J Mol Evol (2001) 52:490–501
DOI: 10.1007/s002390010179

JOURNAL OF MOLECULAR EVOLUTION

Molecular Chaperones Encoded by a Reduced Nucleus: The Cryptomonad Nucleomorph

John M. Archibald,[1] Thomas Cavalier-Smith,[2] Uwe Maier,[3] Susan Douglas[4]

[1] Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4H7, Canada
[2] Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom
[3] Cell Biology and Applied Botany, Philipps-University Marburg, Karl-von-Frisch-Strasse, D-35032 Marburg, Germany
[4] National Research Council of Canada Institute for Marine Biosciences and Program in Evolutionary Biology, Canadian Institute for Advanced Research, 1411 Oxford Street, Halifax, NS, Canada B3H 3ZI

Abstract. Molecular chaperones mediate the correct folding of nascent or denatured proteins and are found in both the organelles and cytoplasm of eukaryotic cells. Cryptomonad algae are unusual in possessing an extra cytoplasmic compartment (the periplastid space), the result of having engulfed and retained a photosynthetic eukaryote. Within the periplastid space is a diminutive nucleus (the nucleomorph) that encodes mostly genes for its own expression as well as a few needed by the plastid. Two plastid-encoded chaperones (GroEL and DnaK) and a nucleomorph-encoded chaperone (Cpn60) have been reported from the cryptomonad, *Guillardia theta.* Here we analyse *G. theta* nucleomorph genes for members of the cytosolic HSP70 and HSP90 families of molecular chaperones, a heat shock transcription factor (HSF), and all eight subunits of the group II chaperonin, CCT. These are presumably all active in the periplastid space, assisting in the maturation of polypeptides required by the cell; we propose a central role for them also in the structure and assembly of a putative relict mitotic apparatus. Curiously, none of the genes for co-chaperones of HSP70, HSP90, or CCT have been detected in the nucleomorph genome; they are either not needed or are encoded in the host nuclear genome and targeted back into the periplastid space. Endoplasmic reticulum (ER) homologs of HSP70 and HSP90 are also not present. Striking differences in the degree of conservation of the various nucleomorph-encoded molecular chaperones were observed. While the *G. theta* HSP70 and HSP90 homologs are well conserved, each of the eight CCT subunits (α, β, γ, δ, ε, η, θ, and ζ) is remarkably divergent. Such differences are likely evidence for reduced/different functional constraints on the various molecular chaperones functioning in the periplastid space.

Key words: Molecular chaperone — Chaperonin — *Guillardia theta* — Nucleomorph

Introduction

Chaperone-assisted protein folding is a universal cellular process. In eukaryotes, molecular chaperones function in most cellular compartments, and in addition to mediating the proper folding of nascent proteins, are involved in the disassembly of oligomeric protein complexes, directing the conformational maturation of signal-transducing molecules, and facilitating the degradation of unstable proteins. Molecular chaperones also play an important role in guiding the translocation of proteins across organellar membranes. HSP70/DnaK, HSP90, and the chaperonins are three of the most ubiquitous and well-studied classes of ATP-dependent molecular chaperones, and together share the common feature of recognizing their target substrates via interactions with hydrophobic regions exposed on non-native proteins (see Bukau and Horwich 1998; Johnson and Craig 1997 for review).

The correct folding and transport of nascent proteins is especially complicated in organisms such as crypto-

Correspondence to: S. Douglas; *email:* susan.douglas@nrc.ca

monads, photosynthetic algae that acquired their plastids secondarily from eukaryotic endosymbionts, because nuclear-encoded gene products destined for the plastid must traverse an extra pair of membranes (Cavalier-Smith 2000). An early model for protein trafficking in these cells proposed that nuclear-encoded polypeptides translated on host ribosomes present on the perichloroplast endoplasmic reticulum (CER) were transported to the plastid by vesicles budding from the periplastid membrane (derived from the plasma membrane of the former symbiont) and fusing with the outer plastid envelope membrane (Gibbs 1979; Gibbs 1981). However, other roles for such vesicles have been suggested (Cavalier-Smith 1999), and it is more likely that nuclear-encoded polypeptides are translocated across the CER and the periplastid membrane (by a still unknown mechanism) directly into the periplastid space (Cavalier-Smith 1999; Ishida et al. 2001; Deane et al. 2000). Molecular chaperones that function in the periplastid space are therefore probably involved in (1) mediating the initial translocation process itself, (2) facilitating the proper folding of newly-imported and newly-synthesized proteins destined to function in the periplastid space, and (3) assisting the translocation of nuclear- and nucleomorph-encoded polypeptides through both membranes of the plastid envelope. They could also facilitate the disaggregation of proteins and couple proteins to nascent polypeptides being translated on periplastid ribosomes.

The gene for one such chaperone, a cytosolic HSP70, has been found in the nucleomorph genome of the cryptomonad alga *Rhodomonas salina* (*Pyrenomonas salina*) (Hofmann et al. 1994; Rensing and Maier 1994). The predicted polypeptide sequence of the HSP70 from *R. salina* contains a putative nuclear localization signal (Rensing and Maier 1994), suggesting that it can be targeted to the nucleomorph itself. Hsp70 genes have also been isolated and sequenced from the plastid (Wang and Liu 1991) and the nucleus (Rensing et al. 1997), but not the nucleomorph, of another cryptomonad, *Guillardia theta.* Genes encoding the organellar chaperonin Cpn60 have been isolated and sequenced from both the plastid (Douglas and Penny 1999) and nucleomorph (Wastl et al. 1999) of *G. theta*, and presumably mediate the folding of plastid-targeted proteins.

The periplastid space—essentially a highly reduced eukaryotic cytoplasm—contains very few of the normal cell constituents. Apart from the plastid and the nucleomorph itself, only starch grains, 80S ribosomes, and smooth periplastid vesicles are ultrastructurally visible. It lacks mitochondria, peroxisomes, and lysosomes or large vacuoles, as well as any visible rough endoplasmic reticulum, Golgi complex, or cytoskeleton. The nucleomorph was generally thought to divide amitotically (McKerracher and Gibbs 1982; Morrall and Greenwood 1982), as electron microscopic studies failed to detect the presence of cytoskeletal elements normally associated with a mitotic apparatus (Gillott and Gibbs 1980; McKerracher and Gibbs 1982; Morrall and Greenwood 1982). However, the recent identification of alpha-, beta-, and gamma-tubulin genes (the major protein components of microtubules) in the *G. theta* nucleomorph genome (Keeling et al. 1999; Zauner et al. 2000) strongly suggests that some microtubule-related structures are present. Other than HSP70 (Hofmann et al. 1994; Rensing and Maier 1994), evidence for molecular chaperones that might function in the periplastid space of cryptomonads is currently lacking. Here we present an analysis of *G. theta* nucleomorph-encoded genes for the cytosolic HSP70 and HSP90 classes of chaperone, eight different subunits of the cytosolic chaperonin, CCT, as well as a heat-shock transcription factor (HSF) known to be involved in the expression of heat shock proteins.

In vivo, the predominant substrates of CCT (and its co-chaperonin prefoldin) are tubulins and actins (Kubota et al. 1994), although approximately 70 different newly-translated polypeptides have also been shown to co-precipitate with the CCT complex (Thulasiraman et al. 1999). Mediating the assembly of the nucleomorph-encoded tubulins may be an important function for CCT in *G. theta.* Missing from the *G. theta* nucleomorph genome are genes for the ER homologs of HSP70 and HSP90 as well as co-chaperones HSP40/dnaJ, hip, hop, and prefoldin, which assist HSP70, HSP90, and CCT, respectively. This suggests that these proteins are no longer needed or are encoded in the host nuclear genome and targeted into the periplastid space.

HSP90 is the most abundant cytosolic protein in eukaryotic cells (Jakob and Buchner 1994) and, together with other proteins, is responsible for the conformational maturation of signal-transducing proteins such as steroid hormone receptors and protein kinases (Bohen et al. 1996). Interestingly, HSP90 is also known to associate with the centrosome and participate in cell cycle control in animals, probably by facilitating interactions with centrosomal proteins (Lange et al. 2000) such as γ-tubulin, with which it co-localizes (Zarzov et al. 1997). HSP90 has been shown to play a role in spindle pole body duplication in *Saccharomyces cerevisiae* (Lange et al. 2000) and to associate with HSP70 to form a multichaperone (Scheufler et al. 2000). The presence and properties of the nucleomorph HSP70 and 90 homologs described here, together with the discovery of gamma tubulin and the centrosomal protein ranbpm (Zauner et al. 2000), raise the possibility that all four proteins interact to form a relict mitotic apparatus of red algal origin in the cryptomonad nucleomorph, even though cytological evidence for nucleomorph mitosis has not yet been found.

