Recurrent paralogy in the evolution of archaeal chaperonins

John M. Archibald, John M. Logsdon Jr and W. Ford Doolittle

Chaperonins are multisubunit double-ring complexes that mediate the folding of nascent proteins [1,2]. In bacteria, chaperonins are homo-oligomeric and are composed of seven-membered rings. Eukaryotic and most archaeal chaperonin rings are eight-membered and exhibit varying degrees of hetero-oligomerism [3,4]. We have cloned and sequenced seven new genes encoding chaperonin subunits from the crenarchaeotes *Sulfolobus solfataricus*, *S. acidocaldarius*, *S. shibatae* and *Desulfurococcus mobilis*. Although some archaeal genomes possess a single chaperonin gene, most have two. We describe a third chaperonin-encoding gene (TF55-γ) from two *Sulfolobus* species; phylogenetic analyses indicate that the gene duplication producing TF55-γ occurred within crenarchaeal evolution. The presence of TF55-γ in *Sulfolobus* correlates with their unique nine-membered chaperonin rings. Duplicate genes (paralogs) for chaperonins within archaeal genomes very often resemble each other more than they resemble chaperonin genes from other archaea. Our phylogenetic analyses suggest multiple independent gene duplications – at least seven among the archaea examined. The persistence of paralogous genes for chaperonin subunits in multiple archaeal lineages may involve a process of co-evolution, where chaperonin subunit heterogeneity changes independently of selection on function.

Address: Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4H7, Canada.

Correspondence: John M. Archibald
E-mail: jmarchib@is2.dal.ca

Received: **1 July 1999**
Revised: **9 August 1999**
Accepted: **9 August 1999**

Published: **13 September 1999**

Current Biology 1999, 9:1053–1056
http://biomednet.com/elecref/09609822009001053

0960-9822/99/$ – see front matter

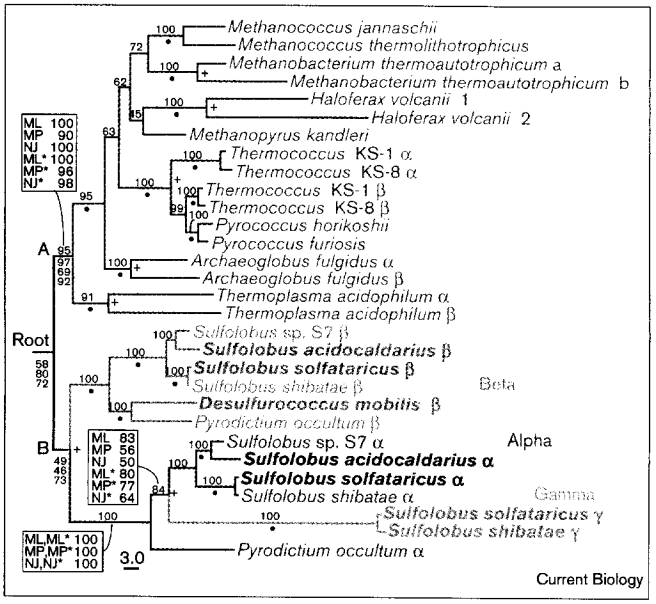
Results and discussion

Archaeal chaperonins were first described in the crenarchaea *Pyrodictium occultum* and *S. shibatae* [5–7] and show remarkable sequence similarity to the eight homologous subunits present in the eukaryotic cytoplasmic chaperonin complex CCT (chaperonin-containing TCP-1, also known as TriC (TCP-1 ring complex)) [8–11]. *Sulfolobus* chaperonins were first described as homo-oligomeric [7] but were later found to have two different subunit species, named TF55-α and TF55-β [12–14]. In a search for possible additional chaperonin genes, we applied degenerate PCR to genomic DNA of *Sulfolobus* and *Desulfurococcus*. Orthologs of TF55-α and -β [12–14] were cloned and sequenced from *S. solfataricus* (TF55-β), *S. acidocaldarius* (TF55-α and -β) and *D. mobilis* (TF55-β). TF55-α as well as a third, previously undescribed, chaperonin gene which we call TF55-γ were sequenced during sequencing of the *S. solfataricus* p2 genome [15]. We confirmed the presence and sequence of TF55-α and γ by PCR cloning and sequencing. A TF55-γ ortholog was also obtained from *S. shibatae* by PCR, and successfully hybridized to *S. shibatae* genomic DNA (data not shown).

