Review
The substrate recognition mechanisms in chaperonins

Paulino Gómez-Puertas¹, Jaime Martín-Benito², José L. Carrascosa², Keith R. Willison³
and José M. Valpuesta²*¹

¹Centro de Astrobiología (CSIC/INTA). Ctra Torrejón-Ajalvir, Km 4, Torrejón de Ardoz, 28850 Madrid, Spain
²Centro Nacional de Biotecnología, CSIC, Campus Universidad Autónoma de Madrid, 28049 Madrid, Spain
³Cancer Research UK Centre for Cell and Molecular Biology, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, Chelsea, London SW3 6JB, UK

Chaperonins are a family of proteins devoted to assisting the folding of other proteins. They are large oligomers assembled into ring structures that enclose a cavity in which folding takes place. For this process to occur, the chaperonin must first recognize and interact with the unfolded polypeptide, then undergo a conformational change upon nucleotide binding that results in the closure of the cavity which in turn mediates the folding reaction inside the cavity. Although this general mechanism seems to apply to every chaperonin studied so far, there exist two different modes of interaction between the chaperonin and the substrate. The first occurs mainly through the interaction between the exposed hydrophobic residues of the unfolded polypeptides and those of the chaperonin substrate binding site, as elucidated for the chaperonin GroEL from E. coli. The second type of mechanism has been described so far only for the cytosolic chaperonin CCT (Chaperonin Containing TCP-1) and here the interaction seems to be of a more specific nature, involving charged and polar residues in both the chaperonin and the substrate, which interacts with CCT in a structured, quasi-native conformation. Copyright © 2004 John Wiley & Sons, Ltd.

Keywords: protein folding; molecular chaperones; chaperonins; GroEL; CCT; thermosome; actin; tubulin

Received 28 July 2003; revised 22 September 2003; accepted 25 September 2003

INTRODUCTION

Although the information required for a protein to attain its native, functional conformation is stored in its amino acid sequence, protein folding is usually a problematic task in the crowded environment of the cell, where a multitude of unwanted interactions occur all the time. Nature has tried to overcome this problem by creating a large number of proteins, termed molecular chaperones, which assist in the folding of other proteins (Mogk et al., 2001), mostly by protecting their aggregation-prone regions and by providing protected environments for proteins to fold by themselves. Most of these proteins are heat-shock proteins and their level of expression is greatly increased upon stress conditions which cause misfolding and/or aggregation. Among the molecular chaperones, one of the best characterized families is the chaperonins, which are found in every known organism (Ellis, 1996), although they are divided in two groups, depending on whether they are found in eubacteria and in the endosymbiotic organelles (group I; Bukau and Horwich, 1998; Ellis and Hartl, 1999), or in archaea and the cytosol of eukarya (group II; Willison, 1999; Gutsche et al., 1999; Cowan and Lewis, 2002). All chaperonins share a similar structure, i.e. they are large oligomers composed of ~60 kDa proteins built up in the shape of a toroid, usually composed of two rings placed back-to-back (Carrascosa et al., 2001). Each oligomer subunit in every chaperonin has also a similar structure, divided into three domains (Braig et al., 1994; Fenton et al., 1994): the equatorial domain, that holds the nucleotide binding site and most of the interactions between the subunits of the same ring and those of the opposing ring; the intermediate domain, which acts as a transmitter of the conformational changes generated upon nucleotide binding between the equatorial domain and the apical domain, where the substrate binding is located.

However, both groups of chaperonins also have important differences. Whereas the ring in all group I chaperonins is composed of seven identical subunits, group II chaperonins can have octameric rings composed of one or two different subunits, nonameric rings composed of three different subunits (Gutsche et al., 1999) and octameric rings composed of eight different subunits, such is the case of the cytosolic chaperonin CCT (Chaperonin Containing TCP-1, also termed TRiC or c-cpn; Frydman et al., 1992; Gao et al., 1992; Lewis et al., 1992). Another important difference is related to the presence in group I chaperonins of a small oligomer, termed cochaperonin, which acts in conjunction with the chaperonin in the folding of proteins by capping the chaperonin cavity (Xu et al., 1997). This cochaperonin is...
absent in group II chaperonins but its function is mimicked by an extra sequence located at the tip of the apical domain named helical protrusion (Klumpp et al., 1997; Ditzel et al., 1997). After the conformational changes generated upon nucleotide binding, it locks the chaperonin cavity acting like an iris (Ditzel et al., 1997; Llorca et al., 1999a, 2001a,b; Gutsche et al., 2000; Schoehn et al., 2000a,b).