Materials and Methods

Cloning and sequencing. DNA of the nucleomorph chromosomes was cloned into pBluescript plasmid vector (Stratagene) and sequenced as

described (Zauner et al. 2000). Genes were identified by BlastP, BlastX, and BlastN computer searches of the public databases (http://www.ncbi.nlm.nih.gov/BLAST). The web tool pSort (http://psort.nibb.ac.jp/) was used to identify targetting signals.

Phylogenetic Analyses

Inferred protein sequences of the *G. theta* α, β, γ, δ, ε, η, θ, and ζ CCT subunit genes were added manually to an alignment of CCT sequences constructed previously (Archibald et al. 2000). The alignment contained 103 sequences (10 representative archaebacteria and 7–19 taxa from each of the CCT subunit families), and represented the full diversity of CCT sequences presently known. 251 unambiguously aligned amino acid sites were used for phylogenetic analysis, taking into account partial sequence data. From this master alignment, smaller alignments containing individual CCT subunits (paralogs) were constructed. Individual subunit alignments used for phylogenetic analyses contained considerably more alignable sites—between 489 and 512 amino acids positions.

Alignments of selected HSP70 and HSP90 protein sequences were kindly provided by A. Roger (Dalhousie University). To obtain additional sequences, particularly cytosolic HSP70s and HSP90s, we searched the public databases by BLAST (Altschul et al. 1990) using the *G. theta* nucleomorph sequences as queries. New sequences were added to the alignments manually based on globally conserved regions.

For HSP90, 12 independent EST sequences were obtained from the red alga *Porphyra yezoensis* (Nikaido et al. 2000). The sequences were assembled using Sequencher (Gene Codes Corp., Ann Arbor, MI), and two portions of nearly overlapping open reading frame were identified. A single EST covered a 500 nt portion of the 5′ end of the gene (GenBank accession number AV432848), and the remaining 11 ESTs (Accession numbers AV431882, AV432126, AV433084, AV436237, AV429660, AV429526, AV429525, AV429680, AV436049, AV429756, and AV429533) were assembled to form a contig of approximately 1500 nucleotides in length (between one and five-fold coverage), encoding 504 amino acids of the carboxy-terminal half of the molecule. No frameshifts were detected in the coding region, and only those portions of the sequence unambiguously assigned as HSP90 coding sequence were used in phylogenetic analyses.

For HSP70, an alignment of 51 eukaryotic HSP70 sequences and 526 unambiguously aligned amino acid positions was used for phylogenetic reconstruction. No more than one taxon was missing data at any one site. For HSP90, a dataset of 58 sequences, including bacterial HSP90s, was used in preliminary analyses, and a final set of 37 eukaryotic (cytosolic and ER) sequences was used for most analyses. The alignment contained 543 amino acid positions.

Using the heat shock transcription factor (HSF) encoded on chromosome 1 of the *G. theta* nucleomorph genome as a query, we searched the public databases by BLAST (Altschul et al. 1990) for additional HSFs from diverse eukaryotes. A large set of animal, fungal and plant HSF-like sequences were obtained, many of which were duplicates or isoforms encoded in the same genome. HSF sequences were aligned using CLUSTAL W (Higgins and Sharp 1988) and the alignment was adjusted manually. 89 amino acid residues of the conserved DNA-binding domain were used for phylogenetic reconstruction. All alignments are available upon request (johna@hades.biochem.dal.ca or susan.douglas@nrc.ca).

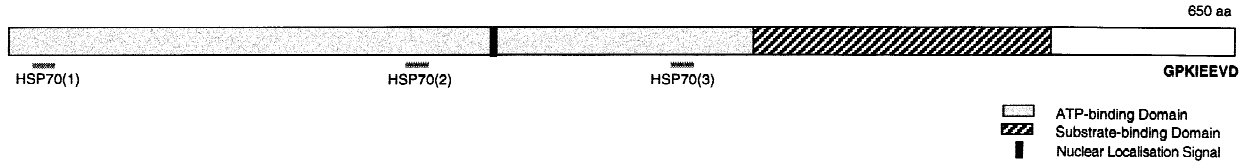
Phylogenetic trees were inferred using maximum likelihood (ML), distance-based, and maximum parsimony (MP) methods of tree reconstruction with the following programs: (1) maximum likelihood, protML using the JTT-F model in MOLPHY (-jf option; Adachi and Hasegawa 1996), quartet puzzling in PUZZLE 4.02 accounting for site-rate heterogeneity with an eight rate category discrete approximation to the Γ distribution plus an additional invariable rate category (Strimmer and von Haeseler 1997), (2) distance, PROTDIST (PAM matrices), NEIGHBOR, and FITCH in PHYLIP 3.57 (Felsenstein 1993), (3) maximum parsimony, PAUP* 4.0 (Swofford 1998). PUZZLE 4.02 was also used to perform Chi-square tests to compare the amino acid composition of the *G. theta* nucleomorph CCT, HSP70, HSP90, and HSF sequences to those of other eukaryotes. Quartet puzzling support values (from PUZZLE 4.02) or RELL values (resampling estimated log-likelihoods, obtained by quick-add searches of 1000 or 2000 trees in protML (options -q -n 1000, 2000)) were used as measures of support for ML trees. Statistical support for MP and distance trees was obtained by bootstrapping with 100 resampling replicates. Support values for ML-distance trees were obtained by bootstrapping (500 replicates) with PUZZLEBOOT (A. Roger and M. Holder, http://members.tripod.de/korbi/puzzle/).

Results

HSP70 family. A single gene encoding HSP70 was detected on the largest chromosome (1) of the nucleomorph genome. The predicted polypeptide is 650 amino acids long (72,485 kDa) and contains the highly conserved motif, GPKIEEVD, at the carboxy terminus (Fig. 1a). It also possesses all three of the conserved consensus patterns for the hsp70 gene family at positions 12–19, 204–217, and 341–355 (Fig. 1A, B). Computer analysis using the search tool pSort showed that HSP70 possesses a bipartite nuclear localisation signal described by Robbins et al. (1991). A perfect heat shock element (TTCnnGAAnnTTC) necessary for the binding of heat shock transcription factor is located at positions −41 to −29 relative to the initiator methionine (Fig. 1B).

A phylogenetic analysis of HSP70 is presented in Fig. 2. Most notably, the *G. theta* sequence clusters with cytosolic HSP70s to the exclusion of the ER homologs, and branches strongly with the nucleomorph sequence from another cryptomonad, *Rhodomonas salina.* Together, the nucleomorph sequences branch at the base of a clade containing the glaucophyte *Cyanophora paradoxa,* the green alga *Chlamydomonas reinhardtii,* and a variety of land plants (moderately supported with protML and distance methods, but not resolved with quartet puzzling). Given that the cryptomonad endosymbiont is thought to be of red algal origin (e.g. Douglas and Turner 1991, Cavalier-Smith et al. 1996), this result is particularly interesting in light of the debate over the origin(s) of primary plastids (see Palmer 2000 for review). While some data support the notion of a common origin of red algae and green plants (e.g. Burger et al. 1999), an analysis of the nuclear RNA polymerase II (*RPB*1) did not (Stiller and Hall 1997). More recent analyses (Moreira et al. 2000), including a concatenation of 13 nuclear genes, showed strong evidence for the monophyly of these two groups. Overall, the topology of the HSP70 tree presented here is very similar to that obtained in a recent comprehensive analysis by Germot and Philippe (Germot and Philippe 1999). Interestingly, the *G. theta* nuclear HSP70 branched near the base of the cytosolic HSP70 tree, and not near the algal/plant grouping, as was observed in previous analyses (Rensing et al. 1997). The two cryptomonad nucleomorph HSP70s are remarkably

a.



b.

