The chaperonins: perspectives from the Archaea

P.A. Lund1, A.T. Large and G. Kapatai
School of Biosciences, University of Birmingham, Birmingham B15 2TT, U.K.

Abstract

Heat-shock protein (Hsp) 60 chaperones are almost ubiquitous and almost always essential. They can be divided on the basis of sequence homology into two broad types: group I (found in bacteria, mitochondria and chloroplasts) and group II (found in Archaea and the eukaryotic cytosol). Of the two, the group I chaperones are the better understood. Data on their structure, mechanism of action and cellular role will be briefly presented. The group II chaperones are less well studied. In eukaryotes they form large complexes with 8-fold symmetry containing eight different subunits, all of which are essential. They appear to have a major role in the folding of actin and tubulin, although they may also act on other substrates. No crystal structures are available for these complexes. The situation in the Archaea is simpler, with organisms containing between one and three genes for these chaperones. A 2.6 Å structure exists for one archaeal group II chaperone complex. Some progress has been made in defining the reaction cycle of the archaeal group II chaperones and this has shown that they have some properties distinct from the group I chaperones. To date, the in vivo role and importance of the archaeal group II Hsp60 chaperones has not been determined.

We have now shown that in the halophilic archaean Haloferax volcanii not all the genes for these proteins are essential. Further analysis of these proteins in the Archaea should be very productive in yielding more information about these important chaperones and their cellular functions.

Introduction

Molecular chaperones are a diverse class of proteins that help various cellular components reach an active and final state that they could not reach in the chaperone’s absence. The term chaperone is now applied to many proteins that play a role in ensuring that a range of substances – from metal ions to nucleic acids – are helped to find their way to a place or a conformation where they can display their correct cellular function. The early use of the term chaperone, however, focused on proteins that helped other proteins to fold. Even in this more limited context, there are several distinct chaperone families, found in nearly all cells. Extensive studies on these has shown that they exhibit a wide range of properties, binding and acting on different substrate proteins by distinct mechanisms.

The best studied chaperones are the heat-shock protein (Hsp) 60 proteins, a group related by sequence and found in nearly all organisms, the only exceptions known to date being two mycoplasma species. These are also known as the chaperonins [1]. Hsp60 genes are highly conserved, indicating a central role in cell viability, to the extent that loss of the hsp60 gene (better known as groEL) from Escherichia coli can be complemented by a mitochondrial homologue [2]. Sequence analysis shows clearly that the chaperonins fall into two major groups, one (referred to as group I) found in bacteria, mitochondria and chloroplasts, and the other (group II) in the cytosol of eukaryotes and in the Archaea. Biochemical and structural studies on these two classes have revealed both strong parallels and significant differences in their properties, which will now be discussed in more detail.

The group I chaperonins of bacteria, mitochondria and chloroplasts

The mycoplasma Ureaplasma urealyticum was the first organism to be discovered that did not possess the gene for a group I chaperonin [3]; the vast majority of organisms contain at least one gene for this protein. The groEL gene in E. coli has long been known to be essential [4] and this is also the case for all other cases so far tested. Intriguingly, a significant number of bacteria contain two or more homologues of this gene. Where this is so, it is generally the case that one of them has been found to be essential. Highly conserved homologues are also found in mitochondria and chloroplasts, and again these are known to be essential in all cases tested [5,6].

These proteins play a key role in helping many other proteins to fold into their active conformation. The complete set of substrates for the GroEL protein in E. coli has not been fully defined but it includes several metabolic enzymes, RNA polymerase subunits, and other proteins involved in transcription and translation [7]. If the levels of GroEL protein are lowered, wholesale misfolding of proteins is seen [8–10]. Expression of the GroEL protein is also strongly induced by heat shock, the signal for this being the presence of unfolded proteins in the cell [11]. E. coli cells depleted of all Hsps show very poor growth and extensive protein aggregation at temperatures above 20°C, but growth and protein solubility can be restored simply by elevated expression...
of the GroEL protein, plus its cofactor GroES, showing the central role of these chaperone proteins [12,13].

