

# Origin and Evolution of Eukaryotic Chaperonins: Phylogenetic Evidence for Ancient Duplications in CCT Genes

John M. Archibald, John M. Logsdon Jr.,<sup>1</sup> and W. Ford Doolittle

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

Chaperonins are oligomeric protein-folding complexes which are divided into two distantly related structural classes. Group I chaperonins (called GroEL/cpn60/hsp60) are found in bacteria and eukaryotic organelles, while group II chaperonins are present in archaea and the cytoplasm of eukaryotes (called CCT/TriC). While archaea possess one to three chaperonin subunit-encoding genes, eight distinct CCT gene families (paralogs) have been characterized in eukaryotes. We are interested in determining when during eukaryotic evolution the multiple gene duplications producing the CCT subunits occurred. We describe the sequence and phylogenetic analysis of five CCT genes from *Trichomonas vaginalis* and seven from *Giardia lamblia*, representatives of amitochondriate protist lineages thought to have diverged early from other eukaryotes. Our data show that the gene duplications producing the eight CCT paralogs took place prior to the organismal divergence of *Trichomonas* and *Giardia* from other eukaryotes. Thus, these divergent protists likely possess completely hetero-oligomeric CCT complexes like those in yeast and mammalian cells. No close phylogenetic relationship between the archaeal chaperonins and specific CCT subunits was observed, suggesting that none of the CCT gene duplications predate the divergence of archaea and eukaryotes. The duplications producing the CCT $\delta$  and CCT $\epsilon$  subunits, as well as CCT $\alpha$ , CCT $\beta$ , and CCT $\eta$ , are the most recent in the CCT gene family. Our analyses show significant differences in the rates of evolution of archaeal chaperonins compared with the eukaryotic CCTs, as well as among the different CCT subunits themselves. We discuss these results in light of current views on the origin, evolution, and function of CCT complexes.

## Introduction

Chaperonin-mediated protein folding is a universal cellular process (reviewed in Bukau and Horwich 1998). Chaperonins are multisubunit double-ring complexes that harbor nascent or denatured polypeptides in their central chamber and facilitate protein folding through the hydrolysis of ATP (Ranson, White, and Saibil 1998; Sigler et al. 1998). Eukaryotic cells possess two distantly related (but clearly homologous) chaperonin classes with different evolutionary histories. Bacterial-type (group I) chaperonins, called cpn60 or hsp60, reside in eukaryotic organelles, while archaeal-type chaperonins (group II; called CCT [chaperonin-containing TCP-1] or TriC [TCP-1 ring complex]) are present in the eukaryotic cytosol (Trent et al. 1991; Frydman et al. 1992; Willison and Kubota 1994; Kubota, Hynes, and Willison 1995a).

Crystal structure comparisons of group I and group II chaperonins reveal remarkable structural conservation (Ditzel et al. 1998). There are, however, significant differences between the two chaperonin types. While group I chaperonins utilize the co-chaperonin GroES/cpn10 in the protein-folding process, no such homolog functions in the group II chaperonin complex. Instead, an extended "apical domain," present in the group II chaperonins but absent in the group I chaperonins, is thought to cap the central cavity in a manner analogous to GroES/cpn10 (Klumpp, Baumeister, and Essen 1997; Horwich

and Saibil 1998; Llorca et al. 1999b). Recent experiments suggest that novel co-chaperonins, unrelated to GroES/cpn10, interact with CCT to assist protein folding (Gebauer, Melki, and Gehring 1998; Geissler, Siegers, and Schiebel 1998; Vainberg et al. 1998; Siegers et al. 1999). Group I and group II chaperonins also differ in the number of subunits present in each chaperonin ring. *Escherichia coli* GroEL, the archetypal bacterial chaperonin, has a double-ring structure with seven subunits per ring (Braig et al. 1994), while archaeal and eukaryotic cytosolic chaperonin complexes are composed of eight- or nine-membered rings (reviewed in Willison and Horwich 1996; Klumpp and Baumeister 1998; Gutsche, Essen, and Baumeister 1999).

The most unusual feature of the group II chaperonins is their hetero-oligomeric composition. Unlike the homo-oligomer GroEL, archaeal chaperonins are often composed of several different (but homologous) subunits. We concluded previously that in the chaperonin complexes of archaea, hetero-oligomerism likely evolved multiple times independently (Archibald, Logsdon, and Doolittle 1999). In archaeal genomes, duplicate chaperonin genes (paralogs) are often more similar to each other than to those in other archaea, suggesting recent (lineage-specific) duplication. Compared with archaeal chaperonins, the eukaryotic CCT is even more hetero-oligomeric. This was first suggested biochemically (Frydman et al. 1992; Lewis et al. 1992), and subsequent sequence comparisons of CCT genes in mouse confirmed the existence of eight distinct subunit species ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\eta$ ,  $\theta$ , and  $\zeta$ ), each thought to occupy a unique position in the eight-membered CCT rings (Kubota et al. 1994; Kubota, Hynes, and Willison 1995a; Liou and Willison 1997). The divergent nature of these genes, as well as the discovery of clear yeast orthologs to each of the mouse subunits (Kim, Willison, and Horwich 1994; Kubota et al. 1994; Stoldt et al. 1996), suggests an ancient paralogy within eukaryotes.

<sup>1</sup> Present address: Department of Biology, Emory University.

Key words: chaperonins, parabasalids, diplomonads, gene duplication, eukaryotic evolution, phylogeny.

Address for correspondence and reprints: John M. Archibald, Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7. E-mail: jmarchib@is2.dal.ca.

*Mol. Biol. Evol.* 17(10):1456–1466. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

*Trichomonas vaginalis* and *Giardia lamblia*, members of the parabasalids and the diplomonads, respectively, are two parasitic unicellular eukaryotes. Originally, based on their lack of mitochondria and ultrastructural simplicity, parabasalids and diplomonads were suggested to represent early-diverging eukaryotic lineages (for recent review see Roger 1999). These lineages (and several others) were called "Archezoa" by Cavalier-Smith (1987) and were proposed to have diverged from other eukaryotes prior to the bacterial endosymbiosis that gave rise to mitochondria (i.e., they were primitively amitochondriate). Consistent with this idea, phylogenies of small-subunit ribosomal RNA (SSUrRNA) and several proteins placed *Trichomonas* and *Giardia* among the deepest branches on the eukaryotic tree (Sogin et al. 1989; Leipe et al. 1993; Hashimoto et al. 1994; Stiller, Duffield, and Hall 1998). However, the discovery of group I (i.e., bacterial/ mitochondrial) chaperonin genes in the nuclear genomes of these and other amitochondriate eukaryotes (Clark and Roger 1995; Roger, Clark, and Doolittle 1996; Roger et al. 1998) suggests that these organisms once possessed mitochondria (or their progenitors) and that mitochondrial absence is a derived feature. Furthermore, confidence in the deepest branches of phylogenetic trees has recently been shaken. Like parabasalids and diplomonads, the microsporidia (other amitochondriate members of Cavalier-Smith's [1987] Archezoa) also branched deeply in SSUrRNA and elongation factor trees (Leipe et al. 1993; Kamaishi et al. 1996). However, other data (Keeling and Doolittle 1996; Germot, Philippe, and Le Guyader 1997; Hirt et al. 1997, 1999) have shown that microsporidia are, in fact, relatives of fungi, and that their deep placement in phylogenetic trees was artifactual due to a fast rate of sequence evolution. Furthermore, Hirt et al. (1999) suggested that support for the placement of *Trichomonas* and *Giardia* among the deepest eukaryotic groups might also be suspect, although no alternate placement for these organisms with respect to other eukaryotes is evident.

Is the presence of eight CCT subunits a universal feature of eukaryotic cells? An early eukaryotic lineage that diverged from other eukaryotes prior to multiple CCT gene duplications might be expected to possess a smaller and/or different complement of CCT genes. The taxonomic diversity needed to address this question is, however, currently lacking. To this end, we sought to (1) increase the phylogenetic diversity of known CCT genes, (2) perform phylogenetic analyses of archaeal and eukaryotic chaperonins to determine when during eukaryotic evolution (and in what order) the gene duplications that gave rise to the CCT subunits occurred, and (3) address specific hypotheses regarding the origin and evolution of CCT from an archaeal-like homo- or moderately hetero-oligomeric chaperonin complex ancestor. Previous comparative sequence analyses (Kubota et al. 1994) have indicated that a completely hetero-oligomeric CCT was present in the common ancestor of animals and fungi. Our results push back the origin of the CCT gene duplications to the common ancestor of animals, fungi, plants, parabasalids, and diplomonads,

and likely to the common ancestor of all extant eukaryotes. While the exact position of the archaeal root to the eukaryotic CCTs is ambiguous, no close phylogenetic relationship between the archaeal chaperonins and specific eukaryotic CCT subunits was observed, suggesting that the eukaryotic CCT complex became hetero-oligomeric independent of the archaeal chaperonins. The gene duplications producing the CCT $\delta$  and CCT $\epsilon$  subunits, as well as those in the CCT $\alpha$ /CCT $\beta$ /CCT $\eta$  clade, represent the most recent duplications of the CCT gene family.

