Published in final edited form as:
Cell. 2014 November 20; 159(5): 1042–1055. doi:10.1016/j.cell.2014.10.042.

The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT

Lukasz A. Joachimiak[§,1,^], **Thomas Walzthoeni**[#,@], **Corey Liu**[‡], **Ruedi Aebersold**[#,£], and **Judith Frydman**[§,^]

[§]Department of Biology and Genetics, Stanford University, Stanford, CA 94305, USA [‡]Stanford Magnetic Resonance Lab, Stanford University, Stanford, CA 94305, USA [#]Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, 8093 Zurich, Switzerland [@]Ph.D. Program in Molecular Life Sciences, University of Zurich/ETH Zurich, 8057 Zurich, Switzerland [£]Faculty of Science, University of Zurich, Zurich, Switzerland

Summary

The eukaryotic chaperonin TRiC (also called CCT) is the obligate chaperone for many essential proteins. TRiC is hetero-oligomeric, comprising two stacked rings of eight different subunits each. Subunit diversification from simpler archaeal chaperonins appears linked to proteome expansion. Here, we integrate structural, biophysical and modeling approaches to identify the hitherto unknown substrate-binding site in TRiC and uncover the basis of substrate recognition. NMR and modeling provided a structural model of a chaperonin-substrate complex. Mutagenesis and crosslinking-mass spectrometry validated the identified substrate binding interface and demonstrate that TRiC contacts full-length substrates combinatorially in a subunit-specific manner. The binding site of each subunit has a distinct, evolutionarily conserved, pattern of polar and hydrophobic residues specifying recognition of discrete substrate motifs. The combinatorial recognition of polypeptides broadens the specificity of TRiC and may direct the topology of bound polypeptides along a productive folding trajectory, contributing to its unique ability to fold obligate substrates.

[^]Correspondence to: jfrydman@stanford.edu and lajoachi@gmail.com.
[1]current address: Pacific Biosciences, 1380 Willow Street, Menlo Park, CA 94025, USA

Author Contributions LAJ and JF conceived the project, designed and interpreted experiments. LAJ cloned, expressed and purified all apical domain proteins, performed SPR experiments. modeling calculations and prepared the TRiC-substrate complexes for XL-MS. LAJ and CL performed all NMR experiments. TW performed XL-MS experiments; TW and RA carried out the XL-MS analysis and interpreted the data. LAJ and JF wrote the manuscript. All authors contributed to the preparation of the manuscript.

The authors declare no competing financial interest.

Supplemental Information
Supplemental Information includes 4 Supplemental Tables and 7 Supplemental Figures and can be found online at X.

Introduction

The health and integrity of the cellular proteome depends on molecular chaperones, which through their distinct substrate specificities and modes of action maintain protein homeostasis (Balch et al., 2008; Kim et al., 2013; Li and Buchner, 2013; Saibil, 2013). Among these, the eukaryotic chaperonin TRiC (for TCP-1 Ring Complex, also called CCT for Chaperonin Containing TCP1) is distinguished by its complex architecture and mechanism, which allow it to fold a subset of essential and topologically complex proteins, including cell cycle regulators, signaling proteins and cytoskeletal components (Bigotti and Clarke, 2008; Kim et al., 2013).

TRiC/CCT is a large hetero-oligomeric ATP-dependent complex consisting of two eight-membered rings stacked back to back (Bigotti and Clarke, 2008; Hartl et al., 2011; Spiess et al., 2004). Each ring creates a central chamber where substrate polypeptides bind and fold. Unlike simpler archaeal chaperonins, TRiC contains eight different paralogous subunits, named CCT1-CCT8, at fixed positions within each ring (Kalisman et al., 2012; Leitner et al., 2012). All subunits are structural homologues that consist of an ATP-binding equatorial domain and a substrate-binding apical domain linked by an intermediate domain (Bigotti and Clarke, 2008; Spiess et al., 2004) (Fig. 1A). Each subunit also contains an apical segment that forms a lid over the cavity. An ATP-driven conformational cycle links TRiC-mediated folding to opening and closure of the lid, encapsulating the substrate in the cavity (Cong et al., 2012; Meyer et al., 2003; Reissmann et al., 2012; Reissmann et al., 2007)

Understanding how TRiC recognizes its substrates has important implications for human health (Balch et al., 2008). TRiC interacts with approximately 10% of the proteome and is essential for viability (Yam et al., 2008). Mutations in CCT5 and CCT4 are linked to sensory neuropathy (Bouhouche et al., 2006). Cancer-linked proteins p53, von Hippel Lindau tumor suppressor (VHL) and STAT3 are also TRiC substrates (Kasembeli et al., 2014; Trinidad et al., 2013) and mutations in the TRiC-binding sites of VHL lead to misfolding tumorigenesis (Feldman et al., 2003; Feldman et al., 1999). TRiC also suppresses aggregation and toxicity of Huntingtin in Huntington's Disease (Behrends et al., 2006; Kitamura et al., 2006; Tam et al., 2006; Tam et al., 2009). TRiC is also important for folding viral proteins, and required for replication of important human pathogens, including HCV and HIV (Inoue et al., 2011; Zhou et al., 2008). In HIV, TRiC interacts with proteins Gag, Vif and p6 (Hong et al., 2001; Jager et al., 2012).

The unique architecture and mechanistic features of TRiC set it apart from other chaperones. The diversification of subunits in TRiC is likely central to understand why many essential proteins, such as actin, Cdc20 and Cdh1, can only be folded with assistance from TRiC (Hartl et al., 2011; Spiess et al., 2004). Despite their extensive conservation in the ATP-binding domains, TRiC subunits have widely divergent functions within the ATP-driven cycle (Reissmann et al., 2012). Additionally, the surface properties of the different subunits result in an asymmetric distribution of electrostatic charges within the folding chamber (Leitner et al., 2012).

The principles driving TRiC substrate recognition are poorly understood. *In vivo*, TRiC folds a subset of cellular proteins, suggesting a degree of specificity; however, its substrates are functionally and structurally diverse, indicating the potential to bind a broad array of proteins. The apical domains of each TRiC subunit are thought to recognize different motifs in substrates (Spiess et al., 2004; Spiess et al., 2006) (Fig. 1A, B). However, to date no precise structural or sequence rules for TRiC-substrate binding have been identified. We here integrate biophysical and computational structural biology approaches with chemical crosslinking and mass spectrometry (XL-MS) to define the basis of TRiC substrate recognition. We find that unique subunit-specific patterns of polar and hydrophobic residues underlie the distinct substrate binding properties of each subunit in the complex. The diversification of TRiC subunits thus provides a modular menu of binding specificities that allows for combinatorial recognition of substrate polypeptides. This likely contributes to TRiC's unique ability to fold structurally diverse and topologically complex substrates. Evolutionary analyses further suggest that diversification of TRiC subunits from its simpler archaeal ancestors enabled the expansion of eukaryotic genomes to acquire proteins with novel folds and functions.

Results

Kinetic analysis of substrate motif recognition by TRiC apical domains

To understand the molecular basis of this recognition specificity, we exploited substrates where the cognate CCT subunit and the relevant substrate motif have been identified (Fig. 1C). The 54 amino acid long HIV protein p6, and the related protein p4 from MPMV, associate directly with subunit CCT3 of TRiC (Hong et al., 2001). A short 6–9 amino acid long hydrophobic motif in VHL, called Box1, contacts subunit CCT1 (Spiess et al., 2006). Importantly, the isolated recombinant apical domains of each TRiC subunit retain the ability to bind substrates and substrate-derived motifs with the specificity of the same subunits within the intact complex (Spiess et al., 2006; Tam et al., 2006; Tam et al., 2009).

We examined the association of TRiC apical domains of CCT1 (herein ApiCCT1) and CCT3 (ApiCCT3) (Spiess et al., 2006) with cognate and non-cognate substrate recognition motifs using purified HIV-p6 (herein p6) and VHL-Box1 (herein Box1) (Fig. 1D). A Surface Plasmon Resonance (SPR)-based assay measured association and dissociation kinetics for ApiCCT-substrate pairs (Fig 1 E,F; Fig. S1). Binding kinetics of immobilized VHL-Box1 and HIV-p6 to their cognate and non-cognate ApiCCT binding partners were monitored by SPR over a range of concentrations (Fig. 1D–F, Fig. S1A–D). Apparent association and dissociation rates (Fig. 1E) and binding constants (Fig. 1F), were calculated from the sensograms (Fig. S1). These indicated that the ratio of association over dissociation rates, i.e. the overall affinity, was higher for the cognate ApiCCT-substrate pairs (Fig. 1E), consistent with the specificity of these motifs for these subunits within the TRiC complex. The measured on-rates, determined at approximately 10^3 M^{-1} s^{-1}, markedly slower than diffusion controlled binding (Fig. 1E, blue bars) but consistent with the relatively slow substrate binding kinetics of TRiC (Melki et al., 1997). Cognate interactions exhibited slower dissociation kinetics than non-cognate interactions (Fig. 1E; Fig. S1A–D). Both association and dissociation rates contribute to substrate specificity for different subunits.

For p6 the difference between cognate and non-cognate interaction was largely driven by dissociation rates, whereas for Box1 cognate and non-cognate discrimination was a result of differential on- and off-rates (Fig. 1E). Of note, even the cognate interactions are relatively weak, with an overall affinity of approximately 0.25–0.5 μM (Fig. 1F). Accordingly, stable TRiC binding to most substrates will depend on multivalent recognition of several elements in the polypeptide by several subunits in the chaperonin.

NMR chemical-shift mapping of ApiCCT3 identifies the substrate-binding interface

We focused on the ApiCCT3 and p6 interaction pair to gain a deeper structural understanding of TRiC-substrate recognition. NMR-based chemical shift (CS) mapping was used to identify the substrate recognition interface in ApiCCT3 (Fig. 2 and Fig. S2). The ^{15}N-^{1}H Heteronuclear Single Quantum Coherence (HSQC) spectrum of ApiCCT3 yielded well resolved and dispersed spectra, accounting for 142 of 167 peaks, covering 85% of the protein sequence (Fig. 2A; Fig S2A and not shown). Standard triple resonance carbon experiments, guided by specific amino acid labeling to anchor the sequence connectivities allowed us to successfully assign >85% of the peaks in the 2D HSQC spectrum, including all the ApiCCT3 residues perturbed upon substrate addition (Fig. S2A–D). Titration of increasing amounts of unlabeled p6 into ^{15}N-labeled ApiCCT3 produced concentration-dependent shifts in a specific subset of peaks (Fig. 2A, B); 5 peaks were strongly perturbed (>0.2 p.p.m.) and another 4 peaks were perturbed weakly (>0.1 p.p.m; Fig 2B). Similar experiments were performed with p6-related protein p4 from M-PMV, which binds CCT3 with lower affinity [(Hong et al., 2001), data not shown]. p4 addition affected the same residues in ApiCCT3 than p6 (data not shown) albeit to a lower extent. In contrast, no perturbations were observed upon addition of Box1 (data not shown).