GroEL has a remarkable structure. It consists of two rings, each of seven subunits, stacked back to back. The X-ray structures have been solved for the protein complex in both an unliganded state and with ADP bound [14,15], and these, together with cryo-electron microscopy studies [16–19], have revealed the presence of very large domain movements during the reaction cycle. Each GroEL subunit consists of three domains, referred to as equatorial, intermediate and apical (see Figure 1). The equatorial domain binds nucleotides, and provides the ring–ring contacts and most of the contacts between adjacent subunits. The intermediate domain links the equatorial domain and the apical domain, and shows large movements in the ATP-driven folding cycle. The apical domain binds substrate and GroES and also shows substantial movement during the folding cycle. In outline, the GroEL protein acts as a two-stroke engine, with one ring binding and then hydrolysing ATP and simultaneously acting as the place where protein folding occurs, followed by an identical reaction on the opposite ring. The unfolded protein substrates bind initially to a hydrophobic patch at the top of the apical domain, and are then displaced by GroES competing for the same sites. The GroES cofactor also has a role in capping the ring to which folding protein is initially bound, thus sequestering it within the folding cavity [20–24].

The chief role of GroEL appears to be to help proteins to fold that would otherwise aggregate. It does this by providing an environment in which they can fold, namely the interior of the GroEL protein complex. Once capped by the GroES cofactor, this cavity is sufficiently large to accommodate an unfolded protein of around 60 kDa [14]. The folding of the protein may be assisted by the hydrophilic nature of the walls of the cavity [25], which may hasten the burying of hydrophobic residues which is the first stage in protein folding and which, once completed, significantly lowers the propensity of proteins to aggregate.

The group II chaperonins of eukaryotes

Early searches revealed a set of cytoplasmic proteins that had low but significant homology with GroEL [26,27]. Investigation of these in more detail revealed that there are eight such proteins, each encoded by a separate gene. These have been referred to by a variety of names but are now usually called the chaperone-containing T-complex polypeptide 1 (CCT) proteins. They are present at low levels in eukaryotic cytosol, and they are not heat-shock inducible. All eight of the genes are essential in yeast [28].

Structural studies on the proteins from eukaryotes show that they also assemble into a double ring structure, reminiscent of the group I chaperonins but with 8-fold, rather than 7-fold, symmetry [29,30]. The position of each subunit in the ring is invariant, with all eight proteins being represented in each ring in a specific order [31]. There is no crystal structure available for the eukaryotic CCT complex, but cryo-electron microscopy studies have revealed its structure in some detail. A clear difference between the GroEL and CCT complexes is the absence of the GroES cofactor in the latter. However, cryo-electron microscopy studies show the presence of a lid on the CCT complex which is actually part of the CCT...
Table 1 Numbers of chaperonin genes present in 16 Archaea whose complete genomes were available on 28th January 2003

Organisms with a * also contain complete groEL and groES genes.

<table>
<thead>
<tr>
<th>Organism name</th>
<th>Number of Group II chaperonins</th>
<th>Organism name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeropyrum pernix K1</td>
<td>2</td>
<td>Crenarchaeota</td>
</tr>
<tr>
<td>Sulfolobus solfataricus P2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sulfolobus tokodai strain 7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Pyrobaculum aerophilum IM2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Archaeoglobus fulgidus DSM4304</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Halobacterium sp. NRC1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Methanobacterium thermoautotrophicum delta H</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Methanococcus jannaschii DSM2661</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Methanosarcina mazei Geoe1*</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Methanosarcina acetivorans C2A*</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Methanopyrus kandleri AV19</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pyrococcus horikoshii Shinkai OT3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pyrococcus abyssi GES</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pyrococcus furiosus DSM 3638</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Thermoplasma acidophilum DSM 1728</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Thermoplasma volcanium GSSI</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

protein, and the X-ray structure of archaeal CCT protein confirms this (see Figure 1, and below).

The main substrates of CCT in eukaryotes are the cytoskeletal proteins actin and tubulin [32–34]. Other proteins have also been shown to bind to CCT, but it is not clear how many of these are genuine in vivo substrates. Other protein cofactors are required for CCT to bind and fold actin and tubulin [35]. There is good evidence that the invariant positions of the subunits around the ring are reflected in precise positioning of bound substrate [36,37].

**Group II chaperonins in Archaea**

The archaeal chaperonins are also members of the group II family, although two cases are also known where a group I homologue is present in the genome sequence (see Table 1). The proteins are strongly induced by heat shock. Indeed, in the thermophilic archaeon *Sulfolobus shibatae*, the chaperonin complex is the most abundant protein in heat-shocked cells, comprising up to 40% of the total cellular protein [38]. For this reason it is referred to as the thermosome, although the designation CCT is also used by analogy with the eukaryotic proteins.