## Materials and Methods

### *Trichomonas* and *Giardia* Genomic DNAs

Genomic DNA from *T. vaginalis* strain NIH-C1 (ATCC#30001) was a gift from M. Müller (Rockefeller University, New York). Genomic DNA was isolated as described previously (Roger, Clark, and Doolittle 1996) from *G. lamblia* cells (strain WB; ATCC#30957) provided by A. Roger and D. Edgell.

### Cloning and Sequencing of *Trichomonas* and *Giardia* CCT Genes

Degenerate PCR primers were designed based on an alignment of published archaeal and eukaryotic chaperonin protein sequences (forward primers: CCT-1-for [5'-TACGGTGAYGGNACNAC-3'], CCT-5-for [5'-GAAATCGGNGAYGGNAC-3'], CCT-9-for [5'-CCAGTCGGTCTNGAYAAARATG-3']; reverse primers: CCT-3-rev [5'-TGGAGCTCCNSCNCNG-3'], CCT-4-rev [5'-CTCTACAGCNCCNSCNC-3'], CCT-7-rev [5'-ACGATGCACATNGHRTCRGTG-3']). PCR reactions were carried out under standard conditions (Gibco BRL Taq polymerase, buffer and dNTP, Ericomp and MJ Research Inc. PTC-100 thermal cyclers), with 40–45 cycles of 92°C for 30 s, 50°C for 30 s, and 72°C for 30–60 s. PCR products of the expected size were isolated (BIORAD, Prep-a-gene) and cloned (TA cloning kit, Invitrogen), or they were cloned directly from low-melt agarose (TA-TOPO cloning kit, Invitrogen). PCR products were sequenced manually (T7 sequencing kit, Pharmacia). Multiple independent PCR and genomic library clones were sequenced using LiCor and ABI automated sequencers.

The *Trichomonas* chaperonin genes presented in this study were obtained using the following primer combinations: *Trichomonas Cct2*: CCT-1-for/CCT-4-rev, CCT-5-for/CCT-4-rev, CCT-9-for/CCT-CCT-7-rev; *Trichomonas Cctd*: p80-4B (5'-CTGCCATTYGTGGCNATG-3')/P80-5 (5'-AGCGATGAACTTNARDAT-3'); *Trichomonas Cctg*: CCT-1-for/CCT-4-rev; *Trichomonas Cctz*: CCT-5-for/CCT-3-rev, CCT-5-for/CCT-4-rev, CCT-5-for/CCT-7-rev. A *Trichomonas* cDNA library clone encoding a protein with significant similarity to CCT $\eta$  was a gift from R. Hirt and M. Embley.

For *Giardia*, a portion of the *Cctd* gene was obtained with degenerate PCR primers (CCT-1-for/CCT-4-rev; see above). A recent sequence survey of the *G. lamblia* genome (Smith et al. 1998) revealed the presence of coding regions with similarity to several additional CCT subunits. Exact-match PCR primers were designed based on prelim-

inary genome sequence data from *Giardia* and, in combination with degenerate primers (above), were used to amplify multiple CCT genes (*Giardia Ccta*: GL.alpha.for.1 [5'-GTAGACATGCTTGTCTGCAG-3']/GL.alpha.rev.1 [5'-GTCGTGTATGCTCTAGTAGC-3']; *Giardia Cctb*: GL.beta.for.1 [5'-CCATAGCTGAGTTATAGATG-3']/GL.beta.rev.2 [5'-TAATCTTGTTCAGAGTCCATG-3'], CCT-1-for/GL.beta.rev.3 [5'-AGGTGCACAGCTTATTATGC-3']; *Giardia Cctg*: CCT-5-for/GL.gamma.rev.1 [5'-TCCGCAGAACCATACGCCAG-3']; *Giardia Ccte*: GL.eps.for.1 [5'-ATGATTAGTATCTCTCAGTG-3']/GL.eps.rev.1 [5'-GCTGAACGATCGTTGTTCATG-3']; *Giardia Cctq*: GL.theta.for.1 [5'-TTCTTCCATGATGAAGGTGC-3']/GL.theta.rev.1 [5'-GACCACGTA CTGCTCTAGAC-3']; *Giardia Cctz*: GL.zeta.for.1 [5'-AGAATTTTCATGCTGCTATC-3']/GL.zeta.rev.1 [5'-TGCTCAGAACGTGGTATCTG-3']). Preliminary sequence data from the *Giardia* Genome Project was obtained from the Josephine Bay Paul Center Web site at the Marine Biological Laboratory (www.bpc.mbl.edu). Sequencing was supported by the National Institute of Allergy and Infectious Diseases using equipment from LI-COR Biotechnology. The sequences presented in this study have been submitted to GenBank under the accession numbers AF226714–AF226726.

#### *Trichomonas vaginalis* Genomic Library Screening

PCR products of *Trichomonas* CCT genes were isolated (BIORAD, Prep-a-gene), labeled with  $\alpha^{32}\text{P}$  (Prime-It II random primer labeling kit, Stratagene), and used as probes to screen a *T. vaginalis* genomic library (Lambda ZapExpress, Stratagene; constructed previously by N. Fast and J. Logsdon). Genomic library clones containing full-length *Ccta*, *Cctd*, and *Cctz* genes, as well as a 5'-truncated clone containing *Cctg*, were obtained.

#### Phylogeny

Based on an alignment of group II chaperonin protein sequences constructed previously (Archibald, Logsdon, and Doolittle 1999), a larger alignment containing diverse bacterial/organellar, archaeal, and eukaryotic cytosolic chaperonin sequences (i.e., group I and group II chaperonins) was constructed and adjusted manually, taking into consideration published alignments (Kim, Willison, and Horwich 1994; Kubota et al. 1994; Waldmann et al. 1995) and crystal structures (Ditzel et al. 1998). Amino acid sequences inferred from the *Trichomonas* and *Giardia* CCT genes were added manually based on globally conserved regions. From this master alignment, smaller alignments containing subsets of sequences (e.g., bacteria + archaea + eukaryotes, archaea + eukaryotes, eukaryotes only) were constructed and used for phylogenetic analysis. For eukaryotes + archaea, the alignment consisted of 355 unambiguously aligned amino acid positions and included 10 archaeal sequences (six euryarchaeotes, four crenarchaeotes) and five sequences from each of the eight eukaryotic CCT paralogs (partial sequences presented here [CCT $\gamma$  from *Trichomonas* and *Giardia*, *Giardia* CCT $\delta$ ] were excluded in order to maximize the number of sites). For bac-

teria + archaea + eukaryotes, the alignment included 15 diverse bacterial sequences, 10 representative archaeal sequences, and the same eukaryotic sequences as above. This alignment contained 227 amino acid sites and corresponded primarily to the universally conserved ATP-binding/ATPase domains. Alignments of individual CCT subunits were used to estimate site-by-site evolutionary rates and the proportion of invariant sites (see below), and they contained the following taxa and number of sites: CCT $\alpha$ —338 sites (*Homo*, *Xenopus*, *Arabidopsis*, *Drosophila*, *Dictyostelium*, *Schistosoma*, *Caenorhabditis*, *Tetrahymena*, *Saccharomyces*, *Trichomonas*, and *Giardia*); CCT $\beta$ —358 sites (*Homo*, *Saccharomyces*, *Schizosaccharomyces*, *Caenorhabditis*, *Plasmodium*, and *Giardia*); CCT $\delta$ —361 sites (*Homo*, *Caenorhabditis*, *Fugu*, *Saccharomyces*, *Schizosaccharomyces*, *Glycine*, and *Trichomonas*); CCT $\epsilon$ —360 sites (*Homo*, *Caenorhabditis*, *Drosophila*, *Plasmodium*, *Saccharomyces*, *Avena*, *Cumcumis*, *Arabidopsis*, and *Giardia*), CCT $\eta$ —360 sites (*Homo*, *Caenorhabditis*, *Saccharomyces*, *Schizosaccharomyces*, *Plasmodium*, *Tetrahymena*, and *Trichomonas*); CCT $\gamma$ —287 and 353 sites (*Homo*, *Caenorhabditis*, *Xenopus*, *Drosophila*, *Arabidopsis*, *Tetrahymena*, *Oxytricha*, *Saccharomyces*, *Schizosaccharomyces*, and *Leishmania*; with and without *Trichomonas*, *Giardia*, respectively); CCT $\theta$ —359 sites (*Homo*, *Caenorhabditis*, *Candida*, *Saccharomyces*, *Schizosaccharomyces*, *Tetrahymena*, and *Giardia*); CCT $\zeta$ —360 sites (*Homo* [zeta1, zeta2], *Caenorhabditis*, *Drosophila*, *Saccharomyces*, *Schizosaccharomyces*, *Trichomonas*, and *Giardia*). Where missing data precluded site rate calculations (e.g., partial *Trichomonas* and/or *Giardia* sequences), amino acid positions were considered slowly evolving if they were present in all taxa in a particular subunit alignment. Conservative amino acid substitutions were also taken into consideration. All alignments are available from J.M.A. (email: jmarchib@is2.dal.ca).