Since Y247 in ApiCCT3 (Fig. 2A, B) was strongly perturbed upon substrate binding, we used ^{19}F-NMR on 3F-tyrosine labeled ApiCCT3 for an orthogonal assessment of the binding interface (Fig. 2C and Fig. S2E). 1D ^{19}F-NMR spectra of 3F–tyrosine labeled ApiCCT3 revealed five discrete peaks, consistent with the five tyrosine residues in ApiCCT3 (Fig. 2C). Systematic tyrosine-to-phenylalanine point mutations assigned each peak to unique tyrosine residues (Fig S2E). Upon addition of p6, one of the peaks exhibited a well-defined 0.2 p.p.m. shift. In good agreement with our chemical-shift mapping, the perturbed peak corresponded to the ^{19}F-tyrosine peak of Y247 (Fig 2C).

Structural model of ApiCCT3 from NMR backbone chemical shifts

Guided by NMR-CS information (Fig. 2D*i*), we used CS-Rosetta and modeling to gain a structural understanding of ApiCCT3 in the substrate bound conformation (Shen et al., 2009) (see Experimental Procedures and Fig. S4D, E). The lowest energy models were comparable to the deposited ApiCCT3 structure without substrate (Pappenberger et al., 2002). Of note, our NMR-derived structural model resolved the apical protrusion, not resolved in the ApiCCT3 crystal structure and shown to be intrinsically disordered in a previous NMR study of an archaeal apical domain obtained without substrate (Heller et al., 2004).

Our NMR-derived model provided insight into the conformational dynamics of the apical domain (Fig. 2D*ii*). The regions of higher mobility in the structure included the helical protrusion (Fig. 2D*ii* and E, red) and to a lesser extent the flexible loop adjacent to Helix 11 (Fig. 2*Dii*, and E, herein PL for Proximal Loop). The highly flexible helical protrusion is involved in formation of the closed lid (Heller et al., 2004), but its role in the open chaperonin conformation is not well understood. Interestingly, Y247, whose chemical shift was strongly perturbed by substrate binding, is at the "hinge" between the flexible lid-forming protrusion and the apical domain. Y247 may participate in both substrate recognition and modulating the conformation of the lid protrusion for subsequent release when the closed lid forms.

Mapping the substrate-induced CS-perturbations (Fig. 2B) onto the ApiCCT3 structure revealed a continuous and extensive interaction surface spanning three sets of secondary structure elements (Fig. 2D*iii* and E,F). The CCT3 substrate interaction interface is primarily defined by a shallow groove formed between Helix 11 and the PL and comprises approximately 700 $Å^2$. The core of the substrate binding site was relatively constrained, consisting of residues on the surface of relatively rigid Helix 11 (herein H11) (residues 296–306). The distal portions corresponded to more flexible elements, including the PL adjacent to H11 (residues 223–232) and the hinge connecting to the flexible lid-forming protrusion including Y247. Supporting this analysis, we obtained low resolution diffracting crystals of the p6-ApiCCT3 complex which, following model building and refinement demonstrated an additional density in the same region of the apical domain identified through NMR (LAJ, Ryan McAndrew, JF and Paul Adams, unpublished data).

Previously characterized chaperone binding sites, such as those of Hsp70 and the bacterial chaperonin GroEL, rely predominantly on the recognition of hydrophobic determinants (Ashcroft et al., 2002; Chen and Sigler, 1999; Hua et al., 2001; Rudiger et al., 1997; Swain et al., 2006). In contrast, the substrate binding site of ApiCCT3 contained a mixture of hydrophobic and polar residues (Fig. 2F). In addition to Y247, H11 contributes hydrophobic (L299, M305), polar (Q301) and charged (R306, D298, H302) residues, while the PL immediately below presents a contiguous stretch of basic residues (H226, R228, R230, R231) and a single hydrophobic (M229) residue (Fig. 2F, schematically represented by the box diagram in Fig. 2F bottom). Such a shallow, extensive binding surface comprising hydrophobic and polar residues is very different from the mostly hydrophobic substrate binding sites of Hsp70 and GroEL.

Mutational analysis links chemical properties of substrate recognition site to binding kinetics

We next designed and purified a large unbiased panel of Alanine substitutions in ApiCCT3, comprising 31 surface-exposed residues in the 167 residue apical domain (Fig. 3 and S3). The global impact of these mutations on p6 binding was determined using SPR (Fig. 3; S3 and data not shown). Of the initial 31, ten mutations had no impact on p6 binding and were not examined further, except for a few mutants, such as Y274A, chosen as a control (Fig. S3). We determined kinetic binding parameters for twenty-two alanine mutants, including the one without effect as a control, by carrying out full titration series using SPR, followed

by kinetic global fitting of the data (Fig. 3B–D, Table S1 and summarized in Fig. 3E,G; Fig. S3). This unbiased mutational analysis of ApiCCT3 independently confirmed the NMR-based identification of the substrate binding site (Fig. 2). Thus, mutation of residues perturbed by NMR mapping dramatically affected p6 binding (Fig. 3E–H), while mutation at control sites (e.g. Y274 to A274) did not (Fig. S3).

Our analysis reveals the kinetic underpinnings of TRiC-substrate recognition (Fig. 3F,H and I). Substrate association (k_{on}) and dissociation rates (k_{off}) were very differently affected by mutations at discrete positions in the binding site, clustered into two distinct regions (Fig. 3F, H, I). The association kinetics were predominantly perturbed by mutations of positively charged residues in the flexible PL (Fig. 3F, I). On the other hand the contribution to the dissociation rate is distributed across both H11 and the PL. Residues on H11, particularly a mix of non-polar, polar and charged side chains, contributed predominantly to the dissociation rate (Fig. 3H, I). In the PL, the arginine residues allow both charge-charge interactions, likely contributing to the association rates, as well as cation-pi and aliphatic chain interactions with nonpolar residues, which likely contribute to the dissociation rates. Thus, the mixed chemical nature of the CCT3 binding site, combining polar and hydrophobic residues, establishes a dual mode of substrate recognition (Fig. 3J). The overall contribution of H11 and the PL to the binding constant is distributed across residues L299, Q301, H302 and R228, R230, R231, respectively. These findings resonate with studies of the interfaces between folded proteins, where a core of hydrophobic residues contributes to dissociation rates and polar interactions at the periphery drive association and orientation (Bogan and Thorn, 1998; Clackson et al., 1998).

NMR identification of the chaperonin-binding determinants in the substrate

NMR CS-mapping identified next the chaperonin binding site in the substrate (Fig. 4). ^{15}N-^{1}H HSQC spectra of ^{15}N-labeled p6 were assigned and CS-NMR used to model the peptide structure and identify determinants recognized by ApiCCT3 (Fig. 4A, Fig. S4A). Adding unlabeled ApiCCT3 to ^{15}N-p6 caused a concentration-dependent chemical shift perturbation in a subset of peaks (Fig. 4A, B). These peaks mapped to the contiguous and highly conserved S41-N45 element at the p6 C-terminus, consisting of both nonpolar and polar residues (S41-N45; Fig. 4C).

To determine the role of the S41-N45 element in chaperonin recognition, we generated and purified the penta-alanine substitution $p6_{SLFGN=>AAAAA}$ (herein $p6_{mut}$). We examined whether the SLFGN=>AAAAA mutation affected ApiCCT3 binding using NMR, by examining the perturbation of the ^{15}N-ApiCCT3 spectra upon titration of unlabeled $p6_{mut}$ (Fig. 4D). Mutation of the S41-N45 motif largely abrogated the p6-induced chemical shift perturbations in the ApiCCT3 spectrum, indicating this element mediates chaperonin binding (Fig. 4D). We next examined the role of S41-N45 in the interaction of p6 with intact TRiC/CCT (Fig. 4E). Purified $p6_{WT}$ or $p6_{mut}$, labeled with an N-terminal biotin tag, were incubated with mammalian cell extracts, which contain the intact hetero-oligomeric TRiC complex (Fig. 4E). Following biotin affinity isolation, the p6 interaction with endogenous TRiC was evaluated by immunoblot analysis (Fig. 4E). As expected, p6 bound TRiC, while $p6_{mut}$ did not (Fig. 4E). These orthogonal approaches support the conclusion that the S41-

N45 motif in p6 is indeed the TRiC binding site. To further corroborate this conclusion, we examined the chemical shift perturbations induced upon incubation of ^{15}N p6 with intact purified TRiC (Fig. S4D). Indeed the same subset of p6 residues was affected upon addition of intact TRiC, indicating that the ApiCCT3 binding site of p6 mediates its interaction with TRiC.

CS-Rosetta was next used to derive the solution structure of p6 (Fig. 4F*i*, Fig. S4B, C). Analysis of the top scoring models showed that p6 in aqueous solution contains a structured helical element at the C-terminus and a flexible polar region (Fig. 4G). The C-terminus of p6, containing S41-N45, adopts a helical conformation (Fig. 4F*i*). The ApiCCT3 interacting residues L42, F43 and N45 map to one face of the helix (Fig. 4F*ii*). In contrast, the N-terminus is highly dynamic as highlighted by the Cα-R.M.S.D. map (Fig. 4F*i*, Fig. S4B, C). CD measurements revealed only a very subtle decrease in helicity when p6 binds the chaperonin, consistent with the weak helicity observed in the NMR-guided structural model of p6 (Fig. S4B, S4E). Interestingly, while the helical chaperonin-binding determinant has significant hydrophobic character, the flexible N-terminus contains a series of acidic residues (Fig. 4G). The structural model of p6 implies that these acidic residues may interact with the basic residues in the ApiCCT3 loop, providing a molecular rationale for the electrostatic-driven association kinetics observed by SPR.

Structural model of the chaperonin-substrate interface

RosettaDock with CS-derived site constraints was employed to obtain a structural model of the ApiCCT3-p6 complex (Fig. 5 and S5). The lowest energy model was fully consistent with the NMR data of both p6 and ApiCCT3 (Fig. 5A, B) and the interface agreed with all our experimental data, including CS perturbations (Fig. 5C, Figs. 1 and 3) and mutagenesis (Figs. 2 and 3; Fig. 5D).

