doi: 10.1061/S0022-2836(02)00190-0 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 318, 1367–1379

JMB

Crystal Structure of the CCTγ Apical Domain: Implications for Substrate Binding to the Eukaryotic Cytosolic Chaperonin

Günter Pappenberger[1,2], Julie A. Wilsher[1], S. Mark Roe[1] Damian J. Counsell[2], Keith R. Willison[2] and Laurence H. Pearl[1]*

[1]*Chester Beatty Laboratories Section of Structural Biology and Cancer Research UK DNA Repair Enzyme Group, The Institute of Cancer Research 237 Fulham Road, London SW3 6JB, UK*

[2]*Cancer Research UK Centre for Cell and Molecular Biology The Institute of Cancer Research, 237 Fulham Road London SW3 6JB, UK*

**Corresponding author*

The chaperonin containing TCP-1 (CCT, also known as TRiC) is the only member of the chaperonin family found in the cytosol of eukaryotes. Like other chaperonins, it assists the folding of newly synthesised proteins. It is, however, unique in its specificity towards only a small subset of non-native proteins. We determined two crystal structures of mouse CCTγ apical domain at 2.2 Å and 2.8 Å resolution. They reveal a surface patch facing the inside of the torus that is highly evolutionarily conserved and specific for the CCTγ apical domain. This putative substrate-binding region consists of predominantly positively charged side-chains. It suggests that the specificity of this apical domain towards its substrate, partially folded tubulin, is conferred by polar and electrostatic interactions. The site and nature of substrate interaction are thus profoundly different between CCT and its eubacterial homologue GroEL, consistent with their different functions in general *versus* specific protein folding assistance.

Keywords: chaperone; chaperonin; protein folding; actin; tubulin

Introduction

Chaperonins are present in all three kingdoms of life and are grouped into two families, based on sequence similarity and structural characteristics: group I chaperonins,[1,2] found in eubacteria (e.g. GroEL in *Escherichia coli*) and eukaryotic organelles of eubacterial descent (e.g. Cpn60 in mitochondria and chloroplasts), and group II chaperonins,[3,4] found in archaea (thermosome) and the eukaryotic cytosol (CCT). Both groups share a common monomer architecture of three domains: an equatorial domain that carries ATPase activity, an intermediate domain, and an apical domain, involved in substrate binding.[5,6] The chaperonin monomers assemble into a characteristic double-toroidal quaternary structure of 2×7, 2×8, or 2×9 subunits.

Chaperonins bind non-native proteins to help them in achieving their native states. For this activity, they undergo a functional cycle that is driven by ATP hydrolysis and involves large movements of the substrate-binding apical domains.[2,7,8] Group I chaperonins (e.g. GroEL) are thought to recognise a broad range of non-native proteins through hydrophobic interactions[9] and provide them with a refolding opportunity inside the central cavity of the torus, protected from the crowded environment of the cell. They rely for their function on a co-chaperonin (e.g. GroES) that binds in a lid-like fashion onto the torus[10] and may be involved in displacing bound substrate into the cavity.

Group II chaperonins lack a GroES-like co-chaperonin. Instead, the crystal structure of the thermosome shows an integrated lid formed by the helical protrusions of the apical domains.[6] Little is known about the natural substrates and biological role of the thermosome, but its heat-shock induction suggests a role similar to that of GroEL.[3] In contrast, CCT is dedicated to the folding of only a few essential cellular proteins, including actin and tubulin.[11–14] CCT is further unique amongst the chaperonins for its complexity: all eight subunits of the torus (CCTα, β, γ, δ, ε, ζ, η, θ) are products of separate genes and have substantially different

Abbreviations used: CCT, chaperonin containing TCP-1; TriC, TCP-1 ring complex; TCP-1, tailless complex polypeptide 1.

E-mail address of the corresponding author: laurence@icr.ac.uk

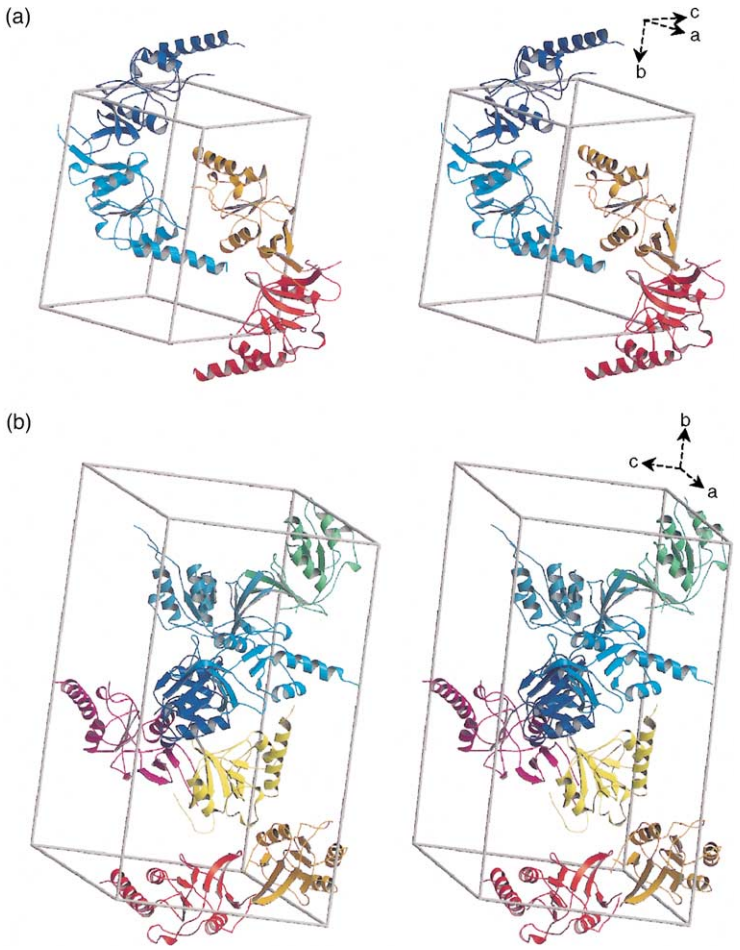

Figure 1. Crystal packing of the mouse CCTγ apical domain. Stereo view of the asymmetric units of the two crystal forms. (a) Triclinic (*P*1) crystal form. A non-crystallographic 2-fold screw axis runs in direction of the *c*-axis. (b) Monoclinic ($P2_1$) crystal form. The complete unit cell is generated using the 2-fold screw axis along the *b*-axis. Both crystal forms contain an essentially identical dimer with approximate 2-fold symmetry as the fundamental unit. See Table 1 for unit cell dimensions.

amino acid sequences.[15] Most divergent are their apical domains, responsible for substrate binding. The subunit divergence might thus be caused by CCT's specialisation towards few substrates.[16] This is now supported by electron microscopic studies of actin and tubulin bound to CCT.[7,8,17] These substrate proteins bind to CCT in defined geometries that involve interactions with specific subunits on CCT. To begin to understand the molecular basis of this substrate specificity, we determined the crystal structure of the CCTγ apical domain from mouse.