Phylogenetic trees were inferred using maximum parsimony (MP), distance-based, and maximum-likelihood (ML) methods of tree reconstruction with the following programs: MP in PAUP\*, version 4.0 (Swofford 1998); distance, PROTDIST (PAM matrices), NEIGHBOR, and FITCH in PHYLIP, version 3.57 (Felsenstein 1993); ML, protML using the JTT-F model in MOLPHY (Adachi and Hasegawa 1996); and quartet puzzling using PUZZLE, versions 4.0 and 4.02 (Strimmer and von Haeseler 1997). Statistical support for MP and distance-based trees was obtained by bootstrapping with either 100 or 1,000 resampling replicates. Quartet puzzling support values (from PUZZLE) or RELL values (resampling estimated log likelihoods; obtained by quick-add ML searches of 100 or 1,000 trees in protML [Adachi and Hasegawa 1996]) were used as measures of support for ML trees. Support values for ML-distance analyses were obtained by bootstrapping (500 replicates) with PUZZLEBOOT, version 1.02 (A. Roger and M. Holder; <http://members.tripod.de/korbi/puzzle/>). PUZZLE was used to calculate ML distance matrices (using an eight rate category discrete approximation to the  $\Gamma$  distribution plus one invariable rate category), to determine the proportion of constant amino acid positions in align-

ments, and to statistically assess the significance of different tree topologies using the Kishino-Hasegawa test (Kishino and Hasegawa 1989). To estimate site-by-site evolutionary rates, discrete  $\Gamma$  distributions approximated with eight variable-site rate categories were calculated over neighbor-joining or Fitch-Margoliash trees using the JTT-F model of amino acid substitution in PUZZLE.

## Results

### Identification of CCT Genes in Putatively Ancient Eukaryotic Lineages

We isolated and sequenced partial or complete coding sequences for multiple *Cct* genes from the parabasalid *T. vaginalis* (*Ccta*, *Cctg*, *Cctd*, *Ccth*, and *Cctz*) and the diplomonad *G. lamblia* (*Ccta*, *Cctb*, *Cctg*, *Cctd*, *Ccte*, *Cctq*, and *Cctz*). For *Trichomonas*, genomic library clones encoding two different *Ccta* and two different *Cctd* genes were obtained. While both *Cctd* genes code for identical proteins (but have many synonymous substitutions), the two *Ccta* genes encode slightly different proteins. The full-length CCT $\alpha$ -1 from *Trichomonas* has an unusually short and divergent 3' end, while the CCT $\alpha$ -2 clone is 5'-truncated and possesses a carboxyl terminus typical of other CCTs. Southern hybridization of a *Ccta* PCR product to *Trichomonas* genomic DNA produced multiple hybridizing bands, confirming the presence of at least two genomic copies of this gene (data not shown). No spliceosomal introns were found in any of the genes presented here, consistent with their complete absence from the protein-coding genes described in these organisms thus far. To further increase the taxonomic sampling of CCT sequences for comparative study and phylogenetic analysis, we searched the public sequence databases by BLAST (Altschul et al. 1997) using the *Trichomonas* and *Giardia* CCT sequences as queries. We obtained complete sets of CCT protein sequences (eight or nine) for humans, mice (Kubota et al. 1994; Kubota, Hynes, and Willison 1995b), yeast, and *Caenorhabditis elegans*, as well as single or multiple CCT sequences for *Plasmodium falciparum*, *Leishmania major*, and a variety of animals, plants, fungi, and ciliates. Several of the *C. elegans* sequences (CCT $\gamma$ , CCT $\eta$  and CCT $\theta$ ; obtained from the Sanger center [http://www.sanger.ac.uk/]) contained unique insertions and/or deletions that were most likely the result of incorrect intron/exon boundary predictions; in each case, we identified alternate splice sites that removed the apparent insertions or added the missing exons. Our data set of group II chaperonins included 35 archaeal sequences and 85 eukaryotic CCTs.

The evolutionary relationship between the *Trichomonas* and *Giardia* sequences and other eukaryotic and archaeal chaperonins was examined by constructing phylogenetic trees. Figure 1 shows an unrooted neighbor-joining tree produced from an alignment of 11 representative archaeal sequences and all 85 eukaryotic CCTs, inferred from an alignment of 260 amino acid positions. Most notably, the *Trichomonas* and *Giardia* sequences form robust clades with each of the eight different CCT paralogs (100% support with all phyloge-

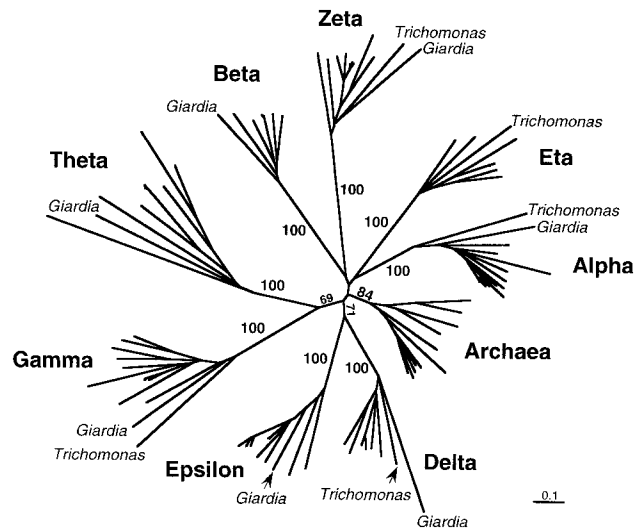


FIG. 1.—Placement of *Trichomonas* and *Giardia* chaperonins into known CCT subunit families. The tree shown is an unrooted neighbor-joining distance tree of 11 archaeal chaperonins and 85 eukaryotic CCT sequences, inferred from an alignment of 260 amino acid positions. The *Trichomonas* and *Giardia* sequences are labeled. In each CCT subunit clade, the *Trichomonas* and/or *Giardia* sequence branches strongly with subunit members from other species. This was also observed in parsimony and maximum-likelihood analyses. Bootstrap support (100 replicates) for the deepest branches is indicated if it is >50%. The scale bar indicates estimated number of amino acid substitutions per site.

netic methods; data not shown). This indicates that (1) the gene duplications producing the paralogs predate the divergence of *Trichomonas* and *Giardia* from other eukaryotes, and (2) multiple CCT paralogs have been retained over a large timescale of eukaryotic evolution. It is likely that both *Trichomonas* and *Giardia* possess all eight CCT paralogs. Indeed, a portion of the one CCT subunit gene not isolated from *Giardia*, *Ccth*, has recently been sequenced by the *Giardia* Genome Sequencing Project (Smith et al. 1998). Figure 1 also shows that the branch leading to the archaeal chaperonins is remarkably short compared with the branches leading to the different CCT subunits. The branch lengths within the various CCT clades also appear variable. To assess the significance of the latter observation, we calculated the percentages of amino acid identity shared between the mouse CCTs and the *Caenorhabditis*, *Saccharomyces*, *Giardia*, and *Trichomonas* sequences, as well as the proportion of constant amino acid residues found in each individual CCT subunit alignment (see *Materials and Methods*). The results (fig. 2) suggest differences in the degree of conservation of the individual CCT subunits. CCT $\theta$  (and to a lesser extent CCT $\gamma$ ) appears to be the least conserved subunit, showing the lowest percentage of identity in all within-ortholog comparisons. Furthermore, only 14.5% of the amino acid residues in the CCT $\theta$  alignment were constant (this number dropped further to 9.5% when the divergent *Plasmodium* CCT $\theta$  sequence was included), compared with 21.3%–37.7% constant residues in the other CCT subunit alignments. To statistically assess differences in the substitution rates of the different CCT paralogs, we performed a molec-