The ApiCCT3-p6 structure provides unprecedented detail on chaperonin-substrate binding (Fig. 5A, B). p6 makes tight packing interactions with unique features in the ApiCCT3 H11 and PL region through the specific presentation of side-chains (Fig. 5A,B). The interface consists of two distinct regions, highlighting the dual non-polar and polar nature of binding and providing a rationale for the bipartite substrate binding mode observed in the kinetic analyses. The interaction core is established by a mix of nonpolar and polar interactions, centered on H11 (Fig. 5B–D; namely L299, H302, M305, Q301 and Y247 in ApiCCT3 and L42, F43 and N45 in p6). A region of mostly electrostatic interactions is centered primarily on the PL, between positively charged residues in ApiCCT3 and acidic residues in p6 (Fig. 5B). These charge-charge interactions both confer specificity for discrete elements in the substrate and serve to orient the substrate upon binding the apical domain (Fig. 5B). Additional nonpolar contacts make close packing interactions with the aliphatic chains of lysine and arginine in the apical domain. As a result, of the 1073 A^2 of buried surface area, 762 A^2 correspond to nonpolar contacts and 311 A^2 to polar contacts.

Integrating conservation across orthologs for ApiCCT3 and p6 sheds light on the potential co-evolution of surfaces employed in chaperone-substrate interaction (Fig. 5E). The core interface residues are conserved in both ApiCCT3 as well as p6 variants across HIV clades. Conservation in this p6 region could respond to the requirement for TRiC interaction and/or

interaction with orthogonal binding partners such as VPR and ALIX, which also bind in this p6 region (Salgado et al., 2009).

We next employed a similar RosettaDock-based analysis to obtain a structural model for VHL-Box1 in a complex with the apical domain of CCT1 (Fig. 5F and Fig. S5AB). The ApiCCT1-Box1 structural model placed Box1 at the same H11/PL region of ApiCCT1 where p6 binds ApiCCT3 (Fig. 5F). Box1 adopts an extended conformation upon binding (Fig. 5G). The side chains of L116 and W117 in Box1, known to be critical for VHL binding to TRiC *in vivo* and *in vitro* (Feldman et al., 2003) pack between H11 and the PL in ApiCCT1 (Fig. 5F, inset; Fig. 5G). Interestingly, comparison of the substrate complexes of CCT1 and CCT3 (Fig. 5B vs Fig. 5G) shows that the substrate binding surface of CCT1 is more hydrophobic than that of CCT3, consistent with the higher hydrophobicity of Box1 over p6. We validated the ApiCCT1-Box 1 structural model using ApiCCT1 alanine-substitution mutagenesis followed by affinity measurements (Fig. 5H and Fig. S5C). A set of 25 mutants in ApiCCT1 was purified and their interaction with Box1 analysed by an affinity ranking SPR approach (Fig. 5H and Fig. S5C). Strikingly, the five ApiCCT1 alanine mutants that most significantly perturbed VHL-Box1 binding mapped to the interface predicted by the structure (Fig. S5C). These data, together with previous analysis of Box1 residues required for TRiC binding (Feldman et al., 2003), validate the structural model for ApiCCT1-Box1. We conclude that the groove formed between H11 and the flexible PL is the general substrate recognition site of TRiC/CCT subunits.

Mapping the contacts between the TRiC hetero-oligomer and full length substrates

To extend our understanding of TRiC substrate recognition to full length substrates we identified contact points between intact TRiC and three full length physiological substrates: actin, tubulin and HIV Gag using chemical crosslinking-mass spectrometry (XL-MS) (Fig. 6, Fig. S6). For actin and tubulin, the heterotypic TRiC crosslinks localized to substrate determinants previously implicated in TRiC binding by peptide arrays and mutagenesis (Table S2 and Fig. 6B) (Hynes and Willison, 2000; Ritco-Vonsovici and Willison, 2000; Rommelaere et al., 1999). Both tubulin and Gag crosslinked to multiple TRiC subunits through specific regions in the polypeptide (Table S2 and Fig. 6B), consistent with a multivalent contact between TRiC and its substrates.

Exploiting the conservation of general architecture of TRiC subunits, we mapped the location of substrate crosslinks to CCT2, CCT6 and CCT7 (Figure 6C green, blue and cyan spheres, respectively) onto the ApiCCT3 structure highlighting its substrate binding surface (Fig. 6C, red surface). Strikingly, the substrate crosslinks are proximal to the apical domain substrate binding interface between CCT3-p6 and CCT1-Box1, validating this region as the general location of the substrate-binding site in all TRiC subunits.

Comparing the chemical properties of the H11/PL region in subunits CCT2, CCT6 and CCT7 (Fig. 6D) shows that the substrate-binding site of each subunit has a distinct pattern of hydrophobic and polar residues. Thus, the dual recognition mode observed for CCT3-p6 is a general feature of TRiC-substrate recognition. Interestingly, analysis of the location of crosslink sites in the substrate primary sequence (Fig. 6E for Gag; see also below, Fig. S6C) indicated that chaperonin contact points within the polypeptide are close to the boundary

between a nonpolar and polar region (Fig. 6E, yellow trace: hydrophobicity; purple trace: polarity). The distinctive combination of polar and hydrophobic elements in both the substrate and each chaperonin subunit H11/PL region may underlie subunit specific interactions (Figs. 6,S6C).

We next mapped the TRiC-crosslink sites onto the folded structures of actin, tubulin and Gag (Fig. 6F). The chaperonin-contact sites are proximal to both a structured hydrophobic region, either helix or strand, and a more unstructured polar loop (Fig. 6F, top panel, N-terminus of Gag; bottom panel, tubulin, not shown for actin). The tubulin crosslinks map to two surface loops at the tips of the N-terminal and C-terminal lobes of the protein (Fig. 6F, bottom panel). The TRiC binding sites overlap with the interface of the tubulin heterodimer, indicating that folded and assembled β-tubulin cannot bind to TRiC (Fig 6F, bottom). The binding site for tubulin assembly factor Rbl2/Cofactor A (CoA), which acts directly downstream of TRiC in tubulin assembly (Tian and Cowan, 2013) also overlaps with the Tub2 site of crosslink to CCT6 (You et al., 2004). The overlapping tubulin binding sites for TRiC and CoA suggest possible mechanisms for Tub2 release from TRiC, and indicate that the chaperonin protects this oligomerization surface from inappropriate intra- and inter-molecular interactions.

To understand how the same subunit can bind distinct substrate motifs, we used the crosslinking information as a physical constraint to generate models of the CCT2-substrate interaction with Gag and tubulin (Fig. 6G, Fig. Fig. S6A–C). The lowest energy models placed both substrate-derived peptides in the CCT2 binding site formed by H11 and the PL, even though the starting distance constraint, i.e. the site of crosslink, was distal from this site. Gag and Tubulin bind in different configurations to the same apical domain of CCT2. Comparing all the data and structural models obtained here for different apical domain-substrate complexes reveals common rules for TRiC-recognition and specificity. The shallow groove created by H11 and the PL allows flexibility in binding, allowing the same apical domains to bind different substrates with no sequence similarity (Yam et al., 2008). H11 and the PL provides the apical domain surface for substrate recognition through a combination of polar and hydrophobic interactions. The specific polar-hydrophobic pattern of both apical domain and substrate serve to provide specificity and orient the substrate to dictate the binding topology.

Most excitingly, the XL-MS analysis provided a topological description of the substrate when bound to TRiC (Fig. 6H). Subunit specific contacts provide anchors that determine a global configuration of substrate polypeptides bound to TRiC; for Gag, the polypeptide is stretched across the open complex while for tubulin the two contact points at the tips of its two lobes are at antipodal positions of the ring. Taken together, these data indicate that binding to TRiC orients and restricts the global topology of the bound substrate, perhaps allowing the domains to start folding while associated with the chaperonin.

Discussion

How TRiC/CCT discriminates between non-native substrates and their folded counterparts is intriguing in view of its obligate requirement for folding a subset of cellular proteins that

share no sequence or structural similarities. By defining the structural basis of substrate recognition we begin to understand how subunit diversification enabled TRiC to balance the plasticity required to recognize a broad array of substrates with the specificity required to assist their folding.

Implications for substrate selection and folding in the hetero-oligomeric chaperonins

Mapping the substrate binding in the apical domains of TRiC to a groove between H11 and the PL has fundamental implications for the mechanics of substrate folding (Fig. 7A). Within the intact complex, the interaction with each subunit relies on a recognition code integrating polar and hydrophobic contributions, which enables combinatorial substrate recognition (Fig. 7A (i)). Polar contacts orient the substrate locally upon binding, while the distribution of subunit-specific substrate interactions stipulates the global topology of the TRiC-bound polypeptide. This may direct folding of TRiC-bound substrates along a preferred pathway (Fig. 7A (ii)). Our mapping of the substrate binding site provides a compelling mechanism of substrate release upon ATP-dependent closure (Douglas et al., 2011)(Fig. 7A (v)). ATP-induced closure of the lid brings the PL region in one apical domain into direct contact with a loop in the neighboring subunit, termed RLS (Release Loop of Substrate) (Fig. 7A (iii)) which mediates substrate release locally, through ATP-induced contacts between adjacent apical domains (Douglas et al., 2011)(Fig. 7A; (iv)). This mechanism of release is well suited to the low affinity of each individual apical domain-substrate interaction, as it permits the local displacement of the substrate from the apical domain by the ATP-induced proximity of the RLS. Since ATP binding and hydrolysis function within the ring is asymmetric (Reissmann et al., 2012)(Fig. 7A, dark grey: high ATP affinity; light gray: low ATP affinity), it is possible that the substrate is released sequentially during the conformational cycle (Fig. 7A, brackets). The particular dissociation rates for a given subunit, will determine when specific regions of the polypeptide are released from their binding site into the folding chamber. By allowing certain subdomains to fold first, TRiC may promote productive folding trajectories.

Principles driving the diversification of recognition and specificity in TRiC subunits

NMR, mutagenesis, modeling and XL-MS indicate that all TRiC subunits contact substrates through the same region in their apical domains. This region is evolutionarily conserved across orthologs (i.e. across all CCT5 from eukaryotes, Fig. 7C and Fig. S7A, B) but diverges across TRiC/CCT paralogs (Fig. 7B), suggesting a distinct and important function in each subunit.

Chaperonin binding sites balance plasticity and specificity in substrate recognition through two modular elements that discretely control binding and specificity: charged and polar residues contribute to enhancing on-rates and hydrophobic residues contributing to decreased off-rates. The polypeptide binding platform in the H11/PL region combines a rigid helical element and a flexible loop. Substrate interaction involves burial of nonpolar residues in the groove formed by the H11 helix-PL loop region. The PL loop is highly variable among subunits, both in terms of chemical properties and length (Fig. S7B; Fig. 7C). Providing conformational flexibility to PL in the polypeptide binding groove may enable recognition of a larger set of substrates. Among all subunits, CCT2 presents the most

nonpolar, classical “chaperone-like”, binding surface (Fig. 7C); interestingly this subunit is crosslinked to all full length substrates examined.

