Supporting Information

Kalisman et al. 10.1073/pnas.1119472109

SI Methods

Optimal Cross-Linking Concentration. In designing the cross-linking procedure, our main concern was finding the minimal BS3 concentration that would still provide sufficient inter-subunit cross-linking. To achieve this, we cross-linked mmCPN, the archaeal group II chaperonin from Methanococcus Maripaludis, with increasing BS3 concentrations. The recombinant mmCPN complex is homo-oligomeric and available in larger quantities than TRiC; it is highly homologous to TRiC in sequence and overall architecture (1) and is expected to cross-link in a similar manner. We assessed the cross-linking performance by running the cross-linked mmCPN on agarose-SDS gel (Fig. S1). As expected, very high cross-linker concentrations caused all the 16 subunits to travel in a single band, indicating that cross-linking probability is high enough to ensure every subunit participates in at least one inter-subunit cross-link. The lowest BS3 concentration for which a clear single band occurred was 2 mM final concentration, and we used this value to cross-link TRiC. Interestingly, Chen et al. (2) found nearly the same value to be optimal under similar criterions on a completely different system, suggesting that this value may be suitable for any large protein complex.

Mass Spectrometry. The cross-linked complex was precipitated by acetone at −80°C, centrifuged at 10,000 rpm, 4 °C for 10 min. The supernatant was discarded and the precipitated protein dried in a SpeedVac. The protein was reconstituted in 15 μL 8 M Urea, 50 mM ammonium bicarbonate, pH 8.0 and 20 μL of 0.25% protease max (Promega) in 50 mM ammonium bicarbonate pH 8.0. The sample was reduced with 50 mM DTT at 55 °C for 30 min and then alkylated with 100 mM propionamide for 30 min at room temperature. Digestion volume was adjusted to keep the Urea concentration to less than 1 M, and trypsin was added at a 1:50 enzyme to protein ratio. Digestion occurred overnight at 37 °C, after which the reaction was quenched by the addition of 10 μL 50% formic acid water. The peptides were further purified and enriched using an in-house packed stage-tips (3).

We used 10% of the digest without further processing for LC-MS/MS. The other 90% of the digest was fractionated by strong cation exchange chromatography (200 × 2.1 mm Poly SULFOETHYLA column; Poly LC) as described by Chen et al. (2). Fractionation followed a monotonically increasing gradient from solvent A (5 mM KH_2PO_4, 10% acetonitrile, pH 3.0) to solvent B (solvent A with 1 M KCl). Solvent B values of 0%, 9%, 18%, 36%, and 100% were reached at times of 0 min, 10 min, 15 min, 18 min, and 22 min, respectively. The flow rate was 200 μL/ min and fractions were collected every 1 min. Fractions eluting at KCl concentrations of 80 mM and higher were retained and desalted using stage-tips for subsequent LC-MS/MS analysis (on a Thermo LTQ-Orbitrap Velos instrument).

Peptides were first loaded onto a POROS trap column at a flow rate of 10 μL/ min and subsequently onto a self packed C18 analytical reversed phase column (150 uM ID × 12 cm length) at a flow rate of 600 nL/ min following a linear gradient going from 2% acetonitrile in 0.1% formic acid to 35% acetonitrile in 0.1% formic acid in 160 min. We used the Orbitrap to detect both the peptides and their fragments at high resolution. A charge state of 3+ and higher was used to trigger the MS/MS analysis followed by dynamic exclusion of 60 s with a 20 ppm exclusion mass window. HCD was performed on the top eight most intense precursor ions meeting the predefined qualities, and the repeat count was two.

Sequences, Alignments, and Modeling. Bovine TRiC sample without cross-linking was digested and subjected to standard proteomic analysis. The analysis confirmed the presence of all the eight different TRiC subunits as well as substantial amounts of the CCT6 isoform2 subunit, residual tubulin (both alpha and beta), and residual actin. Posttranslation modification analysis revealed that the N termini of most subunits are modified and found additional minor changes in other parts of the sequences (see sections ahead for a complete list). We used all the 12 different sequences found in the sample during the MS data analysis (eight TRiC chains, CTT6 isoform2, alpha and beta tubulin and actin) in generating the list of all possible cross-links. However, high confidence cross-links were identified only within the standard eight TRiC sequences.

