Archaeal chaperonins

Andrew T. Large, Peter A. Lund

School of Biosciences, University of Birmingham, Birmingham B15 2TT, United Kingdom

TABLE OF CONTENTS

1. Abstract
2. Introduction
   2.1. The molecular chaperone concept
   2.2. Chaperonins
       2.2.1. Overview
       2.2.2. Nomenclature
       2.2.3. Group I chaperonins. The view from bacteria
       2.2.4. Group II chaperonins. The view from eukaryotes
       2.2.5. Why study the archaeal Group II chaperonins? Some key questions
3. Cellular roles of archaeal chaperonins
   3.1. Evidence for a role in protein folding
   3.2. Other potential roles of archaeal chaperonins
   3.3. The roles of multiple chaperonin proteins in archaea
   3.4. Group I chaperonins in archaea
4. Structure, function and mechanism
   4.1. Crystal structures of archaeal thermosomes
   4.2. Studies on conformational changes
   4.3. ATPase activity and allostery
   4.4. Studies on protein binding/folding
   4.5. What do we know about the substrate binding sites?
   4.6. The role of prefoldin in thermosome-mediated protein folding
5. Conclusions and prospects
6. Acknowledgements
7. References

1. ABSTRACT

Chaperonins are ubiquitous and essential protein folding machines. They have a striking structure, with two rings of seven, eight, or nine protomers forming a “double doughnut” complex, with the cavity in each ring being the likely site for protein folding to take place. The group I chaperonins, found in bacteria and the organelles descended from them, are well characterised in terms of their structure, mechanism, and in vivo roles. The group II chaperonins, found in eukaryotic cytosol and archaea, are less well understood. In this review, we focus on what is known about the archaeal chaperonins, both in terms of their in vivo role and their structure/function relationships, in order to more fully understand their significance in archaea and as models for chaperonin function in general.

2. INTRODUCTION

2.1. The molecular chaperone concept

Some denatured proteins can, under suitable conditions, return spontaneously to their soluble folded form when the denaturant is removed, demonstrating that in aqueous solutions, the folded state is more stable than the unfolded state (1). However, protein folding inside the cell is more complex than the process modelled in these experiments, because (a) proteins are synthesised vectorially on ribosomes, (b) all proteins have to fold under the same conditions, (c) in many cases proteins need to be kept unfolded in order to cross membranes, and (d) protein aggregation into non-functional complexes is an ever-present danger in the highly crowded conditions that exist inside the cell. The molecular chaperone concept emerged
Archaeal chaperonins

Figure 1. Structure of the thermosome of T. acidophilus, in the closed conformation. In both the top and bottom ring, two sub-units have been removed to show the cavity that exists inside the two rings. Alpha and beta sub-units alternate around both rings.

Following the discovery of several proteins which bind other proteins and assist their folding to their final active form, and it bridges the conceptual gap between what goes on in simple in vitro protein folding experiments and in the complex and crowded milieu of the cell (2-4). A useful broad definition of molecular chaperones is that they are proteins that "prevent or reverse incorrect interactions which can occur when reactive macromolecular surfaces are transiently exposed to the intracellular environment" (5). This definition modifies the concept of spontaneous folding by showing that under certain conditions this process requires assistance to prevent the formation of off-pathway intermediates such as aggregates. The action of chaperones is thus on the kinetics rather than the thermodynamics of the protein-folding process. Molecular chaperones do their job by binding to regions of polypeptide chains that would otherwise be likely to aggregate or fold incorrectly in some way. This binding is transient, so that molecular chaperones are not part of the final active protein, and is mediated in different ways depending on the particular interaction and the particular molecular chaperone concerned.

Molecular chaperones are a broad group, and as more are discovered it becomes harder to make meaningful generalization about them, because they vary in size, structure, mechanism, occurrence, role, and importance. However, a number of chaperones which occur in a large proportion of all organisms are known, and these can be classified by sequence similarities into a relatively small number of families. They are often referred to as heat shock proteins (HSPs), as many molecular chaperones are also induced by heat shock, but it is important to remember that many chaperones are not HSPs, and many HSPs are not chaperones. The HSP60s or chaperonins are one family of molecular chaperone that have been intensively studied, and it is this particular family that will be the subject of this review.

2.2. Chaperonins

2.2.1. Overview

Chaperonins are related by sequence similarity, and they have several properties of particular interest. First, they are found in all organisms studied to date with the exception only of a few Mycoplasmas. Second, they are essential in all cases tested, which is not true for any of the other known molecular chaperones. All appear to fold a subset of proteins, some of which are essential for cell function, acting as a key part of a complex network of chaperones that act on many cellular proteins both during and after translation (see Figure 2 in ref. 6, and Figure 1 in ref. 7). Third, many in vitro studies show that chaperonins can bind to a range of unfolded or partially folded substrate proteins and refold them to an active form under conditions where they fail to spontaneously refold. This process always requires nucleotide binding, with ATP being the preferred nucleotide; hydrolysis of the nucleotide is also usually required. They have a remarkable structure: all of them (with some possible exceptions (8)) exist as large ring-shaped oligomers, with several sub-units in each ring enclosing a central cavity (see Figure 1, which shows the structure of an archaeal chaperonin). In most cases, the active structure is a double ring, and thus contains two cavities. Much work over the past two decades has been aimed at understanding the mechanism of these proteins in terms of their structure; more recently, the emphasis has broadened to include learning more about their precise functions within the cell. A large amount of this work has been on the E. coli chaperonin GroEL, and because the chaperonins can be divided very clearly into two phylogenetic types, this work will be briefly discussed below in section 2.2.3 separately from what is known about the eukaryotic and archaeal representatives of this family.

The two phylogenetic groups of chaperonin emerge from alignment of chaperonin sequences from the three domains of life; they are referred to as type I and type II or, more usually, as group I and group II. Group I chaperonins are found in all bacteria (apart from some Mycoplasmas), chloroplasts (9), and mitochondria (10), and related organelles such as hydrogenosomes and mitosomes (11). Their phylogenetic relatedness in bacteria and organelles is one of the arguments supporting the endosymbiotic hypothesis for the origin of eukaryotic cells, and abundant evidence shows that they are also functionally related (12, 13). Group II chaperonins are found in the eukaryotic cytosol and archaea, with the archaeal proteins being phylogenetically distinct from the eukaryotic ones but significantly more similar to them than they are to the group I chaperonins.

As expected from the significant level of sequence similarity between group I and group II chaperonins (typically of the order of 40%), the two types of chaperonin show many structural similarities. Both generally form double ring structures, with each ring made...
Archaeal chaperons

What are the roles of these intriguing protein complexes, and how does the structure explain the mechanism whereby they carry out these roles? Before we consider the archaeal chaperons that are the main focus of this review, it will be useful to look at the chaperons from the perspectives both of bacteria and eukaryotes, since this will give us a conceptual framework on which to base questions about the archaeal chaperons, both in the context of their specific role in archaea and in the broader context of what they can tell us about chaperons in general.

2.2.2. Nomenclature

“Chaperonin” is the term applied to all proteins with shared homology to the archetypal member of the class, the GroEL protein from *E. coli*. Because of the many different guises under which chaperons have been studied, the nomenclature for these proteins is somewhat complicated. The bacterial proteins are often referred to as GroEL (which strictly speaking should be reserved for the *E. coli* protein only) or Cpn60 (for chaperonin) (18). The proteins with which they function are referred to as cochaperons, GroES (in *E. coli*), or Cpn10. The mitochondrial and chloroplasts Cpn60 homologues are usually called Hsp60 and Rubisco sub-unit binding protein (alpha and beta) respectively. The mitochondria cochaperonin is usually called Hsp10; in chloroplasts the cochaperonin situation is more complex as at least two homologous cochaperonin proteins appear to exist, usually called Cpn10 and either Cpn20 or Cpn21. The eukaryotic group II chaperonin was originally called TCP-1, which refers really to one sub-unit of the complex (19); it is now more often called either TRiC (for “TCP-1 containing ring complex”) or CCT (for “chaperonin containing TCP-1”). The archaeal chaperons are most frequently referred to as thermosomes (20), but the terms TF55 (21), rosettasomes (22), archaeosomes (23), and CCT (24) are also used.