The complete *S. solfataricus* TF55-γ gene encodes a protein of 539 amino acids (predicted molecular weight (MW) = 59.258 kDa, pI = 5.10) that has 55.2% and 43.2% identity with *S. solfataricus* TF55-α (MW = 59.659 kDa, pI = 5.12) and *S. shibatae* TF55-β (MW = 59.681, pI = 5.3; [13]), respectively. A comparison of the *S. solfataricus* and *S. shibatae* TF55-γ nucleotide sequences revealed a significant bias towards synonymous (silent) substitutions ($K_A/K_S \cong 0.1$; data not shown), as would be expected if these genes are expressed and are evolving under selection.

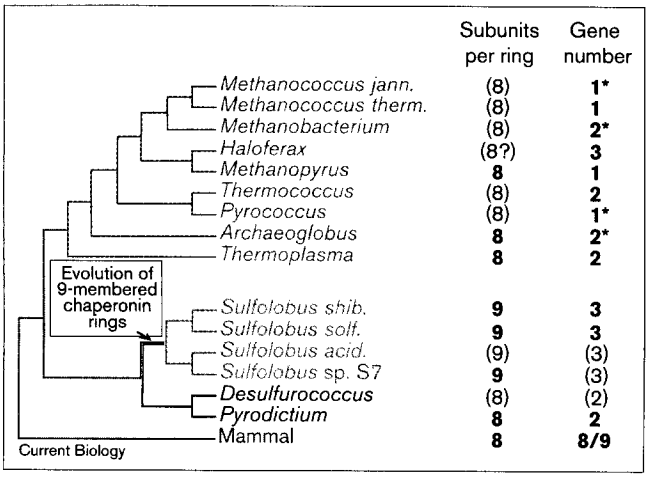
Archaeal genome sequences completed to date contain only one (*Methanococcus jannaschii*, *Pyrococcus horikoshii*) or two (*Methanobacterium thermoautotrophicum*, *Archaeoglobus fulgidus*) chaperonin subunit genes. The finding of a third gene in *Sulfolobus*, quite divergent from the other two, is therefore surprising. To investigate a possible relationship between this third gene and the unique nine-membered structure of the *Sulfolobus* chaperonin complexes, we performed phylogenetic analyses on available archaeal chaperonin sequences; in some analyses, the eukaryotic CCT sequences were used as an outgroup. A remarkable recurring pattern of gene duplication and loss in archaeal chaperonins was observed (Figure 1). The deepest branching separates the two recognized kingdoms (euryarchaeotes and crenarchaeotes) within the Archaea, consistent with the notion that a single chaperonin subunit gene in the last common ancestor of the two kingdoms gave rise to all modern archaeal chaperonin genes. Nevertheless, within both euryarchaeotes and crenarchaeotes, paralogy is rampant: a minimum of seven events of chaperonin gene duplication can be inferred. Within the euryarchaeotes, 'lineage-specific' gene duplications have occurred in *Methanobacterium thermoautotrophicum*, *Haloferax volcanii*, *A. fulgidus*, *Thermoplasma acidophilum* and the *Pyrococcus*/*Thermococcus* clade. Amino acid identities between euryarchaeal paralogs range from 58.3% (*H. volcanii* 1 and 2) to 80.6% (*Thermococcus* KS-1 α

Figure 1



Phylogeny of archaeal chaperonins. The maximum likelihood tree (-lnL 11053.5) of archaeal chaperonin amino acid sequences from an exhaustive protML analysis [26] is shown. Sequences from this study are in bold; sequences are deposited under the accession numbers AF149920–AF149925 and AF181261. Euryarchaeotes are shown in green and the three different crenarchaeal subunits are shown in red, blue and orange. Percentage support values (RELL values from a heuristic (quick-add) protML search [26]) are given above each node; bullets indicate nodes that were constrained in the exhaustive protML analysis. Inset boxes indicate support for nodes of particular interest; values were derived from various tree reconstruction methods (ML, maximum likelihood; MP, maximum parsimony; NJ, neighbor-joining distance). The influence of site-by-site rate variation on the support for these nodes was also tested (see Supplementary material); support values from analyses in which fastest-evolving sites were removed are labeled with an asterisk. Grey branches indicate the region and position of the eukaryote outgroup root, determined from additional phylogenetic analyses (see Supplementary material). Support values for nodes A, B and Root are given in the order ML, MP, NJ from top to bottom. The scale bar indicates 3.0 substitutions per 100 amino acid sites. Plus signs indicate inferred gene duplications (see text).

