperones of plants are localized and function in all of the major subcellular compartments of the cell. Collectively, all of the various forms are encoded by a multigene family in the nucleus. At least 12 members of this family have been found, and sequence and DNA blot analyses provide an emerging description of the diversity of gene structure organization for this family of evolutionarily conserved proteins in spinach. They exhibit not only structural diversity in the organization of coding and noncoding regions but also distinct expression patterns for different tissues and abiotic conditions. The results of phylogenetic analyses are concordant with at least four major evolutionary events that gave rise to stress 70 molecular chaperones in each of four major subcellular compartments of plant cells: the plastid, mitochondrion, cytoplasm, and endoplasmic reticulum. The varied expression patterns also illustrate the complexity of effectively interpreting the role of any one of these stress-related proteins in response to abiotic stress in the absence of context to the other members of the family.

INTRODUCTION

A diverse group of at least 11 different classes of proteins that show transient increased expression in response to a rapid increase in temperature comprises the heat shock proteins (Nover and Scharf, 1997). The heat shock response is common to bacteria, animals, and plants and is considered to be an adaptive mechanism to help a cell, organ, or organism cope with an unfavorable temperature condition (see Morimoto et al., 1994). Many of the different classes of heat shock proteins function in protein metabolism as molecular chaperones (Ellis, 1993). Some function in biogenesis by assisting in folding processes and translocation into organelles; others are involved in assembly and disassembly of oligomers or proteolysis of damaged or misfolded proteins (see Morimoto et al., 1994).

The Hsp70 (proteins of ~70 kD) class of heat shock proteins has been found to serve a variety of functions. Because Hsp70 is a major protein induced by heat shock, it has often been linked to the characteristic of acquired thermotolerance (Lee and Schoffl, 1996). An organism subjected to a short (1 or 2 hr) nonlethal heat treatment will subsequently be able to withstand a higher temperature and survive, whereas it would not be able to without the pretreatment. This survival at higher temperatures is the result of acquired thermotolerance. One way that Hsp70 might function in acquired thermotolerance would be to stabilize heat-labile proteins and supermolecular structures (Minton et al., 1982) and prevent them from becoming irreversibly denatured or damaged.

Hsp70 can prevent irreversible aggregation of nonnative proteins by binding to determinants not normally solvent accessible in native conformations. The binding and release of denatured and nonnative peptides containing residues with nonpolar side chains are nucleotide dependent (Flynn et al., 1989). Thus, Hsp70 proteins are composed of separate nucleotide binding and peptide binding domains. Binding nonpolar regions in nonnative peptides can prevent coalescence into an insoluble aggregate, thereby maintaining the peptide in a form that may be subsequently refolded to a native conformation. Peptide release from Hsp70 is associated with a cycle of ATP hydrolysis and release of ADP, which can be facilitated by cochaperones (Pierpaoli et al., 1997).

The peptide binding and release of heat denatured proteins and prevention of aggregation during heat shock by Hsp70 have a more general application in cells. Proteins are initially synthesized as nascent linear chains without higher order structure. To prevent irreversible aggregation before folding and assembly can be completed for a newly synthesized protein, there is a requirement for constitutive Hsp70 peptide binding and release functions. Indeed, constitutively expressed cognates of the heat shock–induced Hsp70 (HSC70s) are now known for many organisms, including plants (Ingolia and Craig, 1982; Craig et al., 1983; Duck et al., 1989; Marshall et al., 1990; DeRocher and Vierling, 1995). Their role as chaperones of nonnative nascent polypeptides has many facets, including assisting at several steps or
stages along the protein biogenesis pathway, stabilizing folded proteins, and participating in degradative processes of abnormally folded polypeptides.

Heat shock-inducible Hsp70 and the constitutively expressed cognates are encoded by gene families in most organisms. The yeast HSP70 family, also known as the stress 70 family, has been the most thoroughly characterized of all higher organisms (Boorstein et al., 1994). The yeast stress 70 family has 14 known members, and most were discovered before the complete sequencing of the yeast genome (James et al., 1997; Rassow et al., 1997). The stress 70 family of Escherichia coli includes two members, dnaK and hscA (Bardwell and Craig, 1984; Kawula and Lelivelt, 1994; Seaton and Vicker, 1994), that encode proteins of \( \sim 70 \) kD and four additional genes that have stress 70 motifs and encode proteins ranging from 278 to 556 amino acids (Tatusov et al., 1997). At least nine distinct members of the stress 70 family are known for Caenorhabditis spp (Heschl and Baillie, 1990), nine for Drosophila (Rubin et al., 1993; Günther and Walter, 1994), and 10 for humans (Mues et al., 1986; Günther and Walter, 1994). Many stress 70 genes have been characterized from a variety of plants, including both the heat-inducible (Roberts and Key, 1991) and cognate forms for all of the major subcellular compartments (Duck et al., 1989; Denecke et al., 1991; Lin et al., 1991; Watts et al., 1992; Wang et al., 1993; DeRocher and Vierling, 1995), but little is known about the family in any plant. However, without an understanding of all of the members in even one plant species, it is difficult to know fully how they contribute to cell function or recognize whether a new member in a second species is orthologous or paralogous with other known stress 70 genes. Similarly, it remains difficult to understand adequately the evolution of this important family of essential proteins in plants or higher order aspects of their developmental and environmental specialization and stress regulation.

Several classes of heat shock proteins have become important tools in our understanding of the molecular evolution of organisms, especially when it concerns the origin of mitochondria, because of their near-universal occurrence and the relative degree of conservation that facilitates sequence data comparisons (Bui et al., 1996). However, only the evolutionary relationships of the small heat shock proteins (sHsp) of plants have been explored to date. The plant sHsp superfamily is composed of five (Waters, 1995) or six (Waters et al., 1996) families that are hypothesized to form monophyletic groups (Waters, 1995). The different families are considered to have arisen by duplications that occurred before the diversification of flowering plants. The diversification that characterizes this heat shock protein family is thought to reflect a molecular adaptation to the unique environmental and developmental conditions that plants must withstand during their life cycle (Waters, 1995). When different sHsp families are compared, it appears that they exhibit unequal rates of molecular evolution (Waters et al., 1996), as has been suggested for some of the subclasses of the stress 70 family (Ohta, 1994).

Here, we present the results of a comprehensive characterization of the stress 70 multigene family of spinach. We show that spinach has a minimum of 12 stress 70 genes, with most having one or more introns. The genes show differential expression patterns in a variety of tissues and in response to different environmental conditions. Using sequence and gene structural organization data, we show that members of the four major subcellular compartments of plants arose in four separate, independent evolutionary events. Furthermore, the family appears to have arisen from a series of duplications and divergences, of which some are considered to have occurred relatively recently.

**RESULTS**

**Isolation of Clones**

The stress 70 molecular chaperones of spinach are encoded by a gene family, and Table 1 summarizes all of the different members of the family currently known and their corresponding molecular clones. The coding sequences vary in length from 1941 bp for the cytosolic hsc70-1 to 2145 bp for the stromal form (hsc70-9). The number of amino acids encoded ranges from 647 for the cytosolic forms to 715 for the stromal form, which includes a long transit peptide.