In this article we adopt the convention of referring to the eukaryotic chaperons as CCT proteins, the archaeal chaperons as thermosomes, and we reserve the nomenclature Cpn60/Cpn10 for the group I chaperons and cochaperons. The terms GroEL and GroES we will use only for the *E. coli* proteins.

2.2.3. Group I chaperons. The view from bacteria

GroEL is the archetypal bacterial chaperonin. The gene for this protein (together with that of its cochaperonin GroES) was originally discovered in a mutagenic screen for genes required for the growth of bacteriophage lambda (25, 26), but it soon became clear that both have a key role in the *E. coli* cell. Both groEL and groES are essential under all conditions studied (27). Over-expression of GroEL and GroES partially suppresses the loss of viability and decreased protein folding caused by lowered expression of all the heat shock proteins in the cell (28, 29), improves the folding of some heterologously expressed proteins (30), and suppresses the phenotypes of a large number of temperature-sensitive mutants in a wide range of proteins (31). These studies, together with many *in vitro* experiments, confirm that these two proteins can

Figure 2. Cartoon representation of a single subunit of the thermosome from *T. acidophilum*, colour coded to show the domain structure of the protein. Equatorial domain – red; intermediate domain – blue; apical domain – yellow; helical protrusion – green.
Archaeal chaperonins

act to assist the folding of at least some proteins to their native states.

Defining the in vivo substrates of any chaperonin is challenging. The fact that a protein can bind to a chaperonin in vitro is no proof that it does so in vivo, and even if an in vivo association can be demonstrated (for example, by cross-linking, co-immunoprecipitation, or indirectly through genetic studies) this does not prove that the associated protein is actually assisted in its folding by the chaperonin. A satisfactory demonstration requires evidence of in vivo activity of the chaperonin on the particular substrate under study, backed up by in vitro data to show that the chaperonin can indeed fold the substrate under conditions where it does not normally fold. Experiments using co-immunoprecipitation of protein bound to GroEL, followed by proteomic analysis, have shown that GroEL binds a significant sub-set of cellular proteins including several which are essential for growth (32, 33). Some of these have been confirmed by in vitro study to absolutely require GroEL and GroES to fold to their active state (34). Whether this is a complete list of GroEL substrates has been questioned (35, 36), and it is interesting to note that similar experiments with two other bacterial Cpn60 proteins have come up with lists of chaperonin substrates in those organisms which do not substantially overlap with those determined for E. coli GroEL (37, 38). It seems however certain from the above studies that Cpn60 and Cpn10 are essential because they are required for the folding in vivo of several proteins which are themselves required for growth.

An important question is what makes some proteins require GroEL for folding and others not. The substrates identified in the studies cited above are relatively enriched in the (alpha-beta) triosephosphate isomerase (TIM) barrel domain (34), but not all proteins with this domain are GroEL substrates. It may simply be the case that GroEL preferentially binds proteins with exposed hydrophobic residues and thus selects from the population of newly synthesized (or heat shock damaged) proteins those which are most in need of assistance to refold. Theoretical analysis shows that GroEL proteins tend to have on average a lower propensity to fold but a higher tendency to form aggregates than other E. coli proteins (39).

The mode of action of GroEL has been intensively studied (and extensively reviewed; see for example references 36, 40, and 41 for recent reviews) and a great deal is now known about the mechanism. The reaction cycle, briefly, involves binding of the unfolded protein substrate and ATP to one end (the cis end) of the double ring complex, followed by binding of GroES which has the effect of capping the cavity in the ring while displacing the substrate into this cavity. The substrate is then free to fold within the cavity until ATP and substrate, followed by GroES, bind to the opposite (trans) ring, and this cannot happen until the ATP in the cis ring has hydrolysed, which is a relatively slow reaction. Binding of substrate in the trans ring causes the release of GroES and substrate from the cis ring, and the whole process happens again on the trans ring of the complex. A consensus on exactly how this cycle assists protein folding does not yet exist; it is possible that GroEL acts in the cell in more than one way, and that different experimental approaches highlight these differences. Some data support the hypothesis that a major role of GroEL is to unfold proteins which have become misfolded, a process which is known to occur and is probably driven by the ATP-dependent movement of the apical domains (42-44). Other data supports the Anfinsen cage model (45), which envisages the cavity in the ring, turned into a cage by the binding of GroES, as the key structure in the cycle. The cage could have a purely passive role on folding, in that by allowing the protein to fold in an environment where it is protected from other folding proteins, it reduces the danger of aggregation between two unfolded chains. Alternatively, its role may be more active, in that the combination of close confinement and the nature of amino-acyl residue side chains on the internal wall of the cavity may separately or together help favour the folded state (46). The actual mechanism may share features of the different models described above, and may vary for different substrate proteins.

What is clear is that under normal circumstances the reaction is timed by the ATP hydrolysis cycle. GroEL is a weak ATPase which shows significant allostery, with binding and hydrolysis being positively co-operative within a single ring but negatively co-operative between rings (47, 48). The effect of the positive co-operativity may be to favour the rapid release of bound protein into the cavity from multiply bound positions on the apical domain, since substrate proteins bind less tightly to the ATP-bound form of GroEL. The effect of the negative between-ring co-operativity is to block the binding of ATP and substrate to the trans ring until ATP hydrolysis on the cis ring has occurred, which in turn allows the encapsidated substrate protein time to fold in the cavity. Binding and folding thus alternate between the two rings in a manner timed by ATP hydrolysis (49, 50).

Before moving on to consider the group II chaperonins, it should be stressed that that focusing overmuch on E. coli GroEL runs the risk of losing important information about group I chaperonins in general. For example, other group I chaperonins may be able to function as single rings (51), and there is evidence that GroEL homologues in other bacteria have important roles which may be in addition to or even distinct from their roles in protein folding, including mediating attachment of bacteria to other cells and cell signaling (reviewed in 52, 53). In the same way, although work on group I chaperonins has been very illuminating in understanding how the chaperonins may function, caution must always be taken in extrapolating from group I to group II chaperonins.

2.2.4. Group II chaperonins. The view from eukaryotes

The group II chaperonins found in eukaryotes form double ring structures with eight sub-units in each ring, and as noted above they have the same basic fold and domain structure as the group I chaperonins. However, unlike the group I chaperonins, they consist of eight different proteins which are always found in the same
Archaeal chaperonins

position relative to each other in the two rings, and are always in the same phase with the second ring (54, 55). Phylogenetic analysis of these different sub-units shows very deep branching, indicating that that functional specialization of the individual sub-units in the eukaryotic group II chaperonin was a very early event in the evolution of the eukaryotic cell, and that this specialization, moreover, must reflect the involvement of the chaperonins in some aspect of eukaryotic cellular life which is essential to all eukaryotes (56, 57). This is also demonstrated by the fact that the genes for all eight sub-units of the eukaryotic CCT are essential in S. cerevisiae (58), and knock-downs in the genes in C. elegans using RNAi generally results in embryonic lethal phenotypes (www.wormbase.org; also see 59).

It is tempting to speculate that the role of CCT in eukaryotes is the same as that of GroEL in E. coli. However, it is striking that whereas the group I chaperonins are generally strongly induced by heat shock (in both bacteria and in mitochondria), the eukaryotic group II chaperonins are not; moreover their cellular abundance is only about 10% of that of the group I chaperonins (36). If they do indeed help protein to refold, this suggests a more limited range of substrates.