Genome analysis of Archaea shows that they contain between one and three cct genes, the exception being *Methanosarcina acetivorans*, which contains five cct genes. Phylogenetic analysis suggests that cct gene duplication and gene loss is occurring all the time in archaeal lineages, and that the presence of multiple cct genes is a consequence of this and has no selective significance [39,40]. Our recent analysis of cct gene function in *Halofex volcanii* tends to support this view (see below). However, it is interesting that the *Methanosarcina* species that contain large numbers of cct genes also possess group I chaperonins, presumably acquired by horizontal gene transfer from bacteria [41]. It may be that these organisms have particularly specialized or stringent requirements for chaperonin function.

Nothing is yet proven about the in vivo role of the CCT proteins in Archaea, although their strong induction by heat makes it likely that they possess a chaperone function, and such a function has indeed been demonstrated in vitro [42]. The in vivo substrates are not known. As Archaea do not contain actin or tubulin, definition of the substrates of the CCT complex may improve our understanding of how the eukaryotic CCT proteins and the eukaryotic cytoskeleton evolved. An interesting observation is that the CCT complex from *Sulfolobus solfataricus* is involved in binding and processing of 16 S RNA and may thus have a role in ribosomal biogenesis [43].

The X-ray structure of the thermosome from *Thermoplasma acidophilum* has been solved at 2.6 Å resolution [44]. This is useful both in aiding our understanding of the archaeal CCT proteins and also for the insights it gives into the more complex eukaryotic homologues. The protein assembles as two rings with 8-fold symmetry, with the two CCT proteins present (subunit α and subunit β) alternating around the ring. The structure of the subunits shows close similarity to that of the GroEL subunits (see Table 1), the major difference being the presence of the large helical loop protruding from the apical domain, referred to above. This loop may serve as a lid for the complex. CCT proteins can also assemble with 9-fold symmetry in the cases where there are three proteins present [45]. Whereas 9-fold symmetry is an obvious solution to the problem of how to assemble three proteins into a ring...
complex, it must be borne in mind that the presence of three genes does not mean that all three proteins will be expressed at the same level. Indeed, in *H. volcanii* (which has three *cct* genes), we have shown this not to be the case [46].

Crystallographic and cryo-electron microscopy data give somewhat conflicting results about the conformational states of the archaeal CCT complex, but a study using small-angle neutron scattering suggests that the thermosome of *T. acidophilum* exists in an open state when a nucleotide is bound, but when the ATP is hydrolysed to ADP and P, the complex is transiently closed before the phosphate is released [44]. How this correlates with the folding-active state of the complex is currently unknown.

**Genetic analysis of chaperonin function in Archaea**

Phylogenetic data predict that when there are multiple *cct* genes in Archaea, they are likely to be redundant to some extent since they arise as the result of fairly frequent gene duplications and gene losses. An alternative prediction is that, by analogy with the eukaryotic CCT proteins, the individual proteins may need to be precisely positioned to show particular functions. The archaean *H. volcanii* contains three *cct* genes [47,48], and is closely related to *Halobacterium* sp. NRC-1, which contains only two [49]. As *H. volcanii* is genetically tractable, we have been able to test this directly by attempting to make knockout mutations in all three genes. This was done by individually introducing the cloned genes containing a cassette with a selectable marker (methylin resistance) either on a non-replicating plasmid or on a linear piece of DNA, and selecting for recombinants. Resistant colonies were subsequently screened to test whether the targeted gene had indeed been lost. Our data show that both the *cctI* and the *cct3* genes can be knocked out, but so far our attempts to knock out *cct2* have failed. The organisms carrying the gene knockouts do not show any obvious physiological difference from the wild-type, which supports the hypothesis that the genes are functionally redundant. We have purified protein complexes from both knockout strains (see Figure 2) and measured their ATPase activity. Results show that the loss of the CCT3 protein has no effect on the activity of the CCT complex but loss of CCT1 leads to a reduction by approx. 50% in its ATPase activity (results not shown). Thus it is clear that, unlike their eukaryotic counterparts, the archaecal CCT proteins are not all essential. The reason for the apparent essential nature of the CCT2 protein in this organism is currently under investigation.

**Figure 2** | Purified CCT protein complex from *H. volcanii* demonstrating selective loss of CCT1 and CCT3 proteins

Source: wild-type *H. volcanii* (lane 1), *cct1* knockout (lane 2) and *cct3* knockout (lane 3). Bands were assigned and proteins purified as described in [46]. Molecular-mass markers (lane 4) are 97.4 and 66.2 kDa respectively. These markers overestimate the mass of the CCT proteins, which were retarded in the gel because of their high negative charge [50].

**References**

27 Ellis, R.J. (1990) *Science* 250, 954-959

Received 31 January 2003