The shallow nature of the chaperonin binding groove allows a subunit to recognize motifs with different features, providing plasticity in binding, as shown for CCT2 (Fig. 6). The low affinity for a single substrate-apical domain interaction is consistent with an avidity-driven interaction, whereby multiple discrete low affinity contacts to different subunits mediate stable binding to the complex (Fig. 7A). This combinatorial recognition suggests a simple model for the discrimination of folded from non-folded proteins. TRiC will recognize those conformations interacting with more than one subunit but will not interact with proteins where most binding motifs are no longer available.

The evolution of hetero-oligomeric chaperonins

It is intriguing to consider what drove the evolution of such a complex hetero-oligomeric folding machine. TRiC substrates tend to encode complex topologies, and many coevolved with TRiC to the point of being unable to fold in its absence, (e.g., actin). Archaea have simpler chaperonins , ranging from one to five subunits depending on the organism (Bigotti and Clarke, 2008). We considered whether changes in the proteome are linked to subunit diversification. Strikingly, comparing all organisms containing TRiC-like chaperonins we find a positive correlation between subunit diversity and the size of its proteome (Fig. 7E). The possible link between, subunit diversification of TRiC and expansion of the proteome in eukaryotes raises questions on the mechanisms linking protein evolution to changes in chaperone structure and composition. Subunit diversification may increase the probability that a given protein will present two or more binding sites that can combinatorially bind to the chaperonin and benefit from the mechanistic advantages of folding in its chamber. Thus, the complexity of the chaperonin appears functionally optimized for the complexity of the genome, suggesting that the folding machinery contributes to dictate proteome size. A better understanding of the substrates and recognition motifs for the different TRiC and archaeal subunits may provide exciting insights into protein evolution.

Experimental Procedures Summary

Apical domains of TRiC were expressed and purified as described previously (Spiess et al., 2006). A set of alanine mutants were cloned and purified based on the chemical shift pertubation experiments and binding kinetics were analyzed by SPR. All apical domains were soluble and folded, as assessed by CD. For the SPR experiments, peptides were immobilized using maleimide chemistry on a PEG derivatized surface and a dilution series of apical protein were flowed over as analyte. For NMR, $^{15}N/^{13}C/^{2}D$ ApiCCT3 and $^{15}N/^{13}C$ p6 samples were expressed and purified using standard isotope labeling procedures (see Supplemental Experimental Approaches for details). NMR-chemical shift mapping and backbone assignment experiments for ApiCCT3 and p6 were carried out using 300 μM protein sample on a 800 MHz Innova Varian spectrometer outfitted with a cryogenic probe. Biotinylated $p6_{WT}$ or $p6_{mut}$ peptides were incubated with extracts from human HEK293 cells and affinity isolated via the biotin tag (IP). The presence of TRiC in the IP is visualized by Western blot analysis against TRiC antibodies raised against CCT5 and CCT2. DSS-

crosslinked TRiC-substrate samples were treated with trypsin, enriched for crosslinked peptides by size exclusion chromatography and analyzed by tandem mass spectrometry. Crosslinked peptides were identified by xQuest (Rinner et al., 2008). Conservation scores were calculated using Rate4site (Pupko et al., 2002) and mapped onto the models using ConSurf (Ashkenazy et al., 2010). Generation of structural models was guided by experimentally determined backbone chemical shift parameters in CS-Rosetta (Lange et al., 2012; Mao et al., 2014; Shen et al., 2009; van der Schot et al., 2013). RosettaDock was used to model the ApiCCT-substrate complexes with CS-NMR-based site constraints or with XL-MS-based atom pair constraints. Homology model building for the remaining ApiCCT domains was carried out using the Rosetta software package (Chivian and Baker, 2006; Davis and Baker, 2009).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Drs. R. Andino, A. Joachimiak, D Gestaut and K Dalton for comments on the manuscript; and Drs. L Gierasch and D. Agard for advice. We thank Dr. M Eckart from the PAN facility for help with the SPR experiments. This work was supported by an NIH fellowship to LAJ; NIH grants to JF.

References

Ashcroft AE, Brinker A, Coyle JE, Weber F, Kaiser M, Moroder L, Parsons MR, Jager J, Hartl UF, Hayer-Hartl M, et al. Structural plasticity and noncovalent substrate binding in the GroEL apical domain. A study using electrospay ionization mass spectrometry and fluorescence binding studies. J Biol Chem. 2002; 277:33115–33126. [PubMed: 12065585]

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic acids research. 2010; 38:W529–W533. [PubMed: 20478830]

Balch WE, Morimoto RI, Dillin A, Kelly JW. Adapting proteostasis for disease intervention. Science. 2008; 319:916–919. [PubMed: 18276881]

Behrends C, Langer CA, Boteva R, Bottcher UM, Stemp MJ, Schaffar G, Rao BV, Giese A, Kretzschmar H, Siegers K, et al. Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. Mol Cell. 2006; 23:887–897. [PubMed: 16973440]

Bigotti MG, Clarke AR. Chaperonins: The hunt for the Group II mechanism. Archives of biochemistry and biophysics. 2008; 474:331–339. [PubMed: 18395510]

Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. Journal of molecular biology. 1998; 280:1–9. [PubMed: 9653027]

Bouhouche A, Benomar A, Bouslam N, Chkili T, Yahyaoui M. Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5) gene causes autosomal recessive mutilating sensory neuropathy with spastic paraplegia. J Med Genet. 2006; 43:441–443. [PubMed: 16399879]

Chen L, Sigler PB. The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. Cell. 1999; 99:757–768. [PubMed: 10619429]

Chivian D, Baker D. Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. Nucleic Acids Res. 2006; 34:e112. [PubMed: 16971460]

Clackson T, Ultsch MH, Wells JA, de Vos AM. Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. J Mol Biol. 1998; 277:1111–1128. [PubMed: 9571026]

Cong Y, Schroder GF, Meyer AS, Jakana J, Ma B, Dougherty MT, Schmid MF, Reissmann S, Levitt M, Ludtke SL, et al. Symmetry-free cryo-EM structures of the chaperonin TRiC along its ATPase-driven conformational cycle. EMBO J. 2012; 31:720–730. [PubMed: 22045336]

Davis IW, Baker D. RosettaLigand docking with full ligand and receptor flexibility. Journal of molecular biology. 2009; 385:381–392. [PubMed: 19041878]

Douglas NR, Reissmann S, Zhang J, Chen B, Jakana J, Kumar R, Chiu W, Frydman J. Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. Cell. 2011; 144:240–252. [PubMed: 21241893]

Feldman DE, Spiess C, Howard DE, Frydman J. Tumorigenic mutations in VHL disrupt folding in vivo by interfering with chaperonin binding. Mol Cell. 2003; 12:1213–1224. [PubMed: 14636579]

Feldman DE, Thulasiraman V, Ferreyra RG, Frydman J. Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TRiC. Mol Cell. 1999; 4:1051–1061. [PubMed: 10635329]

Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011; 475:324–332. [PubMed: 21776078]

Heller M, John M, Coles M, Bosch G, Baumeister W, Kessler H. NMR studies on the substrate-binding domains of the thermosome: structural plasticity in the protrusion region. J Mol Biol. 2004; 336:717–729. [PubMed: 15095983]

Hong S, Choi G, Park S, Chung AS, Hunter E, Rhee SS. Type D retrovirus Gag polyprotein interacts with the cytosolic chaperonin TRiC. J Virol. 2001; 75:2526–2534. [PubMed: 11222675]

Hua Q, Dementieva IS, Walsh MA, Hallenga K, Weiss MA, Joachimiak A. A thermophilic mini-chaperonin contains a conserved polypeptide-binding surface: combined crystallographic and NMR studies of the GroEL apical domain with implications for substrate interactions. J Mol Biol. 2001; 306:513–525. [PubMed: 11178910]

Hynes GM, Willison KR. Individual subunits of the eukaryotic cytosolic chaperonin mediate interactions with binding sites located on subdomains of beta-actin. J Biol Chem. 2000; 275:18985–18994. [PubMed: 10748209]

Inoue Y, Aizaki H, Hara H, Matsuda M, Ando T, Shimoji T, Murakami K, Masaki T, Shoji I, Homma S, et al. Chaperonin TRiC/CCT participates in replication of hepatitis C virus genome via interaction with the viral NS5B protein. Virology. 2011; 410:38–47. [PubMed: 21093005]

Jager S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, et al. Global landscape of HIV-human protein complexes. Nature. 2012; 481:365–370. [PubMed: 22190034]

Kalisman N, Adams CM, Levitt M. Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:2884–2889. [PubMed: 22308438]

Kasembeli M, Lau WC, Roh SH, Eckols TK, Frydman J, Chiu W, Tweardy DJ. Modulation of STAT3 folding and function by TRiC/CCT chaperonin. PLoS biology. 2014; 12:e1001844. [PubMed: 24756126]

Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annual review of biochemistry. 2013; 82:323–355.