Table 1. Statistics for data collection and refinement

A. Data collection and processing				
Space group	*P*1		$P2_1$	
Unit cell dimensions	$a = 51.74$	$\alpha = 89.97$	$a = 60.24$	
a, *b*, *c* (Å)	$b = 65.02$	$\beta = 103.95$	$b = 234.23$	$\beta = 114.70$
α, β, γ (deg.)	$c = 65.47$	$\gamma = 90.35$	$c = 62.70$	
Molecules in asymmetric unit	4		8	
Solvent content (%)	52		49	
Resolution (highest res. shell) (Å)	29.0–2.20 (2.31–2.20)		29.0–2.80 (2.94–2.80)	
Mosaicity (deg.)	1.05		0.35	
Unique reflections	39,106		33,738	
$I/\sigma(I)$	8.0 (4.2)		6.3 (4.7)	
R_{merge}	0.058 (0.150)		0.051 (0.147)	
Redundancy	1.9 (1.8)		2.1 (2.0)	
Completeness (%)	93.0 (92.6)		85.2 (74.1)	
B. Refinement				
R_{cryst}/R_{free}	0.203/0.234		0.238/0.287	
Model				
Protein atoms	4815		8979	
Water molecules	645		45	
Others	Four glycerol molecules		2 Ca^{2+}	
r.m.s deviation				
Bonds (Å)	0.0097		0.0096	
Angles (deg.)	1.43		1.34	
In most favoured region of Ramachandran plot (%)	92.3		83.7	

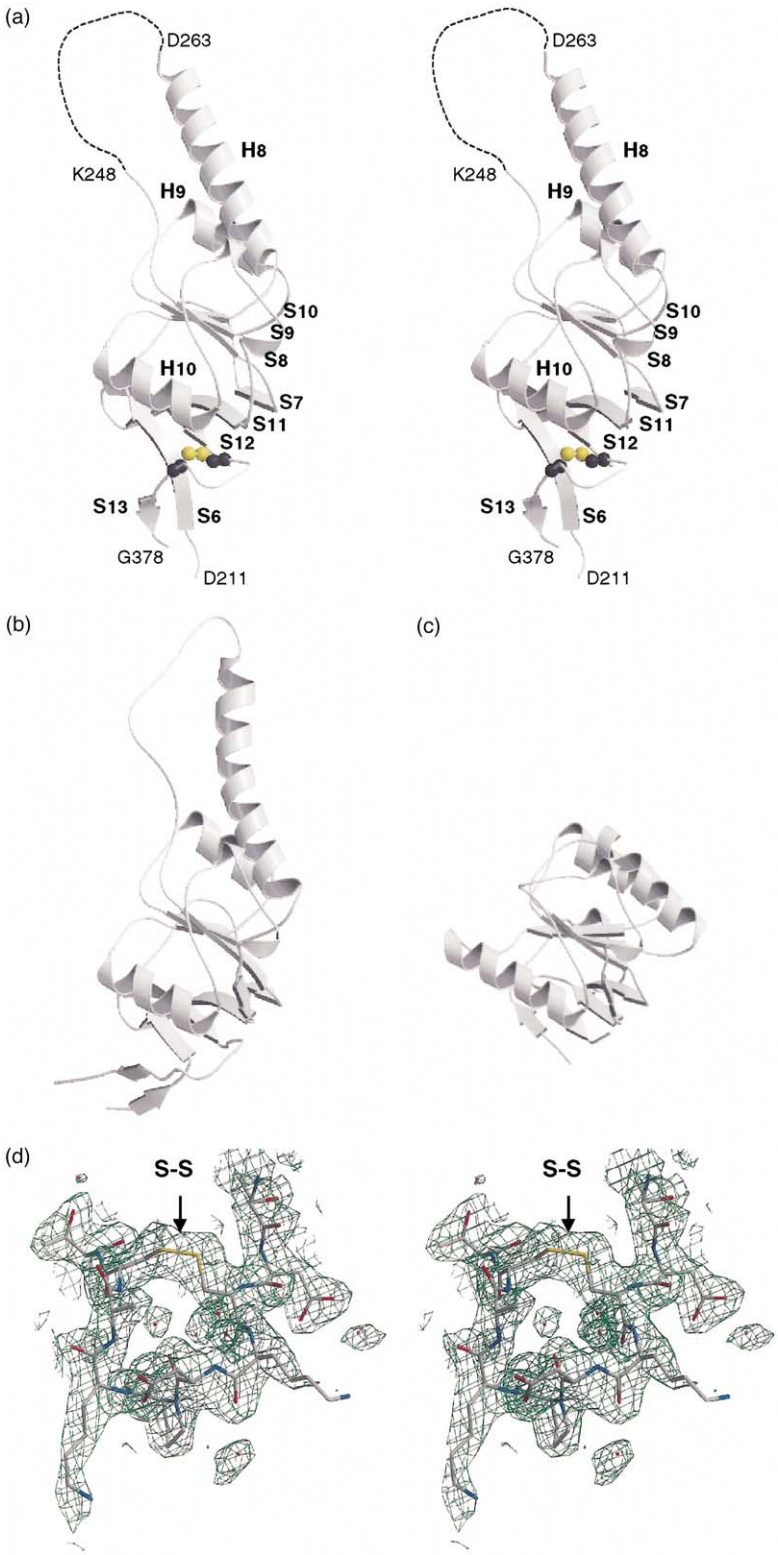

Figure 2. Structural comparison of apical domains from group II and group I chaperonins. (a) Stereo view of a ribbon diagram of the CCTγ apical domain structure. The secondary structure elements are numbered according to Braig *et al.*[46] α-Helices, H8–H10; β-strands, S6–S13. No interpretable electron density was observed for the N-terminal half of the helical protrusion (K248-D263). (b) β-Apical domain of the archaeal group II chaperonin thermosome. Coordinates were obtained from the crystal structure of the complete thermosome.[6] (c) Apical domain of the eubacterial group I chaperonin GroEL.[29] (d) Stereo view of the $2F_o - F_c$ electron density map from the triclinic crystal form, contoured at 1.0σ. The cysteine residues C366 and C372 are found to form a disulphide bridge (indicated), covalently closing the loop around P369 near the C terminus of the domain.

Results and Discussion

Structure of the CCTγ apical domain

The mouse CCTγ apical domain was crystallised from two substantially different crystallisation conditions in triclinic ($P1$) and monoclinic ($P2_1$) crystal forms with four and eight monomers per asymmetric unit, respectively (Figure 1(a) and (b); Table 1). Its structure was solved by molecular replacement with the apical domains from the archaeal homologue thermosome as search models.[18,19] Although the protein is predominantly monomeric in solution, a fraction of 5–10% was

detected as dimer in gel filtration chromatography and both crystal forms have an essentially identical dimer with 2-fold symmetry as the fundamental unit (Figure 1(a) and (b)). The hydrophobic dimer interface (L215, M219, I355, F360, F362, L375, L376) is largely equivalent to the expected interface to the intermediate domain in the holo-subunit.

The overall architecture of the mouse CCTγ apical domain (Figure 2(a)) is similar to that of the thermosome (Figure 2(b)) and both are related to their bacterial counterpart GroEL (Figure 2(c)). Two nearly orthogonal β-sheets form a β-sandwich that is flanked on either side by two long loops (between strands S6 and S7, and helix H10 and strand S11, respectively). The top sheet of the β-sandwich (S8, S9, S10) is covered by two helices (H8, H9), while the bottom sheet (S6, S7, S11, S12, S13) and helix H10 form the interface to the intermediate domain. The secondary structure arrangement in the core region of both group II chaperonins differs only in the region preceding the β-strand S11. Here, the short β-strand linking this region to the central β-sheet in the thermosome is missing in the CCTγ apical domain. This is due to replacement of the thermosomal sequence ^{343}GTAE^{346} by the longer and glycine-rich stretch ^{344}GTGAG^{348} in CCTγ. Further differences of secondary structure arrangement between the two group II chaperonins are at the termini, but those are likely to be due to the absence of the intermediate domain and consequent dimerisation in the CCTγ apical domain.