The structures of several archaeal class II chaperonins are known by either X-ray crystallography (4, 5) or cryo-EM (1). These simpler systems share high sequence identity with all the eight sequences of TRiC (range 30–42%, median 38%), making high-accuracy homology modeling extremely likely. The close-state thermosome structure from Thermococcus strain KS-1 (PDB ID code 1Q3R) was our template of choice, because it is homo-oligomeric and presents more structured residues than any other template. ClustalW generated the multiple sequence alignment between the TRiC and thermosome sequences (Dataset S1). Manual correction of the alignment at four regions involving beta-sheet strands was required. We used SEGMOD (6) to generate the homology models with their internal insertions and deletions. The long stretches of residues missing from the template at the N and C termini were not modeled.

MS Data Analysis. In-silico trypsin digestion gave an initial digest list of all the possible peptides with the following parameters: (*i*) up to four missed cleavages; (*ii*) fixed modification to cystines (proponamide); (*iii*) variable modifications to methionines (oxidation), N termini (acetylation), and lysines (BS3 monolink reacted with water or ammonia). The sequences of the eight TRiC subunits, CCT6-isoform2, tubulin alpha and beta, and actin were used in the generation of the digest list. The list of all the possible cross-linked species was taken as all pairs of lysine residues and N termini from the initial digest list (including self pairs). The raw MS data was converted into MGF format using MassMatrix (7). The mass of each precursor ion was compared to the list of possible cross-linked species. Matching candidates within a mass tolerance of 6 ppm were further analyzed for MS/MS fit. The expected b and y series were compared to the measured MS/MS spectrum with the same 6 ppm tolerance. The b1 and y1 fragments of the cross-linked peptides were not used. The MS/MS-score assigned to a candidate match is

$$S_{\mathrm{MS2}} = \frac{\#\,\text{matching b fragments and y fragments}}{\mathrm{length(peptide\,1)} + \mathrm{length(peptide\,2)}}.$$

After the MS/MS scores of all the candidate matches were assessed, a precursor ion is assigned with a specific cross-link entry in the list if three conditions are fulfilled: (*i*) This entry had the highest MS/MS-score of all the other candidate matches to that ion. If two entries have the same top MS/MS-score, the entire precursor ion is discarded. (*ii*) The MS/MS-score is higher than 0.5 or at least four matching b or y fragments occur on each of the two peptides. (*iii*) The MS1 peak cannot be assigned to a single (non-cross-linked) peptide under similar matching criteria. The resulting list of cross-links is subjected to a final filtering step

to check for identical lysine pairs originating from alternative trypsin digestion sites: only the occurrence with highest MS/MS-score is kept.

Change to Sequences. From the UniProt database (8) we retrieved the following entries: TCPA_BOVIN, TCPB_BOVIN, TCPG_BOVIN, TCPD_BOVIN, TCPE_BOVIN, TCPZ_BOVIN, TCPW_BOVIN, TCPH_BOVIN, and TCPQ_BOVIN as our working sequence base for subunits: CCT1(α), CCT2(β), CCT3(γ), CCT4(δ), CCT5(ϵ), CCT6(ζ), CCT6 isoform 2, CCT7(η), and CCT8(θ), respectively. Proteomic analysis of TRiC without cross-linking identified peptides from all the nine different subunits, with average peptide coverage of 77%. Manual inspection of high-intensity peaks revealed a few minor corrections to the UniProt sequences, as well as posttranslational modifications. We list these alterations bellow. We emphasize that additional sequence-errors/modifications might have escaped our detection because: (*i*) The peptides containing them were not detected in the MS or (*ii*) they only occur in a small fraction of the particles.

A change of the residue type at position 510 in the CCT4(δ) subunit from Q to L resulted in a verified assignment to a high-intensity peak:

$MS1_{error}$ [ppm] = 3.8

```
      bb bbbbbbbb
K500GGISNILEELVVQPLLVSVSALTLATETVR530
                  yyyyyyyyyy  yyy
```

Interestingly, this sequence (L510) exists in NCBI-BLAST's nonredundant sequence database. Another residue type change at position 158 in the CCT4(δ) subunit from E to V also resulted in a verified assignment to a high-intensity peak. This sequence also exists in NCBI-BLAST's nonredundant sequence database:

$MS1_{error}$ [ppm] = −0.3

```
    bb b
G147IEILTDMSRPVELSDRETLLNSAATSLNSK177
          y             y yyyyyyyy
```