**Protein Alignments**

The complete deduced protein sequences for three new spinach stress 70 members (HSC70-3, HSC70-9, and HSC70-10) for three separate compartments are shown in Figure 1. These include the cytosolic, plastid stromal, and mitochondrial matrix. The first conserved motif \([V/A][V/I]GIDLGGTFDVS\) of these and many stress 70 proteins begins eight residues from the N terminus of the cytosolic members (HSC70-1 and HSC-3) but farther downstream in the endoplasmic reticulum (ER) luminal (HSC70-2), mitochondrial (HSC70-10), and plastid stromal (HSC70-9) forms because of their targeting presequences. We prepared a primer, CG24, that is positioned at the center of this motif. A second area of conservation begins at residue 145 in the cytosolic members and extends to residue 183. A short, highly conserved region of \(FDLGGTDFDV\) begins at residue 204. This is the location for primers CG14 and CG144. From residues 240 to 339, there is a region of sequence diversity. Another region of diversity among the members was found beginning at residue 526 and continuing to the C terminus. The C terminus, depending on subcellular localization, tends to be conserved across evolutionarily distant organisms (Boorstein et al., 1994), but it is clearly unique to each of the major subcellular localization sites in plants. For the cytosolic forms, the conserved sequence is \(GPKIEEVD\); for the ER luminal member, it is HDEL; for the mitochondrial matrix member, it...
is PEAHEYEEAKK; and for the stromal member, it is PEG-DVIDADFTDSK. The C-terminal motif is known to be a subcellular localization factor for the ER luminal form (Munro and Pelham, 1987) and may function as an ATPase domain interaction/recognition site for the cytosolic forms (Hightower et al., 1994). Whether the C terminus of the plant organelle forms have functional significance is not known; however, given their conservation, it seems likely.

To date, we have sequenced nearly full-length cDNAs for six members of the spinach stress 70 family. hsc70-1 is 99.2% identical to a cDNA sequence designated as sce70 (Ko et al., 1992), and hsc70-9 is 99.6% identical to a previously described partial cDNA sequence (chsp70) (Wang et al., 1993).

DNA and protein sequence analyses limited to the coding regions for the six members revealed the degree of relatedness for the members representing four different subcellular compartments: cytosol, ER lumen, mitochondrial matrix, and chloroplast stroma. The two mitochondrial members are most similar to each other, as are two of the cytosolic forms (Table 2). The two mitochondrial members have 97.1% nucleotide sequence identity, whereas the cytosolic members (hsc70-1 and hsc70-3) are 84.3% identical. The protein sequences for the two pairs have an identity of 98.8 and 92.9%, respectively. The chloroplast stromal member (hsc70-9) had the least nucleotide and amino acid sequence homology with the cytosolic members. The ER luminal member was more homologous to the cytosolic members than to either the mitochondrial matrix or the chloroplast stromal members. In contrast, the chloroplast and mitochondrial members were not significantly more homologous to each other than to any members localized to other subcellular compartments.

Limiting the sequence comparisons to the portion of the ATP binding domain that is flanked by the highly conserved ATP interaction motifs (Flaherty et al., 1990; McKay et al., 1994), we were able to assess the relatedness of 11 members of the spinach stress 70 family (Tables 1 and 3). This part of the DNA and protein sequences is highly conserved,
Figure 1. Alignment of the Deduced Amino Acid Sequences for Five Members of the Spinach Stress 70 Molecular Chaperone Family Localized to Four Different Subcellular Compartments.
showing minimum identities of 60.4 and 44.5%, respectively. Again, the chloroplast stromal member was found to be the most divergent of the family. A group of related members (hsc70-5, hsc70-6, hsc70-7, and hsc70-8) were most similar to the major cytosolic forms (hsc70-1 and hsc70-3) and most divergent from the chloroplast stromal member. At this time, we have not resolved their subcellular localization, but we suspect that they might be another complement of cytosolic stress 70 proteins based on their sequence and intron structure similarity to the major cytosolic members. Another cytosolic member (hsc70-12) originally designated as HSE is known by its partial cDNA (Ko et al., 1992), and it belongs to the hsc70-1 and hsc70-3 subgroup.

**Gene Structure and Polymerase Chain Reaction**

Genomic clones for two of the major cytosolic forms have been characterized. Both genes have a single large intron that interrupts the coding sequence at the position where residue 72 downstream of the N terminus is encoded (Figure 2A). The intron for both lies squarely in the highly conserved ATP binding domain. The intron for hsc70-1 is 1034 bp, whereas the intron for hsc70-3 is 941 bp. The genomic structure for the ER luminal form had been characterized previously in our laboratory (Anderson et al., 1994) and shown to have seven introns. Two of these reside in the same vicinity of the coding region, but neither of them is positioned in exactly the same spot as the introns in the cytosolic forms. Also in striking contrast, the two introns of the ER luminal member are small (103 and 111 bp) compared with the ~1-kb introns in the two major cytosolic forms. The ER luminal form does possess a large intron, but it resides in the targeting sequence for translocation into the ER (Anderson et al., 1994).

Two genomic clones from the EMBL4 library, α and β, were obtained for hsc70-3. They differed in length by a >500-bp extension at the 5′ end of the hsc70-3β clone (Figure 2A). Sequencing revealed that the two are 99.9% identical over the 5000 bp that include the entire coding sequence, the intron, and the 3′ noncoding region. At position −877 (805 and 1340 of the α and β clones), the sequences abruptly and completely diverge. We isolated two independent β clones and many α clones, making it unlikely that the different sequences arose from a cloning artifact.

Using a polymerase chain reaction (PCR) approach, we have also established the structural organization for a plastid and mitochondrial member. The mitochondrial member has four introns. Two intersperse the targeting peptide sequence, one is in the ATP binding domain, and the last is in the C-terminal domain (Figure 2). In contrast, the plastid stromal member has seven introns. None of these interrupts the targeting sequence. The spinach plastid stress 70 gene structure is relatively conserved compared with that for the pea plastid member, which also has seven introns in approximately the same positions (Casey et al., 1994).

It occurred to us that with two exceptions (Roberts and Key, 1991; Dhankher et al., 1997), almost all of the plant stress 70 members in which the genomic sequence had been obtained possessed one or more introns in the ATP binding domain that is flanked by two highly conserved motifs associated with the interaction of ATP at the binding site. The high conservation of these motifs has been maintained across kingdoms and provides a basis for designing primers that will amplify sequences from a variety of the members of this family. We used two sets of primers to amplify, from genomic DNA, fragments corresponding to a 190 to 200+ amino acid region of the ATP binding domain that encompassed the site(s) of insertion of the intron(s). Our initial primers were based on a limited set of plant stress 70 sequences. As more sequences became available for both the spinach family and other plant members, we redesigned the 3′ primer to be more universally homologous. Figure 2B shows

| Table 2. Comparison of Spinach hsc70 DNA and Protein Sequences |
|-------------------|---------------|---------------|---------------|---------------|---------------|
| hsc70-1           | hsc70-2       | hsc70-3       | hsc70-9       | hsc70-10      | hsc70-11      |
| hsc70-1           | —             | 66.0          | 84.3          | 59.7          | 61.0          | 63.3          |
| hsc70-2           | 64.1          | —             | 65.7          | 58.7          | 60.1          | 60.3          |
| hsc70-3           | 92.9          | 63.7          | —             | 61.1          | 60.7          | 61.1          |
| hsc70-9           | 47.4          | 48.0          | 46.8          | —             | 59.1          | 59.5          |
| hsc70-10          | 50.0          | 50.1          | 48.9          | 53.8          | —             | 97.1          |
| hsc70-11          | 49.7          | 49.7          | 48.6          | 54.6          | 98.8          | —             |

*DNA sequence homologies are above the dashed diagonal, and protein sequence homologies are below the dashed diagonal. Values are expressed as percentages.*
the diversity of fragments that were obtained with the first- and second-generation primers. As we had hoped, different members of the family possessed different numbers and sizes of introns and therefore yielded different-sized fragments. Six fragments representing eight different members of the stress 70 family were verified (Figure 2B).

We cloned and sequenced all of these fragments and used them to screen the cDNA library for representative clones. Surprisingly, a PCR product for hsc70-1 was not obtained. We have no reasonable explanation, because the second-generation primer is a perfect match for this gene class. This means that there are at least six different subclasses of stress 70 genes present in spinach. Also, we fully expected to amplify a fragment of ~600 bp representing a gene possessing no introns in this part of the coding sequence. After many attempts, no such product was produced, so it remains an open question whether spinach possesses a heat shock-inducible member of the stress 70 family completely lacking an intron(s), as do other organisms, including some plants.