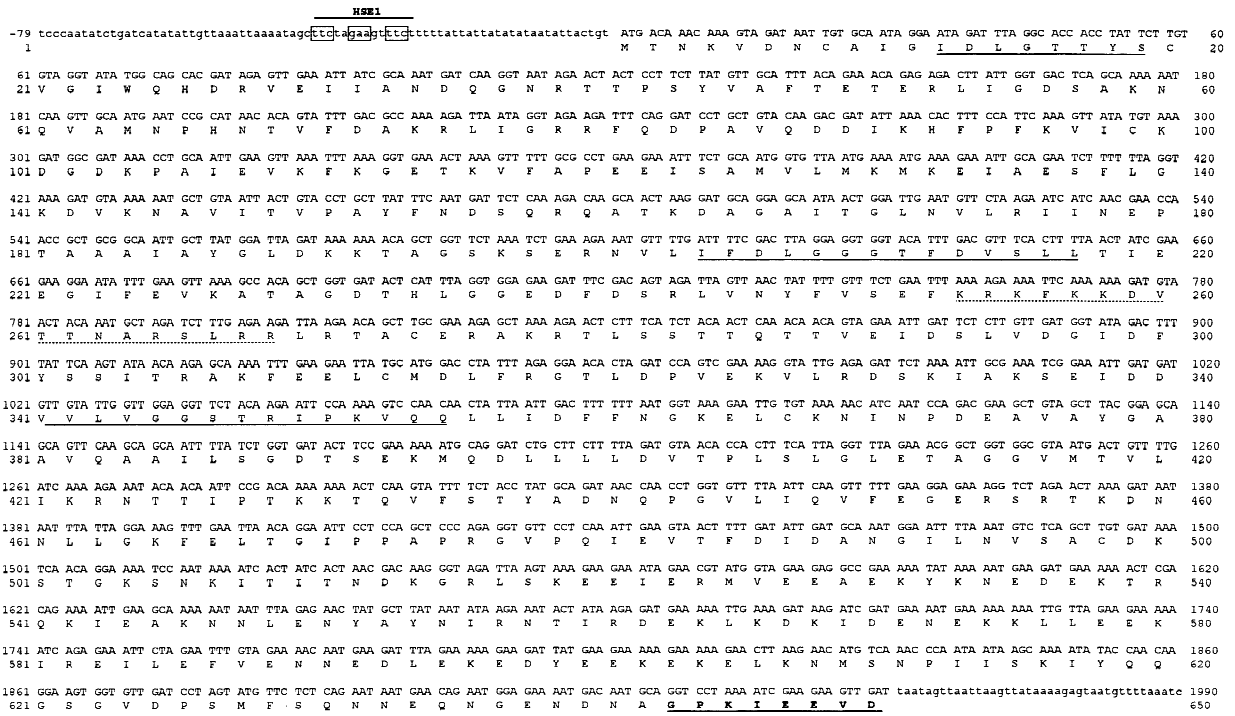


Fig. 1. The *hsp70* gene from the *Guillardia theta* nucleomorph. **A.** Schematic of HSP70 showing domains. **B.** Nucleotide sequence of *hsp70.* The deduced amino acid sequence is shown beneath the nucleotide sequence in upper case letters. The conserved consensus patterns for the *hsp70* family are underlined and the highly conserved motif at the carboxy terminus is shown in boldface and underlined. The bipartite nuclear localisation signal is indicated by dotted underlining and the heat shock element upstream of the coding sequence is boxed.

well conserved, and possess relatively short branches in our phylogenetic trees compared to many of the other protist HSP70s (e.g. the alveolates, *Trichomonas vaginalis* and *Giardia lamblia*). The *G. theta* nuclear HSP70 sequence is a slightly longer branch, which may explain its basal position in the cytosolic HSP70 tree.

HSP90 family. A single gene encoding HSP90 (HSP82) was reported on chromosome 1 of the nucleomorph genome (Zauner et al. 2000). The predicted polypeptide is 684 amino acids long (79,245 kDa) and contains the highly conserved motif, MEAVD, at the carboxy terminus (Fig. 3). It also possesses the consensus pattern for the HSP90 family (YSNKEIFLRE) at positions 23–32 and a nuclear localization signal (KKKKK) at positions 232–236. Two near-perfect heat shock elements (TTGnnGAAnnTTC) are located at positions −66 to −49, and −46 to −34 relative to the initiator methionine.

As was the case for HSP70, the *G. theta* nucleomorph HSP90 is very well conserved and is clearly cytosolic, clustering strongly with the other cytosolic homologs with all phylogenetic methods (Fig. 4). As expected, the *G. theta* sequence forms a highly supported group with the red algal HSP90 from *Porphyra yezoensis* (assembled from multiple EST sequences; see materials and methods), consistent with previous accounts of a red algal ancestry for the cryptomonad nucleomorph genome (Douglas et al. 1991; Cavalier-Smith et al. 1996). Together, these sequences form a weakly supported clade with HSP90s from a variety of land plants. Again, this result is consistent with recent data suggesting the common ancestry of red algae and green plants (Moreira et al. 2000). Most surprising was the observation that the sisterhood of animals and fungi, supported by protein and small subunit ribosomal RNA (SSU rRNA) phylogenies, as well as a 12 amino acid insertion in EF1-α (see Baldauf 1999 for recent review), was not recovered in our analyses. Largely due to limited taxon sampling, few phylogenetic studies have been performed on HSP90. Gupta (1995), using parsimony and distance-based methods of tree reconstruction, first noted that this molecule

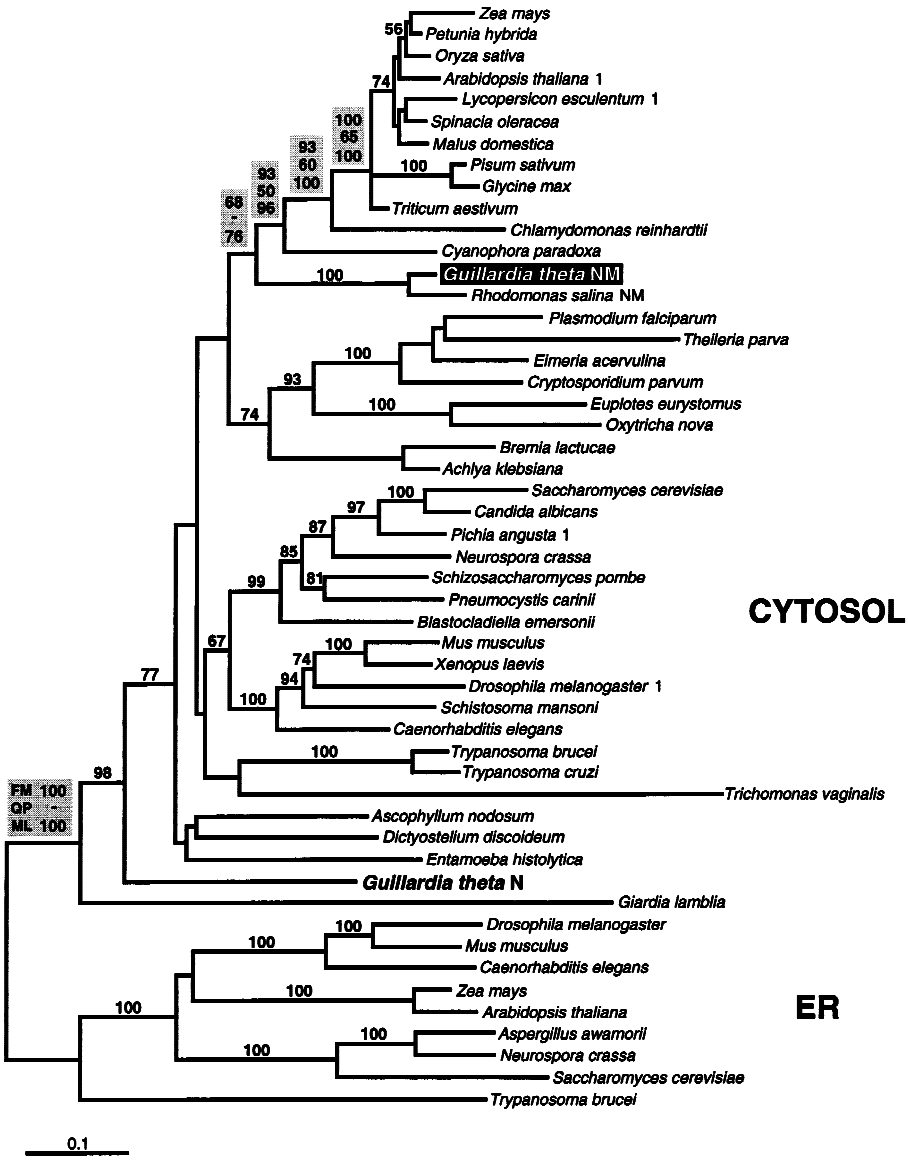


Fig. 2. Phylogenetic analysis of HSP70 from the *Guillardia theta* nucleomorph. The maximum likelihood (ML) tree (lnL = −15731.55) from a heuristic search of 2000 trees in protML (Adachi and Hasegawa 1996) is shown; ML branch-lengths were inferred using a rate heterogeneity model (JTT-F + Γ; see Materials and Methods) in PUZZLE (Strimmer and von Haeseler 1997). 526 unambiguously aligned amino acid positions were used. The *G. theta* nucleomorph and nuclear HSP70 sequences are highlighted. Support values above the branches are ML RELL values and are indicated if >50%. Gray boxes contain support values for nodes of particular interest (FM, distance (Fitch-Margoliash) bootstrap values; QP, quartet puzzling support values, ML, ML RELL values). The scale bar indicates the inferred number of amino acid substitutions per site.

failed to show a relationship between animals and fungi, and instead showed a weak relationship between animals and plants. Our analyses also failed to resolve an animal-fungal connection, despite the use of maximum likelihood methods (including taking into account site-to-site rate variation), a more conservative alignment (543 unambiguously aligned amino acid positions), and increased taxon sampling. We performed phylogenetic analyses with and without a bacterial outgroup, and with the selective removal of long-branch taxa (e.g. various long-branch fungi), but the basal position of the fungal sequences was a consistent pattern in all analyses (data not shown). In our analyses, the animal HSP90s and the HSP90 from *Dictyostelium discoideum* also branch near the base of the cytosolic tree. It is therefore likely that the apparent paraphyly of the animals and fungi is simply due to the misplacement of the root.