Figure 2

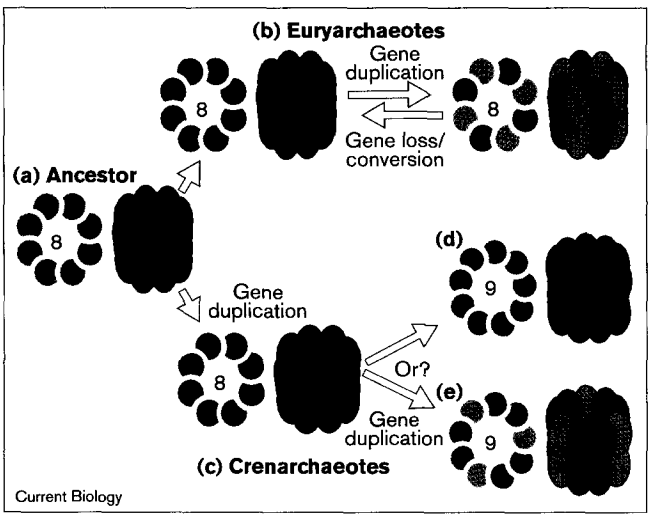


Evolution of chaperonin symmetry and gene number. A cladogram of archaeal relationships based on Figure 1 is shown on the left. Euryarchaeotes are shown in green; crenarchaeotes are shown in orange (*Sulfolobus*) or red (*Desulfurococcus* and *Pyrodictium*) based on the presence of both α and γ subunits or α only, respectively. The subunit number per chaperonin ring and the number of known or inferred chaperonin genes are shown on the right. Subunits per ring: bold values indicate known subunit stoichiometry from electron microscopic studies; values in parentheses are predicted. Gene number: bold values indicate known gene number from sequence data; asterisks indicate that the total gene number is confirmed by complete genome sequence; values in parentheses are predicted. A recent study [27] found a third chaperonin gene in *Haloferax*, the sequence of which has not yet been determined; the subunit number of *Haloferax* chaperonin complexes is also not known. In mouse, nine chaperonin genes are known, but only eight are constitutively expressed; the ninth subunit shows testis-specific expression [28].

and KS-1 β) suggesting that some duplications occurred more recently than others. Interestingly, the complete genome of *Pyrococcus horikoshii* has a single chaperonin gene. As this gene (together with a single gene from *Pyrococcus furiosis*) forms a clade with only one of the two paralogous genes in *Thermococcus* strains K1 and K8, the simplest interpretation is that *Pyrococcus* lost one of the paralogs (Figures 1,2).

Among crenarchaeotes, an early duplication producing α and β genes predated the separation of *Sulfolobus* and *Pyrodictium* (Figure 1). In contrast, the duplication giving rise to the *Sulfolobus* α and γ paralogs took place after this point. This observation is most interesting in the light of observed differences in crenarchaeal chaperonin-complex architectures (Figures 2,3). The α and β subunits in *P. occultum* chaperonins are thought to alternate in each eight-membered ring [5,6], similar to the known arrangement in the euryarchaeote *Thermoplasma acidophilum* [16]. The organization of subunits in the nine-membered chaperonin rings of *Sulfolobus* species remains enigmatic, however. The widespread distribution of eight-membered chaperonin rings outside crenarchaeotes (in all eukaryotes and euryarchaeotes examined thus far; Figure 2) suggests this is ancestral, and that a transition from eight- to nine-membered chaperonin rings occurred during crenarchaeal evolution in an ancestor of *Sulfolobus* (Figures 2,3). Because α and β subunits could not alternate equally in a nine-membered ring, Kagawa *et al.* [13] proposed that *Sulfolobus* chaperonins consist of two homo-oligomeric rings, one of α subunits and the other of β subunits. More recently, Ellis *et al.* [17] examined two-dimensional crystals prepared from *S. solfataricus* and proposed that each ring has threefold symmetry with an $(\alpha_2\beta)_3$ arrangement. Our discovery of a third chaperonin-subunit-encoding gene raises the interesting possibility that *Sulfolobus* chaperonins are in fact nine-membered rings with an $(\alpha\beta\gamma)_3$