Despite these differences, a very general cycle can be outlined for all chaperonins. They seem to adopt two main conformations: an ADP or nucleotide-free conformation in which the chaperonin has an open structure and a high affinity for the unfolded substrate, and an ATP-bound conformation with a closed structure and a low affinity for the substrate. Thus, unfolded polypeptides are bound by the ADP or the nucleotide-free, open conformation of the chaperonin. Subsequently, nucleotide binding induces the conformational changes leading to the closure of the cavity (either by the capping of the cochaperonin in the case of the group I chaperonins or by the extra helical protrusion in the case of group II chaperonins) and the liberation of the substrate inside the cavity, where in an isolated environment and free from unwanted interactions, it can fold using the information imprinted in its own amino acid sequence. Since this mechanism is used by group I chaperonins to assist in the folding of a multitude of proteins, the substrate recognition mechanism must be capable of recognizing the unfolded polypeptide in myriad conformations. This can only be achieved by a very generalized recognition mechanism whereby the hydrophobic residues of the unfolded polypeptide, exposed on the surface of unfolded proteins, are recognized by and interact with the hydrophobic residues of the chaperonin substrate binding region. This mechanism seems to apply to all known chaperonins, but with an important exception, the cytosolic chaperonin CCT, which uses a mechanism that relies on a more specific groove between \( \alpha \)-helices H8 and H9 and an extended loop below, all located in an exposed region that faces the chaperonin channel. The fact that mutation of charged residues within this region did not abolish substrate binding led the authors to suggest that the substrate binding domain of GroEL [and of the rest of group I chaperonins, see the high degree of homology amongst them in Plate 1(A)] is located at the rim of the channel and is formed by a ring of hydrophobic residues, which seem to recognize the unfolded polypeptide [Plate 1(B)]. The hydrophobic interactions between the chaperonin and the substrate seem to be the major driving force in their recognition and binding. Biochemical experiments have shown that GroEL interacts preferentially with the side chains of hydrophobic amino acids (Richarme and Kohiyama, 1994) and other studies, again using a battery of different techniques, have also outlined the importance of the hydrophobic interactions between the chaperonin and the substrate (Hayer-Hartl et al., 1994; Hlodan et al., 1995; Izhaki et al., 1995; Lin et al., 1995; Coyle et al., 1997). However, there are certain exceptions, and other types of interactions such electrostatic ones may play also an important role (Katsumata et al., 1996; Aoki et al., 1997; Hutchinson et al., 1997).

Two atomic structures of the complex between a polypeptide and the GroEL apical domain have confirmed this region as the one involved in substrate binding (Buckle et al., 1997; Chen and Sigler, 1999). Both structural studies have shown that the unfolded polypeptide chain is located in the groove between \( \alpha \)-helices H8 and H9, interacting mostly with the hydrophobic residues that form part of the two \( \alpha \)-helices [Plate 1(C)]. These studies also revealed that the substrate binding domain is flexible enough to accommodate different types of unfolded structures, which confirms numerous studies showing that the unfolded protein can interact with GroEL in different conformations, from structured ones (Landry and Gierash, 1991; Landry et al., 1992), to partially folded conformations (Martin et al., 1991; Robinson et al., 1994) and completely unfolded ones (Badoe et al., 1991; Zahn et al., 1994). The flexibility of the apical domain (Braig et al., 1994) and its ability to recognize and to interact with different types of unfolded structures, are the basis of the high promiscuity of GroEL in the assistance to the folding of other proteins, and explains why GroEL can interact in vivo with at least 10% of newly synthesized polypeptides (Houry et al., 1999). The range of molecular weights of GroEL substrates is large and the chaperonin can accommodate in its cavity, molecules up to 65 kDa in mass. Consequently the unfolded polypeptide, depending on its size, can interact with GroEL through multiple apical domains (Farr et al., 2000).

The folding process continues through a series of steps, some of them which promote a new set of recognition processes. ATP binding induces a conformational change in the apical domain that allows the binding of the GroEL cochaperonin, GroES, which in turn generates a further conformational change in the apical domain. The interaction between GroES and GroEL is of a hydrophobic nature and

**MECHANISM OF SUBSTRATE RECOGNITION BY GROUP I CHAPERONINS**

Most of the work carried out on the structure and function of group I chaperonins has been performed with GroEL, the chaperonin from *E. coli*, and the characterization of the mechanism of substrate recognition is no exception. All kinds of techniques have been used to characterize the mechanism of substrate binding in GroEL and its localization within the structure of the oligomer, which electron microscopy studies have shown to be at the entrance of the GroEL cavity (Braig et al., 1993; Chen et al., 1994). The atomic structure of GroEL obtained by X-ray diffraction revealed the structure of this region, termed the apical domain (Braig et al., 1994). An accompanying mutational study showed that several residues of this domain (A152, Y199, S201, Y203, F204, L234, L237, L259 and V263), almost all of them hydrophobic in nature, are involved in substrate recognition, as their substitution diminishes or even abolishes unfolded polypeptide binding to GroEL (Fenton et al., 1994). This region corresponds to \( \alpha \)-helices H8 and H9 and an extended loop below, all located in an exposed region that faces the chaperonin channel. The fact that mutation of charged residues within this region did not abolish substrate binding led the authors to suggest that the substrate binding domain of GroEL [and of the rest of group I chaperonins, see the high degree of homology amongst them in Plate 1(A)] is located at the rim of the channel and is formed by a ring of hydrophobic residues, which seem to recognize the unfolded polypeptide [Plate 1(B)]. The hydrophobic interactions between the chaperonin and the substrate seem to be the major driving force in their recognition and binding. Biochemical experiments have shown that GroEL interacts preferentially with the side chains of hydrophobic amino acids (Richarme and Kohiyama, 1994) and other studies, again using a battery of different techniques, have also outlined the importance of the hydrophobic interactions between the chaperonin and the substrate (Hayer-Hartl et al., 1994; Hlodan et al., 1995; Izhaki et al., 1995; Lin et al., 1995; Coyle et al., 1997). However, there are certain exceptions, and other types of interactions such electrostatic ones may play also an important role (Katsumata et al., 1996; Aoki et al., 1997; Hutchinson et al., 1997).