At least five patterns of intron arrangement are evident for the various family members (Figure 2B). The mitochondrial forms have a single small intron of 81 to 84 bp in the ATP binding domain. In contrast, the ER luminal member has two small introns located at different sites. The stromal form has a single large intron in the region flanked by the ATP interaction motifs that is positioned more toward the C terminus than it is in any of the other members. The fourth pattern is represented in four members (hsc70-5, hsc70-6, hsc70-7, and hsc70-8). The large intron in this group is positioned more toward the N terminus than that of the ER luminal and plastid members and 6 bp downstream of where the intron in the major cytosolic forms is found (hsc70-1 and hsc70-3). The intron arrangement of this group is most similar in placement and size to that of the two major cytosolic members. The intron inventory of the plastid and mitochondrial members is not limited to the single intron described above.

The plastid member has other introns 3’ of the CG14/144 site, and the mitochondrial member has one intron in the C-terminal half of the coding region and two other introns in the targeting peptide region (Figure 2A).

### Gene Family Complexity

The genomic complexity for the two closely related major cytosolic forms and the stromal and matrix forms was examined by DNA gel blot hybridization using full-length cDNA probes. Stringency was increased during washing until the radiolabel remaining on the blots was reduced to the minimum detectable (Figure 3). Based on the known EcoRI sites for the two cytosolic forms (Figure 2A) and genomic clone restriction fragments (data not shown), a total of five hybridization bands were expected. Fragments for the predicted α and β genomic clones for hsc70-3 were found and confirmed the veracity of the clones. However, at least 12 restriction fragments were detected. Some were very strong, suggesting the existence of additional hybridizing members. hsc70-12 is a cytosolic member most related to hsc70-1 and hsc70-3, and it is probable that some of the additional strong hybridizing fragments belong to its gene(s) or possibly to other yet unidentified members. Alternatively, some of the weaker bands may belong to the other subgroup of members (hsc70-5, hsc70-6, hsc70-7, and hsc70-8), with more limited relatedness to the two major cytosolic forms.

The same membranes were stripped and reprobed for the chloroplast and mitochondrial forms (Figure 3). By reprobing the blots, it was possible to visualize potentially cross-hybridizing fragments as well as to resolve fragments that differed little in length. The patterns for these two organelle forms shared some differences as well as some similarities with the cytosolic forms, despite the high conservation of this family. What was most evident from these blots is that far fewer bands were detected for the two organelle mem-

### Table 3. Comparison of Spinach hsc70 DNA and Protein Sequences from the ATP Binding Region Encompassed by Primers CG24-CG14/144

<table>
<thead>
<tr>
<th>hsc70-1</th>
<th>hsc70-2</th>
<th>hsc70-3</th>
<th>hsc70-4</th>
<th>hsc70-5</th>
<th>hsc70-6</th>
<th>hsc70-7</th>
<th>hsc70-8</th>
<th>hsc70-9</th>
<th>hsc70-10</th>
<th>hsc70-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>69.7</td>
<td>82.9</td>
<td>66.4</td>
<td>72.8</td>
<td>73.8</td>
<td>73.0</td>
<td>73.4</td>
<td>60.6</td>
<td>63.0</td>
<td>63.7</td>
<td></td>
</tr>
<tr>
<td>71.0</td>
<td>68.8</td>
<td>67.5</td>
<td>67.5</td>
<td>64.7</td>
<td>64.9</td>
<td>65.2</td>
<td>61.3</td>
<td>62.5</td>
<td>62.4</td>
<td></td>
</tr>
<tr>
<td>92.0</td>
<td>70.4</td>
<td>73.7</td>
<td>73.7</td>
<td>73.4</td>
<td>73.4</td>
<td>74.1</td>
<td>61.3</td>
<td>61.0</td>
<td>61.7</td>
<td></td>
</tr>
<tr>
<td>69.3</td>
<td>61.9</td>
<td>68.6</td>
<td>71.4</td>
<td>70.9</td>
<td>68.9</td>
<td>69.9</td>
<td>63.2</td>
<td>65.4</td>
<td>62.9</td>
<td></td>
</tr>
<tr>
<td>67.5</td>
<td>62.6</td>
<td>69.4</td>
<td>65.0</td>
<td>93.0</td>
<td>92.1</td>
<td>91.9</td>
<td>61.4</td>
<td>65.1</td>
<td>65.3</td>
<td></td>
</tr>
<tr>
<td>69.4</td>
<td>62.6</td>
<td>70.9</td>
<td>64.3</td>
<td>92.2</td>
<td>95.5</td>
<td>94.7</td>
<td>61.8</td>
<td>64.4</td>
<td>67.1</td>
<td></td>
</tr>
<tr>
<td>69.9</td>
<td>63.6</td>
<td>70.9</td>
<td>64.3</td>
<td>91.7</td>
<td>94.7</td>
<td>94.7</td>
<td>99.0</td>
<td>62.5</td>
<td>65.3</td>
<td></td>
</tr>
<tr>
<td>69.9</td>
<td>63.6</td>
<td>70.9</td>
<td>65.0</td>
<td>91.7</td>
<td>94.7</td>
<td>94.7</td>
<td>99.0</td>
<td>62.5</td>
<td>65.3</td>
<td></td>
</tr>
<tr>
<td>51.0</td>
<td>55.8</td>
<td>50.0</td>
<td>44.5</td>
<td>46.4</td>
<td>46.9</td>
<td>47.3</td>
<td>47.3</td>
<td>60.4</td>
<td>61.4</td>
<td></td>
</tr>
<tr>
<td>54.5</td>
<td>58.9</td>
<td>51.0</td>
<td>51.1</td>
<td>50.2</td>
<td>49.3</td>
<td>51.7</td>
<td>51.7</td>
<td>56.2</td>
<td>98.6</td>
<td></td>
</tr>
<tr>
<td>54.5</td>
<td>58.9</td>
<td>51.0</td>
<td>51.1</td>
<td>50.2</td>
<td>49.3</td>
<td>51.7</td>
<td>51.7</td>
<td>56.2</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

aDNA sequence homologies are above the dashed diagonal, and protein sequence homologies are below the dashed diagonal. Values are expressed as percentages.
bers compared with the cytosolic members, which is a pattern that is similar to that found previously for the ER luminal form (Anderson et al., 1994).

A strong signal at 2.2 kb was observed for the stromal form as well as for the two cytosolic forms. We know from PCR studies and sequence analyses that a 2.2-kb fragment is present in both forms of hsc70-3 but not in hsc70-1. Therefore, this fragment in the hsc70-1 blot is a probable cross-hybridization event. This fragment covers the last 300 residues of the coding region in hsc70-3. However, for the stromal member, the 2.2-kb fragment covers a coding region with an intron that begins at residue 86 and ends at residue 390, which is \( \sim 20 \) residues before the hsc70-3 fragment begins. Therefore, the 2.2-kb bands are unique, nonoverlapping fragments of the same size that are not cross-hybridizing.

**Expression Patterns**

The pattern of expression in different tissues at the RNA level for five of the stress 70 family members was examined. All five members were expressed to varying levels in leaf and cotyledon tissues, with hsc70-12, a cytosolic form, showing the lowest level of expression (Figure 4A). The expression of both the chloroplast stromal (hsc70-9) and the ER luminal (hsc70-2) members was lower in fully expanded cotyledons than it was in young true leaves, perhaps reflecting the declining importance of the photosynthetic and secretory functions of cotyledons as young seedlings shift from heterotrophic to autotrophic functioning. We did not anticipate that all five members would be highly expressed in roots, especially the stromal member (hsc70-9). In contrast, the ER luminal member was not detectably expressed in unimbibed seed. This finding is consistent with previous protein gel blot results showing that dry seed contained no detectable ER luminal stress 70 protein either (Anderson et al., 1994).

hsc70-12 was most highly expressed in dry seed, whereas the expression of the other cytosolic members was at their lowest. The patterns of the three cytosolic members in leaf and dry seed showed an obvious reversal of expression, indicating that they must function in a tissue-specific manner.