Heat shock transcription factor. A single gene encoding a heat shock transcription factor was detected on chromosome 1 of the nucleomorph genome. To our knowledge, the *G. theta* nucleomorph HSF presented here is the first protist sequence to be described. Most noticably, the predicted polypeptide is a mere 185 amino acids long (22,260 Da), much shorter than typical eukaryotic HSFs which are approximately 350 amino acids in length (Fig. 5A). Whether this drastically reduced size is a unique feature of the *G. theta* HSF, or is characteristic of protist HSFs as a whole, is currently unknown. No transactivation domain, typically found in the carboxy-terminal half of eukaryotic HSFs, is detectable in the HSF open reading frame or anywhere else on the completely sequenced nucleomorph genome (Douglas et al. in press). Two of the three consensus patterns for HSF-type DNA-binding domain proteins (HSF1 and HSF2) are well-conserved (Fig. 5C); The HSF3 consensus pattern within the oligomerization domain (Fig. 5B) shows only limited sequence identity although the leu-

b.

```
                HSE1                HSE2
-78 gatatgccacctgatgttctagaagttctattggagaaatttctattgatttaataaaattaattaagaatttatt ATG ATT GAA ACT TAT CAA TTC CAA GCT GAA ATT AAT CAG TTA ATG AGT TTA ATC ATT AAC  60
  1                                                                               M   I   E   T   Y   Q   F   Q   A   E   I   N   Q   L   M   S   L   I   I   N   20

 61 ACT TTT TAT TCT AAC AAA GAA ATT TTT CTT AGA GAG CTT ATT TCC AAT GCT TCT GAT GCA TTA GAT AAA ATT CGT TAT CAA TCG TTA ACT GAC TCT TCT GTT CTA GAT AAT GAA CGA AAA 180
 21 T   F   Y   S   N   K   E   I   F   L   R   E   L   I   S   N   A   S   D   A   L   D   K   I   R   Y   Q   S   L   T   D   S   S   V   L   D   N   E   P   K   60

181 CTT GAA ATT AGA ATA TTA ACT GAT AAA AAT AAC AAA TCT CTG ACT CTT ATC GAT ACT GGT ATA GGT ATG ACA AAA GAC GAC TTA ATC CAG AAT TTG GGT ACT ATC GCA AAA TCA GGC AGA 300
 61 L   E   I   R   I   L   T   D   K   N   N   K   S   L   T   L   I   D   T   G   I   G   M   T   K   D   D   L   I   Q   N   L   G   T   I   A   K   S   G   T   100

301 AAA TCT TTT ATG GAA GCA TTA CAA GCT GGC GCA GAC GTA TCT ATG ATT GGT CAA TTT GGT GTT GGA TTC TAT TCT GCA TAT CTA GTA GCA GAT CGA GTT GTT GTA GAA ACT AAA AAC AAC 420
101 K   S   F   M   E   A   L   Q   A   G   A   D   V   S   M   I   G   Q   F   G   V   G   F   Y   S   A   Y   L   V   A   D   R   V   V   V   E   T   K   N   N   140

421 AAT GAT TCT CAG TAC ATT TGG GAA AGT TCT GCC GGT GGT AGC TTT ACA ATT AAT GAT TCT TCT ATT ACA GAT TTA GCA AGA GGT ACA AAA ATT ACA TTG TTT CTT AAA GAT GAT CAA TTG 540
141 N   D   S   Q   Y   I   W   E   S   S   A   G   G   S   F   T   I   N   D   S   S   I   T   D   L   A   R   G   T   K   I   T   L   F   L   K   D   D   Q   L   180

541 GAA TAT TTG GAA GAA AGA AGA CTA AAA GAT CTA GTA AAA AAA CAT TCT GAA TTT ATA CAA TAT CCA ATA AAT TTA TGG GTA GAA AAA GAA ATT GAA AAA GAG GTT GAT GCT TCT GAT AAT 660
181 E   Y   L   E   E   R   R   L   K   D   L   V   K   K   H   S   E   F   I   Q   Y   P   I   N   L   W   V   E   K   E   I   E   K   E   V   D   A   S   D   N   220

661 GAA GCT ACT AAT GAT GAT AAC AAC AAG GAA GAA AAA AAA AAA AAG AAA ATT AAG GAA ATT TCA CAC GAA TGG CAA TTT CTT AAT AAG AAC AAA CCT ATT TGG ACA AGA AAG CCA GAA GAA 780
221 E   A   T   N   D   D   N   N   K   E   E   K   K   K   K   K   I   K   E   I   S   H   E   W   Q   F   L   N   K   N   K   P   I   W   T   R   K   P   E   E   260

781 ATT TCA AAA GAA GAA TAT TCC TCT TTC TAT AAA ACT ATA TCA AAT GAC TGG GAG GAT CAC CTT GCA GTC AAG CAC TTT TCA GTA GAA GGT CAA TTG GAA TTT AAA GCA TTA TTA TTT GTT 900
261 I   S   K   E   E   Y   S   S   F   Y   K   T   I   S   N   D   W   E   D   H   L   A   V   K   H   F   S   V   E   G   Q   L   E   F   K   A   L   L   F   V   300

901 CCT AAG AGA GCC CCA TTT GAT CTA TTC GAG CCT AGA AAG AAA AAC AAT AAT ATC AAA TTG TAT GTA AAA AGA GTA TTT ATT ATG GAT AAT TGT GAT GAA ATT ATA CCT GAG TAC TTA AAC 1020
301 P   K   R   A   P   F   D   L   F   E   P   R   K   K   N   N   N   I   K   L   Y   V   K   R   V   F   I   M   D   N   C   D   E   I   I   P   E   Y   L   N   340

1021 TTC ATT AAA GGA GTG GTT GAT TCT GAA GAT CTC CCT TTA AAT ATC TCT AGG GAA ACA CTT CAA CAA AAT AAA GTC GTT AAA GTA ATA AGA AAA AAT ATA GTT AAA AAA TGC TTA GAA TTA 1140
341 F   I   K   G   V   V   D   S   E   D   L   P   L   N   I   S   R   E   T   L   Q   Q   N   K   V   V   K   V   I   R   K   N   I   V   K   K   C   L   E   L   380

1141 TTC TTA GAG ATT TCC GAA AAC AAA GAT GAT GTA AAG ATC TTT TAT GAC CAG TAT TCT AAA AAT ATA AAA TTA GGC ATT CAT GAA GAT TCA CAA AAC AGA TCG AAA CTT GCC GAT TTA TTA 1260
381 F   L   E   I   S   E   N   K   D   D   V   K   I   F   Y   D   Q   Y   S   K   N   I   K   L   G   I   H   E   D   S   Q   N   R   S   K   L   A   D   L   L   420

1261 AGG TAT AAA TCT ACC AAA TCA CCT GAT GAA ACT ACT ACG TTA AAA GAA TAT GTG TCA AGG ATG AAA GAA AAT CAA AAT CAA ATT TAT TAC ATA ACG GGT GAA AGT CAA AAA TCA GTA GAA 1380
421 R   Y   K   S   T   K   S   P   D   E   T   T   T   L   K   E   Y   V   S   R   M   K   E   N   Q   N   Q   I   Y   Y   I   T   G   E   S   Q   K   S   V   E   460

1381 AAT TCT CCT TTT CTT GAA AAA CTC AAA CAA AAA GGA TTT GAG GTT TTG TTC ATG ATC GAA CCC ATC GAC GAA TAT TGT GTT CAG CAA TTA AAA GAA TAT GAA GGT AAA AAA TTA GTA TGC 1500
461 N   S   P   F   L   E   K   L   K   Q   K   G   F   E   V   L   F   M   I   E   P   I   D   E   Y   C   V   Q   Q   L   K   E   Y   E   G   K   K   L   V   C   500

1501 GCA ACT AAA GAA GGT TTA GAT TTG GGT GAT TCA GAA AAC GAT AAA AAA GTT AAA GAA AAT GAA AAA GAA CAA TTT GAT GAA TTA TGT AAA GTC ATT AAA GAA ACT TTA AAC GAT AAA GTT 1620
501 A   T   K   E   G   L   D   L   G   D   S   E   N   D   K   K   V   K   E   N   E   K   E   Q   F   D   E   L   C   K   V   I   K   E   T   L   N   D   K   V   540

1621 GAA AAG GTT GTC ATA TCT GAT AGA CTA TCC GAT TCT CCT TGT ATA TTA GTA ACT GGT GAA TAT GGT TGG TCA GCA AAC ATG GAA CGT ATT ATG AAA GCA CAA GCC TTA AGA GAT AGT TCA 1740
541 E   K   V   V   I   S   D   R   L   S   D   S   P   C   I   L   V   T   G   E   Y   G   W   S   A   N   M   E   R   I   M   K   A   Q   A   L   R   D   S   S   580

1741 CTT TCA ACC TAT ATG TCA AGC AGA AAA ACT ATG GAA CTC AAT CCT AAG AAT TCT ATT ATT AAT GAA CTT AAA GAA CGA GTT AAT AAT GAT AGG AAC GAT AAA ACT GTG AAA GAT CTA GTA 1860
581 L   S   T   Y   M   S   S   R   K   T   M   E   L   N   P   K   N   S   I   I   N   E   L   K   E   R   V   N   N   D   R   N   D   K   T   V   K   D   L   V   620

1861 AAC TTA CTA TTC GAT ACT TCA TTA TTA ACT TCA GGT TTT TCT CTT GAT GAA CCT CAT GTT TTC GCA GAA AGA ATC CAT AGA ATG ATT AAA CTA GGG TTG AGT ATC GAC GAA GAA CAT GAC 1980
621 N   L   L   F   D   T   S   L   L   T   S   G   F   S   L   D   E   P   H   V   F   A   E   R   I   H   R   M   I   K   L   G   L   S   I   D   E   E   H   D   660

1981 AAT ATT GAA TGT GAA AAT GAA AAT CAA ATT ATA GAT AAT GAT GAG ATA ATT GAT GAC ATG GAA GCA GTA GAT taagatttaacatattaaactaaatattaaaatatgttaaatagcaaatgtaataaaatttttta 2116
661 N   I   Q   C   E   N   E   N   Q   I   I   D   N   D   E   I   I   D   D   M   E   A   V   D                                                                 684
```