Two atomic structures of the complex between a polypeptide and the GroEL apical domain have confirmed this region as the one involved in substrate binding (Buckle et al., 1997; Chen and Sigler, 1999). Both structural studies have shown that the unfolded polypeptide chain is located in the groove between \( \alpha \)-helices H8 and H9, interacting mostly with the hydrophobic residues that form part of the two \( \alpha \)-helices [Plate 1(C)]. These studies also revealed that the substrate binding domain is flexible enough to accommodate different types of unfolded structures, which confirms numerous studies showing that the unfolded protein can interact with GroEL in different conformations, from structured ones (Landry and Gierash, 1991; Landry et al., 1992), to partially folded conformations (Martin et al., 1991; Robinson et al., 1994) and completely unfolded ones (Badoe et al., 1991; Zahn et al., 1994). The flexibility of the apical domain (Braig et al., 1994) and its ability to recognize and to interact with different types of unfolded structures, are the basis of the high promiscuity of GroEL in the assistance to the folding of other proteins, and explains why GroEL can interact in vivo with at least 10% of newly synthesized polypeptides (Houry et al., 1999). The range of molecular weights of GroEL substrates is large and the chaperonin can accommodate in its cavity, molecules up to 65 kDa in mass. Consequently the unfolded polypeptide, depending on its size, can interact with GroEL through multiple apical domains (Farr et al., 2000).

The folding process continues through a series of steps, some of them which promote a new set of recognition processes. ATP binding induces a conformational change in the apical domain that allows the binding of the GroEL cochaperonin, GroES, which in turn generates a further conformational change in the apical domain. The interaction between GroES and GroEL is of a hydrophobic nature and
occurs through residues of an unstructured region of GroES (the ‘mobile loop’; Landry et al., 1993) and some of the GroEL residues also involved in substrate binding (Fenton et al., 1994; Xu et al., 1997). Therefore GroES binding to GroEL displaces the unfolded polypeptide into the now capped chaperonin cavity, which has been transformed into a much larger and hydrophilic channel as a result of the conformational changes generated upon ATP and GroES binding (Crouy-Chanel et al., 1995; Xu et al., 1997; Wang and Boisvert, 2003). Encapsulated polypeptides are now free to fold in the isolated chamber using the information encoded in their amino acid sequence, in what has been termed the ‘Anfinsen cage’ (Ellis, 1994). However, before the release reaction, it is possible that the conformational changes generated in GroEL might have been used to exert a stretching force on the unfolded polypeptide, thus liberating it from local energy minimum in which it might have been trapped. Thus, it is possible that GroEL assists in the folding of other proteins not only by providing them with a proper environment in the isolated chamber, but also working as an unfoldase and therefore playing a more active role in protein folding (Zahn et al., 1996; Walther et al., 1996; Shtilerman et al., 1999; Wang and Boisvert, 2003).

MECHANISM OF SUBSTRATE RECOGNITION BY GROUP II CHAPERONINS

Contrary to studies with group I chaperonins, there is not much information on the substrate recognition mechanism in group II chaperonins and especially those of archaeal origin, usually named thermosomes because of their thermostability and heat inductibility (Phipps et al., 1991). There is a great deal of structural information on the thermosomes, obtained either from X-ray diffraction (Klumpp et al., 1997; Ditzel et al., 1997) or electron microscopy (Nitsch et al., 1998), which reveals that the overall structure of the thermosomes is similar to that of GroEL. The major differences, also observed when comparing the primary sequence of both type of chaperonins, reside in the apical domains, which as in the case of GroEL, undergo large conformational changes during the thermosome functional cycle (Schoehn et al., 2000a,b). These conformational changes are induced upon ATP binding and hydrolysis and are at the heart of the folding mechanism (Gutsche et al., 2000, 2001). The two main conformations of the thermosomes, as in other chaperonins, are the open, substrate-receptive conformation, and the closed conformation, generated by the movement of the helical protrusions (Ditzel et al., 1997).

Although almost nothing is known about the natural substrates of thermosomes, it is however clear that they are capable of preventing aggregation and eventually assisting in the folding of numerous proteins of mesophilic and thermophilic origin (for a review on this subject, see Gutsche et al., 1999). This suggests a substrate recognition and folding mechanism similar to that of GroEL, based on the interaction between the exposed hydrophobic residues of the unfolded polypeptide and those of the chaperonin substrate binding domain. Indeed there is biochemical evidence pointing to the hydrophobic interactions as the ones driving the binding between the thermosome and the unfolded substrate (Guagliardi et al., 1995). Since the overall structure of the thermosome is similar to that of GroEL, it is reasonable to suggest that the apical domain is the one hosting the substrate binding site (Gutsche et al., 1999). When looking for hydrophobic regions within this domain, a cluster of hydrophobic residues is found near the surface of the nucleotide-free, open conformation of the thermosome, viewed from the centre of the cavity. The atomic structures of the α- and β-thermosome from Thermoplasma acidophilum (Ditzel et al., 1997) have been modelled to fit in an approximate manner into the three-dimensional reconstruction of the structure of nucleotide-free CCT obtained by cryoelectron microscopy (Llorca et al., 2000). (B) A cut section of the surface, viewed from the interior of the cavity, of the apical domains of the structure of Thermoplasma acidophilum, which is thought to emulate the nucleotide-bound, closed conformation. The dark regions correspond to the hydrophobic residues which in the open conformation face the cavity and which almost disappear in the closed conformation. The circles in (A) correspond to the two clusters of hydrophobic residues observed in the apical domains of the thermosomes. Images were generated with GRASP (Nicholls et al., 1991).
homologous region to the GroEL substrate binding domain and at the tip of the helical protrusion [Fig. 1(A)]. These two regions are buried in the atomic structure of the thermosome from *Thermoplasma acidophilum* [Ditzel et al., 1997; Fig. 1(B)]. However, this structure is the closed conformer of the chaperonin and the low-resolution reconstructions of several thermosomes in the absence of nucleotides show an open conformation where unfolded polypeptides are free to interact with the chaperonin substrate binding domain (Nitsch et al., 1998; Schoehn et al., 2000a,b). Docking of the atomic structure of the thermosome monomer into some of these structures show the two hydrophobic domains facing the channel [Nitsch et al., 1998; Schoehn et al., 2000b; Fig. 1(A)]. According to this, the open conformation that occurs in the absence of nucleotide would recognize and bind unfolded polypeptides through the hydrophobic domains. Subsequently, ATP binding would induce the closure of the chaperonin cavity due to the ~70° clockwise movement of the apical domain and the helical protrusion placed at its tip [Fig. 1(B)]. This conformational change would induce the liberation of the unfolded polypeptide and its liberation in a sealed, isolated and now hydrophilic chamber [Ditzel et al., 1997; Fig. 1(B)], allowing the polypeptide to fold by itself.