The RNA levels of all six members increased in response to a heat shock treatment that consisted of transferring plants from 20 to 37\(^\circ\) C and maintaining that temperature for 1 hr (Figure 4B). The majority of the six members were most responsive to heat shock in older, fully expanded leaf tissue and generally the least responsive in hypocotyl tissue. This pattern was most evident with hsc70-6 and almost the
opposite for the mitochondrial matrix member (hsc70-10). hsc70-10 showed stronger induction in the hypocotyl than in the other aerial tissues.

Distinct expression patterns for specific members were observed for three aerial vegetative tissues. Generally, young expanding leaves had higher steady state RNA levels than did fully expanded leaves. The lone exception was the stromal member, with highest RNA levels being in the older leaf tissue. Four of the six members had the highest levels of expression in hypocotyl tissue. Only with the matrix form was the expression level most balanced among all three tissues compared with the other stress 70 family members; however, when heat shocked, it showed the strongest induction in older leaf tissue.

The effect of different abiotic or stress treatments on expression patterns in leaf tissue for five members of the family was evaluated by RNA gel blot analysis (Figure 4C). Water stress to the point of wilting invoked a unique pattern of expression that was most dramatically represented by the ER luminal (hsc70-2) member and for one of the cytosolic forms (hsc70-12). In the absence of stress, the ER luminal member expression level was high, but it became virtually undetected in water-stressed tissue. Its response was mirrored by the opposite trend with hsc70-12, in which expression was strongly induced. Similar to water stress, heat shock (37°C for 2 hr) caused the mRNA level for the ER luminal form to become basically undetectable. One hour after systemic wounding, mRNA levels for the ER luminal form as well as for hsc70-12 and the stromal (hsc70-9) form were also almost undetectable. At 4 hr after wounding, the mRNA levels for the stromal member remained low, whereas those of all of the other members were again strongly expressed. Consistent with other experiments, low temperature stimulated the expression of most stress 70 members examined. The member most responsive to low temperature was hsc70-12, which was strongly induced, whereas the least responsive was the ER luminal member. Consistent with the step-down experiment, expression at 5°C appeared to increase for many of these stress 70 members.

The influence of reduced temperature on the expression of the five members in leaf tissue was plainly demonstrated (Figure 4D). Plants grown at 20°C were subjected to decreasing ambient temperatures to 5°C, followed by increas-

---

**Figure 3.** DNA Gel Blot Hybridization Analysis of Four Members of the Spinach Stress 70 Molecular Chaperone Family. Genomic DNA was digested with four different restriction enzymes as indicated, and 3 µg was separated in an agarose gel and transferred to a Hybond-N nylon membrane. Hybridization and washing were accomplished at high stringency. The membranes used for hsc70-1 and hsc70-3 were stripped and reprobed for hsc70-9 and hsc70-10, respectively. The lengths in kilobases of all detected EcoRI fragments are indicated at left. Symbols indicate whether a fragment corresponds to the mitochondrial (vertical ovals), plastid (horizontal ovals), or one of the two cytosolic (open hexagon and filled box) forms or to a yet unidentified member (filled arrowhead).
Evolution of the Stress 70 Gene Family

E. coli has several genes related to the stress 70 molecular chaperones. One, dnaK, has long been considered the archetype for all stress 70 molecular chaperones (Bardwell and Craig, 1984; Liberek et al., 1991). A second stress 70 member, hscA, was more recently characterized by accident as an open reading frame upstream of a ferredoxin (fdx) gene (Seaton and Vickery, 1994; Lelivelt and Kawula, 1995). We have used these two E. coli members in phylogenetic analyses of the spinach stress 70 family. Using the coding and protein sequences for six members representing four subcellular localization sites, we found that the neighboring trees are rooted with hscA (Hsc66) for both the DNA and protein data sets (Figures 5A and 5B). The first major branching splits the cytosolic and ER luminal members from the organelle members. dnaK appears to be a derivative of the ancestral form that later led to the organellar stress 70s.

5°C with the same photoperiod (12 hr) and irradiance (∼480 μmol m⁻² sec⁻¹) and held for 7 days. The heat shock treatment was the same as given in (B), except that the duration was for 2 hr. Wounding was accomplished by using a pair of forceps to systematically pinch the leaf lamina once every 0.25 cm² at 20°C. Samples were prepared at 1 and 4 hr after the completion of wounding. Film exposure was for 24 hr, except for hsc70-9, which was 48 hr.

(D) Plants grown at 20°C for 3 weeks as given in (C) were transferred to a growth cabinet at 15°C for 1 day. The temperature was lowered to 10°C for 1 day and then lowered again to 5°C and held for 7 days. After 5°C exposure for 7 days, the temperature was increased to 10°C for 1 day, to 15°C for 1 day, and then back to 20°C for another day. Film exposure was the same as given in (C).
The amino acid sequence of DnaK is more closely allied with the mitochondrial forms (Fig-}

Figure 5. Rooted Neighbor-Joining and Parsimony Analyses Using the CLUSTAL Program in DNASTAR and PAUP for the DNA and Protein Sequences of the Spinach Stress 70 Molecular Chaperone Family.

(A) and (B) The neighbor-joining method was used to analyze the truncated coding (DNA) and deduced amino acid (protein) sequences, as noted in Figure 1. Included in the analysis were the two members of the family from E. coli, DnaK and hscA (Hsc66). The scale at the bottom represents the branch distance as the number of changes in character states between neighbors. The number indicates the maximum distance.

(C) and (D) The neighbor-joining method was used to analyze the coding (DNA and deduced amino acids) sequences for the region encompassed by two highly conserved ATP interaction motifs. Eleven of the 12 spinach family members were used in these analyses along with the corresponding sequences for the two E. coli members. The scale at the bottom represents branch distance as the number of changes in character states between neighbors. The number indicates the maximum distance.

(E) and (F) The parsimony method was used to analyze the coding DNA and deduced amino acids) sequences encompassed by the two highly conserved ATP interaction motifs. Ten of the 12 available spinach family members were used in these analyses, along with the corresponding sequences for the two E. coli members. The values on the branches indicate the number of bootstrap replicas supporting the branch. Only bootstrap replication values >50 are shown.

Not unexpectedly, the two pairs of mitochondrial (hsc70-10 and hsc70-11) and cytosolic (hsc70-1 and hsc70-3) members are most similar to each other. The amino acid sequence of DnaK is more closely allied with the mitochondrial forms (Figure 5B), as has been recognized previously (Neumann et al., 1993), than to other forms.

When these analyses were repeated to encompass only the conserved ATP binding domain region for all of the avail-
able spinach sequences, a similar relationship was observed most notably with respect to E. coli dnaK (Figures 5C and 5D). Again, the trees are rooted by hscA (Hsc66). When comparing nucleotide sequences, dnaK appears to branch from an early derived form basal to all of the spinach stress 70 family members known to date. The chloroplast form occupies a branch all to itself, as do the mitochondrial and ER luminal forms. The major cytosolic forms are most closely allied with a group of sequences (hsc70-5, hsc70-6, hsc70-7, and hsc70-8) whose subcellular localization is not known. Based on sequence relationships, we considered that this other group may be a novel subclass of cytosolic members. In contrast to the DNA-based relationships, the picture obtained using the corresponding protein sequences follows that of the more limited comparison with the nearly full-length coding sequences. Again, a major branching divides the organelar forms from the cytosolic and ER luminal forms. DnaK appears to segregate with the organelle members, and this is consistent with the proposed prokaryotic origin of chloroplasts and mitochondria, even though both organelle proteins are now nuclear encoded. Another branching divides the ER luminal form from the remaining members, which includes the major cytosolic forms (HSC70-1 and HSC70-3). Should all of the other members that are closely allied with the major cytosolic forms function in the cytoplasm, then spinach will possess at least three distinct subclasses of stress 70 chaperones for this compartment.