Bovine Polymorphism. The TRiC sample was purified from a tissue pool of several animals and one could expect polymorphisms in the TRiC genes. We identified one polymorphism site (V/I) at position 179 in the CCT6 isoform 2 subunit:

$MS1_{error}$ [ppm] = 3.7

```
      bb bbbbbbb
V160HPQLADVLTEAVVDSVLAVR180
          yyyyyy y yyy
```

$MS1_{error}$ [ppm] = 1.0

```
      bbbbbbbbbb
V160HPQLADVLTEAVVDSVLAIR180
          yyyyy yyyyy
```

We chose to work with the first polymorph (V) in our MS analysis since its MS1 intensity was almost 10-fold higher, indicating higher prevalence.

Posttranslational Modifications in N Termini. Many high-intensity peaks were confidently assigned to modified N termini of the TRiC subunits (Table S2). The MS evidence for the these modifications is as follows:

CCT1(α). $MS1_{error}$ [ppm] = −3.7

```
          bbb
Acetyl-MEGPLSVFGDRSTGEAIR
                   yyy yy
```

CCT2(β). $MS1_{error}$ [ppm] = 1.1

```
          bbb
Acetyl-ASLSLAPVNIFK
            yy y y y
```

CCT3(γ). $MS1_{error}$ [ppm] = −0.8

```
          bbbb bbbb
Acetyl-MMGHRPVLVLSQNTKR
          yy y  yyyyyyyy
```

CCT4(δ). Not determined

CCT5(ϵ). $MS1_{error}$ [ppm] = −0.4

```
          b bb bb
Acetyl-ASVGTLAFDEYGRPFLIIKDQDRK
           yyyyyyyy y yyyyyyyy y
```

CCT6(ζ) and CCT6 isoform2. Not determined

CCT7(η). $MS1_{error}$ [ppm] = 4.1

```
          b
Acetyl-MMPTPVILLKEGTDSSQGIPQLVSNISACQVIAEAVR
                              y yyyyyyyy yy yy
```

CCT8(θ). $MS1_{error}$ [ppm] = −0.7

```
            bbb b
Acetyl-ALHVPKAPGFAQOLK
             y yyyyyyyy
```


1. Zhang J, et al. (2010) Mechanism of folding chamber closure in a group II chaperonin. *Nature* 463:379–383.
2. Chen ZA, et al. (2010) Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* 29:717–726.
3. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2:1896–1906.
4. Ditzel L, et al. (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* 93:125–138.
5. Shomura Y, et al. (2004) Crystal structures of the group II chaperonin from Thermococcus strain KS-1: steric hindrance by the substituted amino acid, and inter-subunit rearrangement between two crystal forms. *J Mol Biol* 335:1265–1278.
6. Levitt M (1992) Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol* 226:507–533.
7. Xu H, Freitas MA (2009) MassMatrix: A database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *Proteomics* 9:1548–1555.
8. UniProt Consortium (2009) The universal protein resource (UniProt) 2009. *Nucleic Acids Res* 37:D169–174.

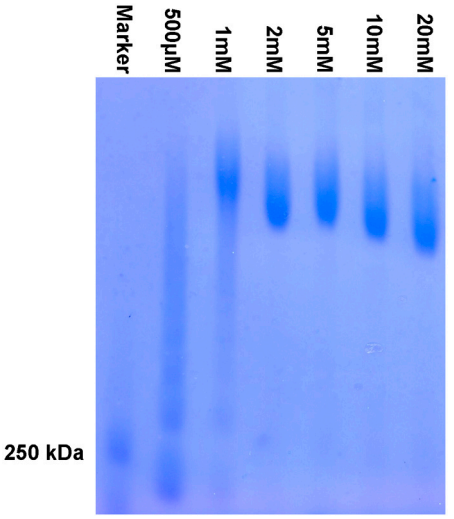

Fig. S1. Agarose-SDS gel of cross-linked mmCPN complex under different cross-linker concentrations. This homo-oligomeric group II chaperonin is very similar to TRiC in both the particle architecture and subunit sequence. Cross-linking occurred under native conditions and SDS was added only after the cross-linking reaction was quenched. High cross-linker concentration leads to a single band, indicating that all 16 subunits are covalently linked to each other by the cross-linker. The lowest cross-linker concentration for which a single band occurs is 2 mM final BS3 concentration. This was our concentration of choice for TRiC.

