The C-terminal tails of GroEL and its mitochondrial and chloroplastic homologs adopt polyproline II helices.

Cristian Segura Rodríguez[1], Rubén López-Sánchez [1] & Douglas Vinson Laurents[1,*]

[1]*Institute for Physical Chemistry "Blas Cabrera", CSIC*

Serrano 119, 28006 Madrid, Spain

*E-mail: dlaurents@iqf.csic.es

Telephone:+34-671499363

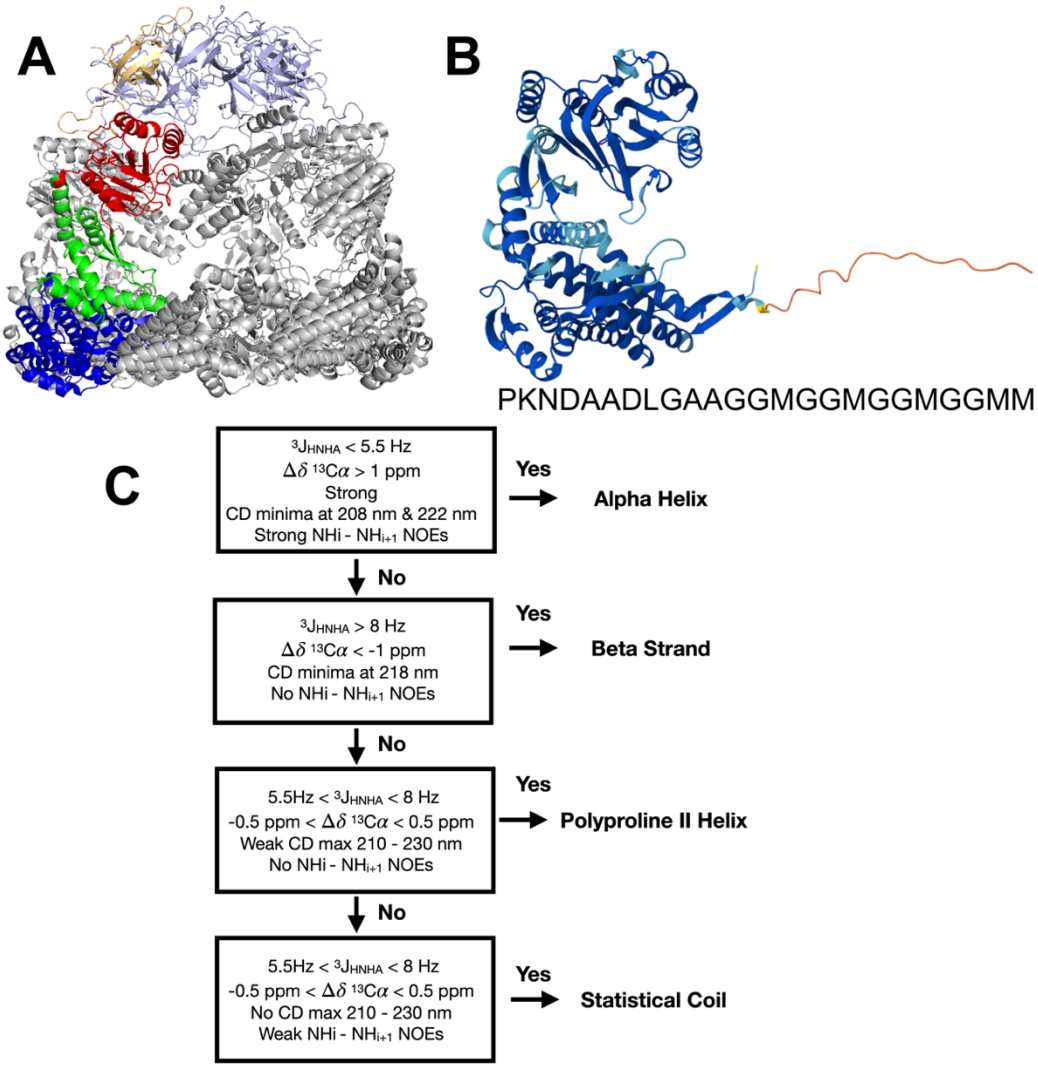
Running title: Polyproline II helix in GroEl's C-terminal tail