CCT subunit	% Identity with Mouse Ortholog <sup>1</sup>				% Constant AA Residues <sup>2</sup>
α alpha	67.0	61.7	52.1	48.9*	27.5*
β beta	66.7	65.6	58.5	—	37.7
δ delta	62.4	58.6	46.4*	51.9	36.8*
ε epsilon	67.5	59.9	57.6	—	37.5
η eta	64.1	62.1	—	53.1	36.9
γ gamma	60.5	60.0	50.3*	47.2*	21.3*
θ theta	55.0	49.5	41.8	—	14.5
ζ zeta 1	67.5	57.5	52.5	55.9	31.6
zeta 2	63.3	53.0	50.8	53.6	

FIG. 2.—Rates of evolution differ among CCT subunits. The table shows percentages of amino acid identities shared between mouse and *Caenorhabditis*, *Saccharomyces*, *Trichomonas*, and *Giardia* CCT protein sequences, along with the proportion of constant amino acid sites in alignments of CCT orthologs. <sup>1</sup> Pairwise amino acid identities were calculated in PAUP\* (Swofford 1998) using a complete alignment (706 positions) of the eight CCT paralogs. <sup>2</sup> The proportions of constant amino acid residues within individual CCT subunit families were obtained using PUZZLE, version 4.0 (Strimmer and von Haeseler 1997), from alignments that included only unambiguously aligned positions, no missing data, and maximum taxonomic diversity (see *Materials and Methods*). A dash indicates that a sequence is unavailable for comparison; an asterisk indicates a partial *Trichomonas* or *Giardia* sequence.

ular-clock likelihood ratio test with  $n - 2$  degrees of freedom in PUZZLE (Strimmer and von Haeseler 1997) on an ML-distance tree of the eight eukaryotic CCTs (40 representative taxa and 355 sites; see *Materials and Methods*). A molecular clock for the CCT paralogs was strongly rejected with  $P < 0.01$  (data not shown).

An alignment of the inferred *Trichomonas* and *Giardia* protein sequences with mouse CCTs is shown in figure 3. All of the sequences possess putative ATP-binding/ATP-hydrolysis sequence motifs similar to those described for other chaperonins (Kubota et al. 1994) and share significant amino acid identity (41%–58.5%) with mouse CCT homologs. The most striking feature of the alignment is the presence of multiple insertions in the *Giardia* CCT sequences that are not found in any CCTs characterized thus far. These insertions generally map to regions of variable length; however, the *Giardia* CCT $\theta$  and CCT $\epsilon$  sequences possess unique insertions (approximately 16 and 9 amino acids, respectively; see fig. 3) in a highly conserved region corresponding to a domain present in the bacterial/organellar chaperonins (positions 339–374 of the *Escherichia coli* GroEL sequence [Ditzel et al. 1998]) but absent from eukaryotic CCTs and archaeal chaperonins. The significance of these insertions (which presumably occurred independently) in terms of chaperonin subunit structure/function is not known.

It has been noted that the multiple CCT subunits are quite divergent from one another, particularly in their polypeptide-binding domains (Kim, Willison, and Horwich 1994). To examine the pattern and degree of conservation in the different CCT subunits more closely, we estimated the rate of evolution at amino acid sites across individual subunit alignments that contained maximal taxonomic diversity (see *Materials and Methods*). When these site rates were mapped onto an align-

ment containing all of the CCT subunits (paralogs), three general categories of amino acid sites were apparent: (1) conserved (slowly evolving) and identical amino acid residues present in multiple subunits (e.g., the ATPase domains), (2) conserved but different amino acid residues present in different subunits, and (3) poorly conserved/fast-evolving residues (i.e., little or no evolutionary constraint) present in one or multiple subunits. The results are presented in figure 3. Most notably, and consistent with a previous report (Kim, Willison, and Horwich 1994), much of the divergence between the different CCT subunits corresponds to their apical domains, the region involved in the binding of substrate. However, we also detected differences in the degree of conservation and amino acid sequence of the putative ATP-binding domains in the different subunits, as well as the presence of highly conserved “paralog-specific” motifs present in the equatorial and intermediate domains (fig. 3).

### Chaperonin Phylogeny

To more rigorously address the question of the evolutionary relationship of the CCT paralogs with the archaeal chaperonins and to determine the position of the bacterial (i.e., group I) root of the group II chaperonin tree, we performed phylogenetic analyses using alignments that contained reduced numbers of taxa and maximal phylogenetic diversity (see *Materials and Methods*). Surprisingly, when the bacterial chaperonin sequences were included as an outgroup (65 taxa, 227-position alignment; see *Materials and Methods*), parsimony, distance-based, and protML analyses produced trees in which the eukaryotic CCT $\zeta$  clade (not archaea) was the deepest branch of the group II chaperonins (data not shown). ML-distance trees (neighbor-joining and Fitch-Margoliash; as above) placed the euryarchaeotes as the deepest branch, but as a paraphyletic group separated from the crenarchaeotes by the CCT $\zeta$  clade of eukaryotes (a similar result was obtained in protML analyses using an alignment from which the fastest-evolving sites had been removed; 24 sites, 203 total sites). The deepest branches in these phylogenies were not well supported, however, suggesting that CCT $\zeta$  (the longest branch of the CCTs; see fig. 1) might be attracted to the long branch of the bacterial outgroup. Clearly, the small number of alignable amino acid positions between the group I and the group II chaperonins (approximately 200 sites, corresponding primarily to the ATP-binding/hydrolysis motifs) provide little phylogenetic signal with which to address the evolutionary history of the archaeal/eukaryotic chaperonin tree. We therefore focused on the group II chaperonin data set and attempted to determine the placement of the archaeal chaperonin root to the eukaryotic CCT tree and the branching order of the various CCT paralogs.

Figure 4A shows an ML tree of representative archaeal chaperonins and eukaryotic CCTs (50 taxa, 355 sites). As in figure 1, strong support for the monophyly of all of the individual CCT subunit clades is recovered.