**MECHANISM OF SUBSTRATE RECOGNITION BY THE CYTOSOLIC CHAPERONIN CCT**

Electron microscopy has shown that CCT shares the general structural mechanism with the rest of the chaperonins, having an open conformation that recognizes and binds the substrate and a closed, ATP-induced one, where the substrate is encapsulated and folded [Plate 2(C and D): Llorca et al., 1998, 1999a]. Apart from that, and even before any structural information on the nature of the interaction between CCT and its substrates was known, some kind of sequence specificity between both proteins was suspected due to the high hetero-oligomeric nature of CCT and its apparent selectivity for actins and tubulins (Kubota et al., 1994). With respect to the first point, CCT is by far the most complex of all chaperonins known so far, since it is composed of eight different, albeit homologous subunits (CCTα, β, γ, δ, ε, ζ, η, θ; CCT1-CCT8 in yeast), whose unique arrangement within a ring has been elucidated (Liou and Willson, 1997). The evolution of CCT from a simpler, less hetero-oligomeric chaperonin suggests a specialization in the function of the cytosolic chaperonin (Archibald et al., 2000), a trend that seems to have been followed by a cochaperone of CCT termed prefoldin (PFD or GimC; Geissler et al., 1998; Vainberg et al., 1998); a small oligomer that transfers the unfolded substrate to CCT via a physical interaction between the two chaperones (Vainberg et al., 1998; Martín-Benito et al., 2002).

With regard to the selectivity for actins and tubulins, and although the number of proteins assisted by CCT is greater than previously thought (Dunn et al., 2001; Valpuesta et al., 2002), it is nevertheless evident that CCT is not a promiscuous chaperonin like GroEL and that it acts upon a limited, albeit large set of proteins. Since its discovery in the early nineties (Frydman et al., 1992; Yaffe et al., 1992; Gao et al., 1992; Lewis et al., 1992), a large body of biochemical and structural information has been accumulated pointing to the substrate recognition mechanism of CCT being different from all the other chaperonins characterised so far. The experiments of Tian et al. (1995a) revealed that although the eukaryotic CCT substrates, actin and tubulin, could be recognized by GroEL in their unfolded forms, this bacterial chaperonin was not able to fold them. In fact, both cytoskeletal proteins have a stringent requirement for CCT to obtain their native conformations (Chen et al., 1993; Rommelaere et al., 1993), which suggests a specific interaction between these two proteins and the cytosolic chaperonin.

**The conformation of the CCT-bound actin and tubulin molecules**

The nature of the CCT-bound actin and tubulin conformations has been studied by many groups and the first thing that emerges is that the two proteins do not adopt any kind of unstructured conformation like the case of the GroEL substrates, but rather some sort of defined, quasi-native conformation (Melki and Cowan, 1994; Tian et al., 1995b; Lewis et al., 1996; Dobrzyinski et al., 1996), ready to undergo the final stage in their folding process upon the ATP-induced action of CCT. The intermediate folding conformation of actin and tubulin is therefore acquired before the interaction with the cytosolic chaperonin, either by themselves or with the help of other chaperones like PFD, and this would explain the slow formation of these intermediates (Melki and Cowan, 1994). The three-dimensional reconstructions of the CCT–actin and CCT–tubulin complexes carried out by cryoelectron microscopy [Plate 2(C) and (E)] reveal that both cytoskeletal proteins interact with CCT in defined conformations, with both molecules crossing the chaperonin cavity and interacting with subunits in opposed regions of the channel (Llorca et al., 1999b, 2000). The two cytoskeletal proteins have the common feature of being composed of two topological domains which are locked together through a nucleotide binding site [Kabsch et al., 1990; Nogales et al., 1998a; Plate 2(A) and (B)]. Docking of the atomic structures of both proteins into their respective volumes of the CCT–substrate complex suggests that the two proteins have reached a conformation close to the native one [Plate 2(C) and (E)], albeit an open conformation susceptible to proteolysis (Grantham et al., 2000) in which the two domains are separated from each other and interact with opposing regions of the CCT cavity (Llorca et al., 2000).