As noted above, it is difficult to identify chaperone substrates unequivocally, and such an identification needs data support from both in vitro and in vivo experiments. On this basis, the two major substrates of the eukaryotic CCT proteins appear to be actin and tubulin (59-62). Mutations in the cct genes lead to cytoskeletal and other defects in a variety of organisms (58, 63, 64), supporting the identification of actin and tubulin as CCT substrates. Interestingly in the light of the fixed arrangement of the CCT protomers in the ring structure, it has been shown that both actin and tubulin take up specific positions on the CCT ring when they bind, indicating that they interact with particular individual subunits in CCT (65, 66). A number of additional substrates have also been found, with varying degrees of experimental support, including G-alpha transducin, cyclin E, CDC20, and others (67). As with GroEL it is not easy to discern any common feature between these different proteins, although it has been noted that the majority are themselves components of oligomers, and some appear to be enriched in WD repeats (68). One important protein that interacts (not as a substrate) with eukaryotic CCT is a hexameric protein complex called prefoldin, which has been shown to have a role in the CCT-assisted folding of actin and tubulin, probably by sequestering these proteins after translation and delivering them to CCT (69, 70). This is discussed further below in section 4.6.

Eukaryotic CCT protein is a large oligomeric structure containing two rings of eight unique proteins each in a fixed position with respect to its partners and with respect to the proteins on the opposite ring (54, 55, 58, 71). No crystal structure is available for this complex but cryo-electron microscopy and single particle analysis shows that the domain structure seen in GroEL and in archaeal thermosomes is also seen in the eukaryotic protein (17). The apical domain shows significant conformational movement in the presence of ATP, with a marked asymmetry induced between the two rings (72). This is very reminiscent of the situation in GroEL, where ATP and GroES binding induces significant conformational changes, increasing the size of the cavity.

ATP hydrolysis also resembles that seen in GroEL in that there is within-ring positive and between-ring negative co-operativity (73), pointing towards a common mechanism of action of the group I and group II chaperonins. However, there is a significant difference: in the eukaryotic CCT, hydrolysis of ATP is sequential rather than concerted in a given ring (74), with hydrolysis spreading from one sub-unit to the adjacent ones. If the concerted hydrolysis of ATP seen in group I chaperonins is important for the simultaneous release of bound substrate from all sub-units, this result with group II chaperonins suggests that a more step-wise release of substrates may occur, which may be important in multi-domain proteins. It has been proposed that substrates bind to CCT with the lid open, and that the lid subsequently closes and encapsidates the substrates in a folding cavity in a way directly analogous to GroEL (75).

2.2.5. Why study the archaeal Group II chaperonins?

Some key questions

There are two compelling reasons to study chaperonins in archaea. First, they are of interest in their own right, in helping us to understand the process of protein folding in archaea. Second, they may hold the key to understanding not only more about chaperonins in general but also the whole process of cellular evolution and the emergence of the eukaryotic domain.

The first point is best illustrated by the fact that, while the archaeal chaperonins are closer phylogenetically to the eukaryotic chaperonins than they are to the bacterial ones, they have a number of features in which they resemble the bacterial proteins. In particular, they are mostly heat shock induced, which the eukaryotic proteins are not. For some archaea, they are almost the only protein synthesized under heat shock conditions (21). Second, although they generally have the same eight-fold symmetry as the eukaryotic proteins, the eight sub-units are not encoded by eight separate genes but by between one and 5, as shown (for all archaeal species for which a complete genome sequence is available) in Table I. The fact that many archaea thrive under conditions which are inimical to the growth of other organisms implies the existence of specialized evolutionary mechanisms to cope with these conditions. High temperature and high salinity are two obvious examples of conditions that will have profound effects on protein folding, and it is reasonable to look at the chaperonins to see how organisms have countered these effects.

Archaeal thermosomes are also excellent models to study general properties of the group II chaperonins. The archaeal protein complexes contain fewer unique sub-units than their eukaryotic counterparts, which simplifies the interpretation of kinetic data. They are easier to prepare
Archaeal chaperonins

in large quantities for use in biochemical and biophysical work, either from their host organism or after over-expression in E. coli. Good quality crystal structures are available. Lack of good genetic tools has to some extent hampered their study to date, but new tools for archaea are being developed all the time and this no longer represents a serious bottleneck.

It is striking that key substrates of the eukaryotic chaperonins are the cytoskeletal proteins. The possession of an organized internal cytoskeleton is an important property of eukaryotic cells (though of course it is now recognized that prokaryotes also have cytoskeletal structures), and is a pre-requisite for phagocytosis (76), which itself is required in many of the models for the evolution of modern-day eukaryotes by endosymbiosis (77). The ability to fold cytoskeletal proteins, which required evolution of precursors of modern-day eukaryotic CCT complexes, must thus have been a key early stage in the evolution of eukaryotic cells, consistent with the deeply branching phylogeny for eukaryotic CCT proteins referred to above. The study of the interactions between thermosomes and the known archaeal homologues of actin and tubulin may therefore be very informative in trying to understand the early stages of eukaryotic cell evolution.

3. CELLULAR ROLES OF ARCHAEAL CHAPERONINS

3.1. Evidence for a role in protein folding

All archaea so far studied contain at least one gene encoding the thermosome, and most have two or more. The actual numbers for all species of archaea for which a complete genome sequence is currently (February, 2008) available are shown in Table 1. Given the essential nature of chaperonins in eukaryotes and bacteria, it can be confidently predicted that they will also be essential in archaea. To date this has been experimentally tested in only one archaea: the halophile Halofex vulcanii, which is one of the more genetically tractable archaea (78). This organism possesses three thermosome genes, referred to in this organism as cct1 to cct3, and genetic studies have shown that while their function is indeed indispensable for growth, two of the three possible combinations of double knockouts of cct genes can be constructed without significant loss of viability under normal growth conditions (79). The presence of multiple genes leads to the obvious question of what the different thermosome proteins are doing in the archaea: do they have significantly different functional roles in the cell, or do they show partial or complete redundancy, as implied by these genetic studies? This question can be addressed in various ways, and is explored further below. But first, we must consider the question of what is known from direct experiments about the role of the archaeal chaperonins.

By analogy with bacterial and eukaryotic systems, it is generally assumed that the chaperonins in archaea assist in the folding of a sub-set of cellular proteins, the identity of which is not currently known. The essential nature of the thermosome in H. vulcanii is consistent with such a role. Several in vitro experiments have shown that purified archaeal chaperonins can indeed bind to unfolded proteins to prevent their aggregation, and in a smaller number of cases, complete assisted refolding of selected substrates has been reported (see Table 2, and further discussion in Section 4.4). In most cases investigated, the archaeal chaperonins have been shown to be induced by heat (21, 24, 80-87), as well as other treatments including high arsenic challenges and reduced pressure in piezophiles that may induce protein misfolding (88, 89). In some cases, they are almost the only genes expressed at any significant level after heat shock, and theoretical studies based on codon usage suggest they are likely to be among the most highly expressed archaeal proteins (90). Again this is entirely consistent with but not proof of a role in protein folding. Definitive proof of such a role must await the identification of particular substrate proteins and the demonstration that in the absence of chaperonin, these substrates fail to fold in vivo.

3.2. Other potential roles of archaeal chaperonins

Other roles have been suggested for archaeal chaperonins. For example, it has been shown that at sufficiently high protein concentrations, the thermosome from Sulfolobus shibatae can form filamentous structures in vitro. These concentrations are exceeded in the archaeal cytoplasm, and electron microscopy has detected such structures in Sulfolobus cells. Their physiologically significance if any is unknown, but the intriguing possibility that they may play a cytoskeletal role has been raised (91). Evidence has also been provided to show that the same protein binds to membranes in vivo and to liposomes in vitro, suggesting a possible role in the maintenance of membrane integrity. Such roles could prove to be a consequence of the normal functioning of chaperonins in protein folding or may indicate quite different aspects of chaperonin function that are not yet understood (92).