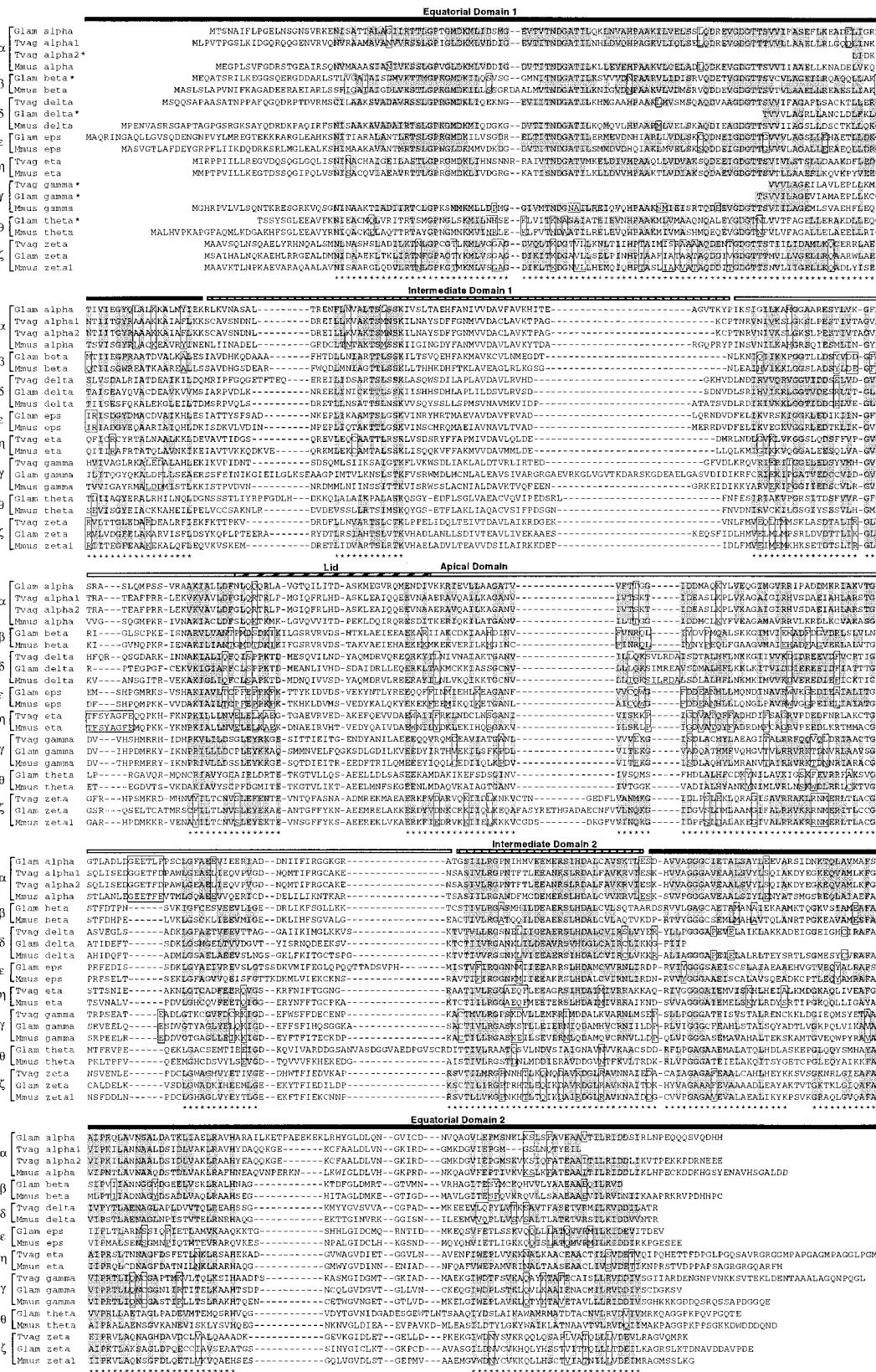


Fig. 3.—Alignment of inferred *Trichomonas* and *Giardia* chaperonin protein sequences with representative mouse CCT homologs (21 sequences, 706 amino acid positions). Functional domains predicted previously (equatorial, intermediate, apical and “lid”; Klumpp, Baumeister, and Essen 1997; Ditzel et al. 1998) are indicated. Amino acid residues estimated to be slowly evolving and amino acid insertions unique to specific CCT subunits are shaded gray; slowly evolving residues and amino acid insertions unique to specific CCT subunits are boxed (see Results). Dashes indicate gaps in the alignment. Stars next to taxon names indicate partial sequences; stars under the alignment indicate amino acid residues used for phylogenetic analysis of group II chaperonins. Abbreviations: Tvag, *Trichomonas*; Glam, *Giardia*; Mmus, mouse.

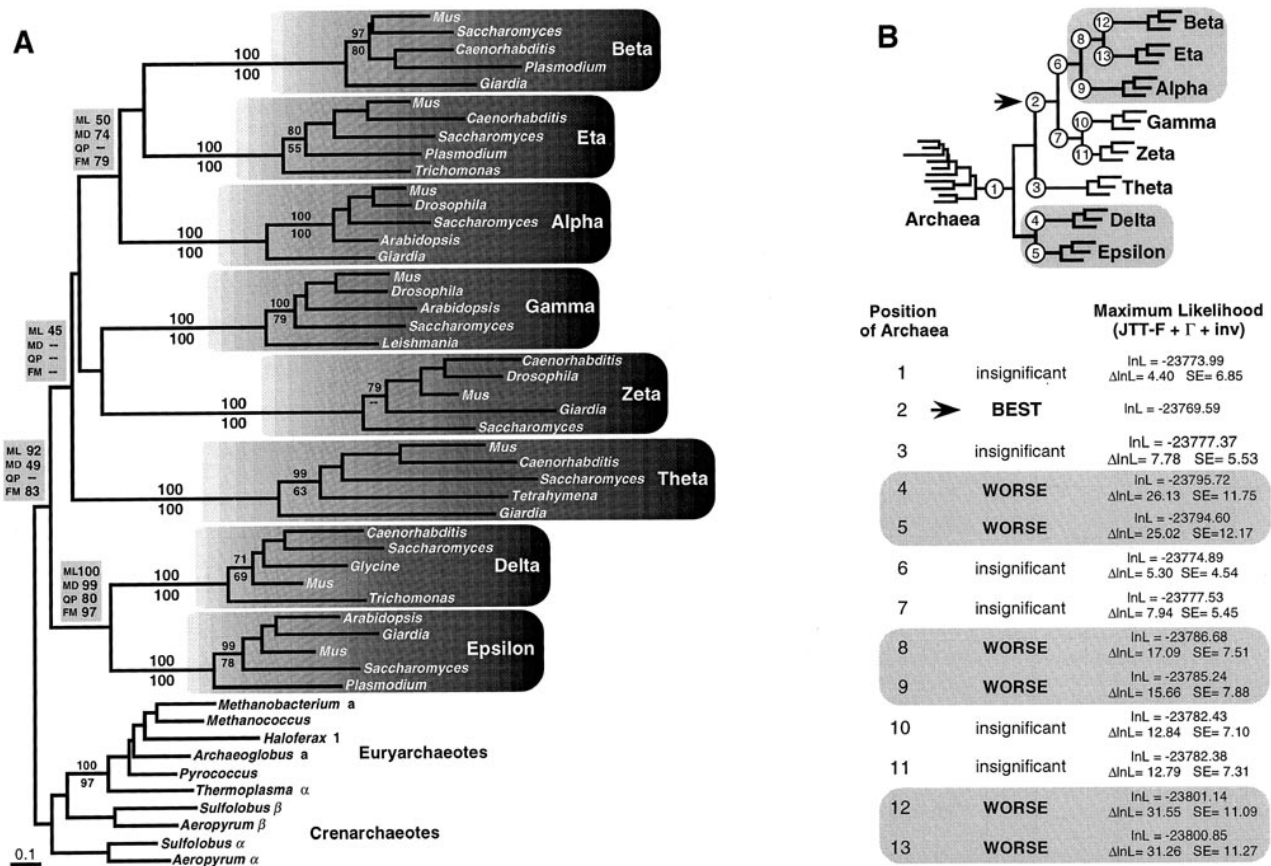


FIG. 4.—Phylogeny of group II chaperonin protein sequences. *A*, The maximum-likelihood (ML) tree (InL = -24,644.31) inferred from a heuristic search of 1,000 trees in protML (Adachi and Hasegawa 1996) using 355 unambiguously aligned amino acid positions. The eukaryotic CCT subunits (paralogs) are highlighted. Bootstrap support for the major branches, as well as for the deepest branch in each CCT clade, is indicated where it is >45%. ML RELL values are given above the branch, ML (with the rate heterogeneity model) distance bootstrap values (500 replicates) are below. Gray inset boxes indicate support values for nodes of particular interest (ML, ML RELL values; MD, ML distance bootstrap values; QP, quartet puzzling support values [10,000 quartet puzzling steps]; FM, distance [Fitch-Margoliash] bootstrap values). Dashes indicate support values <45%. The scale bar indicates the estimated number of substitutions per amino acid site. *B*, Top: schematic of the phylogeny shown in *A*; circles indicate alternate positions of the archaeal chaperonin root to the eukaryotic CCT tree tested by the method of Kishino and Hasegawa (1989) accounting for among-sites rate variation (JTT +  $\Gamma$  + inv model). The optimal placement (node 2) is labeled with an arrow. Bottom: significance of tree topologies in *A* and *B* over alternatives.  $\Delta$ InL values >1.96 SE were considered significantly worse at the 5% level. Shading corresponds to schematic (top).