**Characterization of the CCT substrate-binding regions**

As in other type of studies regarding CCT, most of what is known comes from the experiments performed with actin and tubulin, which have shown that both cytoskeletal proteins interact with specific CCT subunits. The combined results obtained with the docking above mentioned and the immunomicroscopy carried out with CCT–actin and CCT–tubulin complexes (and with complexes of CCT and fragments of the two proteins labelled with antibodies against specific CCT subunits) reveal that actin and tubulin interact...
Plate 1. Structural features of chaperonins. (A) Structure-based alignment of the apical domains of group I and II chaperonins. Using the DALI comparison algorithm (Holm and Sander, 1993), several apical domain sequences of GroEL-like chaperonins (top), thermosomes (middle) and the eight CCT subunits (bottom) have been aligned. The secondary structures depicted for the three types of chaperonins (green boxes correspond to α-helices and red boxes to β-strands) correspond to the elements found in the structures of Escherichia coli GroEL (PDB code 1KID; Buckle et al. 1997), the thermosome of Thermoplasma acidophilum (PDB code 1A6D; Ditzel et al., 1997) and the mouse CCT (PDB code 1GML; Pappenberger et al., 2002). Residues are coloured according to the degree of residue conservation (higher, blue; lower, white). Asterisks indicate the position of the residues of GroEL postulated to be involved in the binding of peptide substrates (Fenton et al., 1994; Buckle et al. 1997). (B) Cut side view (above) and top view (below) of one of the rings of the atomic structure of GroEL (Braig et al., 1994), showing in green the hydrophobic residues located at the entrance of the cavity and involved in substrate binding. (C) Ribbon plot of the structure of the apical domain of GroEL (PDB code: 1KID, interacting with a peptide) (in magenta). Residues shown by Buckle et al. (1997) to be involved in the interaction with the unfolded peptide localize mainly in the two α-helices and are displayed in ball-and-stick representation.
Plate 2. Interaction between CCT and the unfolded actin and tubulin, based in the cryoelectron microscopy studies performed by Llorca et al. (1999b, 2000, 2001b). (A) The atomic structure of actin (Kabsch et al., 1990). The yellow domains correspond to the putative CCT-binding sites. The two topological domains of actin have been historically termed small and large domains, and correspond respectively to the domains coloured blue and red. The small domain is composed mainly of the N-terminal region, although a small stretch of residues of the C-terminal sequence (the last 35 residues) are also located in the small domain. (B) The atomic structure of tubulin, (Nogales et al., 1998a). The color codes for N-terminal domain, C-terminal domain and the putative CCT-binding sites are the same as used for actin in (A). (C) and (E) The three-dimensional reconstructions of the CCT–actin and CCT–tubulin complexes, respectively, in the nucleotide-free, open conformation, viewed from the top (top image) and a from a side, cut view (bottom image). (D) and (F) The three-dimensional reconstructions of the CCT–actin and CCT–tubulin complexes, respectivley, in the presence of nucleotide, which generates the closed conformation of CCT where folding takes place. The atomic structures of actin and tubulin have been docked in all four cases into the masses of the actin and tubulin components of the reconstructed complexes. In the nucleotide-free form, the atomic structures have been opened across their respective hinges to fit their reconstructed mass. In all the images, the actin and tubulin molecules have been coloured respectively as in (A) and (B). Yellow arrows indicate the sequential conformational change in the CCT subunits induced by ATP binding, as proposed by Lin and Sherman (1997). Part of this figure is a modified version of Fig. 8 from Llorca et al. (2001b), The `sequential allosteric ring’ mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. EMBO J. 20: 4065–4075. reproduced by permission from Oxford University Press.
Plate 3. Binding determinants in CCT, actin and tubulin. (A) A model of the atomic structure of CCT in its nucleotide-free, open conformation. The model has been constructed using the atomic structure of the CCT\_apical domain (Pappenberger et al., 2002) and the other subunits modelled according to this structure and their respective sequences, by homology modelling techniques using the program Swiss-Pdb Viewer and the SWISS-MODEL server facilities (Guex and Peitsch, 1997; www.expasy.ch/swissmod/ SWISS-MODEL.html), the DALI comparison algorithm (Holm and Sander, 1993) and fitted into the three-dimensional reconstruction of nucleotide-free CCT obtained by cryoelectron microscopy (Llorca et al., 2000). The CCT subunits are viewed from outside and above the cavity. (B) A model of the atomic structure of CCT in its nucleotide-bound, closed conformation. The model has been constructed fitting the atomic models of the CCT subunits in the atomic structure of the thermosome from Thermoplasma acidophilum (Ditzel et al., 1997). The CCT subunits are viewed from inside the cavity, looking upwards from the equator. In both models, the charged surfaces (negatively charged residues in red, positively charged residues in blue) were generated using GRASP (Nichols et al., 1991). (C) Sequence alignment of the apical domain of the mouse CCT subunits that shows the charged nature of the helical protrusions (marked with a black line) and the regions that flank them (negatively charged residues in red, positively charged residues in blue and hydrophobic residues in yellow). (D) The CCT-binding domains in actin and \(\beta\)-tubulin, according to Llorca et al. (2001a), Hynes and Willison (2000) and Ritco-Vonsovici and Willison (2000). Colouring code as in (C), with the hydrophilic residues in black.
with specific CCT subunits using two modes of interaction (Llorca et al., 1999b, 2000). In the case of actin [Plate 2(C)], the protein, in the open and quasi-native conformation, binds with the small domain (Kabsch et al., 1990) to CCTδ and with the large domain to CCTβ or CCTε. In the case of tubulin [Plate 2(E)], binding to CCT is more complex and involves the interaction of either its N-terminal domain with CCTα and CCTη and its C-terminal domain with CCTβ, CCTγ and CCTθ, or its N-terminal domain with CCTδ and CCTθ and its C-terminal domain with CCTε, CCTζ and CCTβ.