Parsimony analysis produced similar tree topologies, but these were not identical to the neighbor-joining results. The branching patterns for the six nearly full-length sequences were almost identical to those shown by using the neighbor-joining method (data not shown). When the DNA sequences for a region of the ATP binding domain were analyzed, the tree that was rooted by hscA and dnaK formed a branch separate from the spinach sequences. However, the branching topology within the family shows the same overall pattern as that produced by the neighbor-joining method (Figure 5E). For the ATP binding region, parsimony analysis using the protein sequences yielded two equally possible trees. We selected the tree in which DnaK allied closely with the mitochondrial forms because of the correspondence to the neighbor-joining results (Figure 5F).

DISCUSSION

Stress 70 Multigene Family

This study represents one of the most comprehensive analyses of the stress 70 gene family in plants. Previous work with this family in plants has concentrated on only one form or members localized to a single subcellular compartment (Wu et al., 1988; Denecke et al., 1991; Bates et al., 1994; Joshi et al., 1996). In most cases, no more than a few members for any given plant species have been characterized (Wu et al., 1988; Bates et al., 1994; DeRocher and Vierling, 1995; Joshi et al., 1996). Even for Arabidopsis, no more than five distinct members have even been identified (C.H. Wu et al., 1988; S.H. Wu et al., 1994; Koizumi and Sano, 1997), and one of these (Storozhenko et al., 1996) belongs to the distantly related SSE group of yeast (Rassow et al., 1997).

Here, we describe the full-length coding sequences for and the genomic organization of six members of the stress 70 family of spinach, which has members localized in the cytosol, plastid, ER lumen, and mitochondrion. In addition, we present sequence data and genomic organizational information on a total of 12 members of the family. That the stress 70 family of spinach has at minimum 12 members is not surprising. Yeast has 14 members, and 10 or more members are known or indicated for other higher organisms (Günter and Walter, 1994; James et al., 1997; Rassow et al., 1997). Actually, we expected spinach to have more than the 12 identified in this study, because there are indications from genomic DNA gel blot hybridization data that uncharacterized members have not been isolated and therefore are not included in this work. This includes, in addition, any members of the SSE type that spinach might possess.

At the very least, plants should have more stress 70 members because of the presence of an additional organelle compartment that yeast and other eukaryotes do not have. However, this may not be true because there is recent evidence to suggest that a plastid stromal stress 70 member of watermelon can be dually targeted to the interior of the glyoxysome (Wimmer et al., 1997). If multiple targeting of stress 70s in plants were common, this might explain the high degree of sequence identity between one of the major cytosolic forms (HSC70-1) (Li et al., 1994) and the plastid outer envelope–localized Com70 (Ko et al., 1992; Kourtz and Ko, 1997). In addition, multiple targeting would reduce the number of different genes needed to meet the chaperoning needs of the plant cell. The complete sequencing of the Arabidopsis genome in the near future will help to determine how many stress 70 genes are present in plants and to what degree multiple targeting may occur.

Spinach has three cytosolic members (HSC70-1, HSC70-3, and HSC70-12) and two mitochondrial members (HSC70-10 and HSC70-11). Plants are known to possess at least three cytosolic members (Wu et al., 1988; Bates et al., 1994; DeRocher and Vierling, 1995; Joshi et al., 1996), and the three cytosolic forms of spinach are homologous to those known from other plants. This work, however, has identified two other groups consisting of five sequences that are closely related to the major cytosolic forms in sequence and intron organization. The subcellular localization of these members remains unknown, but their close similarity to the major cytosolic forms suggests that they might be additional members that localize to the cytosol. Alternatively, the members of these two groups may have specialized functions at the membrane boundary of the cytosol and other subcellular
compartments (Ko et al., 1992; Mooney and Harmey, 1996; Corpas and Trelease, 1997; Kourtz and Ko, 1997).

None of the spinach genes characterized here corresponds to the SSB subclass of yeast (Boorstein et al., 1994). In fact, a recent search of the GenBank database did not identify any plant stress 70 sequence that corresponded to the SSB subclass. If spinach does not contain the SSB homolog, then perhaps the function provided by the SSB proteins in yeast has been subsumed by one of the three cytosolic forms currently known in plants. In spinach, only one member localized to the ER lumen is known (Anderson et al., 1994). This is in contrast to other species in which the ER luminal form appears to be encoded by a small multigene family (Denecke et al., 1991; Kalinski et al., 1995; Koizumi and Sano, 1997; Wrobel et al., 1997).

Highly conserved signature motifs are present in all of the spinach stress 70 sequences. The sequence centered around residues GIDLGTT is a signature motif near the N terminus that is conserved across evolutionarily distant organisms. This sequence represents one of the ATP interaction sites of the ATP binding domain and has been recognized as a hallmark of this family (Bates et al., 1994). The targeting sequence cleavage site of the ER luminal form is between residues 28 and 29, which is only 12 residues before the GIDLGTT motif. This places the N terminus of this form about the same distance from the GIDLGTT sequence as in the cytosolic forms that lack cleavable targeting sequences. Based on a proposed cleavage site (Marshall and Keegstra, 1992) for the plastid form from pea, the analogous site in the spinach plastid form is between residues 72 and 73, which would leave six residues preceding the GIDLGTT sequence. For the mitochondrial form, using the proposed pattern RX-LFS (Gavel and von Heijne, 1990), the predicted cleavage would be between residues 46 and 47. This would leave 11 residues before the GIDLGTT sequence. Therefore, although the targeting sequences for the noncytosolic forms are all of different lengths, the expected cleavage sites would result in N termini that are of similar lengths before the ATP interaction site (Flaherty et al., 1990).

The C termini of the members are different depending on their subcellular localization, and the sequences are highly conserved and diagnostic for each subcellular site. The conservation of the C terminus of the cytosolic forms became apparent as numerous sequences for the cytosolic form from a variety of organisms were determined (Boorstein et al., 1994; Joshi et al., 1996), and it is considered to be important as an ATPase domain interaction or recognition site (Hightower et al., 1994). The ER luminal forms all end with HDEL or a conservative variation of that sequence, and it is well known as the retention signal for ER luminal resident proteins ( Munro and Pelham, 1987). Although only a few sequences for each of the plastid and mitochondrial stress 70s have been determined for higher plants, it is clear that the C termini of these forms are very conserved and diagnostic for plants (C.L. Guy, unpublished observations). The conserved sequences are PEAYEYEEVKK for the mitochondrial matrix form and PEGVIDADFTDSK for the plastid stromal form. The two spinach mitochondrial forms show a single conservative substitution of an alanine for valine in the C terminus, PEAYEYEEAKK. Neither the matrix form nor the stromal C termini show high identity with the absolute C termini of bacterial DnaKs, cyanobacterial DnaKs, or other eukaryotic forms, but there is significant homology of the matrix form to a penultimate sequence of the DnaK (Bardwell and Craig, 1984): PEAYEYEEAKK versus vAdYEiEiKd (where uppercase letters represent conserved residues and lowercase letters indicate conservative and nonconservative substitutions).