Interestingly, the electron density map shows a disulphide bridge linking the cysteine residues C366 and C372 in the CCTγ apical domain (Figure 2(d)). As disulphide bridges in intracellular proteins are rare, this may be an artefact of subsequent oxidation. However, both cysteine residues are highly conserved in the CCTγ sequences and the corresponding loop in the non-disulphide bonded thermosome is in a similar conformation (Figure 2(a) and (b)). The overall structure may thus bring the two cysteine residues near enough to let them form a bond even under cellular conditions. Reversible disulphide bridge formation in response to oxidative stress has been reported to regulate several intracellular proteins, including the chaperone Hsp33.[20]

Flexibility of the CCTγ helical protrusion

Despite the common structural core of the chaperonin apical domains, the thermosome is distinct from GroEL by virtue of its characteristic helical protrusion (Figure 2(b) and (c)). This helical protrusion is also found in the CCTγ apical domain (Figure 2(a)). It consists of an N-terminal extension of helix H8 (residues D263-K286) that reaches far out from the globular core region of the protein and is connected back to the protein core by a loop of about 15 residues (K248-E262; Figure 2(a) and (b)). In the crystal structure of the assembled thermosome, the non-helical parts of the helical protrusion of neighbouring subunits form a continuous β-sheet[6] and, together with helix H8, build a dome-like lid onto the torus. For the detached thermosome apical domains, the crystal structures show markedly different conformations in the non-helical region of the helical protrusions,[18,19] presumably influenced by extensive crystal contacts.

In the CCTγ apical domain, no interpretable electron density could be found for the non-helical part of the helical protrusion, leaving gaps of 11–25 residues at this position in the models of different molecules. On its N-terminal side, the electron density for the helical protrusion breaks off abruptly at residue K248 in all molecules, and the sequence region ^{248}KKGE^{251} may serve as a flexible hinge allowing various conformations of the helical protrusion. On the C-terminal side of the helical protrusion, in contrast, there is a gradual increase of the temperature factor and thus mobility, as one moves outwards on helix H8. Commonly, the electron density for helix H8 breaks off around residue D263, but in several cases (e.g. molecule D of the monoclinic crystal form) the electron density for helix H8 fades out one to two turns before reaching the outer tip at D263. Although the electron density did not allow us to extend the model in those cases, its course suggests that the chain continues without helical secondary structure. The N-terminal tip of helix H8 may thus be susceptible to localised unfolding by one or two turns. In contrast to the apical domain structures of the thermosome, the dimer interface in the CCTγ apical domain is formed by the end of the molecule opposite to the helical protrusion, and the helical protrusions are involved in only few crystal contacts (Figure 1(a) and (b)). This missing structural support from the crystal packing explains the lack of an ordered structure in the non-helical part of the helical protrusion. It suggests that parts of the helical protrusion have little propensity to form a defined conformation in solution. With this flexibility, the helical protrusions may provide different intersubunit interactions in the various stages of CCT's functional cycle,[19] rather than providing sites for specific substrate binding.

Sequence conservation and signature residues for the CCTγ apical domain

The radiation of the primordial CCT subunit gene into eight different subunits occurred uniquely and very early in eukaryotic evolution.[15,21–23] It has likely been driven by the specialisation of the different CCT subunits towards different functions. This divergence of subunits makes the CCT complex especially suitable for bioinformatic analysis to identify regions that were recruited for new functions. The residues critical for the specific function of a subunit will have a high degree of conservation for this subunit in different species (orthologues), but they will be

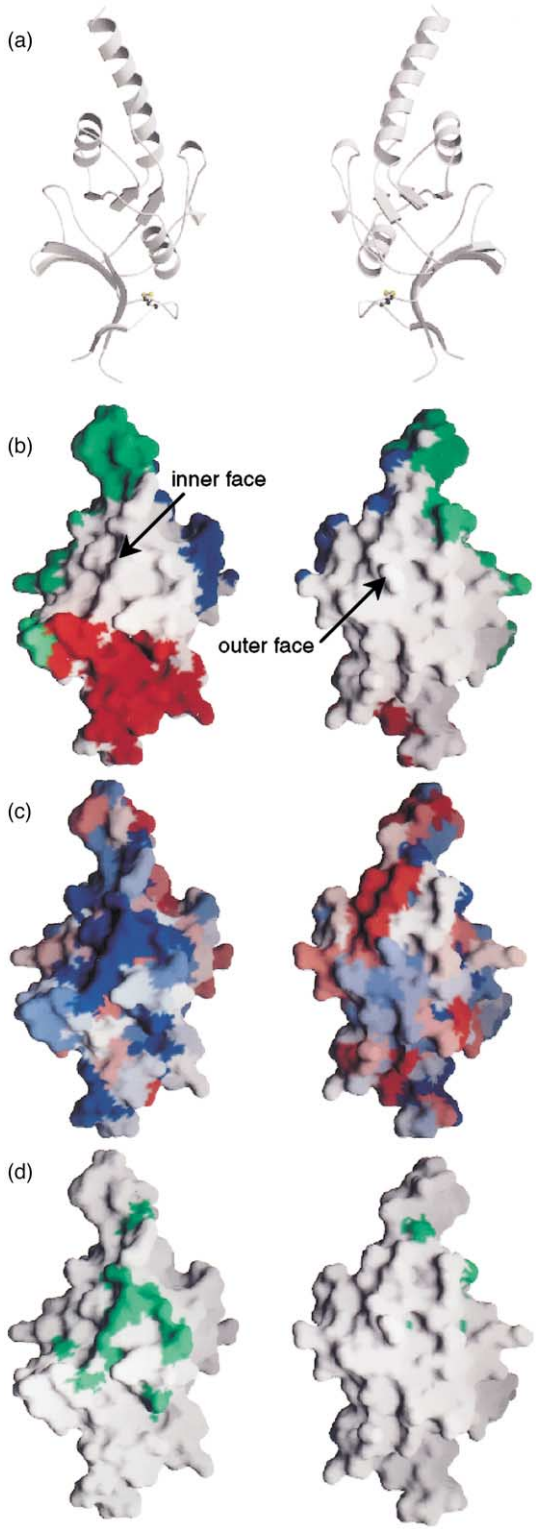

Figure 3. Identification of the substrate-binding region of the mouse CCTγ apical domain. In the left column, the protein regions lining the inside of the torus are facing the viewer. The orientation in the right column is rotated by 180°, showing the outside of the torus. (a) Ribbon diagram to illustrate the orientations of the molecule. (b) Surface regions interacting with neighbouring subunits (green, blue) or the intermediate domain of the same subunit (red) in the complete CCT complex. This has been modelled on the basis of the thermosome crystal structure[6] by superimposing the CCTγ apical domain onto the thermosome β-apical domain and labelling all atoms of the CCTγ apical domain closer than 5 Å to atoms of the surrounding thermosome structure. (c) Residue conservation, mapped onto the surface of the molecule (blue, conserved; red, variable). (d) Signature residues of the CCTγ apical domain (green), mapped onto its surface.

different in other subunits (paralogues). Those specifically conserved residues form a signature for a particular subunit. The wide range of available sequences for CCT apical domains, including some from primitive eukaryotes,[22] provides the sequence conservation patterns with a high statistical significance. In light of the new structural data on the CCTγ apical domain and new sequences for CCT subunits from additional species we revisited our original study on the signature residues in the different apical domains (D.J.C. & K.R.W., unpublished results).

Using the crystal structure of the assembled thermosome as the model,[6] the surface of the CCTγ apical domain was divided into regions involved in intersubunit or interdomain contacts, and into exposed regions facing either the inside or the outside of the torus (Figure 3(a) and (b)). The sequence conservation of CCTγ apical domains from 15 species, when mapped onto the surface of the mouse CCTγ apical domain structure (Figure 3(c)), shows a generally high variability for the residues on the outer face, while the inner face is highly conserved, with many residues being invariant. The proximal portion of the helical protrusion shares this discrimination of conserved inside and variable outside, but its distal region is generally more variable. The putative interdomain and intersubunit contacts show some degree of conservation, but to a much lesser extent than the exposed inner face, despite their obvious importance for subunit assembly. Focusing on the signature residues, the distinction between inner and outer face becomes even more prominent (Figure 3(d)). Those residues that are specifically conserved in the CCTγ apical domains and thus critical for its function, are almost exclusively found at the inner face. This gives clear evidence that it is the inner face which makes the CCTγ apical domain functionally different from the other apical domains. In accord with this interpretation, the inner face is composed predominantly of loop regions (Figures 3(a) and 4(b)). This region is therefore well suited to allow for specific sequence compositions in the different subunits. A recent related study on CCT evolution and subunit characteristics[23] shows the clustering of subunit signature residues at the inner face of CCT apical domains for all subunits.