The three-dimensional reconstructions of the CCT–actin and CCT–tubulin complexes have also shed some light on the localisation of the CCT substrate-binding region [bottom images in Plate 2(C) and (E)]. Albeit at low resolution (25–30 Å), the two three-dimensional reconstructions show actin and tubulin binding to the upper part of the apical domain, just below the helical protrusion in the case of actin, and encompassing part of this protrusion in the case of tubulin, which seems to interact with a much larger area of the chaperonin apical domain (Llorca et al., 2000). Interestingly, this region coincides with the region where Archibald et al. (2001), performing phylogenetic analyses on a large number of CCT sequences from different organisms, have located a cluster of subunit-specific conserved residues that very likely play a role in substrate recognition and interaction (Pappenberger et al., 2002). Most of these residues are charged ones [Plate 3(C)], and not hydrophobic like the ones described in GroEL to be involved in substrate interaction. This is another point that signals that CCT for the substrate recognition mechanism is different from that of other chaperonins, a mechanism based on specific interactions between the substrate and the chaperonin.

Characterization of the CCT-binding sites of actin and tubulin

Several biochemical and structural studies have been carried out to try to characterize the actin and tubulin domains involved in CCT binding, and what emerges from them is that both cytoskeletal proteins interact with the cytosolic chaperonin through several domains acting in a cooperative way. In the case of actin, a biochemical study first pointed to three regions, those encompassed by residues 125–179, 244–285 and 340–375, as involved in CCT binding (Rommelaere et al., 1999). Subsequently, a screening study using actin peptide arrays pointed to three different binding regions (sites I, II and III; Hynes and Willison, 2000; McCormack et al., 2001a). Two of these regions are located at the tips of the small and large domains [Plate 2(A)] and coincide with a docking analysis of the actin molecule bound to CCT in the three-dimensional reconstruction of a CCT–actin complex carried out by cryoelectron microscopy (Llorca et al., 2001a). This analysis, combined with immunomicroscopy experiments performed with complexes formed between CCT and actin or a chimera formed with a fragment of the small domain bound to residues 1–168 of human Ha-Ras, has allowed the identification of the actin sequences and the CCT subunits involved in CCT:actin interaction (Llorca et al., 1999b, 2001a). The results obtained point to four loops exposed on the surface of the native structure [Plate 2(A) and 3(D)]. Two of them (R37–D51 and R62–T66) are placed at the tip of small domain and interact with CCTδ, and the other two (E195–R206 and T229–T250) are located at the tip of the large domain and interact with either CCTβ or CCTε.

In the case of tubulin, the interaction is more complex, and all the studies carried out so far show that tubulin interacts with CCT through several domains distributed along its entire sequence (Dobrzynski et al., 1996, 2000; Rommelaere et al., 1999; Ritco-Vonsovici and Willison, 2000; Llorca et al., 2000, 2001a). A screening study using peptide arrays covering the sequences of α-, β- and γ-tubulin revealed several CCT-binding sites in both the N- and C-terminal domains (Ritco-Vonsovici and Willison, 2000), and most of these sites are corroborated by a docking analysis of the β-tubulin molecule in the three-dimensional reconstruction of a CCT–tubulin complex carried out by cryoelectron microscopy (Llorca et al., 2001b). This analysis, in combination with immunomicroscopy experiments performed with complexes formed between CCT and tubulin or chimeras formed with various tubulin fragments bound to residues 1–168 of human Ha-Ras, has suggested which tubulin sequences and CCT subunits are involved in CCT–tubulin interaction [Llorca et al., 2000; Plate 2(B) and 3(D)]. All these proposed tubulin CCT-binding sites are placed in loops exposed on the surface of the native protein [Nogales et al., 1998a; Plate 2(B)] and bind to CCT using two modes of interaction. Three of these binding sites are located at the N-terminal domain [T33-A57, S126–Q133 and E160–R164; Plate 3(D)] and, according to immunomicroscopy experiments, the sequence T33–A57 would interact with CCTα or CCTθ, and the sequences S126–Q133 and E160–R164 would interact with CCTη or CCTθ (Llorca et al., 2000). In the C-terminal domain, the interaction is even more complex and there are five putative CCT-binding sites, those loops encompassed by sequences T239–K254, P261–H266, S277–V288, V355–P359 and W407–E417, which interact with different CCT subunits in two possible modes of interaction: loops T239–K254 and P261–H266 would interact with CCTζ or CCTγ, loops S277–V288 and V355–P359 would interact with CCTε or CCTβ, and loop W407–E417 would interact with CCTβ or CCTθ (Llorca et al., 2000).

Although previous studies have suggested that the tubulin binding sites have a weak binding affinity, a common consensus exists for a ‘hot spot’ in the so called M loop (residues S277–V288; located between β-strand 7 and α-helix 9 of the tubulin structure; Nogales et al., 1998b) that would possess a higher binding affinity (Dobrzynski et al., 1996, 2000; Rommelaere et al., 1999; Ritco-Vonsovici and Willison, 2000). In fact, a chimera made of a fragment of the first 168 residues of Ha-Ras protein and a tubulin fragment encompassing this loop has a comparable binding affinity to the complete tubulin molecule (Llorca et al., 2000). In the case of actin, a mutational analysis has pointed to the proposed CCT-binding sites being located at the tip of the large domain as the major determinants of CCT binding (McCormack et al., 2001a). This result is reinforced by biochemical and electron microscopy studies carried out with a mutant in a conserved residue located in the putative hinge between the small and large domains that are far apart.
when bound to CCT (McCormack et al., 2001b). The mutation (G150P), which was suspected to induce rigidity in the hinge and prevent the opening of the actin molecule, resulted in the accumulation of the actin mutant on the chaperonin. The electron microscopy analysis of this G150P mutant complexed to CCT reveals that it is tightly packed near the CCT-binding regions present in the large domain of actin drive the interaction with CCT (through CCTβ or CCTε subunits). Owing to the loss of flexibility induced by the presence of proline in the actin hinge, the molecule cannot open its structure to interact with both sides of the CCT cavity and thus accumulates in the chaperonin. CCTβ or CCTε seem to be the subunits with the highest substrate affinity, and this notion is strengthened by immunoprecipitation experiments of CCT–actin complexes, subsequently disrupted with detergent buffers, that show these two CCT subunits are amongst the strongest actin binders (Hynes and Willison, 2000).