Figure 3



Archaeal chaperonin evolution by recurrent paralogy. Schematic representation of chaperonin structures: multimeric chaperonin rings are composed of individual subunits that interact asymmetrically (side-to-side and top-to-bottom). Subunit colors are the same as in Figure 1; hypothesized interactions between rings are based on *T. acidophilum* [16]. **(a)** Hypothetical ancestral state of the chaperonin complex common to euryarchaeotes, crenarchaeotes and, probably, eukaryotes: eight-membered homo-oligomeric rings (see text). **(b)** Chaperonin subunit gene duplications have occurred independently in at least five euryarchaeal lineages (different subunits are indicated by light and dark green). At least one gene loss has also occurred. **(c)** A gene duplication took place early in crenarchaeal evolution. A more recent gene duplication took place in a *Sulfolobus* ancestor; a change from eight- to nine-membered chaperonin rings also occurred. **(d,e)** Two possible nine-membered structures. (d) The $(\alpha_2\beta)_3$ arrangement of Ellis *et al.* [17] inferred from the two-dimensional crystallization of *Sulfolobus* chaperonins. (e) Our prediction of alternating α, β and γ subunits in each *Sulfolobus* chaperonin ring.

arrangement (Figure 3). Indeed, TF55-α and the TF55-γ described here are predicted to have nearly identical biophysical properties, consistent with previous descriptions of a 2:1 TF55-α to -β ratio [12].

Interestingly, a 42 nucleotide (14 amino acid) insertion present in the α and β genes of *Pyrodictium occultum* but absent in all other archaeal sequences provides possible evidence for partial gene conversion (see Supplementary material). Frequent partial gene conversions causing the concerted evolution of paralogous proteins within a genome could conceivably produce a phylogenetic pattern similar to that observed for euryarchaeal chaperonins (multiple lineage-specific paralogs). The highest amino acid identity between any two paralogs in our dataset is only 80.6%, however, suggesting that gene conversion is infrequent. Analyses of silent sites (synonymous codon positions) using GENECONV failed to detect any statistically significant stretches of nucleotide identity (a potential indicator of regions of partial gene conversion) among paralogs within a genome (see Supplementary material).

Even with gene conversion, the persistence of paralogy in so many separate lineages begs for an explanation. If an early archaeal gene duplication produced paralogs with functions that began to diverge soon thereafter, we would expect modern archaea to retain and exhibit such 'deep paralogy': in general, for two paralogs '**a**' and '**b**', **a** genes from different species would be more similar to each other than each is to the **b** gene in the same species. Instead, we have evidence that duplicate genes have arisen — and been retained — independently in five different euryarchaeal lineages. It is possible that, in each of these instances, paralogy is maintained because the hetero-oligomeric chaperonin thus produced has acquired functions that its homo-oligomeric ancestor lacked. The species examined comprise non-thermophilic halophiles (such as *Haloferax*) and both autotrophic (such as *Methanococcus*) and heterotrophic (such as *Archaeoglobus*) thermophiles, however: homo- and hetero-oligomeric chaperonins appear randomly distributed among them with respect to environment and/or lifestyle (Figure 2). There is no reason to suspect that, in archaea, homo- and hetero-oligomeric chaperonins function differently.

Co-evolved interdependence between subunits of a hetero-oligomeric complex seems a more appealing possibility. Ancestrally, chaperonins would have homo-oligomeric rings, the subunits of which are products of a single gene; in archaea, this inference is favored by our phylogenetic analyses. Gene duplication into **a** and **b** paralogs would be followed by sequence divergence, through the fixation, in one or the other paralog, of mutations that are neutral or only slightly deleterious. Duplicate chaperonin genes would thus encode functionally identical subunits that assemble into rings in random proportions determined by their cellular abundance. At this stage, one or the other duplicate could be lost as inconsequentially as it was gained, and partial or complete gene conversion events between recent duplicates might periodically reset the 'divergence clock'. With time, some mutations in the gene for one paralog (say **a**) might increase its ability to bind to the other and/or decrease its ability to bind to itself. At this point, co-evolved changes in the **b** paralog that establish a similar preferential formation of heterodimers with **a** subunits would then make loss of either gene disadvantageous. This last step acts as a 'ratchet', locking hetero-oligomerism into place. Such a process of co-evolved changes leading to preferential subunit–subunit interactions would not only occur within chaperonin rings, but also between chaperonin rings. Such co-evolution could occur without change in overall function of the chaperonin complex and would be selectively neutral. The completely hetero-oligomeric CCT complex found in eukaryotes may in fact be an example of such a process taken to completion. Individual CCT subunits share approximately 30% identity and seem to occupy specific positions relative to the others in each eight-membered ring [4,18–20]. Although six (and probably all)

CCT-subunit-encoding genes are known to be essential in yeast [19–22], separate functions for each subunit have not been clearly demonstrated.