The various CCT subunits have specialised towards specific binding properties for their non-native protein substrates.[7,8,17] The signature region on the CCTγ apical domains is expected to be

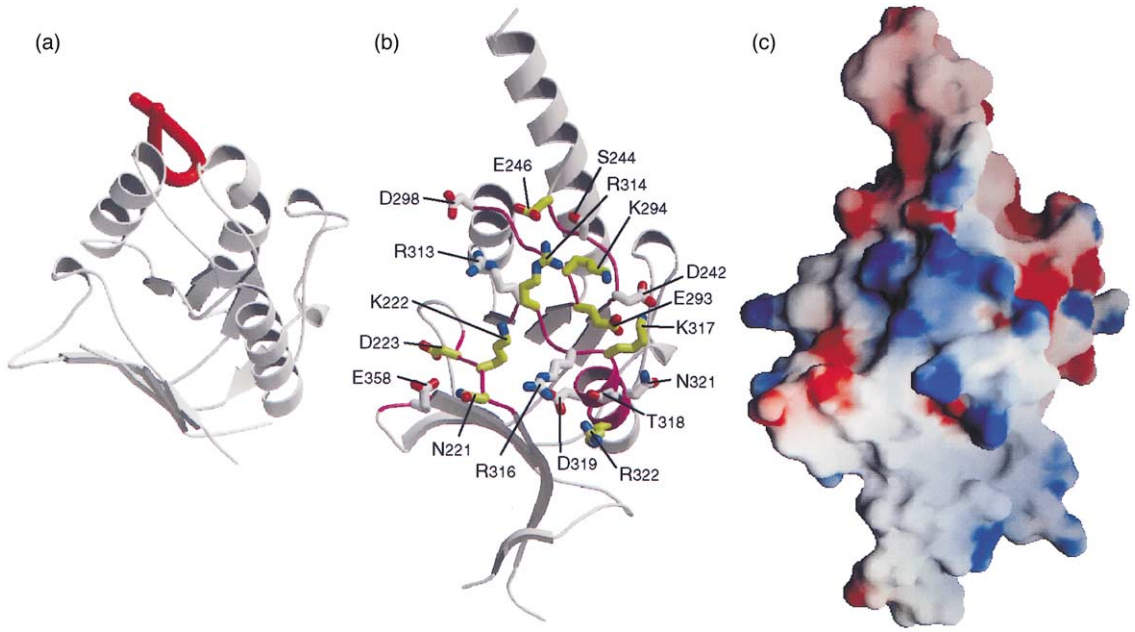

Figure 4. Properties of the substrate-binding regions in group I and group II chaperonins. (a) Structure of the apical domain from the eubacterial group I chaperonin GroEL.[29] A hydrophobic peptide (red) is bound in the groove between helices H8 and H9. (b) Structure of the mouse CCTγ apical domain. All side-chains on the surface of the substrate binding region are displayed in stick representation. Those residues identified as signature residues for CCTγ apical domain are coloured yellow (N221, K222, D223, E246, E293, K294, R314, K317, R322). Only three signature residues are not found in this region and are not displayed (Y247, E272, H302). Backbone stretches that are part of the substrate-binding region are coloured magenta. (c) Electrostatic potential on the surface of the mouse CCTγ apical domain, as calculated by the program GRASP.[44] The view is the same as in (b). The substrate-binding region consists of a large central patch with positive potential (blue), surrounded by several negative patches (red).

involved in the specific binding of its appropriate substrate, partially folded tubulin. Tubulin is one of the most highly conserved proteins,[24] and the absence of covariation in the CCT–tubulin interaction is consistent with the unusually high conservation of the accessible inner face of the CCTγ apical domain. Experimental evidence for the role of the putative substrate-binding region is provided by the electron microscopic reconstructions of tubulin binding to CCT[7,8] that show tubulin attached to the inner wall of the torus and in contact with the CCTγ apical domain. In the AMP–PNP-bound form of the CCT–tubulin complex,[8] CCT adopts the same closed conformation as seen in the thermosome crystal structure.[6] The inner face of the CCTγ apical domain, as predicted from the thermosome structure, is here fully presented to and interacting with the substrate protein.

Implications for substrate binding

The molecular details of substrate interaction are essential for the understanding of the mechanisms of chaperonin action. For the bacterial group I chaperonin GroEL, mutational analysis,[25] biophysical studies[26,27] and structures of peptide complexes[28,29] converge towards a binding site in a hydrophobic groove between the two helices H8 and H9 (Figure 4(a)). These two helices are an inherently flexible region of the GroEL molecule and the resulting plasticity of the hydrophobic groove was suggested to accommodate GroEL's promiscuous interactions with a broad range of non-native proteins.[29,30] However, there are no structural data on the interaction of GroEL with a physiological substrate and, since the GroES co-chaperonin also binds in the same groove between the two helices H8 and H9,[10] there is the possibility that many conclusions from peptide studies on GroEL have more relevance for GroES binding than substrate binding.[30,31]

In contrast to GroEL, the substrate-binding site on the CCTγ apical domain, identified by the sequence conservation patterns, is an exposed surface region facing the inside of the torus (Figure 4(b) and (c)). In the CCTγ apical domain, the groove between the helices H8 and H9 is largely filled by bulky residues (especially Y303) and offers little hydrophobic area for interaction. Interestingly, this region shows an increased flexibility at both the backbone and side-chain level (Figure 5(a) and (b)), similar to the situation in GroEL. The flexibility of helices H8 and H9 is thus likely to be an inherent property of this protein fold and may have been secondarily exploited in the GroEL system for its promiscuous binding properties. In comparison, the loop regions and side-chains constituting the substrate-binding site of the CCTγ apical domain are more rigid (Figure