Finally, although it has been hypothesized that the hydrophobic residues are those responsible for the interaction between the two cytoskeletal proteins and the cytosolic chaperonin (Dobrzynski et al., 1996, 2000; Rommelaere et al., 1999), and even if the role of certain hydrophobic residues in CCT binding cannot be ruled out, it is becoming evident that the interaction of actin and tubulin (and perhaps other substrates) with CCT occurs mainly through the cooperation of polar and electrostatic interactions [Plate 3(A, C and D)]. An analysis of the actin and tubulin residues putatively involved in CCT binding has revealed that a majority of them are charged or polar and a mutational analysis carried out in actin has shown that the stretch of residues F200–S201–T202–T203, mostly of polar nature, and the charged and extremely conserved residue D244, contribute substantially to CCT binding (McCormack et al., 2001a).

**Mechanism of actin and tubulin folding mediated by CCT**

From the results described above and others, a substrate recognition and folding mechanism for CCT which is utterly different than that described for GroEL, begins to emerge. The open, substrate-receptive conformation of CCT [Plate 3(A)] recognizes and binds a substrate that has already acquired a large degree of native folding [Plate 2(C and E)], either by itself or with the help of other chaperones such as PFD. This recognition seems to occur through the upper part of the apical domain [bottom figures in Plate 2(C and E)], including the base of the helical protrusion, in a region that is predominantly populated in all eight different CCT subunits by charged and polar residues [Plate 3(A–C); Pappenberger et al., 2002], which suggests a specific interaction between the chaperonin and the two cytoskeletal proteins. This notion is reinforced by the fact that the proposed CCT-binding regions in actin and tubulin are also predominantly composed of charged and polar residues [Plate 3(C)], and confirmed by immunomicroscopy experiments that show that certain regions of actin and tubulin interact with specific CCT subunits [Plate 2(A and B)]. The interaction between the CCT-binding regions of the two proteins and the substrate-binding regions of the CCT subunits is probably of a weak nature, except perhaps for some CCT-binding regions that may lead to the initial recognition and binding process. Interestingly, these higher-affinity binding sites are located in the C-terminal domain of both actin [the large domain could be considered as such; see legend to Plate 2(A)] and tubulin (residues E195–R206 and T229–I250 in actin and residues S277–V288 and V355–P359 in tubulin) and interact with either CCTβ or CCTε (Llorca et al., 2001a), which have also been shown to have the strongest binding affinity towards the two cytoskeletal proteins. Therefore, it is tempting to suggest that this is the first step in the recognition process, which is then followed by the cooperative binding of the rest of the CCT-interacting sites.

How does folding of actin and tubulin occur? Electron microscopy studies have shown that, upon ATP binding, CCT undergoes a conformational change in the apical domains that together act like an iris and close the chaperonin cavity using the helical protrusions, similar to what happens with the archaeal chaperonins [Plate 2(D and F)]. An important difference with the rest of the chaperonins is that genetic experiments on the ATP binding site of different CCT subunits have revealed that the conformational change in the apical domains is not concerted like in GroEL (Horovitz et al., 2001) but rather of a sequential nature, and probably starting at CCTα (Lin and Sherman, 1997). ATP binding and the subsequent and approximate 70° clockwise movement of the apical domain (Llorca et al., 1999a, 2001) would proceed sequentially throughout the CCT subunits [CCTα → CCTγ → CCTβ → CCTε; this is the order obtained with the mutational analysis of CCT subunits in yeast; Lin and Sherman, 1997; see arrows in Plate 2(C and E)]. This conformational change would push the substrate towards the centre of the cavity [Plate 2(D and F)]. In both actin and tubulin, the small and N-terminal domains would respectively be the first to undergo such forced movement, pushing them towards the large and C-terminal domains, the last to undergo the movement of the apical domain and the ones with strongest binding affinity (Llorca et al., 2001). When this occurs, the N-terminal domain has already been liberated from CCT and interacts with the C-terminal domain. The outcome of all this complex set of movements is that, unlike what happens with GroEL, actin and tubulin are not liberated in the CCT cavity. They remain bound to the apical domains of some of the CCT subunits, through interactions with their C-terminal domains, having acquired in the process their native conformations [Plate 2(D and F); Llorca et al., 2001]. This result from electron microscopy agrees with the finding by Hynes and Willison (2000) of a third CCT-binding domain in actin (site III) at the junction of the small and large domains, which could perhaps be used to stabilize its binding to CCT after the ATP-induced conformational change in the chaperonin. This interaction could be stabilized by the charged residues present in the inner surface of the chaperonin dome [Plate 3(B)]. Nothing is known about the mechanism of substrate liberation, but it is likely to occur subsequent to conformational changes that generate the re-opening of the
CCT cavity and/or with the help of some cofactors. Taken together, the results described here suggest that CCT has evolved from a chaperonin with a more passive folding mechanism, as is the case of GroEL, which liberates the substrate into the cavity allowing its auto-folding, into one with a more active mechanism whereby the ATP-induced movements of its apical domains force the folding of the two cytoskeletal proteins.