Furthermore, the CCT $\delta$  and CCT $\epsilon$  paralogs form a well-supported clade with ML, distance, and parsimony methods (data not shown), as do CCT $\alpha$ , CCT $\eta$  and CCT $\beta$  (although more weakly). For archaea, the clustering of the euryarchaeal sequences together is well supported, while the monophyly of the  $\alpha$  and  $\beta$  paralogs of crenarchaeotes is not. The ML tree shows the crenarchaeal  $\beta$  subunit sequences branching with the euryarchaeotes, suggesting that the  $\alpha/\beta$  paralogy in crenarchaeotes may predate their divergence from euryarchaeotes (this topology was observed with some but not all phylogenetic methods; data not shown). Interestingly, most of the deepest branches of the group II chaperonin tree were poorly resolved, even when the maximum number of alignable amino acid positions was used. The systematic exclusion of individual CCT paralogs from the analyses, most notably CCT $\zeta$  (the longest branch) and CCT $\theta$  (poorly conserved), had little effect on the

support for the relationships among the CCT subunits, suggesting that no particular subset of the data was the cause of the unstructured trees (data not shown). We therefore performed Kishino-Hasegawa tests (Kishino and Hasegawa 1989) in PUZZLE to assess the significance of alternative topologies to the ML tree, taking into account among-sites rate heterogeneity. In these analyses, the optimal topology was slightly different from the protML tree in figure 4A (which was the second-best tree; 0.64 SE difference) and placed the archaeal root between the CCT $\theta$ /CCT $\delta$ /CCT $\epsilon$  and the CCT $\beta$ /CCT $\eta$ /CCT $\alpha$ /CCT $\gamma$ /CCT $\zeta$  clades (fig. 4B). Several other rootings were not considered worse at a 5% level of significance (e.g., the archaea as a sister group to CCT $\gamma$ , CCT $\theta$ , or CCT $\zeta$ ), but were between 1.2 and 1.8 SEs worse than the best tree. Notably, placements of the archaeal root within the CCT $\delta$ /CCT $\epsilon$  and CCT $\alpha$ /CCT $\beta$ /CCT $\eta$  clades were significantly worse topologies,

confirming the results of figure 4A and suggesting that these paralogies are the most recent in the evolution of CCT.

## Discussion

Gene duplications and gene losses make the reconstruction of ancient molecular events difficult. Group II chaperonins are a striking example—lineage-specific gene duplication and gene loss has occurred in archaea, and a remarkably ancient and complex paralogy exists in eukaryotes. Our data bear on several aspects of the origin and evolution of the completely hetero-oligomeric CCT in eukaryotes and on the origin of the eukaryotic cell itself.

The CCT genes presented here from *Trichomonas* and *Giardia*, two of the most divergent eukaryotes presently known, show strong affinity for each of the eight CCT subunit families found in “higher” eukaryotes. The gene duplications producing the different subunits clearly occurred very early in the evolution of the eukaryotic cell, and it is unlikely that the loss of any one of the CCT paralogs could, at this stage, be tolerated. The essential nature of at least six (and likely all) of the eight CCT genes in yeast (Stoldt et al. 1996; Lin et al. 1997) and their seemingly universal distribution in the diverse eukaryotic lineages examined here speak to that constraint. It has been suggested (Willison and Horwich 1996) that CCT evolved from an eightfold symmetric chaperonin complex like that in the crenarchaeote *Pyrodicticum occultum* (Phipps et al. 1991, 1993), based on the near-universal distribution of eight-membered ring structures among group II chaperonins (with *Sulfolobus* being the only exception; Marco et al. 1994). The  $\alpha$  and  $\beta$  subunits of crenarchaeotes (the deepest paralogy in archaea; Archibald, Logsdon, and Doolittle 1999) do not branch preferentially with particular subsets of CCT paralogs, however, as would be expected if the paralogy predated the divergence of crenarchaeotes and eukaryotes: there is no sense in which particular archaeal chaperonin paralogs are more closely related to some CCT paralogs than to others. While the ancestral chaperonin complex in eukaryotes was likely composed of eight-membered rings, it appears that CCT became hetero-oligomeric independent of the chaperonin complexes in archaea.

We attempted to resolve the relative branching order of the CCT paralogs and thus determine the order in which CCT “acquired” so many different subunits. Unlike other paralogous “eukaryote-specific” gene families, such as actins and tubulins, which have very distantly related prokaryotic homologs, the eukaryotic CCTs have relatively close archaeal homologs to serve as an outgroup. The exact placement of the archaeal root on the CCT tree remains unclear, but significantly, our data reject the placement of the root within the CCT $\delta$ /CCT $\epsilon$  and the CCT $\alpha$ /CCT $\beta$ /CCT $\eta$  clades. It is thus likely that CCT underwent intermediate stages of hetero-oligomerism, perhaps similar to the degree observed in present-day archaeal chaperonin complexes, and that the CCT $\delta$ , CCT $\epsilon$ , CCT $\alpha$ , CCT $\beta$ , and CCT $\eta$  subunits rep-

resent more recent divergences in eukaryotic chaperonin evolution.

Kubota et al. (1994) suggested that all CCT subunits should be present in all eukaryotes, and estimated a divergence time of two billion years for the different CCT paralogs based on the assumption that the amino acid substitution rate of each CCT subunit family has been constant. The data presented here are consistent with the former prediction but indicate that a clocklike rate of sequence divergence for each of the eight CCT paralogs is clearly not the case. We observed striking differences in the degree of conservation of the individual CCT subunits, as well as paralog-specific, highly conserved sequence motifs (fig. 3). CCT $\theta$  appears to be the least conserved subunit and may have reduced/different functional constraints. The results of recent biochemical studies (Liou and Willison 1997; Liou, McCormack, and Willison 1998) support this notion: compared with the other CCTs, unique subunit-subunit binding properties were observed for CCT $\theta$  in vitro, as was a much reduced level of CCT $\theta$  mRNA relative to the other CCT genes (Liou and Willison 1997). From this perspective, and in light of our phylogenetic analyses, amino acid identity comparisons which suggest that the eight CCT subunits are approximately equally related to each other (Kubota et al. 1994; Kubota, Hynes, and Willison 1995a) are misleading.

Why are there so many CCT paralogs? It has been suggested that the multiple gene duplications in the CCT gene family were concurrent with (and facilitated) the evolution of the eukaryotic cytoskeleton (Willison and Kubota 1994; Willison and Horwich 1996). Unlike GroEL, which appears to service a broad range of substrates in the bacterial cytoplasm (Houry et al. 1999), CCT is thought to be more “specialized.” Actins and tubulins, the major cytoskeletal proteins of eukaryotic cells, appear to be the predominant substrates of CCT (Willison and Kubota 1994; Kubota, Hynes, and Willison 1995a), although others have been, and continue to be, identified (Farr et al. 1997; Melki et al. 1997; Won et al. 1998; Feldman et al. 1999). Llorca et al. (1999a) have recently provided strong evidence for interactions between  $\alpha$ -actin and the apical domains of specific CCT subunits ( $\delta$ - $\epsilon$  or  $\delta$ - $\beta$ ) within the central chamber of CCT (curiously, our data suggest CCT $\delta$  and CCT $\epsilon$  to be among the most recent CCT duplicates). Such observations suggest coevolution of CCT and its substrates. We recently presented a more neutral model for the evolution of duplicate subunits in archaeal (and, by extension, early eukaryotic) chaperonins (Archibald, Logsdon, and Doolittle 1999), where coevolution between duplicate subunits could also lead to obligatory hetero-oligomerism. In archaea, hetero-oligomeric chaperonin complexes appear to have evolved multiple times independently (recurrent paralogy), a pattern that is inconsistent with a model of coevolution of chaperonin and substrate.

Our analyses of the positions of constant or variable amino acid sites for each CCT subunit family (see *Results*) revealed that many of the CCT subunits possess “signatures” that are invariant, or nearly so, with re-

spect to the other CCT families (fig. 3). Kim, Willison, and Horwich (1994), using an alignment that contained primarily mammalian and yeast CCT homologs, noted that conserved subunit-specific signatures often corresponded to regions of the protein involved in the binding of substrate. The method used here for identifying differences in CCT subunit sequence evolution were consistent with this result but also identified highly conserved subunit-specific motifs that, based on the archaeal thermosome and GroEL crystal structures (Ditzel et al. 1998), correspond to regions of intra- and inter-subunit contacts. We also observed differences in the degree of conservation of the ATP-binding/hydrolysis motifs in CCT $\theta$ , as well as CCT $\gamma$  and CCT $\zeta$  (fig. 3). Finally, genetic studies (Lin et al. 1997; Lin and Sherman 1997) have shown CCT $\zeta$  (CCT-6 in yeast) to be sensitive to mutations not only in its apical domain, but also in subunit-subunit contact regions. Curiously, CCT $\zeta$  was remarkably tolerant to mutations in the ATP-binding/hydrolysis motifs; it is not clear why these motifs should be so highly conserved across all eukaryotic species examined so far.