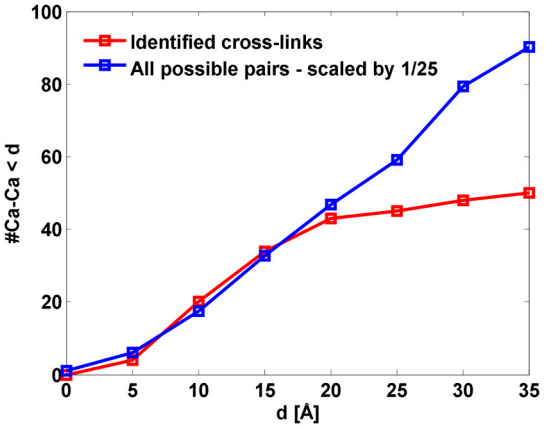

Fig. S2. The cumulative distributions derived from the two distance histograms shown in Fig. 5. The cumulative distribution of the nonspecific (i.e., all possible) lysine pairs (blue) is scaled by 1/25 in order to highlight the similar rise of the slop.

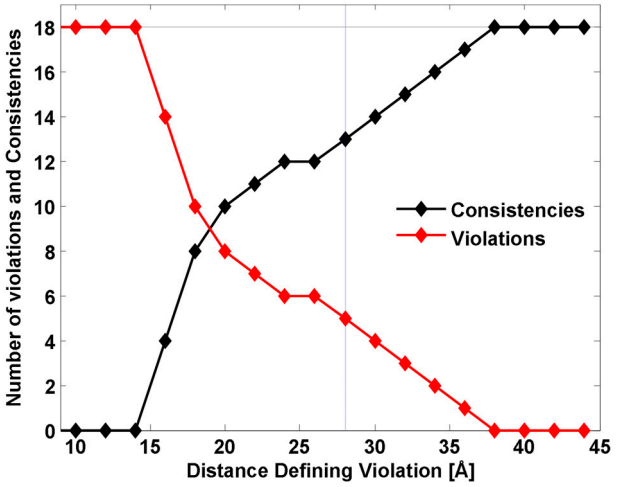

Fig. S3. The dependency of the number of violations on the threshold distance defining a violation for a cross-link. Cross-link set is the 18 inter-subunit cross-links with structured lysine residues. Distances are measured on the OMS full-particle model. In this work the threshold was 28 Å (blue line).

Table S1. Complete listing of top-fitting cross-links identified by our MS analysis