It is clear that CCT is stringent in the folding of both actin and tubulin (Willison, 1999), but what is the contribution of the above described process to the folding pathway of both cytoskeletal proteins? Is it possible that the eukaryotic chaperonin may have evolved to help both cytoskeletal proteins to overcome a kinetic barrier in their folding processes by stabilizing at first an open, quasinative conformation, and subsequently inducing the final steps in folding by placing the two topological domains of the two proteins in the correct position. Both cytoskeletal proteins share common structural features, with two topological domains connected by a linker, and a nucleotide binding site located between the two domains which holds most of the inter-domain bonds. It is tempting to suggest that the critical step in the folding of actin and tubulin may have to do with the correct folding of the nucleotide binding site. In the case of actin, it has been shown that the native conformation depends on the occupancy of the nucleotide binding site (Schüler et al., 2000), and biochemical studies with CCT and tubulin have shown that one of the functions of GTP is to stabilize the tubulin molecule during the folding process (Tian et al., 1995b).

An interesting point is that the putative CCT-binding sites of actin and tubulin are all located in exposed loops of their native structures (Plate 2A and B), and most of them are not present in their respective prokaryote homologues (MreB and FtsA in the case of actin, and FtsZ in the case of tubulin; Llorca et al., 2001a; van den Ent et al., 2001; Andreu et al., 2002). These bacterial proteins successfully fold themselves without the concourse of any chaperone, but they are not capable of generating sophisticated polymer systems like their eukaryotic counterparts. Since the actin filaments and the microtubules are a basic feature of the eukarya kingdom, and CCT seems to have evolved to complete in the early stages of eukaryote evolution (Kubota et al., 1994; Willison and Horwich, 1996; Willison, 1999), it has been hypothesized that CCT evolved from a primordial chaperonin to deal with the folding and regulation problems that appeared in eukaryotes at the same time as actin and tubulin evolved from their respective predecessors (Willison, 1999; Llorca et al., 2000). It is possible that the evolution of a complex CCT-dependent folding pathway for actins and tubulins generated cytoskeletal polypeptides with biophysical properties that differ from those of their bacterial relatives (Llorca et al., 2001a).

Other CCT-interacting proteins

It seems evident now that CCT interacts with more proteins than actin and tubulin (see Willison and Grantham, 2001; Dunn et al., 2001; Valpuesta et al., 2002 for a more comprehensive study on the proteins cited below). Some of these proteins are cytoskeletal proteins such as cofilin, the actin-depolymerizing factor-1 (ADF1), actin-related proteins (ARPs) and the heavy meromyosin subunit (HMM). Other CCT-binding proteins are involved in various and important cellular processes, such as G-protein (signal transduction) and its inhibitor phosducin-like protein (PhLP), Cyclin E (involved in the control of cell cycle), several histone deacetylases (HDACs, involved in the regulation of gene expression; Guenther et al., 2002), and the Von Hippel–Lindau tumour suppressor protein (VHL; Feldman et al., 1999), which is associated with a number of different types of cancer. Other proteins included viral proteins that are assisted in their folding by CCT, and a large number of proteins involved in many different cellular processes, and which have in common to contain WD40 motifs. Some of these proteins have molecular masses higher than 50kDa, and it is therefore unlikely that they are folded by CCT inside the cavity, which suggests that the cytosolic chaperonin may have a function other than its assistance in their folding.

Other CCT-interacting proteins include some molecular chaperones such as Hsp70 (Feldman et al., 1999), Hop/p60 (Gebauer et al., 1998). However, there is only structural information available on the interaction between CCT and PFD, which as discussed previously, binds to the cytosolic chaperonin to transfer unfolded actin and tubulin (Vainberg et al., 1998). Both CCT and PFD have evolved from their respective predecessors towards a hetero-oligomerization, probably forced by a specialization in their function. Thus, both the thermosome and the archaeal PFD are composed of a small number of different subunits (1–3 in the case of the thermosome, 2 in the case of PFD) and interact in vitro with any kind of unfolded polypeptide, whereas CCT and the eukaryotic PFD are composed of eight and six different subunits, respectively, and interact with a limited number of substrates (Martín-Benito et al., 2002). This evolution towards hetero-oligomerization seems to be accompanied by the interaction between specific CCT and PFD subunits. This interaction takes place at the level of the apical domains probably in the same region where substrate binding occurs. The notion that a specific interaction occurs between both chaperones is strengthened by the evolution of the putative chaperonin binding site of PFD, from a domain composed mostly of hydrophobic residues in the archaeal PFD, to a domain highly charged in the eukaryotic counterpart (Martín-Benito et al., 2002).

In conclusion, additional high-resolution structural and functional information on the interaction between CCT and its various substrates is needed before a clear picture of the substrate recognition mechanism and the folding cycle of this exciting molecular machine can be presented.

Acknowledgements

This work was supported by grants from the MCYT (J.M.V. and J.L.C.) and Fundación ‘Ramón Areces’ (P.G.-P). K.R.W. is supported by Cancer Research UK.


Copyright © 2004 John Wiley & Sons, Ltd.


Wang J, Boisvert DC. 2003. Structural basis for GroEL-assisted protein folding from the crystal structure of (GroEL-KMgATP)$_{14}$ at 2.0 Å resolution. *J. Mol. Biol.* 327: 843–855.