ABSTRACT The chaperonin GroEL and its mitochondrial and chloroplastic homologs mHsp60 and Cpn60 are large barrel-like oligomeric proteins. Chaperonins facilitate folding by isolating nascent chains in their hollow interior and undergoing conformational transitions driven by ATP hydrolysis. Due to their vital importance, the structure of GroEL and its homologs have been extensively studied by X-ray crystallography and CryoEM, revealing one or two rings each of which contains seven subunits. Each subunit has three folded domains and a twenty-four residue C-terminal extension. Whereas this C-terminal tail has been reported to bind and stimulate the folding of the client protein, it appears to be invisible or blurry, which suggests disorder. The objective of this study is to characterize conformational preferences in the C-terminal tails of GroEL, mHsp60 and representative Cpn60s using circular dichroism and nuclear magnetic resonance spectroscopies. The tails of GroEL and mHsp60 consist of two segments. The first is rich in residues typical of intrinsically disordered proteins and the second segment consists exclusively (GroEL) or almost entirely (mHsp60) of Gly and Met residues. The spectroscopic results evince that these C-terminal extensions are not wholly disordered but adopt polyproline II helices whose populations are higher in the second Gly/Met-rich segment. Whereas the C-terminal segments of chloroplastic chaperonins are Gly-poor, they are rich in proline and also adopt polyproline II helix conformations. These results provide insight into the function of chaperonin C-terminal tails.

Introduction

Chaperonins are protein folding machines found in all kingdoms of life[1,2]. Eubacteria chaperonins, as exemplified by the well characterized protein GroEL from *E. coli* consist of seven subunits, each containing three folded domains. These oligomerize to form a homo-heptamer with a hollow interior. Unfolded client proteins enter the barrel and bind to its interior surface. The open barrel is then capped by GroES, a smaller homoheptameric protein that acts like a lid. Conformational changes in the complex, which are influenced by ATP hydrolysis, GroES binding and the dimerization of the $GroEL_7 \cdot GroES_7$ complex, aid the proper folding of client proteins[2]. GroEL homologs are present in mitochondria (called Hsp60, sequence identity versus GroEL ~60%) with a similar structure[3], and chloroplasts (called Cpn60, sequence identity to GroEL ~ 60%)[4]. In addition to the ordered domains, both GroEL and Hsp60 contain disordered C-terminal tails of about 25 residues in length which appear invisible or diffuse by X-ray crystallography and CyroEM and are predicted to be disordered by AlphaFold (**Figure 1**). Nevertheless, these tails contribute to substrate folding[5] and form a barrier that blocks substrates from escaping through the bottom of the GroEL barrel[6]. In GroEL and Hsp60, these tails consist of several charged, short or polar residues followed by a ~19 residue segment with a remarkably high glycine and methionine content (**Sup. Table 1**). Recently, a structural class of proteins containing glycine-rich polyproline II helical bundle domains has emerged[7]. Moreover, the C-terminal Gly/Met-rich segments of GroEl and mHsp60 resemble Abductin, an elastic protein present in the hinge ligament of bivalves[8], which adopts polyproline II (PPII) helix conformations[9]. This led us to wonder whether the C-terminal tails of GroEL and mHsp60 might also adopt PPII helical conformations, and here we will test this hypothesis.

Figure 1. The "disordered" C-terminal segment in GroEL/mHsp60 and spectral criteria for its conformational characterization.



A. Structure of mHsp60/mHsp10 complex: PDB=6MRD (ref. #3). Six mHsp10 subunits are tinted blue/white, the seventh is colored gold. Six mHsp60 subunits are colored gray, the seventh is colored red, green and blue for its apical, intermediate and equatorial domains, respectively.

B. AlphaFold2 structural model (https://alphafold.ebi.ac.uk/entry/P0A6F5) of *E. coli* GroEL (one subunit). Residues whose conformation is predicted with high confidence are colored blue (the three folded domains) and those with low/very low confidence are colored orange and red; namely the C-terminal segment, whose amino acid sequence is written.

C. Criteria from NMR data and CD spectral features which can distinguish α-helices, β-strands, polyproline II helices and statistical coil ensembles. These criteria are based on those of Ma *et al.*[10] but substituting the $^1H\alpha\Delta\delta$ for $^{13}C\alpha\Delta\delta$, as the latter are more reliable[11].