The S. solfataricus thermosome has been shown to bind specifically to the 16S rRNA, and to have an activity which cleaves it at a precise position, associated with the maturation of the ribosome (93). In addition, the thermosome has been shown to be associated with a complex which is thought to be the archaeal exosome (a protein complex responsible for RNA processing and degradation; 94), although the association is not tight. The possibility thus exists that chaperonins have a role in RNA processing as well as in protein folding, and indeed this has been argued for group I chaperonins as well (95, 96); further analysis of this phenomenon is needed.

3.3. The roles of multiple chaperonin proteins in archaea

Even though many archaea have more than one chaperonin gene, several lines of evidence suggest that the proteins they encode do not possess significantly different functions. First, several archaea have only a single chaperonin gene, showing that the presence of multiple chaperonins is not a general requirement for archaea. Second as has already been noted, two of the three cct genes in H. volcanii were capable of supporting growth in
Archaeal chaperonins

Table 1. Numbers of group II chaperonin genes (and, when present, group I chaperonin genes) in all archaea for which a complete genome sequence is available

<table>
<thead>
<tr>
<th>Organism</th>
<th>Number of chaperonin genes</th>
<th>Type (E, C, N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeropyrum pernix K1</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Archaeoglobus fulgidus DSM 4304</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Haloarcula marismortui ATCC 43049</td>
<td>4</td>
<td>E</td>
</tr>
<tr>
<td>Halobacterium sp. NRC-1</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Haloquadratum watsoni DSM 16790</td>
<td>3</td>
<td>E</td>
</tr>
<tr>
<td>Hyperthermus butylicus DSM 5456</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Metallophaga sedula DSM 5348</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Methanocarchaeobacter smithii ATCC 35061</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Methanocaldococcus jannaschii DSM 2661</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Methanococcosidicarchaeon burtonii DSM 6242</td>
<td>3</td>
<td>E</td>
</tr>
<tr>
<td>Methanococcus aerolevus Nankai-3</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Methanococcus maripaludis C5</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Methanococcus maripaludis C7</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Methanococcus maripaludis S2</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Methanococcus vanmiedii SB</td>
<td>1 (+ 1 group I)</td>
<td>E</td>
</tr>
<tr>
<td>Methanococcus marisnigri JR1</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Methanopyrus kandleri AV19</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Methanosaccetta thermophila PT</td>
<td>3</td>
<td>E</td>
</tr>
<tr>
<td>Methanosarcina acetivorans C2A</td>
<td>5 (+ 1 group I)</td>
<td>E</td>
</tr>
<tr>
<td>Methanosarcina barkeri str. Funaro</td>
<td>3 (+ 1 group I)</td>
<td>E</td>
</tr>
<tr>
<td>Methanosarcina mazei G01</td>
<td>3 (+ 1 group I)</td>
<td>E</td>
</tr>
<tr>
<td>Methanospirillum stadtmanae DSM 3091</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Methanospirillum hungatei JF-1</td>
<td>2 (+ 1 group I)</td>
<td>E</td>
</tr>
<tr>
<td>Methanothermobacter thermolithotrophicus str. Delta H</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Nanoarchaeum equitum Km4-M</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Natrononomas pharaonis DSM 23160</td>
<td>4</td>
<td>E</td>
</tr>
<tr>
<td>Picrophilus torridus DSM 9790</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Pyrobaculum aerophilum str. IM2</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Pyrobaculum arsenaticum DSM 13514</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Pyrobaculum caldivolcanii ICM 11548</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Pyrobaculum islandicum DSM 4184</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Pyrococcus abyssi GE5</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Pyrococcus furiosus DSM 3638</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Pyrococcus horikoshii OT3</td>
<td>1</td>
<td>E</td>
</tr>
<tr>
<td>Staphylothermus marinus F1</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Sulfolobus acidocaldarius DSM 639</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Sulfolobus solfataricus P2</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Sulfolobus tokodaii str. 7</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>Thermococcus kodakensis KOD1</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Thermoplasma pendens IIEK 5</td>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>Thermoplasma acidophilum DSM 1728</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>Thermoplasma volcanium GSS1</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>uncultured methanogenic archaeon RC-1</td>
<td>4</td>
<td>E</td>
</tr>
</tbody>
</table>

1Numbers were determined by Blast searches against individual genomes. C, crenarchaea; E, euryarchaea; N, nanoarchaea.

the absence of the other two (although growth in these strains was more stress sensitive; 79). In addition, phylogenetic evidence shows that gene duplication and gene loss are frequent events for the archaeal chaperonins, so that in general chaperonin genes within the same archaea species tend to be more closely related to each other than they are to the same genes in different species (97; see Figure 3). This is in marked contrast to the deep branching of chaperonin sub-units seen in eukaryotic phylogenies (56). Detailed phylogenetic analysis has also shown that gene conversions have occurred frequently between duplicated chaperonin genes, particularly in the areas most likely to be associated with binding protein substrates, which implies that the substrate binding spectrum of the archaeal chaperonins has stayed fairly wide, arguing against functional specialisation (98).

However, more recent phylogenetic studies have raised the possibility that some divergence of function of the different thermosome sub-units may occur in archaea with three or more genes (99). Interestingly, the limited in vivo data currently available does support this. Thus, although two of the three H. volcanii cct genes can support growth when expressed on their own, the third cannot. Moreover, of the four thermosome subunits encoded by the halophile Haloarcula marismortui, only two could be expressed in H. volcanii and only one of these could functionally replace loss of all the H. volcanii cct genes (100). The precise nature of this divergence of function, and its adaptive significance, remains unknown. It has also been observed in some archaea with multiple cct genes that these genes show differential expression, and that the relative stoichiometry of the different sub-units making up the thermosomes can change with different temperatures (81, 101-103). This, combined with the fact that different thermosome subunits show differential thermal stability due to differences in their C-terminal regions (104) also argues for at least some divergence of function between the proteins encoded by the duplicated genes.
Figure 3. Phylogenetic relationship of 72 archaeal chaperonins, analyzed as described in reference 91. The two major groups correspond precisely to the euryarchaea (top) and crenarchaea. Numbers at internal nodes illustrate bootstrap support from 1000 repetitions.
Table 2. Comparison of chaperone activities of thermosomes from various Archaea