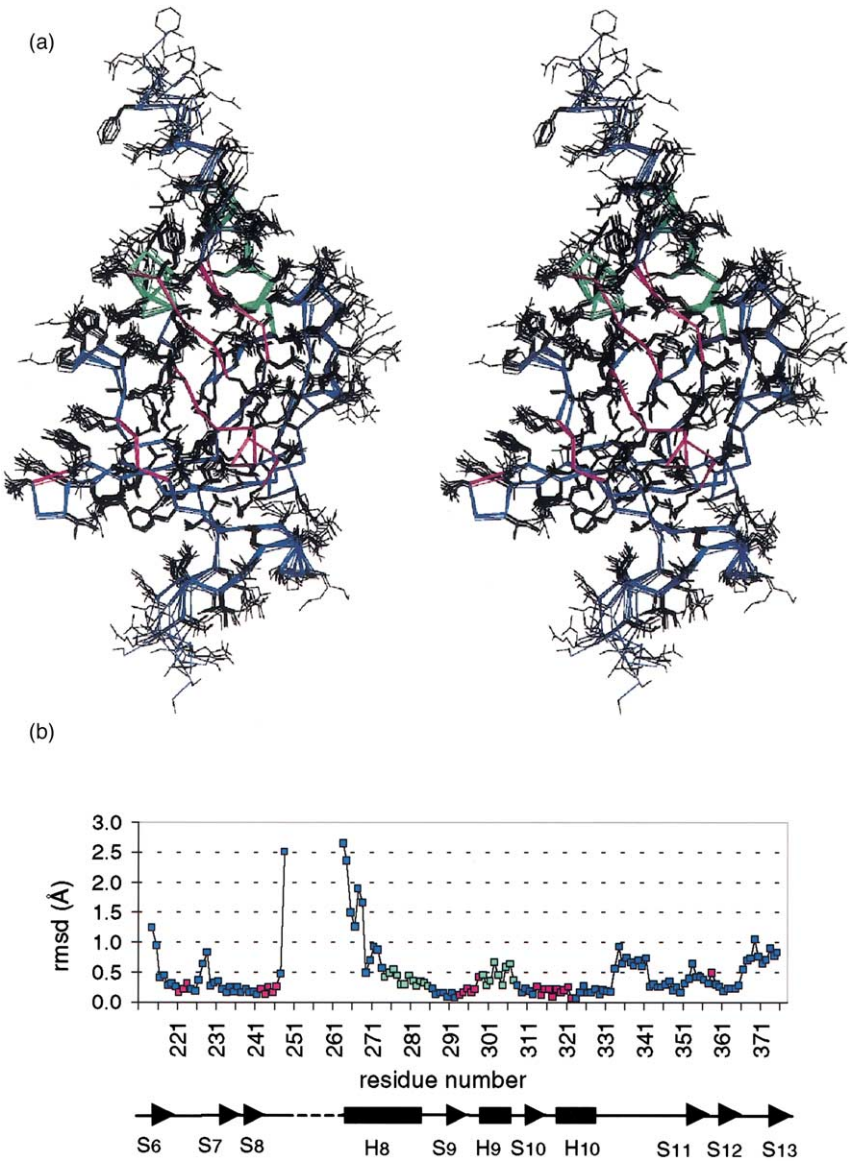

Figure 5. Conformational variability of the mouse CCTγ apical domain. (a) Stereo view of the superposition of all 12 molecules from the two crystal forms, showing the C^{α}-traces (blue) and the side-chains (black). Backbone regions corresponding to the substrate binding site of the CCTγ apical domain are rendered magenta, while the equivalent regions for the GroEL substrate binding site are rendered green. The view is the same as in Figure 4. All C^{α} atoms were used for the superposition, except those near the termini and in the helical protrusion. (b) Average root-mean-square deviation of the C^{α} atom positions of three independently refined CCTγ apical domain molecules (molecules A and C of the triclinic crystal form, and molecule G of the monoclinic crystal form). Colour code is as for (a). A diagram of the location of the secondary structure elements is given below the residue number.

5(a) and (b)), consistent with CCT's more specific interactions with its substrates.

As well as their location and their flexibility, the physical properties of the substrate binding sites in GroEL and CCTγ are fundamentally different: nearly all residues in the substrate binding region of CCTγ are charged. In addition to the signature residues defined by sequence conservation (N221, K222, D223, E246, E293, K294, R314, K317, R322), this also applies to the further side-chains on this surface (D242, S244, D298, R313, R316, T318, D319, N321, E358; Figure 4(b)). As the identification of our signature residues was based on very strict criteria, it is likely that some further residues in this region are involved in substrate binding despite not attaining signature residue status. The differences between the substrate-binding sites in GroEL and CCT in terms of their location, flexibility and physical properties could indicate that they employ different molecular mechanisms to achieve their particular cellular roles.

Judged from the properties of the substrate-binding site, the interaction of CCTγ with its substrate is driven by polar and electrostatic rather than hydrophobic interaction. Similarly, the tubulin-specific chaperone Rbl2p, binding β-tubulin immediately after its release from CCT, was also found to lack hydrophobic surface areas and thus it presumably binds β-tubulin *via* polar interactions.[32] Further, biochemical studies of CCT–substrate interactions indicate that polar surface regions on the substrate proteins are involved in binding to CCT.[33] This is consistent with the fundamental functional difference between CCT and GroEL, their substrate specificity. CCT has to recognise a specific partially folded protein rather than the non-specific property of a protein being unfolded. The specific substrate binding of CCT may thus involve polar interactions and

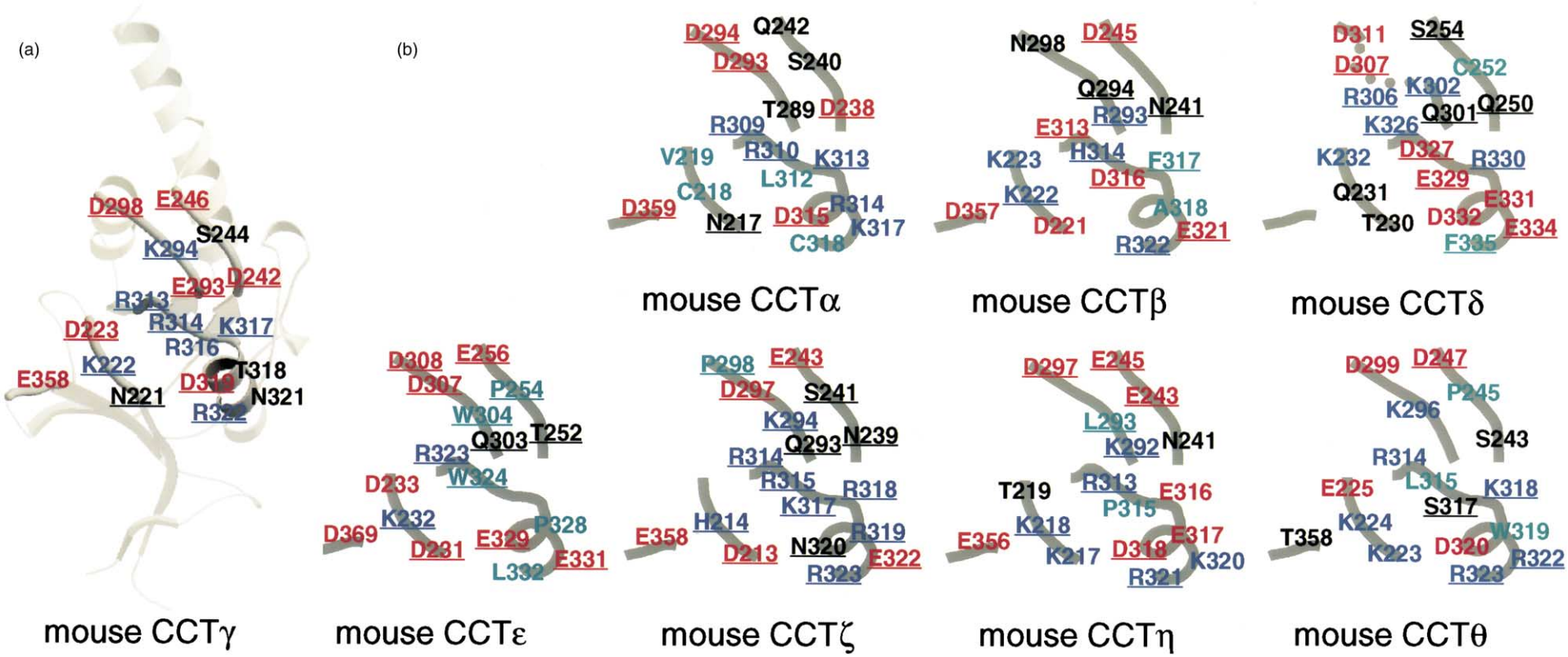

Figure 6. Comparison of the apical domains from all eight CCT subunits, with focus on the region lining the inside of the CCT torus. (a) Structure of the mouse CCTγ apical domain, with the surface residues of the inner face indicated in their approximate positions. (b) Model of the composition of the inner face in the other CCT apical domains from mouse. The position of the loops comprising the inner face has been taken from the CCTγ apical domain structure, with exception of the loop Q301-D311 in CCTδ, where there is an insertion of five residues that has been indicated by a dotted line. The sequence alignment used to determine the signature residues was also used to identify those residues in the other CCT subunits that correspond to the residues on the inner face of the CCTγ apical domain. These residues are indicated in their approximate positions and colour-coded according to their physical properties (green, hydrophobic; black, polar; red, negatively charged; blue, positively charged). Underlined residues are conserved in each subunit in a number of organisms from yeast to human (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, *Homo sapiens*).