Kitamura A, Kubota H, Pack CG, Matsumoto G, Hirayama S, Takahashi Y, Kimura H, Kinjo M, Morimoto RI, Nagata K. Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. Nat Cell Biol. 2006; 8:1163–1170. [PubMed: 16980958]

Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, et al. Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:10873–10878. [PubMed: 22733734]

Leitner A, Joachimiak LA, Bracher A, Monkemeyer L, Walzthoeni T, Chen B, Pechmann S, Holmes S, Cong Y, Ma B, et al. The molecular architecture of the eukaryotic chaperonin TRiC/CCT. Structure. 2012; 20:814–825. [PubMed: 22503819]

Li J, Buchner J. Structure, function and regulation of the hsp90 machinery. Biomed J. 2013; 36:106–117. [PubMed: 23806880]

Mao B, Tejero R, Baker D, Montelione GT. Protein NMR structures refined with Rosetta have higher accuracy relative to corresponding X-ray crystal structures. Journal of the American Chemical Society. 2014; 136:1893–1906. [PubMed: 24392845]

Melki R, Batelier G, Soulie S, Williams RC Jr. Cytoplasmic chaperonin containing TCP-1: structural and functional characterization. Biochemistry. 1997; 36:5817–5826. [PubMed: 9153422]

Meyer AS, Gillespie JR, Walther D, Millet IS, Doniach S, Frydman J. Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. Cell. 2003; 113:369–381. [PubMed: 12732144]

Pappenberger G, Wilsher JA, Roe SM, Counsell DJ, Willison KR, Pearl LH. Crystal structure of the CCTgamma apical domain: implications for substrate binding to the eukaryotic cytosolic chaperonin. J Mol Biol. 2002; 318:1367–1379. [PubMed: 12083524]

Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics. 2002; 1(18 Suppl):S71–S77. [PubMed: 12169533]

Reissmann S, Joachimiak LA, Chen B, Meyer AS, Nguyen A, Frydman J. A gradient of ATP affinities generates an asymmetric power stroke driving the chaperonin TRIC/CCT folding cycle. Cell Rep. 2012; 2:866–877. [PubMed: 23041314]

Reissmann S, Parnot C, Booth CR, Chiu W, Frydman J. Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. Nat Struct Mol Biol. 2007; 14:432–440. [PubMed: 17460696]

Rinner O, Seebacher J, Walzthoeni T, Mueller LN, Beck M, Schmidt A, Mueller M, Aebersold R. Identification of cross-linked peptides from large sequence databases. Nature methods. 2008; 5:315–318. [PubMed: 18327264]

Ritco-Vonsovici M, Willison KR. Defining the eukaryotic cytosolic chaperonin-binding sites in human tubulins. J Mol Biol. 2000; 304:81–98. [PubMed: 11071812]

Rommelaere H, De Neve M, Melki R, Vandekerckhove J, Ampe C. The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. Biochemistry. 1999; 38:3246–3257. [PubMed: 10079067]

Rudiger S, Buchberger A, Bukau B. Interaction of Hsp70 chaperones with substrates. Nat Struct Biol. 1997; 4:342–349. [PubMed: 9145101]

Saibil H. Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol. 2013; 14:630–642. [PubMed: 24026055]

Salgado GF, Marquant R, Vogel A, Alves ID, Feller SE, Morellet N, Bouaziz S. Structural studies of HIV-1 Gag p6ct and its interaction with Vpr determined by solution nuclear magnetic resonance. Biochemistry. 2009; 48:2355–2367. [PubMed: 19254034]

Shen Y, Vernon R, Baker D, Bax A. De novo protein structure generation from incomplete chemical shift assignments. Journal of biomolecular NMR. 2009; 43:63–78. [PubMed: 19034676]

Spiess C, Meyer AS, Reissmann S, Frydman J. Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. Trends Cell Biol. 2004; 14:598–604. [PubMed: 15519848]

Spiess C, Miller EJ, McClellan AJ, Frydman J. Identification of the TRiC/CCT substrate binding sites uncovers the function of subunit diversity in eukaryotic chaperonins. Mol Cell. 2006; 24:25–37. [PubMed: 17018290]

Swain JF, Schulz EG, Gierasch LM. Direct comparison of a stable isolated Hsp70 substrate-binding domain in the empty and substrate-bound states. J Biol Chem. 2006; 281:1605–1611. [PubMed: 16275641]

Tam S, Geller R, Spiess C, Frydman J. The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. Nat Cell Biol. 2006; 8:1155–1162. [PubMed: 16980959]

Tam S, Spiess C, Auyeung W, Joachimiak L, Chen B, Poirier MA, Frydman J. The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. Nat Struct Mol Biol. 2009

Tian G, Cowan NJ. Tubulin-specific chaperones: components of a molecular machine that assembles the alpha/beta heterodimer. Methods Cell Biol. 2013; 115:155–171. [PubMed: 23973072]

Trinidad AG, Muller PA, Cuellar J, Klejnot M, Nobis M, Valpuesta JM, Vousden KH. Interaction of p53 with the CCT complex promotes protein folding and wild-type p53 activity. Mol Cell. 2013; 50:805–817. [PubMed: 23747015]

van der Schot G, Zhang Z, Vernon R, Shen Y, Vranken WF, Baker D, Bonvin AM, Lange OF. Improving 3D structure prediction from chemical shift data. Journal of biomolecular NMR. 2013; 57:27–35. [PubMed: 23912841]

Yam AY, Xia Y, Lin HT, Burlingame A, Gerstein M, Frydman J. Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. Nat Struct Mol Biol. 2008; 15:1255–1262. [PubMed: 19011634]

You L, Gillilan R, Huffaker TC. Model for the yeast cofactor A-beta-tubulin complex based on computational docking and mutagensis. Journal of molecular biology. 2004; 341:1343–1354. [PubMed: 15321725]

Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, et al. Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe. 2008; 4:495–504. [PubMed: 18976975]

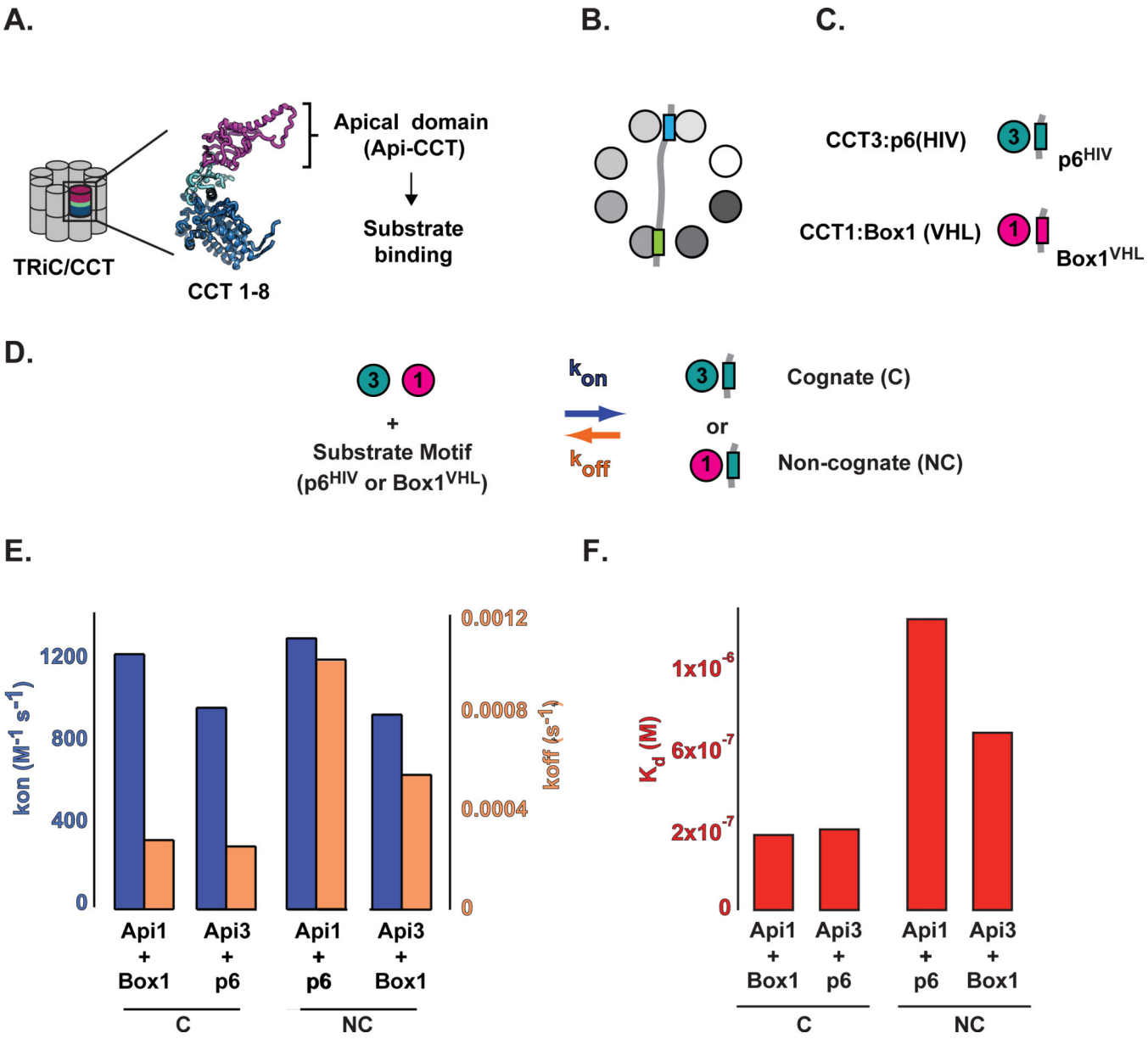

Figure 1. Kinetic analysis of substrate motif recognition by TRiC apical domains

(A) TRiC/CCT subunit domain architecture. **(B)** Substrate polypeptides bind multiple TRiC subunits through discrete motifs. **(C)** Previously characterized subunit-substrate motif interactions. Top: CCT3 binds retroviral proteins p6 from HIV and p4 from MMPV. Bottom: CCT1 binds Box1, a short linear element from VHL. **(D)–(H)** Kinetic analysis of cognate (C) and non-cognate (NC) interactions between purified apical domains and substrate derived motifs. **(D)** Summary of apparent kinetic parameters for cognate and non-cognate interactions, k_{on}, blue bars; k_{off} orange bars. **(E)**. Summary of apparent binding constants for cognate and non-cognate interactions, k_{on}, blue bars; k_{off} orange bars.