The chloroplastic chaperonins, which are key for RuBisCO folding[12], contain two types of subunits, called α and β, have been reported to form heptamers[13]; which then combine to form heterodecatetramers[14]. Like the distinct subunits of cytoplasmatic chaperonins[4], it is possible that the α and β subunits of Cpn60 underwent divergent evolution to optimize function[15]. The chloroplastic subunits also have C-terminal extensions predicted to be disordered by AlphaFold, but unlike the tails of GroEL and mHsp60, they are glycine-poor but are often rich in proline residues (**Sup. Table 1**). Characterizing these segments is a second goal of this study.

Results

The Gly-rich stretch of GroEL and mHsp60 contains a high percentage of PPII helix.

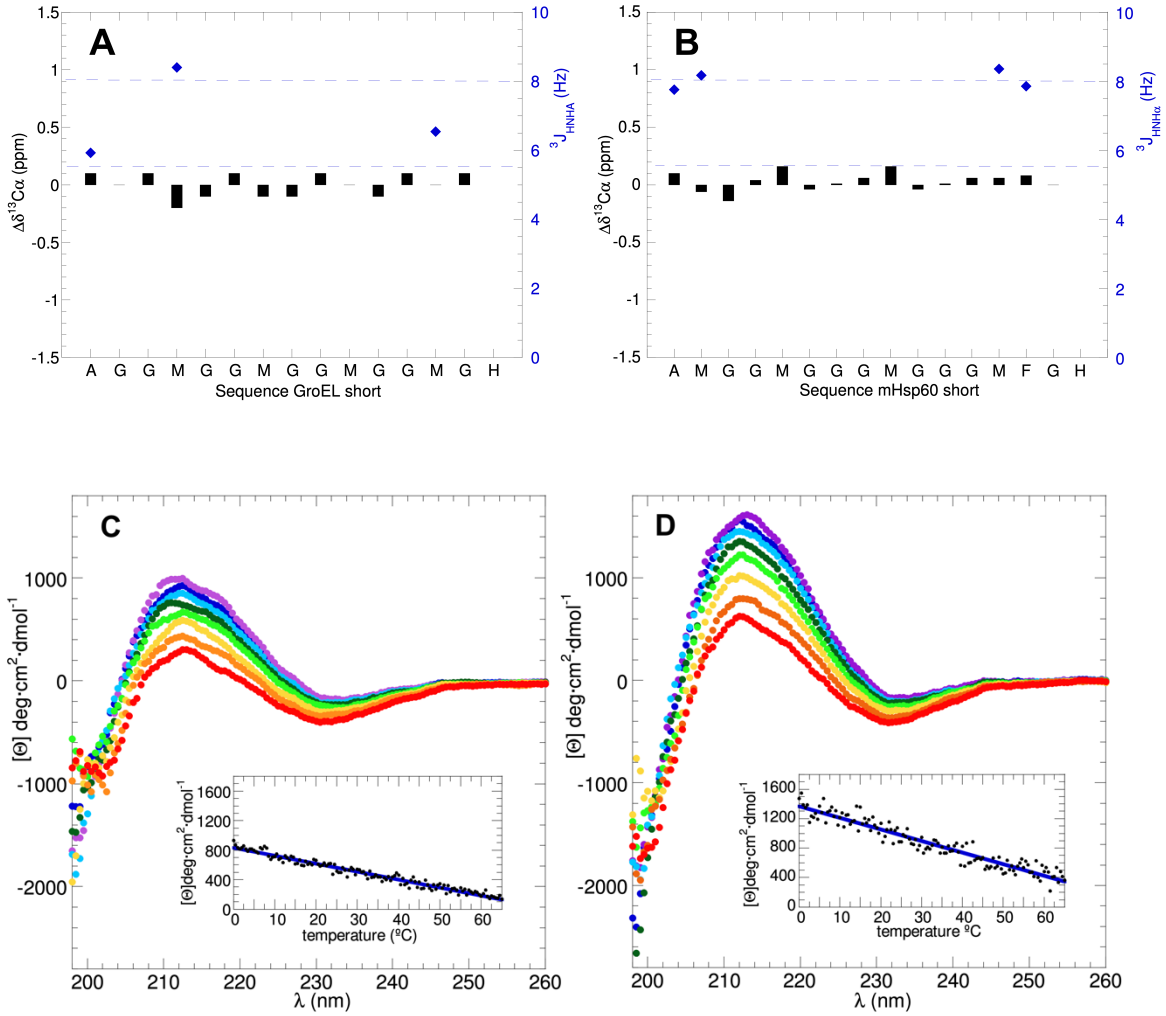
To test for the presence of polyproline II helix, we studied short peptides corresponding to just the Gly/Met-rich stretch right at the C-terminus of GroEL and mHsp60. The NMR spectra of these peptides, called GroELCtS and mHsp60CtS, show poor ^{1}HN and ^{13}Cα chemical shift dispersion and degenerate ^{1}Hα glycine signals (**Sup. Fig. 1**). Moreover the ^{13}Cα conformational chemical shifts are very small. These chemical shift data indicate an absence of α-helix and β-strand conformations (**Figure 2A,B**). The measured $^{3}J_{HNH\alpha}$ couplings are consistent with extended conformations.