<table>
<thead>
<tr>
<th>Organism</th>
<th>Protein used in assay</th>
<th>Substrate used</th>
<th>Chaperone activity found</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeropyrum pernix</td>
<td>R</td>
<td>Rhodanese (E); alcohol dehydrogenase (E)</td>
<td>Prevention of thermal aggregation</td>
<td>122</td>
</tr>
<tr>
<td>Haloferax volcanii</td>
<td>N</td>
<td>-</td>
<td>Not detected</td>
<td>79</td>
</tr>
<tr>
<td>Methanococcus jannaschii</td>
<td>N</td>
<td>Luciferase (B)</td>
<td>Inhibits refolding</td>
<td>166</td>
</tr>
<tr>
<td>Methanococcus maripaludis</td>
<td>R</td>
<td>Rhodanese (M)</td>
<td>Nucleotide dependent folding</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Citrate synthase (E)</td>
<td>Prevention of thermal aggregation</td>
<td></td>
</tr>
<tr>
<td>Methanococcus thermonolithotrophicus</td>
<td>R</td>
<td>Citrate synthase (A)</td>
<td>Nucleotide dependent folding</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glucose dehydrogenase (A)</td>
<td>Nucleotide dependent folding</td>
<td></td>
</tr>
<tr>
<td>Methanopyrus kandleri</td>
<td>N/R</td>
<td>-</td>
<td>Not reported</td>
<td>121</td>
</tr>
<tr>
<td>Methanosarcina mazei</td>
<td>R (reconstituted complexes)</td>
<td>-</td>
<td>Not detected</td>
<td>107</td>
</tr>
<tr>
<td>Pyrococcus hortikushii</td>
<td>R</td>
<td>Citrate synthase (E)</td>
<td>Prevention of thermal aggregation</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isopropylmalate dehydrogenase (B)</td>
<td>Prevention of thermal aggregation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Green Fluorescent protein (B)</td>
<td>Nucleotide dependent folding</td>
<td></td>
</tr>
<tr>
<td>Pyrococcus furiosus</td>
<td>R</td>
<td>Malate dehydrogenase (B)</td>
<td>Prevention of thermal aggregation; nucleotide dependent folding</td>
<td>124</td>
</tr>
<tr>
<td>Pyrococcus furiosus</td>
<td>R</td>
<td>Lysozyme (E)</td>
<td>Prevention of thermal aggregation</td>
<td>125</td>
</tr>
<tr>
<td>Pyrodictium occultum</td>
<td>N/R</td>
<td>Citrate synthase (E); yeast alcohol dehydrogenase (E)</td>
<td>Prevention of aggregation but forms dead end complexes</td>
<td>167</td>
</tr>
<tr>
<td>Sulfolobus shibatae</td>
<td>N</td>
<td>Dihydrofolate reductase fusion protein (E)</td>
<td>Inhibits refolding and nucleotide dependent folding (all cases)</td>
<td>21</td>
</tr>
<tr>
<td>Sulfolobus solfataricus</td>
<td>N</td>
<td>Malic enzyme; alcohol dehydrogenase; glutamate dehydrogenase. (All from Sulfolobus solfataricus.)</td>
<td>Nucleotide dependent folding</td>
<td>135</td>
</tr>
<tr>
<td>Thermoplasma acidophilum</td>
<td>N</td>
<td>Lactate dehydrogenase (B)</td>
<td>Nucleotide dependent folding</td>
<td>131</td>
</tr>
<tr>
<td>Thermococcus KS-1</td>
<td>α homo-oligomer</td>
<td>Green Fluorescent protein (B)</td>
<td>Nucleotide dependent folding</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>β homo-oligomer</td>
<td>Green Fluorescent protein (B)</td>
<td>Nucleotide dependent folding</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β homo-oligomer</td>
<td>Citrate synthase (A)</td>
<td>Nucleotide dependent folding</td>
<td></td>
</tr>
</tbody>
</table>

1R, recombinant; N, native; 2B, bacterial; E, eukaryotic; A, archaean.

3.4. Group I chaperonins in archaea

Intriguingly, a small number of archaea contain genes which are clearly more homologous to the group I chaperonins than to the group II chaperonins (105-109). These include all the Methanosarcina species so far sequenced, plus the closely related species such as *Methanohalophilus portucalensis* (where the group I genes have shown to be heat shock induced, 100), *Methanospirillum hungatei*, and *Methanoregula boonei*. In all these cases, only one gene for a group I chaperone is found, always in an operon with a *cpn10* cochaperonin. Intriguingly, in all the above cases, there is always more than one copy of a group II chaperonin; the highest number (in *Methanosarcina acetivorans*) is five. The genes for group I chaperonin in these species are all very similar to each other, and it has been proposed that they were obtained by lateral gene transfer from bacteria, presumably in the species that was ancestral to all these organisms (105). It has also been argued that the phylogenetic evidence is not consistent with this, at least for *M. acetivorans* (109), raising the intriguing possibility that both kinds of chaperonin may have originally evolved in a single organism (of which the Methanosarcinas are the direct descendants) and group I or group II were subsequently lost from, respectively, all other members of the archaeal and bacterial clades. The closest related bacterial Cpn60 proteins appear to be those found in the Thermotogales – a group of organisms which themselves show evidence for a large number of genes shared with the archaea (110). The *cpn60* gene from *M. mazei* cannot complement for loss of the *groEL* gene of *E. coli*, possibly because of its very low ATPase activity (about 5% of *E. coli* GroEL) but the *cpn10* gene can partially replace the *E. coli* groES gene, and the complete system can under appropriate conditions display chaperone activity in vitro (111).

A second and quite distinct *cpn60-cpn10* operon is found in the archaeon *Methanococcus vanielii* (see Table 1). This operon has not been studied experimentally, but phylogenetic analysis suggests that it is likely to have
Archaeal chaperonins

been acquired by horizontal gene transfer from a Fusobacterium species (PAL, unpublished).

4. STRUCTURE, FUNCTION AND MECHANISM

4.1. Crystal structure of thermosomes

Thermosomes have eight fold or (occasionally) nine-fold symmetry. Sometimes these oligomers are regular with an alpha-beta repeat, as in Thermoplasma acidophilum for which the structure is available (15), whereas in other cases the stoichiometry and sub-unit arrangement is unknown but is unlikely to be regular (e.g., 79).

Structures of the thermosome of T. acidophilum both without and with nucleotide (ADP aluminium trifluoride) have been determined at a resolution of 2.6Å (15; Protein Database accession codes 1a6d (no nucleotide) and 1a6e (with nucleotide)). These structures (see Figure 1, which shows 1a6d, and Figure 2, which shows a single sub-unit from the same structure) consist of two stacked eight-membered rings, as had been previously shown by cryo-EM (112), with alternating alpha and beta subunits. The overall shape is spherical rather than cylindrical, 158Å high and 164Å in diameter. The central cavity has a diameter of from 86Å between the opposing equatorial domains to 54Å at the top of the chamber. In the structures of both the apo- and nucleotide-bound forms the hydrophobic central cavity is blocked by the lid domain, which is referred to as the closed conformation. The intra-ring contacts appear to be similar to those of group I chaperonins but the inter-ring contacts are not conserved. Whereas in GroEL the rings are offset and each subunit makes contact with two subunits in the opposite ring, the subunits of the thermosome are aligned such that each subunit makes contact with one subunit in the opposite ring forming alpha-alpha or beta-beta pairs between the two rings. Although the lid remains closed, nucleotide binding induces a domain arrangement that is distinct from the apo-form. Pivots connecting the equatorial, intermediate and apical domains allow a displacement of up to 4.8Å on the outer surface of the apical domain despite the lid remaining unchanged. More recently several crystal structures of the thermosome from Thermococcus KS1 with and without nucleotide have become available (113; Protein Database accession codes 1q2v, 1q3q, 1q3r and 1q3s). They show broadly the same structure as that seen for the T. acidophilum thermosome, with the structure still in the tightly closed conformation. Thus although these structures are informative, they do not provide much evidence about the large scale conformational changes that take place in thermosomes during the ATPase cycle.

4.2. Studies on conformational changes

Large conformational changes occur in the group I chaperonins during the course of the protein folding reaction cycle, and similar large changes have been observed in both eukaryotic and archaeal chaperonins. In these studies, the rings are referred to as being either “open” or “closed”; with group II chaperonins this refers to whether the helical protrusions at the tops of the apical domains of the sub-units are pointing roughly up and away from the rest of the protein complex (“open”) or whether they are pointing inwards, making contact with each other and sealing the cavity (“closed”). It is likely that the open state is the one which can bind substrate protein, while in the closed state substrate protein may (by analogy with group I chaperonin mechanisms) become entrapped in the central cavity. However, the precise role of both ATP binding and hydrolysis in mediating the transition between the open and closed states is still unclear.

Cyro-EM studies with eukaryotic CCT have shown a cylindrical structure for the apo-CCT. In samples incubated with ATP (but not ADP) the proportion of an asymmetric bullet-shape form increased significantly in which one ring is open (more fully than in the apo-CCT) and one ring is closed; in the closed ring large conformational changes in the equatorial and apical domains were observed enabling full closure of the lid (17, 72). The first clear evidence for an open form in a thermosome came from cryo-EM studies of T. acidophilum thermosome (both native alpha-beta and recombinant alpha only) (114). In addition, three conformations of both the native and the alpha-only thermosome from Sulfolobus shibatae have been studied by cryo-EM – an open form, a fully closed form and a bullet-shaped form (open at one end, closed at the other; 115, 116). Lid closure is therefore conserved in complexes with both eight- (Thermoplasma) and nine-fold (Sulfolobus) symmetry. In both cases the samples were prepared both with and without ATP, but this had no effect on the distribution of the conformations seen.