Our neutral explanation for the persistence of hetero-oligomerism is consistent with the observation that eight-membered rings of euryarchaeal chaperonins are made in some species from single protein subunits and in other species from two (Figures 2,3). Also, heterologous expression studies suggest that proper formation and function of homo-oligomeric chaperonins *in vivo* probably depends on the extent of functional (sequence) divergence between paralogous subunits. The α-only and β-only (homo-oligomeric) chaperonins from *Thermococcus* strain KS-1, in which the α and β paralogs are 80.6% identical, show ATPase activity and protein-folding ability when expressed in *Escherichia coli* [23]. In contrast, homo-oligomeric chaperonin complexes from *Sulfolobus sp.* S7 (α and β have only 55.5% identity) are unstable, prone to dissociation into monomers and show no ATPase activity [24]. In *P. occultum* (α and β are 61.8% identical), α-only and β-only thermosomes were microscopically indistinguishable from their native hetero-oligomeric counterparts, yet exhibit reduced thermal stability and are deemed only partly functional [25]. The remarkable evolutionary pattern of archaeal chaperonin subunits could provide a general framework for understanding the origin and evolution of hetero-oligomerism in multisubunit protein complexes.

Supplementary material

Supplementary material including a protein sequence alignment of archaeal chaperonins and additional methodological details is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

We thank members of the Sulfolobus solfataricus p2 genome sequencing project, D. Faguy, H.-P. Klenk and M. Schenk for *Sulfolobus* and *Desulfurococcus* genomic DNAs, A. Stoltzfus for mol2con.pl, O. Feeley for aa-dna-align.pl and Doolittle laboratory members for helpful discussion and critical review. This work was supported by a grant awarded to W.F.D. by the Medical Research Council (MRC) of Canada. J.M.A. is supported by an MRC studentship awarded to W.F.D. and by an MRC Doctoral Research Award. J.M.L. is supported by post doctoral fellowships from the MRC and NIH.