The far UV-CD spectra of GroELCtS and mHsp60CtS, recorded over a range of temperatures from 0 to 65ºC, are shown in **Fig. 2C,D**. At low temperatures, the spectra show a maximum near 212 nm which is characteristic of polyamide PPII helices[16,17]. Upon heating, this maximum weakens in intensity. The CD spectral bands are stronger for Hsp60CtS than GroELCtS. This could be attributed to a greater number of glycine residues in the former, as the intrinsic PPII forming propensity of Gly is higher than Met[18,19]. Based on their maximum CD signals of 1600 deg·cm^{2}·dmol^{-1} and 1000 deg·cm^{2}·dmol^{-1}, the populations of PPII helix at 0 ºC are estimated to be 47% and 43% for Hsp60CtS and GroELCtS, respectively (see **Experimental Procedures**). At 37ºC, the physiological relevant temperature, the %PPII helix would be 43% and 41% for Hsp60CtS and GroELCtS, respectively.

Next, the signal at 217 nm was recorded as the sample was heated at 1 ºC per minute up to 65 ºC. Instead of the sigmoidal transition typically observed for folded protein, GroELCtS and mHsp60CtS show a linear loss of signal upon heating (**Fig. 2C,D**

insets). This is consistent with a non-cooperative PPII helix to statistical coil transition, which has been previously observed for other isolated PPII helices[20] or partly populated PPII helices in intrinsically disordered proteins[21].

Figure 2. NMR spectral parameters and CD spectra of *GroEL and mHsp60* C-terminal segments.



A. GroELCtS ^{13}Cα conformation chemical shifts (black bars, left y-axis) and $^{3}J_{HNH\alpha}$ (blue diamonds, right y-axis). As in (**B**), the latter are measured for non-overlapping, non-Glycine residues.

B. mHsp60CtS ^{13}Cα conformation chemical shifts (black bars, left y-axis) and $^{3}J_{HNH\alpha}$ (blue diamonds, right y-axis).

C. Far UV-CD spectra of GroELCtS recorded at 0 ºC (violet), 5ºC (blue), 10ºC (celeste), 17ºC (dark green), 25ºC (light green), 37ºC (yellow), 50ºC (orange), 65ºC (red). *Inset:* Terminal denaturation of the GroELCtS. The fit of a linear equation (blue line) to the data (black points) is $[\Theta]_{217nm}$= 835 – 10.8 (T ºC) R= 0.98.

D. Far UV-CD spectra of mHsp60CtS recorded at 0 ºC (violet), 5ºC (blue), 10ºC (celeste), 17ºC (dark green), 25ºC (light green), 37ºC (yellow), 50ºC (orange), 65ºC (red). *Inset:* Heat denaturation of the C-terminal segment of mHsp60CtS. The fit of a linear equation (blue line) to the data (black points) is $[\Theta]_{217nm}$= 13200 – 15.8(T ºC) R= 0.96.

Complete C-terminal tails contain smaller populations of polyproline II helix.