SAXS (Small Angle X-ray Scattering) and to a lesser extent SANS (Small Angle Neutron Scattering) have been used to look at the structure of both eukaryotic and archaeal chaperonins in solution. These methods have the advantage of looking at the structure in solution rather than in a crystal or vitrified ice and so are less prone to artifacts, although the resolution is lower. Eukaryotic CCT in solution has an open cylindrical structure in the absence of nucleotide and a rounder, more compact structure following the closing of the lid, which requires hydrolysable ATP (75). The height of the open structure is rather more than the thermosome crystal structure at 203Å, but closure of the lid reduces the maximum dimension to just 164Å, almost the same as that of the closed CCT structure. A similar study using recombinant thermosome from Thermococcus sp. strain KS1 showed the same conformational changes on the hydrolysis of ATP (117, 118). Asymmetric complexes in which one ring is open and the other closed were also observed after incubation with ADP and beryllium fluoride, which is thought to mimic the ATP bound state (118). However, these two studies suggest a different trigger for the closure – in the case of eukaryotic CCT it is brought about by the transition state of ATP hydrolysis whereas the authors of the studies in Thermococcus believe that ATP binding is the trigger in the thermosome.

SANS measurements with T. acidophilum thermosome (119, 120) have shown much the same picture. In this case, conformational change (lid closure) did not happen at room temperature but required heating the ATP-thermosome to a physiological temperature (50°C).
Table 3. Comparison of ATPase specific activities of thermosomes from various Archaea

<table>
<thead>
<tr>
<th>Organism</th>
<th>Protein used in assay</th>
<th>(kcat)</th>
<th>Km uM</th>
<th>Temperature</th>
<th>Dependent cations</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeropyrum pernix</td>
<td>R</td>
<td>2.21</td>
<td>10.04</td>
<td>90</td>
<td>Mn²⁺ K⁺</td>
<td>122</td>
</tr>
<tr>
<td>Haloferax volcanii</td>
<td>N</td>
<td>118</td>
<td>-</td>
<td>55</td>
<td>3M KCl Mg²⁺</td>
<td>123</td>
</tr>
<tr>
<td>Methanococcus jannaschii</td>
<td>N</td>
<td>73</td>
<td>-</td>
<td>80</td>
<td>Mg²⁺ K⁺</td>
<td>166</td>
</tr>
<tr>
<td>Methanococcus maripaludis</td>
<td>R</td>
<td>59.4</td>
<td>Kᵦ=43' Kᵦ=296²</td>
<td>37</td>
<td>Mg²⁺ K⁺</td>
<td>128,132</td>
</tr>
<tr>
<td>Methanococcus thermometeriophihicus</td>
<td>R</td>
<td>74</td>
<td>-</td>
<td>70</td>
<td>Mg²⁺</td>
<td>126</td>
</tr>
<tr>
<td>Methanopyrus kandleri</td>
<td>N/R</td>
<td>120</td>
<td>150</td>
<td>95</td>
<td>1M NH₄ K⁺</td>
<td>121</td>
</tr>
<tr>
<td>Methanosarcina mazei</td>
<td>R (reconstituted complexes)</td>
<td>2.5</td>
<td>-</td>
<td>37</td>
<td>Mg²⁺ Increased 2-4 fold in 200mM NH₄</td>
<td>107</td>
</tr>
<tr>
<td>Pyrococcus horikoshii</td>
<td>R</td>
<td>38</td>
<td>-</td>
<td>90</td>
<td>Mg²⁺ K⁺</td>
<td>133</td>
</tr>
<tr>
<td>Pyrococcus furiosus</td>
<td>R</td>
<td>-</td>
<td>80</td>
<td></td>
<td>Co²⁺ or Mn²⁺</td>
<td>124</td>
</tr>
<tr>
<td>Pyrococcus furiosus</td>
<td>R</td>
<td>50</td>
<td>-</td>
<td>88</td>
<td>Mg²⁺ K⁺</td>
<td>125</td>
</tr>
<tr>
<td>Pyrodictium occultum</td>
<td>N</td>
<td>300</td>
<td>-</td>
<td>90</td>
<td>Mg²⁺ K⁺</td>
<td>167</td>
</tr>
<tr>
<td>α</td>
<td>R</td>
<td>24-60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>R</td>
<td>30-60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>αβ</td>
<td>R</td>
<td>36-54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfolobus shibatae</td>
<td>N</td>
<td>27</td>
<td>-</td>
<td>75</td>
<td>-</td>
<td>21</td>
</tr>
<tr>
<td>Sulfolobus solfataricus</td>
<td>N</td>
<td>0.1</td>
<td>-</td>
<td>80</td>
<td>K⁺, Mg²⁺</td>
<td>135</td>
</tr>
<tr>
<td>Thermoplasma acidiphilum</td>
<td>N</td>
<td>3.6</td>
<td>Kᵦ=35 Kᵦ=530</td>
<td>58</td>
<td>K⁺ Mg²⁺</td>
<td>143,131</td>
</tr>
<tr>
<td>Thermococcus KS-I</td>
<td>N</td>
<td>68,182</td>
<td>-</td>
<td>60,80</td>
<td>K⁺ Mg²⁺</td>
<td>103</td>
</tr>
<tr>
<td>Thermococcus KS-I</td>
<td>-</td>
<td>-</td>
<td>80</td>
<td>K⁺ Mg²⁺</td>
<td></td>
<td>144</td>
</tr>
<tr>
<td>α</td>
<td>R</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>R</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: R, recombinant; N, native; (kcat) min⁻¹, turnover number; Temp., temperature of measurements in degrees Celsius (C); Ref., reference(s). Units of activity are given as moles ATP hydrolysed per mole chaperonin complex per minute.¹The values given for M. maripaludis are determined for each ring separately (see section 4.3 on co-operativity and allostery).

Whereas the ADP-Pi-thermosome remained closed, the final species of the ATPase cycle (with only ADP bound) resembled the open apo-thermosome.

4.3. ATPase activity and allostery

All chaperonins are ATPases. ATPase activity has been measured for thermosomes (both native and recombinant) from at least 14 different archaea (see Table 3). Most of these activities are magnesium dependent and often also require the presence of potassium ions. There are exceptions to this rule. For example Methanopyrus kandleri thermosome requires 1M ammonium present in addition to magnesium (121), that from Aeropyrum pernix requires manganese rather than magnesium (122), and CCT from Haloferax volcanii (an extreme halophile) requires 3M potassium for maximal activity but is inhibited by sodium ions (123). The thermosome from Pyrococcus furiosus is active with manganese or cobalt and is unusual in having comparable ADPase activity in addition to ATPase, which was also shown to be the case for several other thermosomes (124), although a later report of the same complex (125) does not confirm this. As these authors point out, other archaeal enzymes are known which can, unusually, use ADP as well as ATP (including hexokinase and phosphofructokinase), and this may be a high temperature adaptation, since ATP is less stable than ADP at high temperatures. Comparing the ATPase activity from the various complexes is difficult as the temperature optima for activity and growth varies widely according to organism, but they generally have a turnover of between 20 – 200 moles ATP per mole protein per min at their optimum temperature. Some chaperonin complexes have also been found to be able to hydrolyse alternative nucleoside triphosphates (UTP, CTP or GTP) (126, 127).