In this sense, it seems appropriate to view a particular subunit's function in terms of its contribution to the proper formation of the hetero-oligomeric CCT particle as well as to the binding of substrate(s). We argue that a fairly rigid and ordered arrangement of subunits in the hetero-oligomeric CCT would have had to precede (or be concurrent with) the evolution of subunit-specific roles for interactions with substrates and that a pattern of "recurrent paralogy" should be a necessary intermediate in the evolution of complete hetero-oligomerism. We base this argument on the fact that the specific functional roles of individual CCT subunits in protein folding described so far are context-dependent (i.e., they demand an ordered arrangement of subunits; Llorca et al. 1999a). Interestingly, the evolutionary pattern of the group II chaperonins bears a strong resemblance to that of the proteasome, a barrel-shaped proteolytic complex found in archaea, in the eukaryotic cytosol, and in some bacteria. Archaea possess single  $\alpha$  and  $\beta$  subunits (Baumeister et al. 1998), while eukaryotes possess seven  $\alpha$  and seven  $\beta$  paralogs (subunits), one for each position in the seven-membered  $\alpha$  and  $\beta$  rings (Groll et al. 1997). In both chaperonins and proteasomes, the evolution of single hetero-oligomeric particles, instead of multiple distinct homo-oligomeric ones, suggests that coevolution between duplicate subunits has been a significant factor in shaping their architectures.

It is clear that gene duplication and gene loss have been, and still are, prominent forces in archaeal and eukaryotic chaperonin evolution. A "recent" CCT gene duplication in mammals (CCT $\zeta$ -1, CCT $\zeta$ -2; Kubota et al. 1997), a probable *Sulfolobus*-specific paralogy in crenarchaeotes (Archibald, Logsdon, and Doolittle 1999), and the presence of multiple copies of CCT paralogs in *Trichomonas* (and, undoubtedly, many other eukaryotes) indicate that chaperonin gene duplication is an ongoing process. We recently presented phylogenetic evidence for "recent" gene loss in the euryarchaeal *Pyrococcus* species (Archibald, Logsdon, and Doolittle

1999). An even more striking case can be inferred for yeast CCTs. Genome sequence analyses in *Saccharomyces* suggest that a whole-genome duplication may have occurred after its divergence from *Kluyveromyces* (Wolfe and Shields 1997); given the fact that CCT was already completely hetero-oligomeric at this time (i.e., yeast had at least eight CCT genes), the presence of exactly eight CCT genes in the present-day yeast genome (Stoldt et al. 1996) indicates that multiple CCT duplicates have been lost.

The evolutionary forces influencing the retention of duplicate chaperonin genes are less obvious. What is becoming clear is that many of the complex paralogies unique to eukaryotic genomes (e.g.,  $\alpha$ - and  $\delta$ -DNA polymerases [Edgell, Malik, and Doolittle 1998],  $\alpha$ - and  $\beta$ -tubulins [Keeling and Doolittle 1996], and RNA polymerases I, II, and III [Stiller, Duffield, and Hall 1998]) were present early in eukaryotic evolution. The CCT gene family examined here is the most extreme example thus far. We suggest that a tendency toward highly paralogous gene families (and more "complex" macromolecular machinery) in eukaryotes compared with prokaryotes may reflect fundamental differences in the ways in which prokaryotic and eukaryotic genomes evolve. Larger genomes with multiple linear chromosomes should reduce the probability of gene conversion between recent (unlinked) duplicates and, in general, more easily accommodate duplicate genes (offsetting the effects of random gene loss). Furthermore, chromosomal or whole-genome duplications provide a ready mechanism for doubling the number of paralogs present in a genome. Inherent differences in the mechanisms and frequency of gene/chromosome/genome duplication and gene conversion/loss could influence the retention of duplicate genes as much as the positive selection for new paralog-specific functions.

## Acknowledgments

We thank M. Müller for *Trichomonas* genomic DNA, A. Roger and D. Edgell for *Giardia* cells, M. Embley and R. Hirt for a *Trichomonas* cDNA clone encoding CCT $\eta$ , N. Fast for a *Trichomonas* genomic library, A. Stoltzfus for mol2con.pl, and A. Roger and members of the Doolittle lab for helpful discussion and critical review of the manuscript. M. Leroux is also thanked for helpful discussions on chaperonin evolution. Preliminary sequence data from the *Giardia lamblia* Genome Project was obtained from the Josephine Bay Paul Center Web site at the Marine Biological Laboratory (www.bpc.mbl.edu). Sequencing was supported by the National Institute of Allergy and Infectious Diseases using equipment from LI-COR Biotechnology. This work was supported by a grant awarded to W.F.D. by the Medical Research Council (MRC) of Canada. J.M.A. was supported by an MRC studentship awarded to W.F.D., and by an MRC Doctoral Research Award. J.M.L. was supported by postdoctoral fellowships from MRC and NIH.

## LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**:1–150.
- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG, W. MILLER, and D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- ARCHIBALD, J. M., J. M. LOGSDON JR., and W. F. DOOLITTLE. 1999. Recurrent paralogy in the evolution of archaeal chaperonins. *Curr. Biol.* **9**:1053–1056.
- BAUMEISTER, W., J. WALZ, F. ZUHL, and E. SEEMULLER. 1998. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**:367–380.
- BRAIG, K., Z. OTWINOWSKI, R. HEGDE, D. C. BOISVERT, A. JOACHIMIAK, A. L. HORWICH, and P. B. SIGLER. 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**:578–586.
- BUKAU, B., and A. L. HORWICH. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell* **92**:351–366.
- CAVALIER-SMITH, T. 1987. Eukaryotes with no mitochondria. *Nature* **326**:332–333.
- CLARK, C. G., and A. J. ROGER. 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. USA* **92**:6518–6521.
- DITZEL, L., J. LÖWE, D. STOCK, K. O. STETTER, H. HUBER, R. HUBER, and S. STEINBACHER. 1998. Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93**:125–138.
- EDGE, D. R., S. B. MALIK, and W. F. DOOLITTLE. 1998. Evidence of independent gene duplications during the evolution of archaeal and eukaryotic family B DNA polymerases. *Mol. Biol. Evol.* **15**:1207–1217.
- FARR, G. W., E. C. SCHARL, R. J. SCHUMACHER, S. SONDEK, and A. L. HORWICH. 1997. Chaperonin-mediated folding in the eukaryotic cytosol proceeds through rounds of release of native and nonnative forms. *Cell* **89**:927–937.
- FELDMAN, D. E., V. THULASIRAMAN, R. G. FERREYRA, and J. FRYDMAN. 1999. Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TRiC. *Mol. Cell* **4**:1051–1061.
- FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.
- FRYDMAN, J., E. NIMMESGERN, H. ERDJUMENT-BROMAGE, J. S. WALL, P. TEMPST, and F. U. HARTL. 1992. Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J.* **11**:4767–4778.
- GEBAUER, M., R. MELKI, and U. GEHRING. 1998. The chaperone cofactor Hop/p60 interacts with the cytosolic chaperonin-containing TCP-1 and affects its nucleotide exchange and protein folding activities. *J. Biol. Chem.* **273**:29475–29480.
- GEISLER, S., K. SIEGERS, and E. SCHIEBEL. 1998. A novel protein complex promoting formation of functional alpha and gamma-tubulin. *EMBO J.* **17**:952–966.
- GERMOT, A., H. PHILIPPE, and H. LE GUYADER. 1997. Evidence for loss of mitochondria in microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol. Biochem. Parasitol.* **87**:159–168.
- GROLL, M., L. DITZEL, J. LÖWE, D. STOCK, M. BOCHTLER, H. D. BARTUNIK, and R. HUBER. 1997. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**:463–471.
- GUTSCHE, I., L. O. ESSEN, and W. BAUMEISTER. 1999. Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *J. Mol. Biol.* **293**:295–312.
- HASHIMOTO, T., Y. NAKAMURA, F. NAKAMURA, T. SHIRAKURA, J. ADACHI, N. GOTO, K. OKAMOTO, and M. HASEGAWA. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol.* **11**:65–71.
- HIRT, R. P., B. HEALY, C. R. VOSSBRINCK, E. U. CANNING, and T. M. EMBLEY. 1997. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr. Biol.* **7**:995–998.
- HIRT, R. P., J. M. LOGSDON JR., B. HEALY, M. W. DOREY, W. F. DOOLITTLE, and T. M. EMBLEY. 1999. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl. Acad. Sci. USA* **96**:580–585.
- HORWICH, A. L., and H. R. SAIBIL. 1998. The thermosome: chaperonin with a built-in lid. *Nat. Struct. Biol.* **5**:333–336.
- HOURY, W. A., D. FRISHMAN, C. ECKERSKORN, F. LOTTSPEICH, and F. U. HARTL. 1999. Identification of *in vivo* substrates of the chaperonin GroEL. *Nature* **402**:147–154.
- KAMAISHI, T., T. HASHIMOTO, Y. NAKAMURA, F. NAKAMURA, S. MURATA, N. OKADA, K. OKAMOTO, M. SHIMIZU, and M. HASEGAWA. 1996. Protein phylogeny of translation elongation factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes. *J. Mol. Evol.* **42**:257–263.
- KEELING, P. J., and W. F. DOOLITTLE. 1996. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.* **13**:1297–1305.
- KIM, S., K. R. WILLISON, and A. L. HORWICH. 1994. Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. *Trends Biochem. Sci.* **19**:543–548.
- KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**:170–179.
- KLUMPP, M., and W. BAUMEISTER. 1998. The thermosome: archetype of group II chaperonins. *FEBS Lett.* **430**:73–77.
- KLUMPP, M., W. BAUMEISTER, and L. O. ESSEN. 1997. Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell* **91**:263–270.
- KUBOTA, H., G. HYNES, A. CARNE, A. ASHWORTH, and K. WILLISON. 1994. Identification of six Tcp-1-related genes encoding divergent subunits of the TCP-1-containing chaperonin. *Curr. Biol.* **4**:89–99.
- KUBOTA, H., G. M. HYNES, S. M. KERR, and K. R. WILLISON. 1997. Tissue-specific subunit of the mouse cytosolic chaperonin-containing TCP-1. *FEBS Lett.* **402**:53–56.
- KUBOTA, H., G. HYNES, and K. WILLISON. 1995a. The chaperonin containing t-complex polypeptide 1 (TCP-1). Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. *Eur. J. Biochem.* **230**:3–16.
- . 1995b. The eighth Cct gene, Cctq, encoding the theta subunit of the cytosolic chaperonin containing TCP-1. *Gene* **154**:231–236.
- LEIPE, D. D., J. H. GUNDERSON, T. A. NERAD, and M. L. SOGIN. 1993. Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.* **59**:41–48.
- LEWIS, V. A., G. M. HYNES, D. ZHENG, H. SAIBIL, and K. WILLISON. 1992. T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. *Nature* **358**:249–252.