Stress 70 Gene Structure

All of the genomic clones of spinach stress 70 genes have at least one intron, and several have more than one intron. The two major cytosolic forms have a single large intron at the same position as that for a maize cytosolic heat-shock-inducible Hsp70, indicating that this structural arrangement is conserved across distantly related plants (Rochester et al., 1986). A closely related subclass that includes hsc70-5, hsc70-6, hsc70-7, and hsc70-8 also possesses an intron that is positioned six bases 3' of where the major cytosolic form's intron resides. In contrast, the positioning, size, and number of introns for the ER luminal, mitochondrial matrix, and the plastid stromal forms are distinctly different and characteristic for each form. The ER luminal form has seven introns dispersed throughout the coding region from the targeting peptide and in both ATP binding and peptide binding domains (Anderson et al., 1994). The matrix form has four: two in the targeting sequence region and one each in the ATP and peptide binding domains. The stromal form has seven introns, with one being in the ATP binding domain. This overall arrangement of introns in the various members is striking in that they do not correspond with each other in placement or size, and their placement does not seem to define domain boundaries. So, it appears that as the gene family evolved, the exon-intron structure became set for the particular subcellular localization site for the encoded protein. However, the exon-intron structure is not conserved for all plants, as indicated by the fact that a pea stromal form (Casey et al., 1994) has seven introns, as does spinach, whereas a Chlamydomonas homolog has only five introns (Drzymalla et al., 1996).

Stress 70 Gene Family Complexity

Genomic DNA gel blot analyses for four members of the family provide additional information on the complexity of the family. Previous studies with the ER luminal form (Anderson et al., 1994) did not uncover any evidence that it is encoded by a multigene family, as appears to be the case in a number of species (Denecke et al., 1991; Kalinski et al., 1995; Koizumi and Sano, 1997; Wrobel et al., 1997). Similar to the
chaperones are required at low temperatures in spinach and this suggests that the functions of the stress 70 molecular
and pressed at low temperatures; in the case of
nonstressful temperature as well as in response to heat
pressed to varying levels in leaf and hypocotyl tissue at a
individual members from different plants.

Expression Patterns

Comprehensive studies that include many family members
can better reveal the functional diversification that now ex-
ists within the stress 70 multigene family. Numerous studies
using a variety of stress 70 members localized to different
compartmental in different species have consistently dem-
onstrated differential expression patterns with respect to
specific tissues, developmental stage, and environmental pa-
rameters (Rochester et al., 1986; Duck et al., 1989; Denecke et
al., 1991; Roberts and Key, 1991; Anderson et al., 1994;
DeRocher and Vierling, 1995; Kalinski et al., 1995; Dhankher
et al., 1997). Almost always, such analyses have been con-
ducted with one member or at best three members. There-
fore, although there is much information about the individual
expression patterns, there is little information regarding the
response of multiple members of the family to any environ-
mental or developmental parameter. Although we have not
completed an analysis with all members of the family, we
show results for as many as six members localized to four
different subcellular compartments. What is evident from the
results is that the members of the family show specificity or
differential expression in response to developmental or envi-
ronmental parameters, as expected (DeRocher and Vierling,
1995). However, there appears to be substantial coordinate
regulation of many members that is dependent on the par-

ER luminal member of spinach, genomic gel blot analyses
for the matrix and the stromal forms suggest that these
forms are not encoded by large multigene families. In con-
trast, the complexity of the cytosolic forms is much greater
and suggests that the members that localize to the cytosol
are encoded by a multigene family. Many of the restriction
fragments can be accounted for, but there remain at least
five fragments (0.8, 0.9, 2.6, 7.0, and 10.5 kb) that must be-
long to hsc70-12 or the hsc70-4 to hsc70-8 group or some as
yet unidentified member. Yeast has six members belonging to
to two different subclasses that localize to the cytosol (Boorstein
et al., 1994; Rassow et al., 1997), if the SSE types are ex-
cluded. If the spinach subclasses of hsc70-4 and hsc70-5 to
hsc70-8 function in the cytosol, then spinach will have at
least eight members that function in this compartment.
However, we expect that there are additional members as
yet to be identified.

Evolution of Spinach Stress 70 Molecular
Chaperone Family

Several studies using stress 70 family sequences have been
conducted. Gupta and Golding (1993) used 57 full-length
sequences from a wide variety of organisms. Their analyses
indicated that the E. coli protein MreB is a truncated variant
of Hsp70 and is most likely derived from the same ancestral
gene that gave rise to Hsp70 before the addition of the
C-terminal peptide binding domain. They proposed a model
to explain the evolution of the Hsp70 gene family, conclud-
ing that it occurred in six steps, of which the first four took
place in the ancestor of all present-day life forms. Two sub-
sequent studies used 36 (Boorstein et al., 1994) and 33
(Rensing and Maier, 1994) sequences, respectively. Boorstein
et al. (1994) found that the Hsp70 family in eukaryotes con-

result is that the members of the family show specificity or
differential expression in response to developmental or envi-
ronmental parameters, as expected (DeRocher and Vierling,
1995). However, there appears to be substantial coordinate
regulation of many members that is dependent on the par-

Stress 70 Gene Family 551

in particular may require a degree of enhanced expression
for some members in a fashion analogous to that required
during heat shock. The finding of continued or increased ex-
pression of the stress 70 molecular chaperones at low tem-
peratures is consistent with the identification of a cold
shock–regulated homolog of DnaK in E. coli (Leivelt and
Kawula, 1995) and mutations in yeast stress 70 members
that result in cold-sensitive phenotypes (Craig and Jacobsen,
1985; Schilke et al., 1996).

The six members included in these analyses were ex-
pressed to varying levels in leaf and hypocotyl tissue at a
nonstressful temperature as well as in response to heat

heat

in particular may require a degree of enhanced expression
for some members in a fashion analogous to that required
during heat shock. The finding of continued or increased ex-
pression of the stress 70 molecular chaperones at low tem-
peratures is consistent with the identification of a cold
shock–regulated homolog of DnaK in E. coli (Leivelt and
Kawula, 1995) and mutations in yeast stress 70 members
that result in cold-sensitive phenotypes (Craig and Jacobsen,
1985; Schilke et al., 1996).

Expression Patterns

Comprehensive studies that include many family members
can better reveal the functional diversification that now ex-
ists within the stress 70 multigene family. Numerous studies
using a variety of stress 70 members localized to different
compartmental in different species have consistently dem-
onstrated differential expression patterns with respect to
specific tissues, developmental stage, and environmental pa-
rameters (Rochester et al., 1986; Duck et al., 1989; Denecke et
al., 1991; Roberts and Key, 1991; Anderson et al., 1994;
DeRocher and Vierling, 1995; Kalinski et al., 1995; Dhankher
et al., 1997). Almost always, such analyses have been con-
ducted with one member or at best three members. There-
fore, although there is much information about the individual
expression patterns, there is little information regarding the
response of multiple members of the family to any environ-
mental or developmental parameter. Although we have not
completed an analysis with all members of the family, we
show results for as many as six members localized to four
different subcellular compartments. What is evident from the
results is that the members of the family show specificity or
differential expression in response to developmental or envi-
ronmental parameters, as expected (DeRocher and Vierling,
1995). However, there appears to be substantial coordinate
regulation of many members that is dependent on the par-
partmental in different species have consistently dem-

result is that the members of the family show specificity or
differential expression in response to developmental or envi-
ronmental parameters, as expected (DeRocher and Vierling,
1995). However, there appears to be substantial coordinate
regulation of many members that is dependent on the par-

derture. The finding of continued or increased expression
of the stress 70 molecular chaperones at low temperatures
is consistent with the identification of a cold

shock–regulated homolog of DnaK in E. coli (Leivelt and
Kawula, 1995) and mutations in yeast stress 70 members
that result in cold-sensitive phenotypes (Craig and Jacobsen,
1985; Schilke et al., 1996).