Fig. 3. The *hsp90* gene from the *Guillardia theta* nucleomorph. **A.** Schematic of HSP90 showing domains. **B.** Nucleotide sequence of *hsp90.* The deduced amino acid sequence is shown beneath the nucleotide sequence in upper case letters. The conserved consensus pattern for the *hsp90* family is underlined and the highly conserved motif at the carboxy terminus is shown in boldface and underlined. The nuclear localisation signal is indicated by dotted underlining and the twin heat shock elements upstream of the coding sequence are boxed.

cine zipper motif is preserved. A nuclear localization signal (KRKK) is present at the carboxy-terminal end of the DNA-binding domain.

A phylogenetic analysis of the DNA-binding domain of selected HSFs (Fig. 5D) reveals rampant gene duplication in the evolutionary history of this protein family. Our database survey identified over a dozen distinct HSF-like sequences present in *Arabidopsis thaliana*, as well as multiple paralogs in a variety of other plants (data not shown). Due to its small size (approximately 100 amino acids), the DNA-binding domain contains little phylogenetic signal. However, our analyses do show a well-supported separation of the animal and fungal HSFs from the plants and the single protist sequence (86% and 96%, quartet puzzling, and ML-distance bootstrap, respectively), consistent with an insertion/deletion present in the DNA-binding domain of the animal and fungal HSFs (Fig. 5C). Our analyses suggest that an early gene duplication occurred in land plant evolution, as two distinct, but weakly supported, clusters of HSF sequences are present (Fig. 5D). Multiple duplications have also occurred in animal evolution. The single *G. theta* HSF sequence still retains several conserved motifs characteristic of other DNA-binding proteins (see above), but was a fairly long branch in our phylogenetic trees (Fig. 5D).

The Chaperonin CCT. The genes for CCT subunits α, β, γ, δ, ε, η, θ, and ζ are distributed among all three chromosomes of the *G. theta* nucleomorph genome. The predicted amino acid sequences are clearly alignable with other eukaryotic homologs throughout most of their sequence, but many are shorter at the amino and carboxy terminal regions (data not shown). Unlike HSP70 and HSP90, none of the CCT subunit genes contained recognizable heat shock promoter elements.

A phylogenetic analysis of the CCT subunits and their archaebacterial homologs is presented in Fig. 6. The most striking feature of the *G. theta* nucleomorph CCTs is their divergent nature. All eight sequences branch with their respective orthologous groups with high statistical

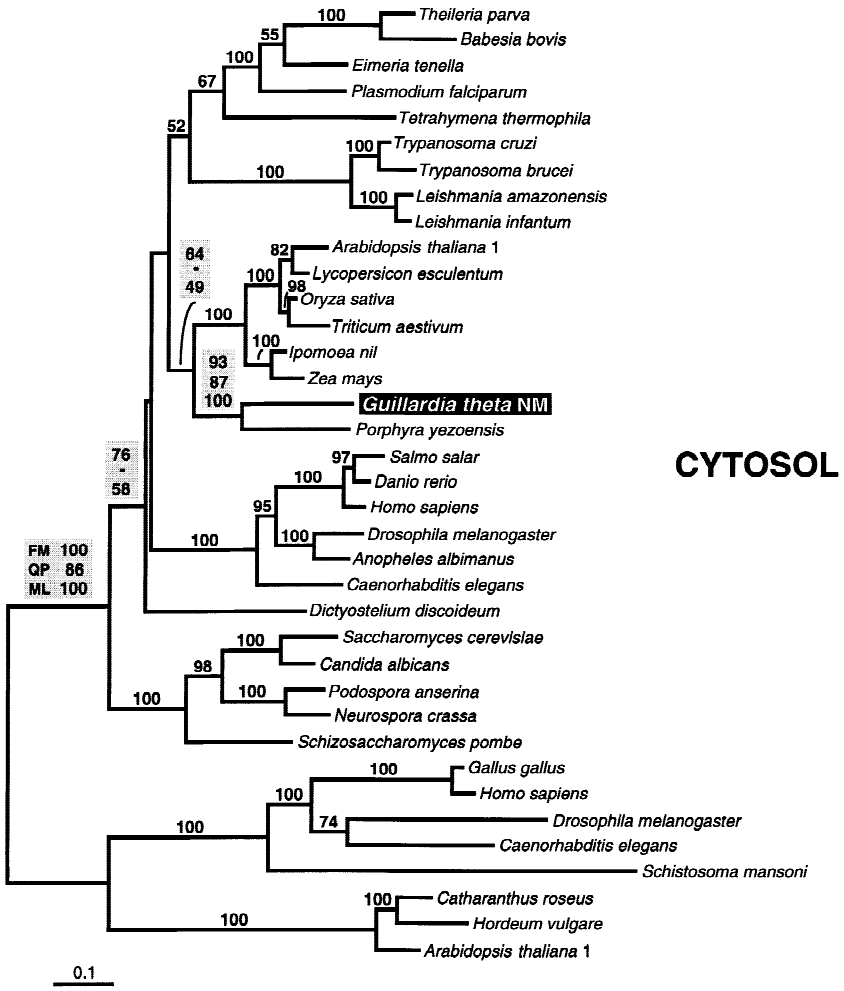


Fig. 4. Phylogenetic analysis of HSP90 from the *Guillardia theta* nucleomorph. The tree shown (lnL = −15585.36) is the topology of the best maximum likelihood (ML) tree from a heuristic search of 1000 trees in protML (Adachi and Hasegawa 1996), using 543 unambiguously aligned amino acid sites. ML branch-lengths were inferred under a rate heterogeneity model (JTT-F + Γ + inv; see Materials and Methods) in PUZZLE (Strimmer and von Haeseler 1997). Cytosolic and ER forms of HSP90 are indicated, and the *Guillardia theta* nucleomorph sequence (highlighted) strongly clusters with the cytosolic homologs. Support values above the branches are ML RELL values, and are given if >50%. Gray boxes contain support values for nodes of particular interest (FM, distance (Fitch–Margoliash) bootstrap values; QP, quartet puzzling support values, ML, ML RELL values). The scale bar indicates the inferred number of substitutions per amino acid site.

support and with all methods (data not shown), but are characterized by remarkably long branch lengths. To investigate this further, we examined the amino acid composition of the *G. theta* CCTs and found that, compared to CCT sequences from a diverse selection of other eukaryotes, the nucleomorph sequences possessed highly biased amino acid compositions. We used PUZZLE 4.O2 (Strimmer and von Haeseler 1997) to perform Chi-square tests for amino acid composition bias on each of the eight CCT subunits and found that the *G. theta* sequence failed in each case. Notably, the sequences appeared to be biased towards asparagine residues in cases where a unique substitution had occurred in the *G. theta* sequence, likely reflecting the extreme A-T bias of the nucleomorph genome. Analysis of codon usage of nucleomorph genes as a whole indicated a bias of greater than 80% towards codons ending in A or T (Douglas et al. in press). While significant differences have been observed in the rates of evolution among the different CCT subunits themselves (Archibald et al. 2000), no particular CCT subunit in *G. theta* appeared noticeably more divergent than the others (Fig. 6). This is in contrast to the pattern observed for the *G. theta* tubulins, where gamma-tubulin was found to be extraordinarily divergent relative to the moderately divergent alpha- and beta-tubulins (Keeling et al. 1999).