- LIN, P., T. S. CARDILLO, L. M. RICHARD, G. B. SEGEL, and F. SHERMAN. 1997. Analysis of mutationally altered forms of the Cct6 subunit of the chaperonin from *Saccharomyces cerevisiae*. *Genetics* **147**:1609–1633.
- LIN, P., and F. SHERMAN. 1997. The unique hetero-oligomeric nature of the subunits in the catalytic cooperativity of the yeast Cct chaperonin complex. *Proc. Natl. Acad. Sci. USA* **94**:10780–10785.
- LIU, A. K., E. A. MCCORMACK, and K. R. WILLISON. 1998. The chaperonin containing TCP-1 (CCT) displays a single-ring mediated disassembly and reassembly cycle. *Biol. Chem.* **379**:311–319.
- LIU, A. K., and K. R. WILLISON. 1997. Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes. *EMBO J.* **16**:4311–4316.
- LLORCA, O., E. A. MCCORMACK, G. HYNES, J. GRANTHAM, J. CORDELL, J. L. CARRASCOSA, K. R. WILLISON, J. J. FERNANDEZ, and J. M. VALPUESTA. 1999a. Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature* **402**:693–696.
- LLORCA, O., M. G. SMYTH, J. L. CARRASCOSA, K. R. WILLISON, M. RADERMACHER, S. STEINBACHER, and J. M. VALPUESTA. 1999b. 3D reconstruction of the ATP-bound form of CCT reveals the asymmetric folding conformation of a type II chaperonin. *Nat. Struct. Biol.* **6**:639–642.
- MARCO, S., D. URENA, J. L. CARRASCOSA, T. WALDMANN, J. PETERS, R. HEGERL, G. PFEIFER, H. SACK-KONGEHL, and W. BAUMEISTER. 1994. The molecular chaperone TF55. Assessment of symmetry. *FEBS Lett.* **341**:152–155.
- MELKI, R., G. BATELIER, S. SOULIE, and R. C. WILLIAMS JR. 1997. Cytoplasmic chaperonin containing TCP-1: structural and functional characterization. *Biochemistry* **36**:5817–5826.
- PHIPPS, B. M., A. HOFFMANN, K. O. STETTER, and W. BAUMEISTER. 1991. A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaeobacteria. *EMBO J.* **10**:1711–1722.
- PHIPPS, B. M., D. TYPKE, R. HEGERL, S. VOLKER, A. HOFFMANN, K. O. STETTER, and W. BAUMEISTER. 1993. Structure of a molecular chaperone from a thermophilic archaeobacterium. *Nature* **361**:475–477.
- RANSON, N. A., H. E. WHITE, and H. R. SAIBIL. 1998. Chaperonins. *Biochem. J.* **333**:233–242.
- ROGER, A. J. 1999. Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**(Suppl.):S146–S163.
- ROGER, A. J., C. G. CLARK, and W. F. DOOLITTLE. 1996. A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proc. Natl. Acad. Sci. USA* **93**:14618–14622.
- ROGER, A. J., S. G. SVARD, J. TOVAR, C. G. CLARK, M. W. SMITH, F. D. GILLIN, and M. L. SOGIN. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc. Natl. Acad. Sci. USA* **95**:229–234.
- SIEGERS, K., T. WALDMANN, M. R. LEROUX, K. GREIN, A. SHEVCHENKO, E. SCHIEBEL, and F. U. HARTL. 1999. Compartmentation of protein folding *in vivo*: sequestration of non-native polypeptide by the chaperonin-GimC system. *EMBO J.* **18**:75–84.
- SIGLER, P. B., Z. XU, H. S. RYE, S. G. BURSTON, W. A. FENTON, and A. L. HORWICH. 1998. Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* **67**:581–608.
- SMITH, M. W., S. B. ALEY, M. SOGIN, F. D. GILLIN, and G. A. EVANS. 1998. Sequence survey of the *Giardia lamblia* genome. *Mol. Biochem. Parasitol.* **95**:267–280.
- SOGIN, M. L., J. H. GUNDERSON, H. J. ELWOOD, R. A. ALONSO, and D. A. PEATTIE. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* **243**:75–77.
- STILLER, J. W., E. C. DUFFIELD, and B. D. HALL. 1998. Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **95**:11769–11774.
- STOLDT, V., F. RADEMACHER, V. KEHREN, J. F. ERNST, D. A. PEARCE, and F. SHERMAN. 1996. The Cct eukaryotic chaperonin subunits of *Saccharomyces cerevisiae* and other yeasts. *Yeast* **12**:523–529.
- STRIMMER, K., and A. VON HAESELER. 1997. PUZZLE. Zoologisches Institut, Universität München, Germany.
- SWOFFORD, D. L., 1998. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). Sinauer, Sunderland, Mass.
- TRENT, J. D., E. NIMMESGERN, J. S. WALL, F. U. HARTL, and A. L. HORWICH. 1991. A molecular chaperone from a thermophilic archaeobacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* **354**:490–493.
- VAINBERG, I. E., S. A. LEWIS, H. ROMMELAERE, C. AMPE, J. VANDEKERCKHOVE, H. L. KLEIN, and N. J. COWAN. 1998. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **93**:863–873.
- WALDMANN, T., A. LUPAS, J. KELLERMANN, J. PETERS, and W. BAUMEISTER. 1995. Primary structure of the thermosome from *Thermoplasma acidophilum*. *Biol. Chem. Hoppe Seyler* **376**:119–126.
- WILLISON, K. R., and A. L. HORWICH. 1996. Structure and function of chaperonins in archaeobacteria and eukaryotic cytosol. Pp. 107–136 in R. J. ELLIS, ed. *The chaperonins*. Academic Press, San Diego.
- WILLISON, K. R., and H. KUBOTA. 1994. The structure, function, and genetics of the chaperonin containing TCP-1 (CCT) in eukaryotic cytosol. Pp. 299–312 in R. I. MORIMOTO, A. TISSIERES, and C. GEORGOPOULOS, eds. *The biology of heat shock proteins and molecular chaperones*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- WOLFE, K. H., and D. C. SHIELDS. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**:708–713.
- WON, K. A., R. J. SCHUMACHER, G. W. FARR, A. L. HORWICH, and S. I. REED. 1998. Maturation of human cyclin E requires the function of eukaryotic chaperonin CCT. *Mol. Cell. Biol.* **18**:7584–7589.

GEOFFREY MCFADDEN, reviewing editor

Accepted June 13, 2000