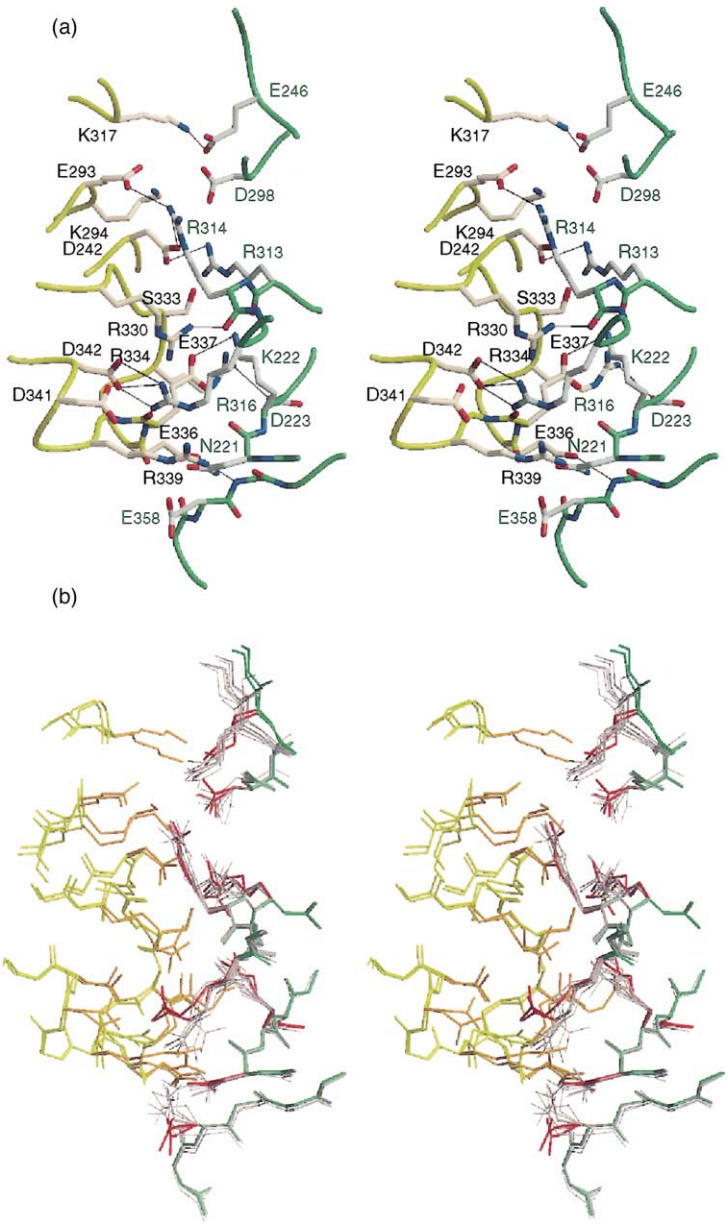

Figure 7. Nature of crystal packing interactions at the substrate-binding region of mouse CCTγ apical domain. (a) Stereo view of a crystal contact in the monoclinic crystal form, between the substrate-binding region of molecule F (green, right) and the loop R330-D342 as well as a peripheral region of the substrate-binding site of molecule A (gold, left). All side-chains or backbone regions with atoms closer than 3.5 Å to atoms of the neighbouring molecule are displayed in stick representation. Interacting atoms are connected by black lines. Charged side-chains are displayed in stick representation if there is a charge interaction with the neighbouring molecule of less than 4.5 Å distance. (b) Conformational changes upon interaction at the substrate-binding region. All 12 molecules of both crystal forms are superimposed and displayed in the same orientation as in (a). Ten molecules show no crystal contact at the substrate-binding site (grey), while two molecules (green, backbone; red, side-chains, molecules F and H of the monoclinic crystal form) interact *via* this region with neighbouring molecules in the crystal (golden, backbone, orange, side-chains; molecules A and C of the monoclinic crystal form). Besides several changes in side-chain conformations, there is a 1.5 Å backbone shift around E246 (upper right corner) upon interaction at the substrate-binding site.

electrostatic complementarity as recognition mechanisms. The calculated electrostatic surface potential for the CCTγ apical domain (Figure 4(c)) suggests that the non-native tubulin will provide several negatively charged side-chains complementary to the central positively charged patch on the inner face (K222, K294, R313, R314, R316, K317). Based on sequence alignments of all apical domains, the substrate-binding region retains its predominantly charged property in the other apical domains, but the distribution of positive and negative charges varies strongly (Figure 6(a) and (b)). A notable exception is CCTζ, which, like CCTγ retains all six positive charges at the central positive patch (Figure 6(b)). The electron microscopic reconstruction shows that CCTγ and CCTζ interact with the same region of tubulin.[7] This confirms the importance of the central positive patch on the inner face for the specificity of tubulin binding. It is noteworthy that the thermosome apical domains possess the same patch of six positively charged residues on their inner face. CCTγ and CCTζ may have retained this feature from the primordial CCT subunit,[22] while the other subunits diverged further during specialisation of their functions. Further studies on the function of the thermosome are necessary to decide if those residues happened to be recruited in CCTγ and CCTζ for a specific function, or whether they also fulfil a specific and possibly even similar role in the thermosome.

A model of how a stable protein–protein interaction can take place involving the substrate binding site of the CCTγ apical domain is provided by a crystal packing interaction in the monoclinic crystal form. Most of the substrate-binding site of

molecule F is found in intimate contact with the long loop linking helix H10 and β-strand S11 of molecule A (R330-D342) and a peripheral region of molecule A's substrate-binding site (Figure 7(a); a similar interaction is found between molecules H and C of this crystal form). Apart from asparagine 221 on chain F (N221:F) and serine 333 on chain A (S333:A), all side-chains in intermolecular interactions are charged. With few exceptions, most notably the stacking of the delocalised electron systems of the arginine residues R330:A, R316:F, and R339:A, there is a clear dominance of electrostatic interactions at the substrate-binding region and charge complementarity of the binding partners underlies their close contact. The positive residues in the centre of the substrate-binding site of molecule F find negatively charged interaction partners on molecule A (K222:F-E337:A, R313:F-D242:A, R314:F-E293:A, R316:F-D342:A) and several negatively charged side-chains on the perimeter of the inner face of molecule F interact with positive residues on the binding partner (E246:F-K317:A, D298:F-K294:A, D223:F-R334:A).

The role of CCT in the folding of actin and tubulin is not limited to simple binding and release, but rather involves conformational rearrangements of both CCT and substrate that have yet to be elucidated in detail. With crystallographically independent structures of 12 molecules of the CCTγ apical domain at our disposal, we can assess the ability of the substrate-binding site to adapt to the interactions fortuitously provided by the crystal contact (Figure 7(b)). Several side-chains of the substrate-binding site show characteristic altered orientations if involved in interactions (e.g. R316, E358), but most backbone conformation is largely unaffected by the interactions. There is, however, a notable movement of the backbone around E246 by 1.5 Å. The side-chains of R313 and E246 are within 4 Å of each other in the absence of a crystal contact, but move apart as both find alternative interaction partners in the crystal contact. This backbone shift is propagated C-terminal of E246 until the gap in the model at the ill-defined N-terminal half of the helical protrusion after K248. Interactions at the substrate-binding site can thereby alter the conformation of the flexible helical protrusion region, which is involved in intersubunit contacts. This may be the first glimpse of an allosteric response that spreads the news of tubulin binding around the torus of CCT.