Discussion

The presence or absence of functionally well-characterized proteins in the periplastid space of cryptomonads may suggest which features it might still share with a "typical" eukaryotic cytosol, and which features have been severely reduced, modified, or lost entirely. As was the case for the *G. theta* tubulin genes (Keeling et al. 1999), the nucleomorph-encoded molecular chaperones described here have important implications for our current understanding of the cell biology of the cryptomonad periplastid complex.

The nucleomorph genome of *G. theta* encodes several of the major families of molecular chaperones necessary for protein folding, transport and degradation. Consistent with its dramatically reduced size, single intronless genes are present for HSP70, HSP90, the heat shock transcription factor HSF, and each of the CCT subunits. This is in contrast to vertebrates and plants, where multiple genes for heat shock proteins and their transcription factors

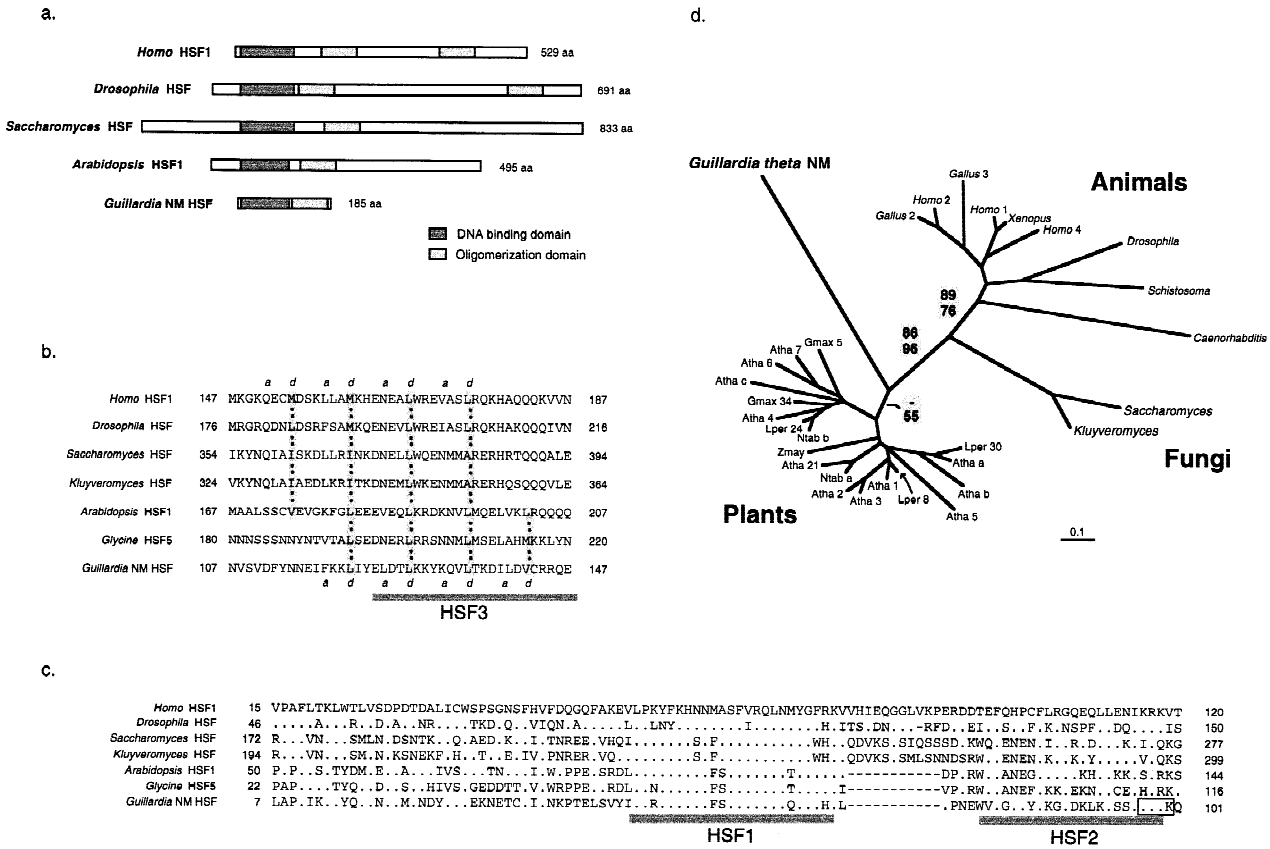


Fig. 5. The heat shock transcription factor gene from the *Guillardia theta* nucleomorph. **A.** Schematic of HSF showing domains. **B.** The oligomerization domain of HSF showing the leucine zipper motif (shaded vertical bars) and HSF3 consensus pattern. **C.** The DNA binding domain of HSF showing HSF1 and HSF2 consensus patterns and nuclear localisation signal (boxed). **D.** Phylogenetic analysis of 31 HSFs using 89 unambiguously aligned amino acid positions in the DNA-binding domain. The tree shown was inferred with a distance (Fitch–Margoliash) algorithm from a maximum likelihood distance matrix calculated with a rate heterogeneity model (JTT–F + Γ + inv; see Materials and Methods). Statistical support for the major branches are shown (quartet puzzling support values (top) and ML-distance bootstrap values (bottom). The scale bar represents the inferred number of amino acid substitutions per site. Abbreviations for plant HSFs: Gmax, *Glycine max;* Atha, *Arabidopsis thaliana;* Lper; *Lycopersicon peruvianum;* Zmay, *Zea mays;* Ntab, *Nicotiana tabacum.* Accession numbers for plant HSFs: Gmax 5, S59539; Gmax 34, S59538; Atha 1, g7428781; Atha 2, CAB63800; Atha 3, CAA74397; Atha 4, CAB16764; Atha 5, AAF16564; Atha 6, CAB63802; Atha 7, CAB63803; Atha 21, AAC31792; Atha a, AAC31222; Atha b, CAB41311; Atha c, AAB84350; Lper 8, P41153; Lper 24, P22335; Lper 30, P41152; Zmay, S57633; Ntab a, BAA83711; Ntab b, BAA83710.

exist. Heat shock elements are found upstream of both heat shock protein genes (but in none of the CCT subunit genes), indicating that this pathway of gene regulation is present. Although both heat shock proteins are similar in size to those found in other eukaryotes, the predicted HSF polypeptide is only about half the size of HSFs from other eukaryotes, and the CCT subunits have slightly reduced amino- and carboxy-termini. Apparently the nucleomorph genome has dispensed with everything but the bare essentials—active sites and functional domains—but retained some of the same regulatory mechanisms.

The activation of heat shock genes is known to be rapid yet transient and is mediated by HSF in response to elevated temperatures (dubbed a "molecular thermostat") and chemical or physiological stress. The *G. theta* HSF, although shorter than homologs from other organisms, possesses both a DNA-binding domain and an oligomerization domain containing a leucine zipper motif (Fig. 5), suggesting that the homo-trimerization required for binding to HSEs in the DNA can occur. The nuclear localization signal (KRKK) confirms that HSF is active in a nuclear environment, in this case most likely the nucleomorph. Under physiological conditions and during recovery from stress, HSP70 negatively regulates HSF transcriptional activity by binding to the HSF transactivation domain (Shi et al. 1998). This HSP70/HSF complex may then interact with HSP90 (or the HSP90 multichaperone complex) resulting in the inability of HSF to form trimers and activate transcription (Zuo et al. 1998). Under conditions of stress, the majority of cellular chaperone is bound to denatured protein, leaving less to bind to and repress HSF, thus allowing transcription of HSP genes to be stimulated. It is also thought that intramolecular interaction between HSF amino- and carboxy-terminal coiled coil domains keeps the protein in an in-