Group I chaperonins undergo two ATP induced allosteric transitions – one at low and one at higher ATP concentrations with respective midpoints of about 16 and 160 µM (47). The first transition reflects the positive co-operativity of ATP binding within a single ring, while the second is a consequence of the negative co-operativity between the rings. In most studies of ATPase in thermosomes no distinction is made between the two classes of ATP-binding site (i.e. in the first and second ring) and co-operativity has often not been considered. The thermosome from T. acidophilum (an alpha-beta hetero-oligomer) does not appear to have intra-ring positive co-operativity (127). However, intra-ring positive co-operativity has been observed with the thermosome from M. maripaludis (homo-oligomer) and in eukaryotic CCT, but is weaker than in GroEL (128, 73). All studies show inter-ring negative co-operativity, which appears to be a universal property of all chaperonins. The values for the saturation midpoints of the two rings (Kᵦ and Kᵦ) are remarkably similar for thermosomes from T. acidophilum (35 and 530µM respectively), M. maripaludis (43 and 296µM) and eukaryotes (bovine testis) (7.6 and 533µM). Inter-ring negative co-operativity and intra-ring positive co-operativity has also been demonstrated with the thermosome from Pyrococcus furiosus (125) although the saturation midpoints for each ring were not determined.
Archaeal chaperonins

In GroEL the intra-ring positive co-operativity is independent of GroES (47), but in Group II chaperonins the built-in lid (helical protrusion) appears to play a role in the allostery of the first ring. This was shown with the *M. maripaludis* thermosome (128) where it was shown that a lidless mutant was still able to hydrolyse ATP but with the Michaelis-Menten kinetics of an enzyme lacking allosteric regulation. The lid segments establish positive co-operativity in the ring and synchronize the ATP-induced conformational change of subunits within one ring, ATP hydrolysis (as opposed to binding alone) being required for this change. The lid also influences inter-ring communication as the lidless variant lacks the second allosteric transition at higher ATP concentrations.

The allosteric transitions in GroEL are believed to be concerted, i.e. they happen simultaneously in all subunits in the ring (129). Recently a detailed cryo-EM study has shown sequential allosteric transitions in eukaryotic CCT, where the conformational change proceeds in a domino-like effect around the ring (74). This sequential mechanism is possibly related to the sequential domain by domain folding and release of the multi-domain substrate proteins, actin and tubulin, and it can be blocked by a mutation in a single CCT sub-unit (130). It is unknown whether the transitions in thermosomes are concerted or sequential, but it is not easy to imagine how a sequential co-operativity would function in a homo-oligomeric complex, such as the *M. maripaludis* thermosome (128).

In summary, the allosteric regulation of the ATPase activity in group II chaperonins promotes binding and hydrolysis of ATP in all eight (or nine) subunits and the closure of the lid of one ring (presumably encapsulating the protein to be folded), whilst making the second ring less able to hydrolyse ATP and close. As the second ring is kept in an open conformation the substrate binding sites would be available to bind unfolded protein and so be primed for the next round of folding. This would enable the chaperonins to function as a "two-stroke" protein-folding machine (118), which is consistent with the observation in the structural studies described above of asymmetric complexes in which one ring is open and the other closed.

4.4. Studies on protein binding/folding

As discussed above, the *in vivo* role of archaeal Group II chaperonins is generally assumed to be to assist in protein-folding based, though direct evidence for this is lacking. Can such an activity be demonstrated *in vitro*? Chaperone activity can be measured in several ways: the most convincing is the demonstration of ATP-dependent refolding where the recovery of the protein is significantly greater in the presence of the chaperone than in its absence. Other assays are also often used, including prevention of the thermal aggregation of a substrate protein, binding of unfolded protein, or inhibition of spontaneous refolding of a chemically denatured substrate protein, and while these are useful assays they are not in themselves sufficient to prove true chaperone activity. A survey of the protein folding abilities of thermosomes (both native and recombinant) from 14 different archaea shows a rather incomplete picture (see Table 2) but does in some cases at least show good evidence for *in vitro* chaperone activity.

For ATP dependent refolding assays, well-characterized model bacterial or eukaryotic proteins are often used, as with the studies with the thermosome from *T. acidophilum* (131), *M. maripaludis* (132), *P. horikoshii* (133), *P. furiosus* (124) and Thermococcus KS1 (134). The substrates chosen for *M. thermolithotrophicus* (126) were commercially available enzymes from another thermophilic archaeon, *T. acidophilum*, whilst for *S. solfataricus* the authors went one step further and purified proteins from this organism (135) which could be argued to give the most physiologically meaningful results.

The most detailed studies to date were with the thermosome of Thermococcus KS1 using a heat-stable mutant of green fluorescent protein (GFP). GTP, UTP and CTP were found to mediate refolding in place of ATP, but this was effectively blocked by ADP (which could not block ATP dependent folding) showing that the chaperonin had a much lower affinity for the alternative nucleoside triphosphates than for ATP or ADP (127). A mutant of the alpha sub-unit homo-oligomer with Gly65 substituted for a much larger side-chain (cysteine or serine) has been constructed that is incapable of ATP dependent protein folding but exhibited an increase in the binding affinity for unfolded proteins in the presence of ATP, another mutant, 112T, showed enhanced folding ability (134). The double mutant (G65C and 112T, called trap-alpha chaperonin) successfully functioned as a trap for unfolded GFP. Beryllium fluoride (BeFx) is able to replace the γ-phosphate of ATP in the chaperonin complex and in conjunction with ATP can form a stable chaperonin-ADP-BeFx complex (136). The α and β homo-oligomers in the presence of ATP and BeFx could refold GFP but not release it. Electron micrographs showed that the complexes adopted a symmetric closed conformation with the folded proteins retained in the central cavity.

The only example reported to date of a thermosome from a mesophilic archaeon possessing *in vitro* chaperonin activity is from *Methanococcus maripaludis*. This complex could bind (but not release) unfolded citrate synthase, but was capable of refolding denatured rhodanese to its active form in the presence of ATP, or (to a lesser degree) a non-hydrolysable ATP analogue (132).

4.5. What do we know about the substrate binding sites?

Although the sites for protein substrate binding are well-mapped in GroEL, their location is less clear in the group II chaperonins. In eukaryotic CCT, actin appears to bind just below the helical protrusion whereas tubulin binds a wider area of the apical domain and the base of helical protrusion (66). Although CCT has not been crystallized, a structure for the CCTγ apical domain has been determined at a resolution of 2.2 and 2.8Å (137). This report implicates polar and electrostatic interactions in substrate binding due to the nature of residues at the putative binding site, and
Archaeal chaperonins

this is also suggested by some cryo-EM (138) and biochemical studies (139). However, other evidence contradicts this view. Hydrophobic interactions have been shown to be involved in the binding of actin and tubulin (140), WD-40 proteins (141), and VHL tumour suppressor (142). In this latter report the substrate binding sites of yeast CCT were localized in a helical region of the apical domain that are in the same relative position as the known substrate binding sites in GroEL, with the VHL substrate binding subunits CCT1 and CCT7.

Our experimental understanding of the basis of substrate binding in thermosomes lags behind that of eukaryotic CCT. Phylogenetic evidence suggests that tracts of the nature (146). Structures of the α and β apical domains of the T. acidophilum thermosome (from isolated domains expressed in E. coli (147, 148)) reveal two clusters of hydrophobic regions, one in the helical protrusion itself and one in the apical domain below the helical protrusion, which are buried in the native closed structure of the whole complex but predicted to be exposed in the open structure (114, 149). These are thus good candidates for potential protein binding sites, since this means that the transition from the open to the closed state would lead to the loss of protein substrate binding and the encapsidation of the released protein in the ring cavity. It has been shown that the entire helical protrusion can be deleted from the thermosome of Thermosoccus KS-1 without affecting the ability of this complex to bind unfolded protein, making the second cluster of hydrophobic residues the more likely candidate for a substrate binding site (113). These also are in the same region as (though not identical to) the residues identified in yeast CCT7 for binding of VHL tumour suppressor (142), and correspond to the known substrate binding region in GroEL. However, direct experimental proof that they bind substrate in thermosomes is not yet available.