Materials and Methods

Protein expression and purification

The construct used in this study comprises residues E210-S380 of the mouse CCTγ subunit plus six C-terminal histidine residues. It was heterologously expressed from a pET11d vector in *E. coli* BL21(DE3)-pLysS. Bacteria were grown in LB medium containing 50 μM ampicillin in an Erlenmeyer flask at 37 °C and 225 rpm shaking. Expression was induced by addition of 0.5 mM IPTG once the cultures reached an $A_{600\ nm}$ of 0.5 and growth was continued at 16 °C overnight. The cells were harvested by centrifugation and resuspended in lysis buffer (20 mM Tris–HCl (pH 8.0), 0.5 mM imidazole, Complete™ EDTA-free protease inhibitor (Roche)) at 1/40th of the original culture volume. The cells were disrupted by sonication and the crude lysate was clarified by centrifugation at 12,000*g* for 40 minutes. This and all further steps during purification were performed at 4 °C. The supernatant was loaded onto a column of 60 ml Talon® resin (Clontech), pre-equilibrated with 20 mM Tris–HCl (pH 8.0), 0.5 mM imidazole, at a flow-rate of 1 ml min^{-1}. The column was washed with about ten column volumes of 20 mM Tris–HCl (pH 8.0), 0.5 mM imidazole. The recombinant protein was eluted with 20 mM Tris–HCl (pH 8.0), 300 mM imidazole, 500 mM NaCl and the fractions containing the mouse CCTγ apical domain were pooled. This pool was separated on a Superdex 75 HighLoad™ 26/60 gel-filtration column (Pharmacia), pre-equilibrated in 20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 0.5 mM EDTA and run at 2 ml min^{-1} in the same buffer. Fractions corresponding to monomeric mouse CCTγ apical domain were pooled and concentrated to about 20 mg ml^{-1} *via* Centriprep YM-10 columns. The purified protein was aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

Crystallisation

The mouse CCTγ apical domain was crystallised in the triclinic *P*1 space group by equilibration of a 2 μl drop of 6 mg ml^{-1} protein in 50 mM Tris (pH 8.0), 300 mM NaCl, 5 mM $Mg(OAc)_2$, 0.2 mM EDTA, 10% (v/v) glycerol against a reservoir of 100 mM Tris (pH 8.0), 200 mM NaCl, 10 mM $Mg(OAc)_2$ in a hanging drop setup at 14 °C. After one week, clusters of plates appeared. By microseeding on day 4 after setup, those clusters were improved to single thick plates of up to 200 μm × 200 μm × 30 μm. For data collection the crystals were transferred into cryobuffer of 50 mM Tris (pH 8.0), 4 mM $Mg(OAc)_2$, 30% glycerol in four steps of increasing glycerol concentration. Crystals of the monoclinic $P2_1$ space group were obtained by mixing 1 μl of protein solution (30 mg ml^{-1} protein, 8 mM Tris (pH 8.0), 400 mM NaCl, 0.4 mM EDTA, 20% glycerol) with 1 μl of buffer (100 mM sodium cacodylate (pH 6.5), 14% (w/v) PEG 8000, 40 mM $Ca(OAc)_2$, 40% glycerol) in a microbatch setup under mineral oil at 14 °C. Crystals appeared after several weeks as brick-shaped blocks of up to 200 μm × 200 μm × 400 μm and were frozen directly from the crystallisation buffer.

Data collection and structure determination

Diffraction data were collected on a CCD detector (ADSC) at beamline 9.6 at the SRS in Daresbury for the triclinic crystal form, and at beamline 14.4 at the ESRF in Grenoble for the monoclinic crystal form. Both data collections were performed in a cold nitrogen stream at 100 K. Diffraction data were processed with MOSFLM[34] and sorted, merged, scaled and truncated using the CCP4 suite of programs.[35] Data collection statistics are provided in Table 1.

The structure of the triclinic crystal form was solved by molecular replacement using the program AMoRe[36] with both apical domain structures from the thermosome as search models (RCSB PDB codes 1ASS, 1E0R)[18,19] in both full and truncated forms (without the helical

protrusion). Due to the four molecules in the asymmetric unit, no clear distinction for the correct solution was obvious in the initial rotation search, and owing to its triclinic crystal form, no translational search could be used to improve this distinction. However, a few solutions scored consistently high for all four search models used. Further, two pairs of orientations from this set of rotation solutions were found to be in accordance with the strong non-crystallographic 2-fold symmetry along the *c*-axis (Figure 1(a)), as indicated in the self-rotation function. Consistency of the search results for the four models was again used to determine the translational part of the solution. The final arrangement of the four molecules was validated by their good packing in the crystal and its consistency with weaker non-crystallographic 2-fold symmetries from the self-rotation functions that relate the monomers in the dimers (Figure 1(a)). Model building with the program O[37] was started from the thermosome β-apical domain structure (RCSB PDB code 1E0R)[19] lacking the helical protrusion, the residues corresponding to ^{344}GTGAG^{348}, and all but few conserved side-chains. The monoclinic crystal form was solved by molecular replacement with AMoRe[36] using a preliminary model of the CCTγ dimer from the triclinic crystal form as search model. Three dimers could be unambiguously located, and the $F_o - F_c$ electron density map clearly showed positive electron density for a fourth dimer. This fourth dimer (molecules G and H) was docked manually into the electron density. Both crystal forms were refined using CNS[38] with 5% of the reflections omitted to calculate R_{free}. Water molecules were added at later stages of refinement and validated by a drop in R_{free} and visual inspection in O.[37] No interpretable electron density was observed for the N-terminal half of the helical protrusion. The termini are ordered to different degrees in the different monomers. Two to seven residues at the N terminus and two to five residues at the C terminus plus the His-tag lack interpretable electron density and were omitted from the model. In the triclinic unit cell, strong non-crystallographic symmetry restraints were used at later refinement stages to couple the pairs of molecules related by the 2-fold axis along the *c*-axis. Molecules A and B as well as molecules C and D of the triclinic crystal form are thus virtually identical in large regions of their structure. In the monoclinic unit cell, non-crystallographic symmetry restraints were used to relate the core regions of all eight molecules throughout the refinement. Refinement statistics are provided in Table 1.

The quality of the structures was assessed with the programs WHATIF[39] and PROCHECK.[40] Figures were drawn with MOLSCRIPT,[41] BOBSCRIPT,[42] Raster3D[43] and GRASP.[44]

Sequence analysis

The conservation score was derived with the program AMAS,[45] using an alignment of CCTγ apical domains from 15 species (*Homo sapiens*, *Mus musculus*, *Xenopus laevis*, *Drosophila melanogaster*, *Lepeophtheirus salmonis*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Trichomonas vaginalis*, *Leishmania major*, *Tetrahymena pyriformis*, *Thalassiosira weissflogii*, *Oxytricha granulifera*, *Giardia lamblia*).

Residues were defined as signature residues if they are both (i) absolutely conserved for CCTγ apical domains from all species (excluding *T. vaginalis*, the sequence of which was found to be atypical in several stretches) and (ii) sufficiently divergent between the eight CCT subunits. As a quantitative criterion for divergence, a similarity score of less than or equal to 50% was used, as derived from the program AMAS[45] (with the predefined AMAS matrix "ch.pt") on a global alignment of all eight CCT subunits from three species (*M. musculus*, *C. elegans*, *S. cerevisiae*). Forty-one of the 172 residues of the CCTγ apical domain are absolutely conserved in the 14 above-mentioned organisms (excluding *T. vaginalis*). Twenty-four of those residues have been classified as signature residues. Of those, ten are completely buried in the structural core of the CCTγ apical domain and two are found in the N-terminal half of the helical protrusion not resolved in this structure. This leaves 12 signature residues on the surface of the CCTγ apical domain structure: N221, K222, D223, E246, Y247, E272, E293, K294, H302, R314, K317, R322. The protein sequence alignments used for the sequence conservation and signature residue analysis can be obtained directly from the authors.†

Protein Data Bank accession codes

Coordinates for both crystal forms have been deposited in the RCSB Protein Data Bank with accession codes 1GML for the triclinic crystal form, and 1GN1 for the monoclinic crystal form.