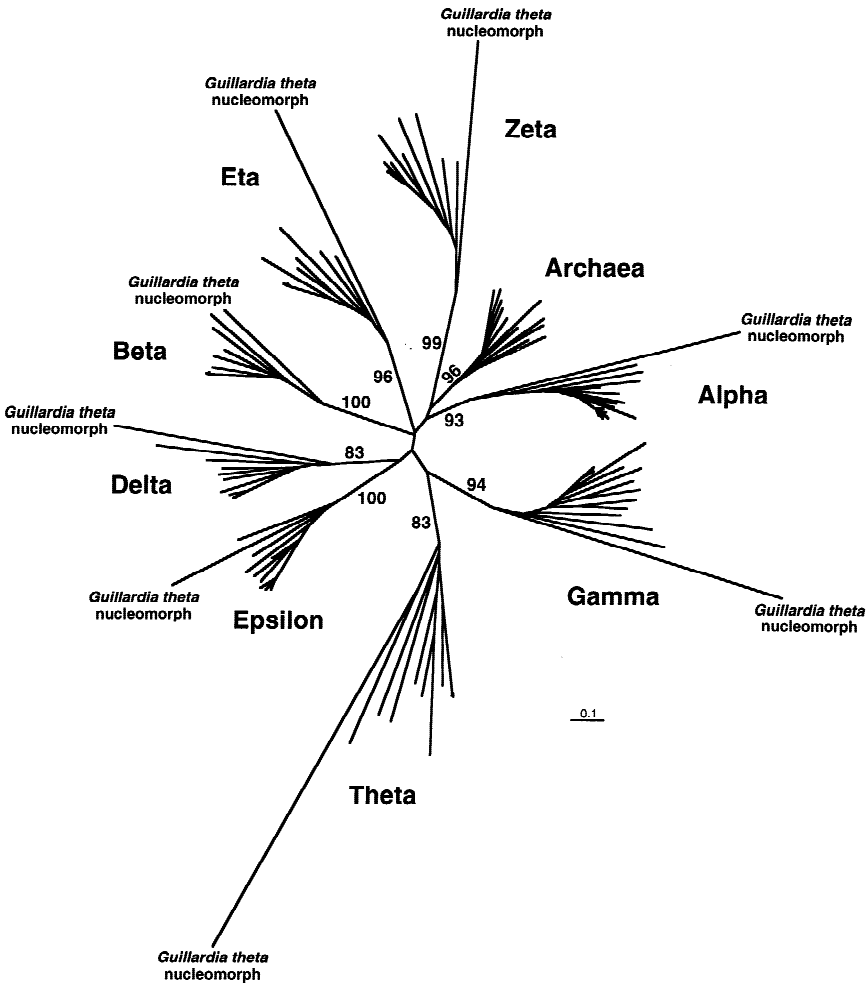


Fig. 6. Phylogenetic analysis of CCT subunits. The tree shown is a comprehensive neighbor-joining (NJ) phylogeny of 103 archaeal and eukaryotic CCTs, constructed from an alignment of 251 unambiguously aligned amino acid residues. Distance (NJ) bootstrap support for the monophyly of the archaeal chaperonins and the different CCT subunits are given (calculated from 100 resampling replicates). The archaeal chaperonins and the eight different CCT subunits (paralogs) are highlighted, as are the *G. theta* nucleomorph CCT subunit sequences. The scale bar indicates the inferred number of amino acid substitutions per site.

active state (Zuo et al. 1994). Interestingly, no recognizable transactivation domain or carboxy-terminal coiled coil domain is present in the nucleomorph HSF indicating that it is regulated by a different mechanism. Without the carboxy-terminal domain, the *G. theta* HSF may be permanently active at some level.

The nucleomorph-encoded HSP70 and HSP90 proteins from *G. theta* possess the hallmarks of typical cytosolic heat shock proteins. HSP70 proteins consist of a highly conserved amino-terminal ATPase domain of 44 kDa and a carboxy-terminal domain of 25 kDa that contains a conserved substrate-binding domain of 15 kDa and a less-conserved domain of 10 kDa with unknown function (Fig. 1A). The ATPase activity of HSP70 is stimulated by HSP40—interestingly, no *hsp40* gene could be found on the nucleomorph chromosomes, suggesting that this component might be imported from the host compartment. The HSP70/HSP40/substrate complex is stabilized through binding of another protein cofactor, Hip, to the ATPase domain (for review, see Frydman and Höhfeld 1997). The carboxy terminus of HSP70 contains eight residues that are highly conserved among all eukaryotes (GPTIEEVD; GPKIEEVD in *G. theta;* Fig. 1) and are responsible for interacting with an adaptor protein, Hop, that facilitates the assembly of the HSP70–HSP90 multichaperone (Scheufler et al. 2000). HSP70 may pass newly synthesized substrates to CCT (Bukau et al. 2000). HSP70 has also been found in the centrosome in animals as well as in the cytoplasm and the nucleus, where it binds HSF. The presence of a nuclear localization signal in the *G. theta* HSP70 suggests that it has the potential to localize to the nucleomorph.

HSP90 is known to interact with HSP70 via an adapter that binds the C-terminal sequence MEEVD (Scheufler et al. 2000); this sequence is MEAVD in the *G. theta* HSP90 (Fig. 3). In animals, HSP90 is a conserved core centrosomal component, and in yeast it plays a role in spindle pole duplication. Although a spindle apparatus and microtubules have not been seen in dividing cryptomonad nucleomorphs (McKerracher and Gibbs 1982; Morrall and Greenwood 1982; Meyer 1987), they could easily have been overlooked because of low contrast in the rather dense nucleomorph and their probably exceedingly transient duration in such a small nucleus: only a few very short microtubules would be needed for a fully functional spindle. We suggest that nucleomorph-encoded HSP90, HSP70, gamma tubulin, and ranbpm proteins interact to form a simplified cen-

trosome and nucleate the assembly of alpha and beta tubulin to form a minute spindle responsible for segregating the nucleomorph chromosomes. This could be tested by immunocytochemistry. Because nucleomorphs divide like most fungal nuclei, by a closed division with no nuclear envelope breakdown (McKerracher and Gibbs 1982; Morrall and Greenwood 1982; Meyer 1987), the putative spindle will probably be intranuclear as in their red algal ancestors (Schornstein and Scott 1982). Our demonstration of a nuclear localization sequence on Hsp90 raises the possibility that the centrosome may lie within the nucleomorph envelope, not outside it opposite a polar fenestra as in red algae (Schornstein and Scott 1982). This would be consistent with the apparent absence of a polar fenestra during nucleomorph division (McKerracher and Gibbs 1982; Morrall and Greenwood 1982; Meyer 1987), and its nondetection by electron microscopy in the periplastid space.

We have demonstrated drastic differences in the rates of protein sequence evolution among the various molecular chaperones encoded in the *G. theta* nucleomorph genome. While each of the eight *G. theta* nucleomorph CCT sequences fall into the previously described CCT subunit families with high statistical support, they are characterized by extremely long branches and highly biased amino acid compositions (see results). In stark contrast, the nucleomorph HSP70 and HSP90 sequences were remarkably well conserved. The phylogenies presented in Figs. 2 and 4 show that these sequences had very short branch lengths compared to other cytosolic homologs. In fact, the *G. theta* HSP70 nucleomorph sequence had a shorter branch than the HSP70 encoded in the host nuclear genome. In contrast to the eight CCT subunits, neither HSP70 or HSP90 showed biased amino acid compositions in Chi-square tests compared to their respective cytosolic homologs.

Such radical differences in the degree of conservation of HSP70 and HSP90 on one hand, and the eight CCT subunits on the other, presumably reflect differing degrees of functional constraint. However, in the absence of detailed information on the exact functions of these chaperones in the periplastid space, it is difficult to know what the reasons for such differences might be. It does seem significant that HSP70 is known to be extensively involved in protein translocation (in addition to general protein folding), a process that should be evolutionarily conserved in highly membranous cells such as cryptomonad algae. In contrast, the chaperonin CCT appears to be exclusively involved in the folding of newly-translated proteins (Willison and Horwich 1996). A reduction in the number of CCT substrates present and functioning in the periplastid space could explain the remarkable divergence of the CCT subunits. Extreme divergence in the evolution of the CCT substrates themselves (due to decreased/different functional constraints) could also influence the evolution of the CCTs—in effect, the co-evolution of chaperonin and substrate. It is interesting that one confirmed set of CCT substrates likely to function in the periplastid space of *G. theta*, the tubulins, are also divergent to varying degrees (Keeling et al. 1999). A homolog of the cytoskeletal element actin, the other well-characterized substrate of CCT, is not present in the *G. theta* nucleomorph genome. It remains to be seen whether this cytoskeletal component is imported or has been lost.

The complete lack of co-chaperones encoded in the nucleomorph is quite unexpected—if gene transfer from the nucleomorph to the host genome is more or less a random process, we would not expect all of the major chaperones to be nucleomorph-encoded, and all co-chaperones to be in the host nucleus. It is thus likely that some of the co-chaperones and interacting factors have been lost altogether during the extreme reduction of the nucleomorph genome. In the case of CCT, the loss of co-chaperonin(s), if possible, would surely influence the evolutionary constraints on the various functional regions of the molecules. A fuller understanding of the reasons for such a drastic increase in evolutionary rates awaits a more detailed knowledge of the full range of in vivo CCT substrates in a "typical" cell cytosol and the ability to compare these substrates to those still functioning in the periplastid space itself. Data on which co-chaperones are truly nuclear-encoded and functioning in the periplastid space, and which ones have been lost, will be essential. The cryptomonad endosymbiont may ultimately tell us about the minimal chaperone system necessary for maintaining basic eukaryotic cellular processes.