References

1. Bukau B, Horwich AL: **The Hsp70 and Hsp60 chaperone machines.** *Cell* 1998, **92**:351-366.
2. Ranson NA, White HE, Saibil HR: **Chaperonins.** *Biochem J* 1998, **333**:233-242.
3. Kubota H, Hynes G, Willison K: **The chaperonin containing t-complex polypeptide 1 (TCP-1). Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol.** *Eur J Biochem* 1995, **230**:3-16.
4. Willison KR, Kubota H: **The structure, function, and genetics of the chaperonin containing TCP-1 (CCT) in eukaryotic cytosol.** In *The Biology of Heat Shock Proteins and Molecular Chaperones.* Edited by Morimoto RI, Tissieres A, Georgopoulos C. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 1994: 299-312.
5. Phipps BM, Hoffmann A, Stetter KO, Baumeister W: **A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria.** *EMBO J* 1991, **10**:1711-1722.
6. Phipps BM, Typke D, Hegerl R, Volker S, Hoffmann A, Stetter KO, Baumeister W: **Structure of a molecular chaperone from a thermophilic archaebacterium.** *Nature* 1993, **361**:475-477.
7. Trent JD, Nimmesgern E, Wall JS, Hartl FU, Horwich AL: **A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1.** *Nature* 1991, **354**:490-493.
8. Willison KR, Dudley K, Potter J: **Molecular cloning and sequence analysis of a haploid expressed gene encoding t complex polypeptide 1.** *Cell* 1986, **44**:727-738.
9. Frydman J, Nimmesgern E, Erdjument-Bromage H, Wall JS, Tempst P, Hartl FU: **Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits.** *EMBO J* 1992, **11**:4767-4778.
10. Lewis VA, Hynes GM, Zheng D, Saibil H, Willison K: **T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol.** *Nature* 1992, **358**:249-252.
11. Rommelaere H, Van Troys M, Gao Y, Melki R, Cowan NJ, Vandekerckhove J, *et al.*: **Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and seven related subunits.** *Proc Natl Acad Sci USA* 1993, **90**:11975-11979.
12. Knapp S, Schmidt-Krey I, Hebert H, Bergman T, Jornvall H, Ladenstein R: **The molecular chaperonin TF55 from the thermophilic archaeon *Sulfolobus solfataricus*. A biochemical and structural characterization.** *J Mol Biol* 1994, **242**:397-407.
13. Kagawa HK, Osipiuk J, Maltsev N, Overbeek R, Quaite-Randall E, Joachimiak A, Trent JD: **The 60 kDa heat shock proteins in the hyperthermophilic archaeon *Sulfolobus shibatae*.** *J Mol Biol* 1995, **253**:712-725.
14. Nakamura N, Taguchi H, Ishii N, Yoshida M, Suzuki M, Endo I, *et al.*: **Purification and molecular cloning of the group II chaperonin from the acidothermophilic archaeon, *Sulfolobus* sp. strain 7.** *Biochem Biophys Res Commun* 1997, **236**:727-732.
15. Sensen CW, Charlebois RL, Chow C, Clausen IG, Curtis B, Doolittle WF: **Completing the sequence of the *Sulfolobus solfataricus* P2 genome.** *Extremophiles* 1998, **2**:305-312.
16. Ditzel L, Lowe J, Stock D, Stetter KO, Huber H, Huber R, Steinbacher S: **Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT.** *Cell* 1998, **93**:125-138.
17. Ellis MJ, Knapp S, Koeck PJ, Fakoor-Biniaz Z, Ladenstein R, Hebert H: **Two-dimensional crystallization of the chaperonin TF55 from the hyperthermophilic archaeon *Sulfolobus solfataricus*.** *J Struct Biol* 1998, **123**:30-36.
18. Liou AK, Willison KR: **Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes.** *EMBO J* 1997, **16**:4311-4316.
19. Lin P, Sherman F: **The unique hetero-oligomeric nature of the subunits in the catalytic cooperativity of the yeast Cct chaperonin complex.** *Proc Natl Acad Sci USA* 1997, **94**:10780-10785.
20. Lin P, Cardillo TS, Richard LM, Segel GB, Sherman F: **Analysis of mutationally altered forms of the Cct6 subunit of the chaperonin from *Saccharomyces cerevisiae*.** *Genetics* 1997, **147**:1609-1633.
21. Stoldt V, Rademacher F, Kehren V, Ernst JF, Pearce DA, Sherman F: **The Cct eukaryotic chaperonin subunits of *Saccharomyces cerevisiae* and other yeasts.** *Yeast* 1996, **12**:523-529.
22. Rademacher F, Kehren V, Stoldt VR, Ernst JF: **A *Candida albicans* chaperonin subunit (CaCct8p) as a suppressor of morphogenesis and Ras phenotypes in *C. albicans* and *Saccharomyces cerevisiae*.** *Microbiology* 1998, **144**:2951-2960.
23. Yoshida T, Yohda M, Iida T, Maruyama T, Taguchi H, Yazaki K, *et al.*: **Structural and functional characterization of homo-oligomeric complexes of alpha and beta chaperonin subunits from the hyperthermophilic archaeum *Thermococcus* strain KS-1.** *J Mol Biol* 1997, **273**:635-645.
24. Yoshida T, Yohda M, Suzuki M, Yazaki K, Miura K, Endo I: **Characterization of homo-oligomeric complexes of alpha and beta chaperonin subunits from the acidothermophilic archaeon, *Sulfolobus* sp. strain 7.** *Biochem Biophys Res Commun* 1998, **242**:640-647.
25. Minuth T, Frey G, Lindner P, Rachel R, Stetter KO, Jaenicke R: **Recombinant homo- and hetero-oligomers of an ultrastable chaperonin from the archaeon *Pyrodictium occultum* show chaperone activity *in vitro*.** *Eur J Biochem* 1998, **258**:837-845.
26. Adachi JM, Hasegawa M: *MOLPHY: Programs for molecular phylogenetics.* Tokyo: Inst. Stat. Math.; 1996.
27. Thompson DK, Daniels CJ: **Heat shock inducibility of an archaeal TATA-like promoter is controlled by adjacent sequence elements.** *Mol Microbiol* 1998, **27**:541-551.
28. Kubota H, Hynes GM, Kerr SM, Willison KR: **Tissue-specific subunit of the mouse cytosolic chaperonin-containing TCP-1.** *FEBS Lett* 1997, **402**:53-56.