Acknowledgements

We thank Julie Grantham for providing the plasmid construct of the CCTγ apical domain and the staff at SRS and ESRF for help with the data collection. We gratefully acknowledge the support of the Cancer Research UK (K.R.W.), The Institute of Cancer Research Structural Biology Initiative (L.H.P.), a Cancer Research UK postdoctoral fellowship to G.P./K.R.W. and a Marie Curie Fellowship of the European Community programme "Quality of Life and Management of Living Resources" under contract number QLK3-CT-2000-51423 to G.P./L.H.P.

References

1. Bukau, B. & Horwich, A. L. (1998). The Hsp70 and Hsp60 chaperone machines. *Cell*, **92**, 351–366.
2. Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A. & Horwich, A. L. (1998). Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* **67**, 581–608.
3. Gutsche, I., Essen, L. O. & Baumeister, W. (1999). Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. *J. Mol. Biol.* **293**, 295–312.
4. Willison, K. R. & Grantham, J. (2001). The roles of cytosolic chaperonin CCT in normal eukaryotic cell growth. In *Molecular Chaperones: Frontiers in Molecular Biology* (Lund, P., ed.), pp. 90–118, Oxford University Press, Oxford.
5. Braig, K., Otwinowski, Z., Hedge, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. *et al.* (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature*, **371**, 578–586.

† http://www.icr.ac.uk/pappenberger/

6. Ditzel, L., Löwe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R. *et al.* (1998). Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell*, **93**, 125–138.
7. Llorca, O., Martin-Benito, J., Ritco-Vonsovici, M., Grantham, J., Hynes, G. M., Willison, K. R. *et al.* (2000). Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J.* **19**, 5971–5979.
8. Llorca, O., Martin-Benito, J., Grantham, J., Ritco-Vonsovici, M., Willison, K. R., Carrascosa, J. L. *et al.* (2001). The "sequential allosteric ring" mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. *EMBO J.* **20**, 4065–4075.
9. Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. & Hartl, F. U. (1999). Identification of *in vivo* substrates of the chaperonin GroEL. *Nature*, **402**, 147–154.
10. Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL–GroES–$(ADP)_7$ chaperonin complex. *Nature*, **388**, 741–750.
11. Yaffe, M. B., Farr, G. W., Miklos, D., Horwich, A. L., Sternlicht, M. L. & Sternlicht, H. (1992). TCP1 complex is a molecular chaperone in tubulin biogenesis. *Nature*, **358**, 245–248.
12. Lewis, V. A., Hynes, G. M., Zheng, D., Saibil, H. & Willison, K. (1992). T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. *Nature*, **358**, 249–252.
13. Gao, Y., Thomas, J. O., Chow, R. L., Lee, G. H. & Cowan, N. J. (1992). A cytoplasmic chaperonin that catalyzes β-actin folding. *Cell*, **69**, 1043–1050.
14. Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J. S., Tempst, P. & Hartl, F. U. (1992). Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J.* **11**, 4767–4778.
15. Kubota, H., Hynes, G., Carne, A., Ashworth, A. & Willison, K. (1994). Identification of six Tcp-1-related genes encoding divergent subunits of the TCP-1-containing chaperonin. *Curr. Biol.* **4**, 89–99.
16. Kim, S., Willison, K. R. & Horwich, A. L. (1994). Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. *Trends Biochem. Sci.* **19**, 543–548.
17. Llorca, O., McCormack, E. A., Hynes, G., Grantham, J., Cordell, J., Carrascosa, J. L. *et al.* (1999). Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature*, **402**, 693–696.
18. Klumpp, M., Baumeister, W. & Essen, L. O. (1997). Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell*, **91**, 263–270.
19. Bosch, G., Baumeister, W. & Essen, L. O. (2000). Crystal structure of the β-apical domain of the thermosome reveals structural plasticity in the protrusion region. *J. Mol. Biol.* **301**, 19–25.
20. Aslund, F. & Beckwith, J. (1999). Bridge over troubled waters: sensing stress by disulfide bond formation. *Cell*, **96**, 751–753.
21. Kubota, H., Hynes, G. & Willison, K. (1995). The chaperonin containing t-complex polypeptide 1 (TCP-1). Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. *Eur. J. Biochem.* **230**, 3–16.
22. Archibald, J. M., Logsdon, J. M., Jr & Doolittle, W. F. (2000). Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol. Biol. Evol.* **17**, 1456–1466.
23. Archibald, J. M., Blouin, C. & Doolittle, W. F. (2001). Gene duplication and the evolution of group II chaperonins: implications for structure and function. *J. Struct. Biol.* **135**, 157–169.
24. Doolittle, R. F. (1995). The origins and evolution of eukaryotic proteins. *Phil. Trans. Roy. Soc. ser. B*, **349**, 235–240.
25. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. (1994). Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*, **371**, 614–619.
26. Tanaka, N. & Fersht, A. R. (1999). Identification of substrate binding site of GroEL minichaperone in solution. *J. Mol. Biol.* **292**, 173–180.
27. Kobayashi, N., Freund, S. M., Chatellier, J., Zahn, R. & Fersht, A. R. (1999). NMR analysis of the binding of a rhodanese peptide to a minichaperone in solution. *J. Mol. Biol.* **292**, 181–190.
28. Buckle, A. M., Zahn, R. & Fersht, A. R. (1997). A structural model for GroEL–polypeptide recognition. *Proc. Natl Acad. Sci. USA*, **94**, 3571–3575.
29. Chen, L. & Sigler, P. B. (1999). The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell*, **99**, 757–768.
30. Feltham, J. L. & Gierasch, L. M. (2000). GroEL–substrate interactions: molding the fold, or folding the mold? *Cell*, **100**, 193–196.
31. Shewmaker, F., Maskos, K., Simmerling, C. & Landry, S. J. (2001). The disordered mobile loop of GroES folds into a defined β-hairpin upon binding GroEL. *J. Biol. Chem.* **276**, 31257–31264.
32. Steinbacher, S. (1999). Crystal structure of the post-chaperonin β-tubulin binding cofactor Rbl2p. *Nature Struct. Biol.* **6**, 1029–1032.
33. Hynes, G. M. & Willison, K. R. (2000). Individual subunits of the eukaryotic cytosolic chaperonin mediate interactions with binding sites located on subdomains of β-actin. *J. Biol. Chem.* **275**, 18985–18994.
34. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter Protein Crystallog.* **26**.
35. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
36. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
37. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjelgaard, M. (1991). Improved methods for binding protein models to electron density maps and the localization of errors in these models. *Acta Crystallog. sect. A*, **42**, 110–119.
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
39. Vriend, G. (1990). WHATIF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56.
40. Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
41. Kraulis, J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallog.* **24**, 946–950.

42. Esnouf, R. M. (1999). Further additions to MOLSCRIPT version 1.4, including reading and contouring of electron density maps. *Acta Crystallog. sect. D*, **55**, 938–940.
43. Merrit, E. & Bacon, D. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
44. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
45. Livingstone, C. D. & Barton, G. J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *CABIOS*, **9**, 745–756.
46. Braig, K., Adams, P. D. & Brünger, A. T. (1995). Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nature Struct. Biol.* **2**, 1083–1094.

Edited by W. Baumeister

(Received 15 November 2001; received in revised form 27 February 2002; accepted 1 March 2002)