4.6. The role of prefoldin in thermosome-mediated protein folding

A screen in S. cerevisiae for mutations that gave a synthetic lethal phenotype when combined with tubulin mutations yielded several genes that were subsequently shown to encode a protein that interacted with both tubulin and actin and delivered them to eukaryotic CCT (150-152). These proteins were called prefoldins, and it was shown in in vitro studies that their presence in refolding experiments with actin and CCT gave much improved yields of folded actin (153). Archaea were also shown to possess such proteins (where they are also sometimes referred to as GimC proteins). For both eukaryotes and archaea, the prefoldin was shown to be a hexameric complex, and in the case of most archaea this was shown to be encoded by two genes, usually called pdfα and pdfβ (154), the products of which are present in a 1:2 ratio in the final complex. Unusually, M. janaschii contains a third prefoldin homologue that does not assemble with the other two, but which has the ability to form long filaments (155). Solution of the structure of the complex from M. thermoautotrophicus showed it had an intriguing shape, with long coiled-coil motifs making it resemble a jellyfish (156). Subsequent work on both the eukaryotic and the archaeal proteins confirmed that its role appeared to be to bind a range of substrate proteins, possibly cotranslationally, and to deliver them to the CCT or thermosome, although the modes of substrate binding appear to be slightly different between the eukaryotic and archaeal versions (157-165). While it is very likely that this improves the efficiency of protein folding in vivo, there is as yet no genetic data on the importance of prefoldin activity in archaea; in yeast, knock-out mutations are viable although they have a variety of cytoskeletal defects.

5. CONCLUSIONS AND PROSPECTS

Thermosomes are essential proteins in archaea which almost certainly act to fold a subset of cellular proteins in an ATP-mediated fashion, analogous to that used by the well-studied GroEL protein of E. coli. It is also possible that they may have other cellular roles. The details of the relationship between structure and function in these proteins is becoming clearer, and their relatively simple composition compared to the eukaryotic CCT complex makes them a very attractive model for understanding group II chaperonin function in general. It is likely that experiments over the next few years, both in vitro and in vivo, will lead to a much better understanding of protein folding as mediated by chaperonins in archaea and eukaryotes. Given the evolutionary link between these two domains, and the central and unique role of the cytoskeleton in the eukaryotic cell, understanding this process, including the details of both its similarities and differences between archaea and eukaryotes, may be very informative in seeking to understand the evolutionary paths of these different cell types.

Many questions remain to be answered about these proteins. From a mechanistic point of view, the details of the reaction cycle need to be fully elucidated, and linked to the structures of the different intermediates of the chaperonin, and this needs to include an understanding of the contribution of the different sub-units that exist in most thermosomes. From the point of the view of events in the cell, a key experimental target is the identification of the cellular substrates of the thermosome, whether these are the same in all archaea, and broadly what features they share. A full description of protein folding pathways in archaea will require the enumeration of the number and identity of proteins that are folded by thermosomes, as well as how many of them are targeted to the thermosome by prefoldin, and whether all folding is post-translational or whether thermosomes can interact with nascent, ribosome-bound peptides. The question of whether thermosomes interact with the known archaeal homologues of actin and tubulin may have important evolutionary ramifications. The strong induction of thermosomes by heat shock and other stresses may imply a key role for them in protection or rescue of
proteins under these conditions, and it will be important to
determine whether stress increases the substrate range of
the thermosome.

6. ACKNOWLEDGEMENTS

Our work on archaeal chaperonins is funded by
the Leverhulme trust and the Biotechnology and Biological
Sciences Research Council, UK. We are grateful to Dr
Klaus Fütterer in our Department for Figure 1, and to Dr
Mario Fares, Head of the Evolutionary Genetics and
Bioinformatics Laboratory Department of Genetics,
University of Dublin, Trinity College, for Figure 3.

7. REFERENCES

1. Anfinsen, C. B.: Principles that govern the folding of

2. Pelham, H. R.: Speculations on the functions of the
major heat shock and glucose-regulated proteins. Cell 46,
959-961 (1986)

3. Ellis, R. J.: Proteins as molecular chaperones. Nature
328, 378-379 (1987)

4. Rothman, J. E.: Polypeptide chain binding proteins:
catalysts of protein folding and related processes in cells.

5. Ellis, R. J.: Molecular chaperones: assisting assembly

6. Hartl, F. U., & M. Hayer-Hartl: Molecular chaperones
in the cytosol: from nascent chain to folded protein. Science
295 1852-1858 (1990)

7. Macario, A. J. L., & E. Conway de Macario: Molecular
chaperones: multiple functions, pathologies, and potential
applications. Front Biosci 12, 2588-2600 (2007)

8. R. Qamra, V. Srinivas & S. C. Mande: Mycobacterium
tuberculosis GroEL homologues unusually exist as lower
oligomers and retain the ability to suppress aggregation of

9. Ellis, R. J.: Molecular Chaperones: The Plant
Connection. Science 250, 954-959 (1990)

10. Reading, D. S., R. L. Hallberg & A. M. Myers:
Characterization of the yeast HSP60 gene coding for a

evolutionary origin for mitochondria and hydrogenosomes.
Proc Natl Acad Sci USA 93, 9651-9656 (1996)

12. Hemmingsen, S. M., C. Woolford, S. M. van der Vies,
K. Tilly, D. T. Dennis, C. P. Georgopoulos, R. W. Hendrix,
& R. J. Ellis: Homologous plant and bacterial proteins
chaperone oligomeric protein assembly. Nature 333, 330-
334 (1988)

Cowan: A single-ring mitochondrial chaperonin (Hsp60-
Hsp10) can substitute for GroEL-GroES in vivo. J

Joachimiak, A. L. Horwich & P.B. Sigler: The crystal
structure of the bacterial chaperonin GroEL at 2.8 A.

Huber & S. Steinbacher: Crystal structure of the
thermosome, the archaeal chaperonin and homolog of CCT.
Cell 93, 125-138 (1998)

16. Shomura, Y., T. Yoshida, R. Iizuka, T. Maruyama, M.
Yohda & K. Miki: Crystal structures of the group II
chaperonin from Thermococcus strain KS-1: steric
hindrance by the substituted amino acid, and inter-subunit
rearrangement between two crystal forms. J Mol Biol
335, 1265-1278 (2004)

17. Llorca, O., M. G. Smyth, J. L. Carrascosa, K. R.
Willison, M. Radermacher, S. Steinbacher & J. M.
Valpuesta: 3D reconstruction of the ATP-bound form of
CCT reveals the asymmetric folding conformation of a type

nomenclature Mol Microbiol 8, 787 (1993)

unequal crossing over within the mouse T/t complex. Proc
Natl Acad Sci USA 77, 6077-6080 (1980)

20. Phipps, B. M., D. Typke, R. Heger, S. Volker, A.
Hoffmann, K. O. Stetter & W. Baumeister: Structure of a
molecular chaperone from a thermophilic archaebacterium

A. L. Horwich: A molecular chaperone from a thermophilic
archaeabacterium is related to the eukaryotic protein t-

Quaite-Randall, A. Joachimiak & J. D. Trent: The 60 kDa
heat shock proteins in the hyperthermophilic archaean

23. Quaite-Randall, E., J. D. Trent, R. Josephs & A.
Joachimiak,: Conformational cycle of the archaeosome, a
TCP1-like chaperonin from Sulfolobus shibatae. J Biol

24. Kuo, Y. P., D. K. Thompson, A. St Jean, R. L.
Charlebois & C. J. Daniels: Characterization of two heat
shock genes from Haloferax volcanii: a model system for
transcription regulation in the Archaea. J Bacteriol 179,
6318-6324 (1997)
Archaean chaperons


Archaeal chaperons


Archaeal chaperonins


Archaeal chaperonins


Archaeal chaperonins


Archaeal chaperons


Archaal chaperonins

**Key Words:** Archaea, Molecular Chaperone, Chaperonin, Groel, Cct, Thermosome, Protein Folding, Review

**Send correspondence to:** Peter A. Lund, School of Biosciences, University of Birmingham, Birmingham B15 2TT, United Kingdom, Tel: 44 121 414 5583, Fax: 44 121 414 5925, E-mail: p.a.lund@bham.ac.uk