Evolution of Spinach Stress 70 Molecular
Chaperone Family

Several studies using stress 70 family sequences have been
conducted. Gupta and Golding (1993) used 57 full-length
sequences from a wide variety of organisms. Their analyses
indicated that the E. coli protein MreB is a truncated variant
of Hsp70 and is most likely derived from the same ancestral
gene that gave rise to Hsp70 before the addition of the
C-terminal peptide binding domain. They proposed a model
to explain the evolution of the Hsp70 gene family, conclud-
ing that it occurred in six steps, of which the first four took
place in the ancestor of all present-day life forms. Two sub-
sequent studies used 36 (Boorstein et al., 1994) and 33
(Rensing and Maier, 1994) sequences, respectively. Boorstein
et al. (1994) found that the Hsp70 family in eukaryotes con-


eukaryotes consists of four monophyletic groups, each corresponding to a
particular subcellular localization. They concluded that the
major eukaryotic stress 70 groups arose before divergence of
eukaryotes 2 billion years ago. In contrast, Rensing and Maier
(1994) found that the tree divided into three major clusters:
chlorobionts, protozoans, and animal cognates. The mito-
chondrial Hsp70 was concluded to be a monophyletic group
of prokaryotic origin (Boorstein et al., 1994) in one study,
whereas in the other, the mitochondrial and chloroplast stress
70s split into two groups, with DnaK types in each.

Taking a narrower view by using only the spinach se-
quences and two members from E. coli, our results are con-
sistent with a model that the plastid and mitochondrial forms
arose from two independent endosymbiotic events in which
the mitochondrial form was contributed by a lineage that
had diverged from the line that gave rise to the plastid endo-
symbiont. Interestingly, the cyanobacterial DnaK proteins
cluster with the higher plant chloroplast stress 70 proteins,
which is consistent with the belief that they are the progeni-
tors of chloroplasts (Wang et al., 1993; Rensing and Maier,
1994). More recent studies in our laboratory with 37 plant
and cyanobacterial sequences support the linkage of the
plastid stromal form with cyanobacterial DnaKs, whereas the
mitochondrial matrix form is more closely allied with eu-
bacterial DnaK of E. coli. This result was further supported by an
analysis using >150 stress 70 sequences from bacteria, fungi,
animals, and plants (C.L. Guy, unpublished data). However,
with the spinach stress 70 family, intron–exon structure and
arrangement also support the evolution of the family because
the members localized to each of four subcellular compartments exhibit distinct organizational patterns. It may be that the diversity in the structural organization of the stress 70 genes in spinach has a functional significance in gene regulation processes or that it is simply the consequence of sequence divergence over evolutionary time.

We see a remarkable concordance of the phylogenetic relationships of the stress 70 family and that of the sHsp superfamily (Waters, 1995; Waters et al., 1996). The similarity seen in the general branching patterns for these two classes of heat shock protein gene families representing protein members localized to four different subcellular compartments could be merely coincidental but more likely reflects the evolution of the functional specialization of the different subcellular compartments. If the two classes of heat shock proteins are necessary for the proper operation of a subcellular compartment in plants (and most would agree that this is true), then it seems that the two classes of proteins could reflect the more general evolution of the plant cell structure and function. Already, heat shock proteins have been used by others to address questions of mitochondrial origins (Bui et al., 1996). Therefore, the evolution of different classes of heat shock proteins, and any set of proteins necessary for the proper functioning of multiple subcellular compartments in plants, could hold important clues to help unravel eukaryotic evolution leading to present-day plant cells.

Domain analysis has shown that the N terminus domain is more conserved than the C terminus domain (Boorstein et al., 1994), and the nonlinear rate of divergence between the ATPase and the peptide binding domains indicates a functional divergence between two or more groups of related genes (Hughes, 1993). In a study of the Hsp70 genes of Giardia, Gupta et al. (1994) found that the cytosolic and ER homologs form distinct groups, indicating that the two groupings arose as a result of a gene duplication very early in the evolution of eukaryotic cells and presumably accompanied the evolution of the ER or the nucleus. Giardia is a protozoan that lacks mitochondria. Based on other evidence, it is considered to be the earliest diverging lineage of eukaryotes; hence, the assertion that duplication and divergence of the Hsp70s was an early event in eukaryotic evolution. Our results also support this conclusion, but the ER luminal stress 70 form appears to have arisen well after the endosymbiosis events that contributed the plastid and the mitochondrial members (Figure 6).

That gene duplication was central to the multiplicity of the Hsp70 sequences in eukaryotes was further supported by the analyses of Ohta (1994). In that study, the sequences of three inducible, three cognate, and two testis-specific orthologous forms from mammals were analyzed along with other types of genes. The Hsp70 cognates showed very low nonsynonymous substitutions, whereas the inducible and testis-specific forms showed higher rates of nonsynonymous substitutions, indicating divergence and the acquisition of new function.

Sequence analyses and intron–exon organization provide evidence for at least three distinct subclasses that have evolved since the divergence of the ER luminal form from the cytosolic forms. One subclass contains three distinct cytosolic members, but the subcellular localization of the other two subclasses remains to be determined. Given the high degree of sequence relatedness with the cytosolic forms, we suspect that these two subclasses may also localize to the cytosol. These two subclasses contain members that are expressed at low levels under nonstressed conditions (Figure 4B, hsc70-6) but appear to be responsive to stressful conditions (Q.-B. Li and C.L. Guy, unpublished data). If they indeed do function in the cytosol, they may have completed some degree of divergence and may have specialized functions that are different from that of the major cytosolic forms. Several reports have identified stress 70 members tightly associated with the outer membranes of chloroplasts (Schnell et al., 1994; Kourtz and Ko, 1997) and mitochondria (Mooney and Harmey, 1996) and more recently with peroxisomes (Corpas and Trelease, 1997). To date, these membrane-bound stress 70 forms are thought to function in protein translocation into the organelles. Perhaps this function can be performed by the major cytosolic forms or may require a separate and distinct subclass of the stress 70 chaperones. Frequently, individual members of a gene family with a high degree of identity are found tightly linked (Koes et al., 1987). The chromosomal positions of the spinach stress 70 members are not known.

In a larger analysis with all available plant stress 70 sequences, the cytosolic grouping is complex and the branching pattern does not necessarily follow established taxonomic relationships. This appears to be caused by the fact that the cytosolic stress 70 proteins in plants are encoded by genes that probably belong to two or more subclasses. Quite likely, not all of the sequences represented in this larger analysis are orthologous. In fact, the cytosolic grouping appears to include several paralogous sequences.

Sequences from the same species are not always the most closely related to each other, as was observed with some of the ER forms. Examples of this were seen with sequences from pea and tomato. This suggests that the duplication events that gave rise to the various cytosolic forms occurred much earlier, and divergence to some degree has taken place. However, within the spinach family, we find discrete members that show almost no sequence variation even in noncoding and intron sequences. One example includes the two mitochondrial members hsc70-10 and hsc70-11. This lack of variation indicates to us that duplication events occurred recently. Another example would be the finding of the hsc70-3α and hsc70-3β sequences. These two members show eight differences over 5100 bp (99.8% identity), including coding and noncoding regions, but abruptly diverge and show no sequence identity for >800 bp. The divergence occurs at position 877, and we think this is the insertion site of a duplication event. If this is correct, then using the number of substitution events in the homologous region and an average substitution rate of 3.65 × 10^{-9} substitutions per site per year estimated from a number of angiosperm genes (Curtis and Clegg, 1984; Gaut et al.,...
Stress 70 Gene Family

In contrast, the GS3A and GS3B genes of pea are almost as conserved but probably are not a recent duplication, based on their presence in ancient relatives of Pisum spp (Walker et al., 1995).