Peptide 1*	Peptide 2*	Subunit1		Subunit2		MS/MS score	Ch	MS1 error [ppm]	Distance in model [Å]†
Cross-links between different subunits									
KKGDDQSR	AGMSSL**K**G	G	K528	Z	K530	0.94	4+	−1.4	N/A
VDNII**K**AAPR	G**K**SAYQDR	B	K522	D	K24	0.94	4+	−1.9	16.5
KGDDQSR	AINS**K**AEVAR	G	K529	Z	K10	0.88	4+	−2.3	N/A
T**K**MK	**K**GESQTDIEITR	A	K245	G	K249	0.88	4+	−1.9	14.8
EAT**K**AAR	NTEPLIQTA**K**TTLGSK	B	K135	E	K170	0.83	4+	−2.0	19.0
EAESLIA**K**K	G**K**SAYQDR	B	K119	D	K24	0.82	4+	−0.7	26.2
DL**K**R	M**K**IMVVK	A	K317	D	K324	0.82	4+	−1.0	17.6
KKGDDQSR	TLNP**K**AEVAR	G	K528	Z	K10	0.78	5+	−2.2	N/A
LAC**K**EAVR	TSLGP**K**GMDK	A	K126	D	K58	0.78	3+	−2.5	21.0
TCLGP**K**SMMK	AINS**K**AEVAR	G	K44	Z	K10	0.75	4+	3.8	30.8
KYAR	AGMSSL**K**G	G	K197	Z	K530	0.75	3+	−0.2	N/A
AVDDGVNTF**K**VLTR	H**K**SETDTSLIR	Q	K400	Z	K199	0.72	4+	−0.8	16.2
LC**K**TVGATALPR	HPDMox**K**K	Q	K326	Z	K222	0.72	4+	−3.7	14.4
AVDDGVNTF**K**VLTR	A**K**R	Q	K400	Z	K317	0.71	4+	−1.2	32.7
TPG**K**EAVAMoxESYAK	G**K**SAYQDR	B	K431	D	K24	0.68	4+	−1.3	36.2
TCLGP**K**SMMK	YKPSV**K**GR	G	K44	Z	K430	0.67	4+	−3.5	19.9
G**K**SAYQDR	G**K**SAYQDR	D	K24	D	K24	0.63	4+	−0.3	22.4
LC**K**TVGATALPR	A**K**R	Q	K326	Z	K317	0.6	4+	−0.3	29.9
TSLGP**K**GMDK_	**K**SR	D	K58	E	K25	0.54	4+	−1.3	34.1
KGDDQSR	AGMSSL**K**G	G	K529	Z	K530	0.53	3+	−1.7	N/A
VPEEDL**K**R	**K**ETEGDVTSVK	H	K320	Q	K225	0.53	4+	−1.5	13.6
Nac-ALHVP**K**APGFAQMLK	TLNP**K**AEVAR	Q	K7	Z	K10	0.5	4+	1.8	N/A
IDDIVSGH**K**K	AGMoxSSL**K**G	G	K527	Z	K530	0.5	4+	3.2	N/A
T**K**MK	IVLLDSSLEY**K**K	A	K245	G	K248	0.5	4+	0.8	17.8
KVQSGNINAAK	AGMSSL**K**G	G	K21	Z	K530	0.47‡	3+	−2.6	N/A
Nac-ALHVP**K**APGFAQMLK	AGMSSL**K**G	Q	K7	Z	K530	0.46‡	3+	2.0	N/A
EGA**K**HFSGLEEAVYR	TLNP**K**AEVAR	Q	K20	Z	K10	0.32‡	5+	0.1	15.9
Cross-links within the same subunit									
KEDKVEQR	**K**LLEK	H	K145	H	K153	1.0	4+	−1.7	12.8
KNLK	DN**K**QAGVFEPTIVK	A	K481	A	K499	1.0	4+	−0.1	9.8
K§ED**K**VEQR	**K**LLEK	H	K148	H	K153	0.92	4+	−5.4	10.3
V**K**K	**K**VCGDSDKGFVVINQK	Z	K272	Z	K280	0.89	5+	−4.4	12.2
G**K**SAYQDR	D**K**PAQIR	D	K24	D	K32	0.87	4+	−2.0	15.0
KEDKVEQR	LLE**K**CAMoxTALSSK	H	K145	H	K157	0.86	4+	−3.3	16.1
KIIELK	**K**VCGDSDK	Z	K273	Z	K280	0.86	3+	−1.1	12.3
G**K**SAYQDR	TSLGP**K**GMDK	D	K24	D	K58	0.83	4+	−3.0	49.3
QI**K**K	M**K**IMVVK	D	K294	D	K324	0.82	3+	−0.6	9.1
IVNA**K**IACLDFSLQK	NT**K**AR	A	K233	A	K368	0.8	4+	−0.5	12.2
IAILTCPFEPP**K**PK	T**K**HK	E	K259	E	K263	0.78	4+	−1.4	8.7
VE**K**AK	M**K**IMVVK	D	K246	D	K324	0.75	4+	1.2	14.0
LV**K**AER	**K**FIEDR	Z	K261	Z	K265	0.75	4+	−2.3	6.1
H**K**SETDTK	SVTLLV**K**GPNK	Z	K199	Z	K377	0.74	4+	−1.2	6.3
IDDIVSGH**K**K	**K**GDDQSR	G	K527	G	K529	0.71	4+	−2.7	N/A
AHEILPDLVCCSA**K**NLR	D**K**R	Q	K152	Q	K406	0.7	4+	−0.2	6.2
IV**K**K	GSN**K**LVIEEAER	D	K215	D	K398	0.69	4+	−0.9	8.4
LAC**K**EAVR	SL**K**FATEAAITILR	A	K126	A	K515	0.68	4+	0.1	8.5
LEDT**K**LIK	LGFAGLV**K**EISFGTTK	E	K223	E	K360	0.67	4+	−0.2	16.3
AYILNLV**K**QIK	M**K**IMVVK	D	K291	D	K324	0.67	4+	1.1	11.7
LNS**K**WDLR	LC**K**TVGATALPR	Q	K318	Q	K326	0.65	4+	0.5	12.5
E**K**LVK	**K**FIEDR	Z	K258	Z	K265	0.64	4+	−1.7	10.2
A**K**R	HTLTQI**K**DAIR	Z	K317	Z	K388	0.64	4+	−2.3	20.2
ED**K**VEQR	YNFFTGCP**K**AK	H	K148	H	K366	0.61	4+	−1.8	18.4
EIAVTV**K**K	ED**K**VEQR	H	K144	H	K148	0.6	3+	−0.3	4.9
YI**K**NPR	IGDEYFTFITEC**K**DPK	G	K234	G	K367	0.59	4+	−0.1	13.9
KFIEDR	**K**IIELK	Z	K265	Z	K273	0.58	4+	−3.3	11.8
TLNP**K**AEVAR	AGMSSL**K**G	Z	K10	Z	K530	0.56	3+	1.3	N/A
LV**K**AER	V**K**K	Z	K261	Z	K272	0.56	4+	−0.8	15.7
KFIEDR	V**K**K	Z	K265	Z	K272	0.56	4+	−0.3	9.9
LV**K**AER	GIDPFSLDALA**K**HGILALR	Z	K261	Z	K307	0.56	4+	3.3	19.0
VAEIEHAE**K**EK	M**K**EK	B	K272	B	K276	0.53	5+	−1.5	8.7
EAESLIA**K**K	EAVAMESYA**K**ALR	B	K119	B	K441	0.5	3+	1.4	11.4
EIDI**K**K	**K**TDNNR	G	K196	G	K317	0.5	4+	−3.2	16.4
LISQQ**K**AFFAK	LLE**K**CAMTALSSK	H	K172	H	K157	0.42‡	4+	−3.7	10.2
GAS**K**EILSEVER	VE**K**IPGGIIEDSCVLR	G	K381	G	K203	0.36‡	4+	−4.0	8.4
Lower confidence cross-links that were not part of the set of 63 used in Fig. 3									
DL**K**R	LGGTIDDCELVEGLVLTQ**K**VANSGITR	A	K317	D	K235	0.48	4+	−3.4	13.6
VL**K**R	LGGTIDDCELVEGLVLTQ**K**VANSGITR	A	K313	D	K235	0.48	4+	0.6	15.1