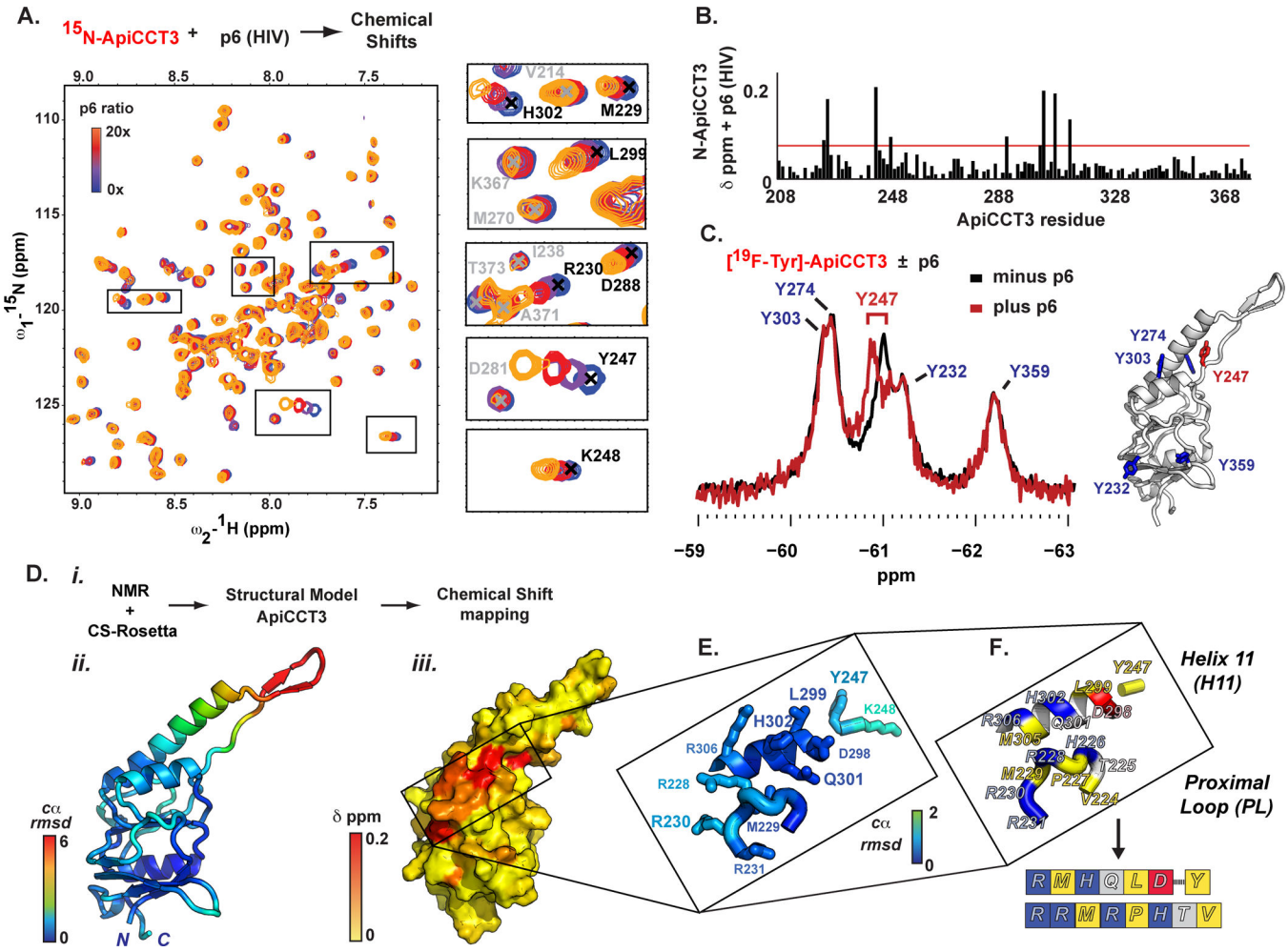

Figure 2. NMR chemical-shift mapping of ApiCCT3 identifies the substrate-binding interface

(A) 2D NMR spectra of ^{1}H-^{15}N ApiCCT3 with increasing amounts of unlabeled p6. Blue, purple, red and orange, respectively show perturbations at 0, 5, 10 and 20-fold excess p6 peptide. Insets show perturbed peaks. **(B)** Chemical shift perturbation due to p6 binding mapped to the ApiCCT3 sequence. **(C)** ^{19}F tyrosine spectra of ApiCCT3 (red trace). Addition of p6 (black trace) produces a discrete perturbation of Y247. Right panel: ApiCCT3 structure with Y247 in red and the remaining tyrosines in blue. **(D)** ***(i)*** Integration of NMR chemical shift parameters with CS-Rosetta. ***(ii)*** C-α variability across a subset of low energy ApiCCT3 models highlights mobile regions. ApiCCT3 shown as cartoon and colored according to C-α r.m.s.d. ***(iii)*** P6 induced chemical shifts map to a contiguous ApiCCT3 surface. Surface representation of ApiCCT3 is colored according to the chemical shift from yellow (0) to red (0.2 p.p.m.). **(E–F)** Substrate binding region on ApiCCT3, defined by a Helix 11 (H11), a loop (PL) and a residue at the hinge of the flexible loop (Y247): **(E)** Cartoon representation colored according to flexibility (Cα-r.m.s.d. as in E); **(F)** Colored according to chemical properties of side chains: basic (blue), acidic (red), polar (white/gray) and non-polar (yellow). Top: cartoon representation of binding site; bottom: simplified box schematic depicting amino acids on the substrate binding site: Upper line: H11 and Protrusion Hinge residue; Lower Line: PL sequence.

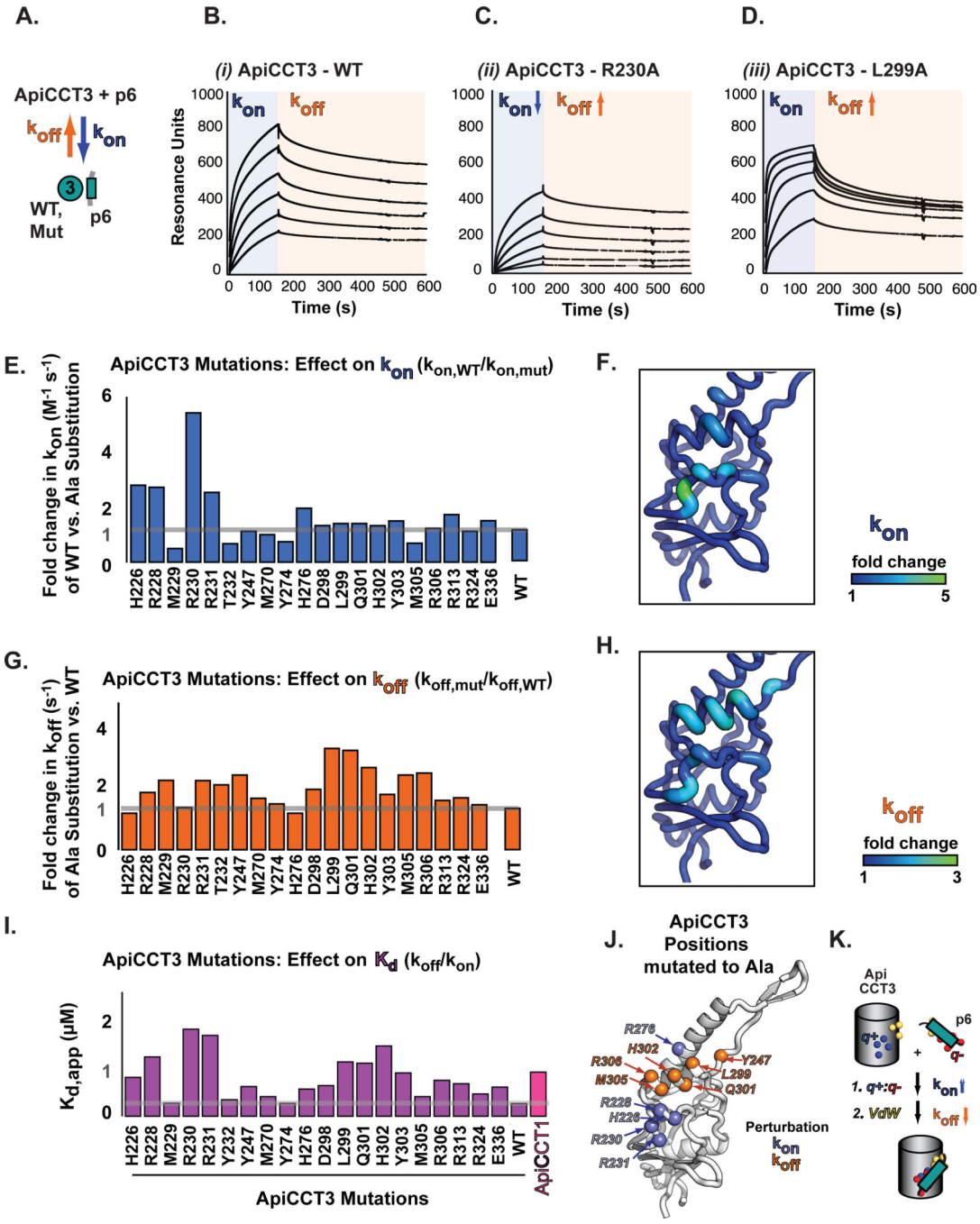

Figure 3. Mutational analysis of CCT3 links chemical properties of substrate recognition site to binding kinetics

(A) Kinetic scheme of ApiCCT3:p6 interaction (k_{on} in blue, k_{off} in orange). **(B–D)** Exemplary SPR titrations of ApiCCT3 variants interacting with an immobilized p6 peptide, performed as in Fig. 1. **(B)** ApiCCT3-WT, **(C)** ApiCCT3-R230A and **(D)** ApiCCT3-L299A. Association phase in blue and dissociation phase in orange. Each titration (black curves) included an analyte dilution series of 50, 25, 12.5, 6.25 and 3.125 μM apical domain. **(E–H)** Summary of fitted kinetic parameters for ApiCCT3 mutants (see also Table S2 for details).

(E, G) Bar graphs showing the fold change relative to WT apparent rates for **(E)** association rate k_{on} (blue, $k_{on,WT}/k_{on,mut)}$ and **(G)** dissociation rate k_{off} (orange, $k_{off,mut}/k_{off,WT}$) from p6 substrate. **(F, H)** Fold change in rates mapped onto the ApiCCT3 structure highlight the clustering of surfaces based on Ala mutants perturbing binding kinetics through **(F)** association and **(H)** dissociation rates. ApiCCT3 structure is colored according to fold-change in kinetic contribution, blue is equal to WT and green is perturbed as indicated. (**I**) Calculated apparent K_d for WT ApiCCT3 and each alanine mutant. The apparent K_d of p6 for ApiCCT1, also measured by SPR, is included for comparison. **(J)** Residues that contribute to association and dissociation rates shown as spheres on a cartoon model of ApiCCT3, perturbations affecting association rates (blue) cluster in the PL; those affecting dissociation rates (orange) cluster on H11 and the hinge at the apical protrusion. Inset, surface representation of ApiCCT3-p6 binding site colored according to amino acid chemical property: basic in blue; acidic in red: polar in white and nonpolar in yellow. **(K)** Differential kinetic contribution of two regions in ApiCCT3 to substrate binding: charge-charge interactions between basic and acidic residues in blue and red respectively, control association rates (R228, R230, R231 also contribute to off rates) while a mix of nonpolar, polar and Van der Waals interactions, shown in yellow, control dissociation rates.