METHODS

Plant Material

Spinach (Spinacia oleracea cv Bloomsdale) seedlings were grown from seeds under controlled environmental conditions previously described by Guy and Haskell (1987). Low-temperature treatment was conducted with a 12-hr photoperiod, as described previously (Guy and Haskell, 1987).

cDNA Library Construction and Screening

Preparation of total RNA and isolation of poly(A)+ RNA from 2-day cold-acclimated leaf tissue and the construction of a cDNA library were described previously (Neven et al., 1993). The cDNA library was screened by in situ plaque hybridization using polymerase chain reaction (PCR) products as hybridization probes. These probes were prepared from total cDNA by PCR using primers to highly conserved regions in the N-terminal ATPase domain (Flaherty et al., 1990). The 5' (CG24) and 3' (CG14 or CG144) oligonucleotides were GGTATTGA(CT)CTCGGTACCAC(CT)TAC and TCCACCACCAAGGTC-GAAGAC or AA(AG)AGTTCC(AC)CCTCCAAG(AG)TCAA, respectively. Positive plaques were excised in vivo.

Genomic Library Construction and Screening

DNA was extracted from 20 g of leaf tissue, as described by Ausubel et al. (1989). A genomic library was constructed by ligating partial EcoRI-digested genomic DNA (>2 kb) into the EMBL4 phagemid vector (Stratagene, La Jolla, CA). Approximately 500,000 plaques were screened using randomly primed α-32P-dCTP–labeled hsc70-1, hsc70-2, and hsc70-3 full-length cDNAs.

PCR Amplification of Genomic DNA and Library cDNA

Synthetic primers were constructed to conserved regions of hsp70 genes. Genomic DNA and library cDNA were amplified using Taq...
polymerase, according to the directions of the manufacturer (Perkin-
Elmer Cetus). The 5' and 3' primers were CG24 and CG14 or CG144,
respectively. PCR included an initial denaturation of 3 min at 94°C,
followed by 40 cycles of 1 min at 94°C, 1 min at 55 to 60°C, 2 to 4 min
at 72°C (according the PCR product length), and a final 10-min ex-
tension at 72°C. The fragments were purified and cloned into the
pGEM-T (Promega) vector (Li and Guy, 1996).

DNA Sequencing and Comparative Analysis

DNA sequencing was accomplished by using the Taq dideoxy termi-
nator cycle sequencing kit (Perkin-Elmer, Norwalk, CT) on an auto-
mated sequencer (Applied Biosystems, Foster City, CA). DNA
sequences were further analyzed using DNASTAR (DNASTAR Inc.,
Madison, WI). Homology searches were done with BLAST (Altschul
et al., 1990).

DNA and RNA Gel Blotting

Total DNA and RNA were extracted from leaf tissue, as described by
Sambrook et al. (1989). DNA gel blotting was performed using total
DNA digested with a 10-fold excess of restriction endonuclease. Di-
gested DNA was electrophoresed in 0.8% agarose gel in Tris–acetate–
EDTA buffer. The gels were then pressure blotted onto Hybond-N nylon
membranes (Amersham, Arlington Heights, IL) and UV cross-linked
(1.2 × 10^5 J/cm^2). RNA gel blotting was performed as described by
Sambrook et al. (1989). Total RNA was electrophoresed in 1.2%
formaldehyde agarose gels in Mops–sodium acetate–EDTA buffer.

Denatured RNA gels were then pressure blotted onto Hybond-N ny-
lon membranes and UV cross-linked. Prehybridization of DNA and
RNA blots was accomplished using 50% formamide, 5 × sodium
chloride–sodium phosphate–EDTA, 5 × Denhardt’s solution (1 ×
Denhardt’s solution is 5 g of Ficoll, 5 g of PVP, and 5 g of BSA in 500
mL of water), 0.2% SDS, and 10 μg/mL salmon sperm DNA at 42°C
for 4 hr. An α-32P-dCTP randomly labeled DNA probe (the entire
DNA as designated or cloned PCR DNA for hsc70-6; see Table 1)
was added to the prehybridization buffer and allowed to hybridize
with the filter for ∼16 hr at 42°C. After hybridization, the DNA gel
blots were each washed once in 2 ×, 1 ×, and 0.5 × sodium chlor-
ide–sodium citrate (SSC) plus 0.2% SDS for 15 min at 42°C. RNA gel
blots were each washed once in 6 ×, 2 ×, 1 ×, and 0.5 × SSC plus
0.2% SDS for 15 min at 42°C, followed by a 10-min wash in 0.1 ×
SSC plus 0.2% SDS at 60 to 65°C, according to the amount of radio-
isotope remaining on the blot.

DNA and Protein Sequence Analysis

Initial DNA or amino acid sequence alignments were performed using
CLUSTAL in the DNASTAR sequence analysis software. Alignments
were further refined manually. All sequences were truncated to in-
clude only that portion that could be aligned unambiguously. The
aligned sequences were analyzed by the neighbor-joining method
(Saitou and Nei, 1987) and parsimony using PAUP 3.1.1 (Swofford,
1993). The PAUP parameters were set for multistate, unordered,
and unweighted analyses, keeping all characters with branch and bound.
Bootstrapping was done with 100 replicates.

ACKNOWLEDGMENTS

We thank Kenton Ko and Tom Leustek for providing cDNA clones
HSE and Chsp70 for spinach stress 70 genes and Wesley Higgins for
assistance with phylogenetic analyses. We also thank Dale Haskell,
Chun Zhang, and Dong-Yul Sung for their constructive comments and
discussion during preparation of the manuscript. We especially
appreciate the efforts of three anonymous reviewers whose reviews helped us substantially improve this work. Facilities and assistance
for DNA sequencing and computer services were provided by the
core laboratories of the Interdisciplinary Center for Biotechnology
Research. This research was supported by Grant No. 93-17540
from the National Science Foundation. This is Florida Agricultural Ex-
periment Station Journal Series No. R-06139.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.
403–410.

Structural organization of the spinach endoplasmic reticulum-
luminal 70-kilodalton heat-shock cognate gene and expression of
70-kilodalton heat-shock gene during cold acclimation. Plant
Physiol. 104, 1359–1370.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman,
Molecular Biology. (New York: Wiley Interscience).

Drosophila and Escherichia coli heat-inducible dnaK gene are homologous. Proc. Natl. Acad. Sci. USA 81,
848–852.

cytosolic hsp70 gene family in Zea mays. Plant Mol. Biol. 25,
909–916.

cular evolution of the HSP70 multigene family. J. Mol. Evol. 38,
1–17.

Bui, E.T., Bradley, P.J., and Johnson, P.J. (1996). A common evo-
Acad. Sci. USA 93, 9651–9656.

Casey, R., Creissen, G., Domoney, C., Forster, C., and Matta, N.

Corpas, F.J., and Trelease, R.N. (1997). The plant 73-kDa peroxi-
somal membrane protein (PMP73) is immunorelated to molecular

of Saccharomyces cerevisiae hsp70 result in reduced growth

Drosophila heat-shock cognate genes during heat shock and


The Organization and Evolution of the Spinach Stress 70 Molecular Chaperone Gene Family
Charles L. Guy and Qin-Bao Li
Plant Cell 1998;10:539-556
DOI 10.1105/tpc.10.4.539

This information is current as of June 28, 2017

References
This article cites 70 articles, 35 of which can be accessed free at:
/content/10/4/539.full.html#ref-list-1

Permissions

eTOCs
Sign up for eTOCs at:
http://www.plantcell.org/cgi/alerts/ctmain

CiteTrack Alerts
Sign up for CiteTrack Alerts at:
http://www.plantcell.org/cgi/alerts/ctmain

Subscription Information
Subscription Information for The Plant Cell and Plant Physiology is available at:
http://www.aspb.org/publications/subscriptions.cfm