Acknowledgments. We thank NSERC and DFG for funding for this project. J.M.A was supported by a Doctoral Research Award from the Medical Research Council of Canada. Sequence data reported here are available from GenBank (accession numbers AF165818, AJ010592, AF083031). This is NRCC publication number 42331.

References

Adachi JM, Hasegawa M (1996) MOLPHY: programs for molecular phylogenetics. Inst Stat Math, Tokyo

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Archibald J, Logsdon JM Jr, Doolittle WF (2000) Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. Mol Biol Evol 17:1456–1466

Baldauf SL (1999) A search for the origins of animals and fungi: comparing and combining molecular data. American Naturalist 154(suppl.):S178–S188

Bohen SP, Kralli A, Yamamoto KR (1996) Hold 'em and fold 'em: chaperones and signal transduction. Science 268:1303–1304

Bukau B, Deuerling E, Pfund C, Craig EA (2000) Getting newly synthesized proteins into shape. Cell 101:119–122

Bukau B, Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. Cell 92:351–366

Burger G, Saint-Louis D, Gray MW, Lang BF (1999) Complete sequence of the mitochondrial DNA of the red alga *Porphyra pur-*

purea: cyanobacterial introns and shared ancestry of red and green algae. Plant Cell 11:1675–1694

Cavalier-Smith T (1999) Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Euk Microbiol 46: 347–366

Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. Trends Plant Sci 5:174–182

Cavalier-Smith T, Couch JA, Thorsteinsen KE, Gilson P, Deane JA, Hill DRA, McFadden GI (1996) Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. Eur J Phycol 31:315–328

Deane JA, Fraunholz M, Su V, Maier U-G, Martin W, Durnford DG, McFadden GI (2000) Evidence for nucleomorph to host nucleus gene transfer: light-harvesting complex proteins from cryptomonads and chlorarachniophytes. Protist 151:239–252

Douglas SE, Murphy CA, Spencer DF, Gray MW (1991) Molecular evidence that cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151

Douglas SE, Penny SL (1999) The plastid genome from the cryptomonad alga, *Guillardia theta:* complete sequence and conserved synteny groups confirm its common ancestry with red algae. J Mol Evol 48:236–244

Douglas SE, Turner S (1991) Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. J Mol Evol 33:267–273

Douglas SE, Zauner S, Fraunholz M, Beaton M, Penny S, Deng L, Wu X, Reith M. Cavalier-Smith T, Maier U-G (2001) The highly reduced genome of an enslaved algal nucleus. Nature (in press)

Felsenstein J (1993) PHYLIP (phylogeny inference package). University of Washington, Seattle

Frydman J, Höhfeld J (1997) Chaperones get in touch: the Hip-Hop connection. Trends Biochem Sci 22:87–92

Germot A, Philippe H (1999) Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. J Euk Microbiol 46:116–124

Gibbs SP (1979) The route of entry of cytoplasmically-synthesized proteins into chloroplasts of algae possessing chloroplast ER. J Cell Sci 35:253–266

Gibbs SP (1981) The chloroplast endoplasmic reticulum: structure, function, and evolutionary significance. Int Rev Cytol 72:49–99

Gillott MA, Gibbs SP (1980) The cryptomonad nucleomorph: its ultrastructure and evolutionary significance. J Phycol 16:558–568

Gupta RS (1995) Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. Mol Biol Evol 12:1063–1073

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73:237–244

Hofmann CJB, Rensing SA, Hauber MM, Martin WF, Muller SB, Couch J, McFadden GI, Igloi GL, Maier U-G (1994) Smallest eukaryotic genomes encode a protein gene: towards an understanding of nucleomorph functions. Mol Gen Genet 243:600–604

Ishida K, Green BR, Cavalier-Smith T (2001) Endomembrane structure and the protein-targeting pathway to chloroplasts in the heterokont alga *Heterosigma akashiwo* (Raphidophyceae, Chromista). J Phycol (in press)

Jakob U, Buchner J (1994) Assisting spontaneity: the role of Hsp90 and small Hsps as molecular chaperones. Trends Biochem Sci 19:205–211

Johnson JL, Craig EA (1997) Protein folding in vivo: unraveling complex pathways. Cell 90:201–204

Keeling PJ, Deane JA, Hink-Schauer C, Douglas SE, Maier U-G, McFadden GI (1999) The secondary endosymbiont of the cryptomonad *Guillardia theta* contains alpha-, beta-, and gamma-tubulin genes. Mol Biol Evol 16:1308–1313

Kubota H, Hynes G, Carne A, Ashworth A, Willison K (1994) Identification of six Tcp-1-related genes encoding divergent subunits of the TCP-1-containing chaperonin. Curr Biol 4:89–99

Lange BHM, Bachi A, Wilm M, Gonzalez C (2000) Hsp90 is a core centrosomal component and is required at different stages in the centrosome cycle in *Drosophila* and vertebrates. EMBO J 19:1252–1262

McKerracher L, Gibbs SP (1982) Cell and nucleomorph division in the alga *Cryptomonas.* Can J Bot 60:2440–2452

Meyer S (1987) The taxonomic implications of the ultrastructure and cell division of a stigma-containing *Chroomonas* sp. (Cryptophyceae) from Rocky Bay, Natal, South Africa. S Afr J Bot 53:129–139

Moreira D, Le Guyader H, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. Nature 405:69–72

Morrall S, Greenwood AD (1982) Ultrastructure of nucleomorph division in species of Cryptophyceae and its evolutionary implications. J Cell Sci 54:311–328

Nikaido I, Asamizu E, Nakajima M, Nakamura Y, Saga N, Tabata S (2000) Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis.* DNA Res 7:223–227

Palmer JD (2000) A single birth of all plastids? Nature 405:32–33

Rensing SA, Maier U-G (1994) Phylogenetic analysis of the stress-70 protein family. J Mol Evol 39:80–86

Rensing SA, Obrdlik P, Rober-Kleber N, Müller SB, Hofmann CJB, van de Peer Y, Maier U-G (1997) Molecular phylogeny of the stress-70 protein family with reference to algal relationships. Eur J Phycol 32:279–285

Robbins J, Dilworth SM, Laskey RA, Dingwall C (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell 64:615–623

Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, Hartl FU, Moarefi I (2000) Structure of TPR domain–peptide complexes: critical elements in the assembly of the Hsp70–Hsp90 multichaperone machine. Cell 101:199–210

Schornstein KL, Scott J (1982) Ultrastructure and cell division in the unicellular red alga *Porphyridium purpureum.* Can J Bot 60:85–97

Shi Y, Mosser DD, Morimoto RI (1998) Molecular chaperones as HSF1-specific transcriptional repressors. Genes Dev 12:654–666

Stiller JW, Hall BD (1997) The origin of red algae: implications for plastid evolution. Proc Natl Acad Sci USA 94:4520–4525

Strimmer K, von Haeseler A (1997) PUZZLE. Universität Muenchen, Zoologisches Institut

Swofford DL (1998) PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA

Thulasiraman V, Yag CF, Frydman J (1999) In vivo newly translated polypeptides are sequestered in a protected folding environment. EMBO J 18:85–95

Wang SL, Liu X-Q (1991) The plastid genome of *Cryptomonas* Φ encodes an hsp70-like protein, a histone-like protein, and an acyl carrier protein. Proc Natl Acad Sci USA 88:10783–10787

Wastl J, Fraunholz M, Zauner S, Douglas SE, Maier U-G (1999) Ancient gene duplication and differential gene flow in plastid lineages: the GroEL/Cpn60 example. J Mol Evol 48:112–117

Willison KR, Horwich AL (1996) Structure and function of chaperonins in archaebacteria and eukaryotic cytosol. In: Ellis RJ (ed) *The chaperonins.* Academic Press, San Diego, pp. 107–136

Zarzov P, Boucherie H, Mann C (1997) A yeast heat shock transcription factor (Hsf1) mutant is defective in both Hsc82/Hsp82 synthesis and spindle pole body duplication. J Cell Sci 110:1879–1891

Zauner S, Fraunholz M, Wastl J, Penny S, Beaton M, Cavalier-Smith T, Maier U-G, Douglas SE (2000) Chloroplast protein and centrosomal genes, a novel tRNA intron, and odd telomeres in an unusually compact eukaryotic genome—the cryptomonad nucleomorph. Proc Natl Acad Sci USA 97:200–205

Zuo J, Baler R, Dahl G, Voellmy R (1994) Activation of the DNA-binding ability of human heat shock transcription factor 1 may involve the transition from an intramolecular to an intermolecular triple-stranded coiled-coil structure. Mol Cell Biol 14:7557–7568

Zuo J, Guo Y, Guettouche T, Smith DF, Voellmy R (1998) Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. Cell 94:471–480