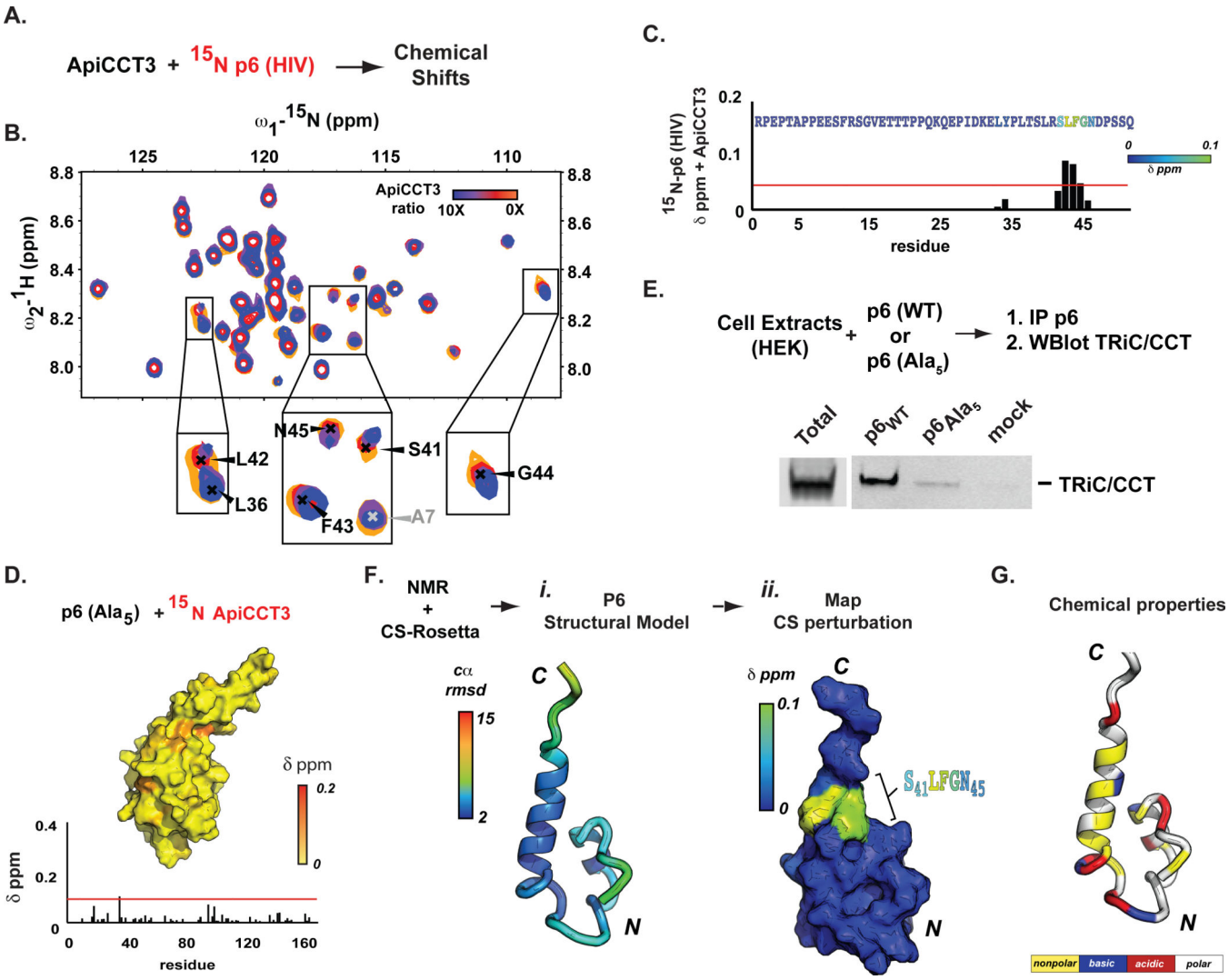

Figure 4. NMR identification of the chaperonin-binding determinants in the substrate

(A) NMR-based chemical shift mapping strategy identifies the p6 interacting surface. **(B)** 2D NMR spectra of ^{1}H-^{15}N p6 with increasing amounts of unlabeled ApiCCT3. orange, red, purple and blue, respectively show perturbations at 0, 2.5, 5 and 10-fold excess ApiCCT3. **(C)** Chemical shift perturbation mapped onto the p6 sequence colored by ApiCCT3 perturbation. **(D)** Mutant $p6_{mut}$ ($p6_{SLFGN=>AAAAA}$) does not induce chemical shift perturbations in the ^{15}N ApiCCT3 spectrum observed with p6WT. Perturbations mapped on the ApiCCT3 sequence and on the ApiCCT3 surface, exactly as in Fig. 2F. **(E)** $p6_{mut}$ mutation abrogates the binding to endogenous TRiC observed for $p6_{WT}$. Biotinylated $p6_{WT}$ or $p6_{mut}$ were incubated with extracts from human HEK293 cells, affinity isolated via the biotin tag (IP). TRiC in the IP is visualized by Western blot. Total: 1% input to the IP. **(F)** Structural model of p6 from CS-NMR. **(i)** C-α variability of p6: cartoon representation colored according to variability across a subset of low energy models, as indicated. **(ii)** Surface representation of p6 colored according to chemical shift perturbation by ApiCCT3 addition, as indicated. **(G)** Chemical properties of p6, colored according to amino acid properties: blue is basic, red is acidic, white is polar and yellow is nonpolar.

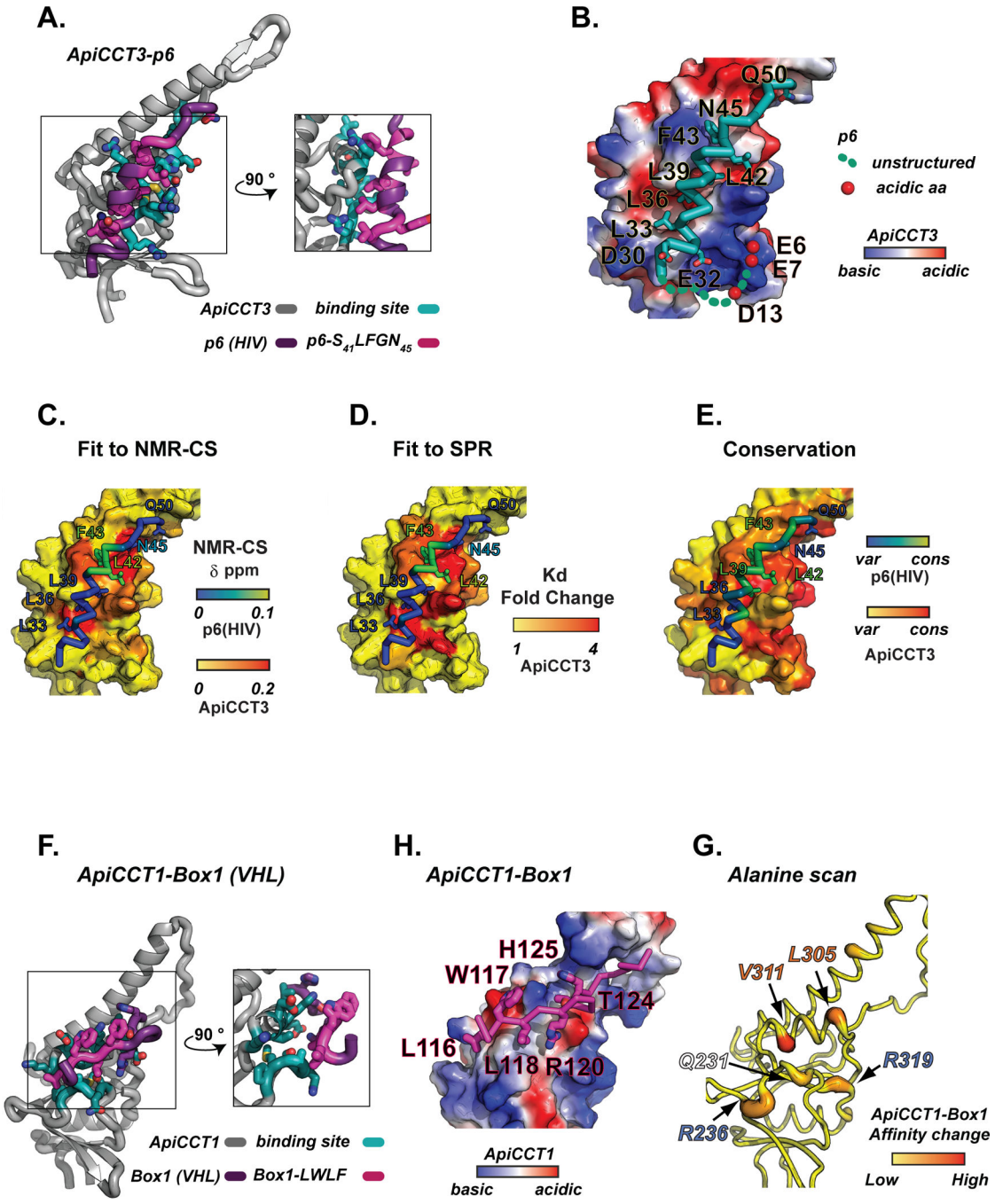

Figure 5. Structural model of the chaperonin-substrate interface

(A–E) Structural model of ApiCCT3-p6. **(A)** ApiCCT3 and p6 colored in grey and purple, respectively. Interfacial residues in stick representation are colored in teal and magenta. Inset, zoom-in of the interface at 90° rotation. **(B)** ApiCCT3 substrate binding interface in surface representation colored according to electrostatic potential; blue, red and white are positive, negative and neutral residues respectively. Bound p6 is shown in teal with interfacial residues in stick representation. **(C–E)** Agreement between experimental data and structural model of ApiCCT3 (surface) and p6 (ribbon): **(C)** Reciprocal chemical shift

perturbations in p6 (blue-green) and ApiCCT3 (yellow-red) upon complex formation **(D)** Mutagenesis perturbation of binding measured by apparent K_d **(E)** Evolutionary conservation among orthologs of CCT3 (yellow-red) and p6 (blue-green). **(F–H)** Structural model of ApiCCT1-Box1. **(F)** ApiCCT1-Box1 complex in cartoon representation colored in grey and purple, respectively. Interfacial residues shown as sticks and colored in teal and magenta. Inset, zoom in of the interface at 90° rotation. **(G)** Electrostatic charge distribution of ApiCCT1 substrate binding site bound to the Box peptide. Surface representation of the ApiCCT1 substrate binding site colored according to electrostatic charge potential; blue, red and white are positively, negatively and neutrally charged residues respectively. Bound Box1 in magenta with interfacial residues shown in stick representation. **(H)** Mutagenesis analysis of CCT1 residues required for Box 1 binding. Affinity ranking SPR experiments assessed binding of Alanine ApiCCT1 mutants for Box1. ApiCCT1 residues that perturb binding are highlighted in red/orange on the cartoon putty representation.