Peptide 1*	Peptide 2*	Subunit1		Subunit2		MS/MS score	Ch	MS1 error [ppm]	Distance in model [Å]†
AVDDGVNTF**K**VLTR	D**K**R	Q	K400	Q	K406	0.47	4+	0.4	10.5
LAC**K**EAVR	V**K**SLK	A	K126	A	K512	0.46	4+	2.1	11.9
H**K**SETDTSLIR	E**K**LVK	Z	K199	Z	K258	0.44	4+	3.2	56.1
LNS**K**WDLR	LNS**K**WDLR	Q	K318	Q	K318	0.38	3+	2.5	117.2

The specific cross-linked lysine residues are listed together with their MS parameters. Abbreviations: A-CCT1(α); B-CCT2(β); G-CCT3(γ); D-CCT4(δ); E-CCT5(ϵ); Z-CCT6(ζ); H-CCT7(η); Q-CCT8(θ); Mox is oxidized methionine; Nac is the acetylated N-terminal of polypeptide.

*The most likely cross-linked lysine is underlined.

†The C^{α}–C^{α} distance within the lysine pair when measured on the OMS arrangement model of TRiC (Fig. 4). N/A means that at least one of the lysine residues corresponds to an unstructured part of the 1Q3R template.

‡Cross-link was included in spite of its low MS/MS score as there are at least four identified MS/MS fragmentation sites on each peptide.

§Lysine side-chain with attached hydrolyzed BS3.

Table S2. N-terminal modifications of the TRiC subunits

Subunit	First methionine missing	N-terminal acetylation
CCT1(α)	No	Yes
CCT2(β)	Yes	Yes
CCT3(γ)	No	Yes
CCT4(δ)	Not determined	Not determined
CCT5(ϵ)	Yes	Yes
CCT6(ζ)	Not determined	Not determined
CCT6 isoform 2	Not determined	Not determined
CCT7(η)	No	Yes
CCT8(θ)	Yes	Yes

Not determined: Neither the unmodified nor the modified forms were identified in any peptide.

Other Supporting Information Files

Dataset S1 (DOC)