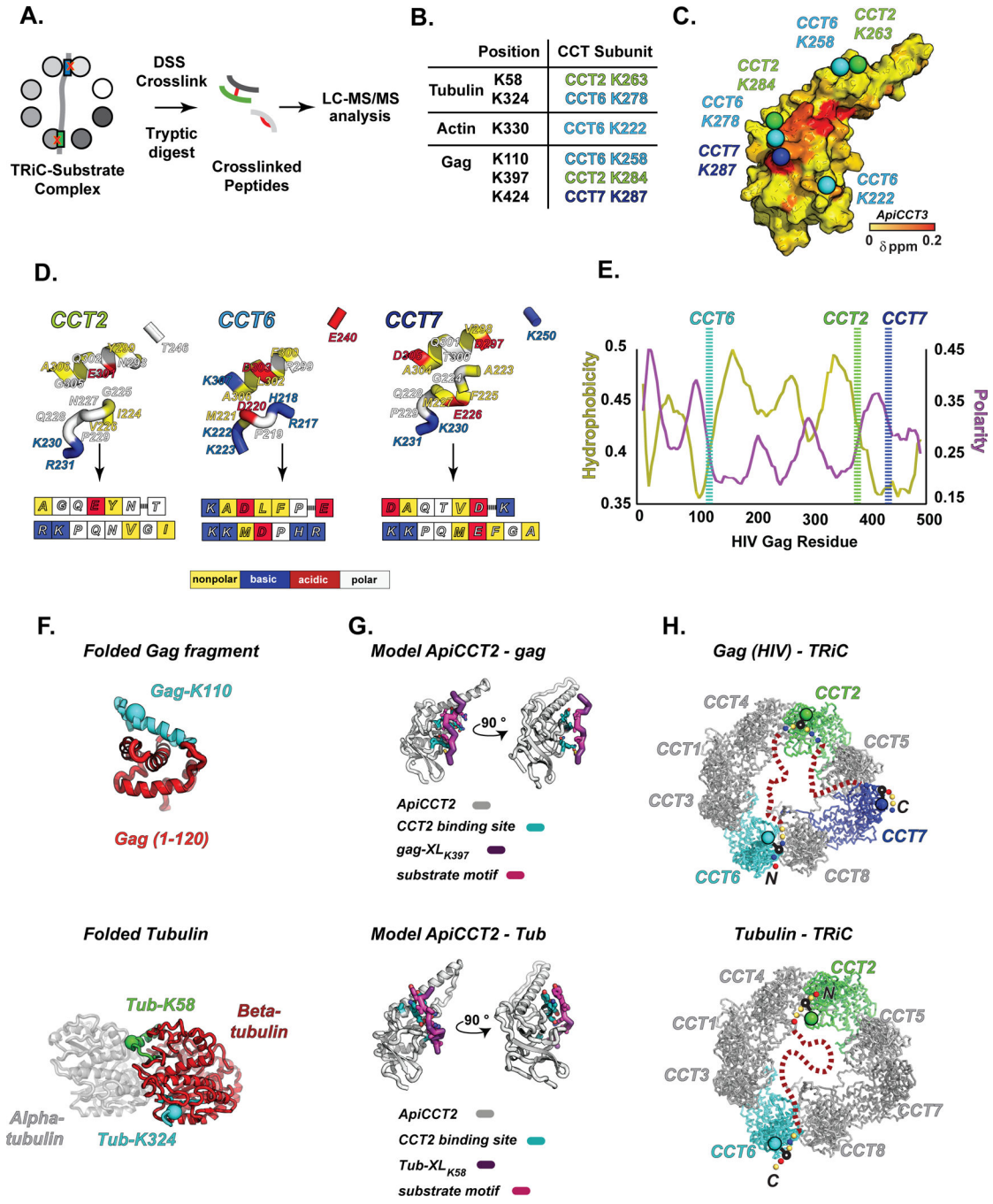

Figure 6. Mapping the contacts between the TRiC hetero-oligomer and full length substrates

(A) XL-MS approach to map TRiC-substrate contacts. **(B)** XL-MS-derived contact points between tubulin, actin and Gag and specific sites on TRiC. **(C)** Substrate crosslink sites mapped onto the surface representation of ApiCCT3. The substrate binding interface in ApiCCT3 is highlighted as in Fig. 2F. The sites of substrate crosslinks shown as spheres are colored green, cyan and blue for CCT2, CCT6 and CCT7, respectively. **(D)** Comparison of H11/PL substrate binding region of subunits CCT2, CCT6 and CCT7 colored according to amino acid properties; blue, red and white correspond to basic, acidic and neutral amino

acids. Bottom: schematic representation of binding region in box format as in Fig. 2H, colored as indicated. **(E)** TRiC crosslink sites mapped onto the Gag protein sequence. Vertical lines show sites where Gag crosslinks to the apical domains of CCT6 (cyan), CCT2 (green) and CCT7 (blue). The plot also maps the hydrophobicity (yellow) and polarity (purple) of Gag. **(F)** Gag and tubulin crosslinks to TRiC mapped onto the folded substrate structures. (Top) Cartoon representation of N-terminal domain of Gag in red with flexible region containing the site of CCT6 crosslink (K110, as spacefill) in cyan. (Bottom) Cartoon representation of tubulin heterodimer: β-tubulin in red, and α tubulin in grey. Loops containing the two β-tubulin-TRiC crosslink sites (in spacefill: K58 in green to CCT2; K324 in cyan to CCT6) are shown in green and cyan. **(G)** XL-MS-derived structural models of CCT2 apical domain-substrate interaction with substrate elements from crosslink sites of Gag (top) and tubulin (bottom). Apical domain-peptide complexes are colored in grey and purple, respectively. Interfacial residues in stick representation are colored teal for the apical protein and magenta for the peptide. **(H)** Putative topological description of the TRiC-bound substrate (dashed line) for Gag (top) and Tubulin (bottom). The open state of TRiC is shown in grey, with subunits CCT2, CCT6 and CCT7 colored green, cyan and blue respectively. The α lysines involved in the crosslink are shown as spheres. The immediate proximity of the crosslink site on the peptide sequence is colored according to amino acid properties, yellow, white, blue and red are nonpolar, polar, basic and red, respectively.

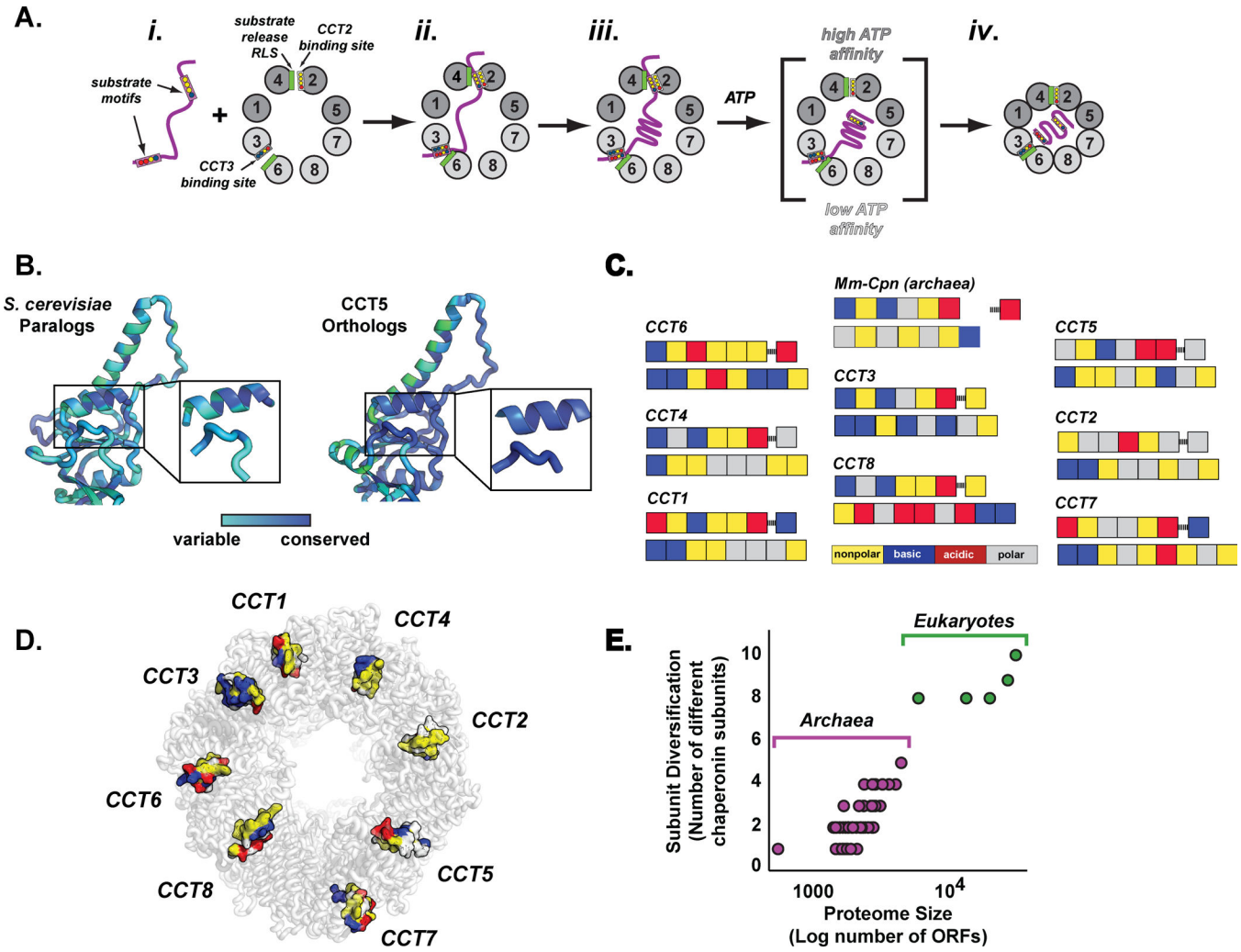

Figure 7. The polypeptide-binding site of TRiC/CCT: functional and evolutionary implications

(A) Role of subunit specific substrate recognition in the context of the TRiC folding cycle. See Discussion for description. Subunits with high ATP affinity in dark gray, with low ATP affinity in light gray. **(B)** Evolutionary divergence across H11/PL substrate binding sites across *S. cerevisiae* paralogs (left) and conservation of binding site across orthologs (CCT5, right panel). **(C)** Substrate binding site properties of the different yeast TRiC subunits schematically shown in box format as in Fig. 2H. Upper line corresponds to H11/Apical Hinge residues, lower line corresponds to PL. The substrate binding site of archeal chaperonin from *M. maripaludis* is included for comparison. **(D)** The substrate binding site interface of the open TRiC conformation. TRiC is shown in semi-transparent gray cartoon and substrate binding sites in surface representation colored according to chemical properties, as indicated. **(E)** Group II chaperonin subunit complexity increases with proteome size. Number of subunits in archeal (purple) and eukaryotic (green) organisms graphed versus proteome size, as number of ORFs in that organism.