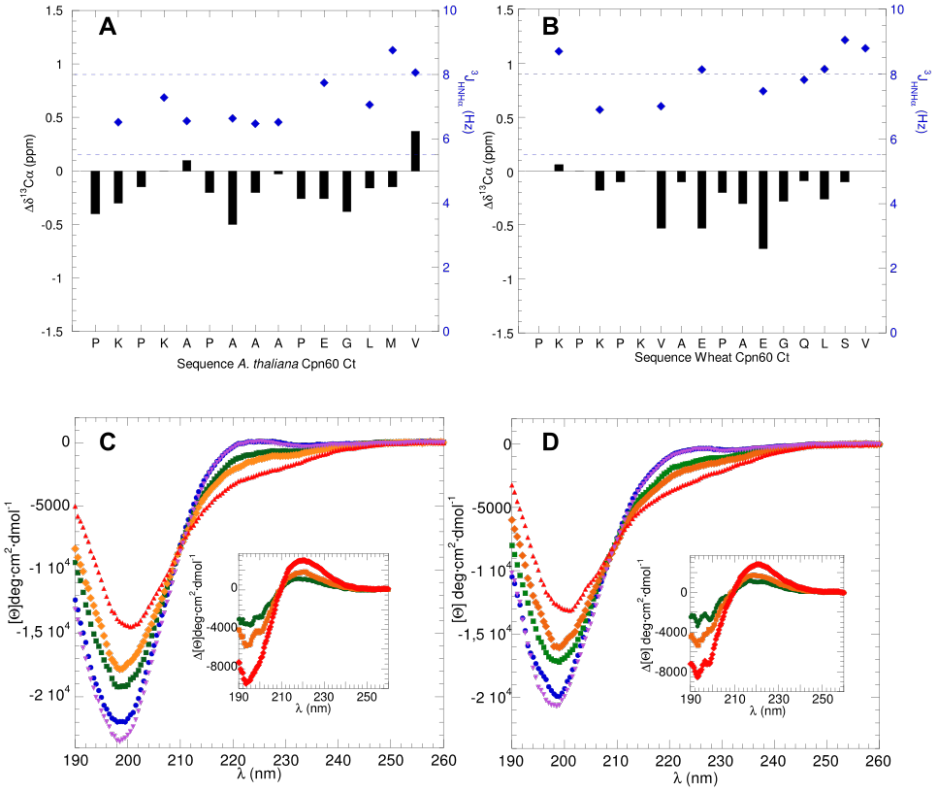
To further corroborate and assess the position of the PPII helix, we also characterized peptides, called GroELCtC and mHsp60CtC, corresponding to the complete segments of GroEL and mHsp60 that are invisible to X-ray crystallographic and CryoEM analyses. The assigned NMR spectra of GroELCtC and mHsp60CtC shown respectively in **Sup. Figures 2 & 3** reveal a lack of ^{1}HN and ^{13}Cα chemical shift dispersion. This is characteristic of an absence of α-helical and β-strand structure[11,22]. Whereas two different ^{1}Hα signals were seen for H-bonded glycine residues in a folded protein consisting of six PPII helices[23], the two glycine ^{1}Hα signals are degenerate in all GroELCt and mHsp60Ct peptides (**Sup. Figures 1, 2 & 3**). This is relevant because the seven C-terminal segments extending from the equatorial domains towards the barrel interior might associate to form a structured bundle of polyproline II helices. The lack distinct glycine ^{1}Hα signals suggests they do not. The small ^{13}Cα conformational chemical shifts and $^{3}J_{HNH\alpha}$ values in the range of 5.5 to 8.0 Hz[24,20,25] rule out the presence of α-helices or β-strands (**Sup. Figure 4 A&B**). The presence of weak maxima in the CD spectra of GroELCtC and mHsp60CtC (**Sup Fig. 4 C&D**) indicate the presence of polyproline II helix. These peaks become weaker upon heating manifesting a partial PPII helix to coil transition which has been observed previously for other PPII helices[26,27]. The difference spectra, which were calculated by subtracting spectra recorded at two different temperatures (**Sup. Fig 4 C&D insets**), are very similar to the CD spectra of peptides[28,29], collagen[30] and proteins[17] known to adopt PPII helices.

Chloroplast homologs of GroEL, which are Proline-rich, also adopt PPII conformations.

The NMR spectra of *A. thaliana* and wheat Cpn60α C-terminal segments, shown in **Sup. Fig. 5**, reveal a poor ^{1}HN chemical shift dispersion which is consistent with a lack of well ordered α-helical or β-sheet structures. There are no detectable $^{1}HN_{\underline{i}}$-$^{1}HN_{i+1}$ or $^{1}HN_{\underline{i}}$-$^{1}HN_{i+3}$ NOE signals which would be characteristic of α-helices. The $^{3}J_{HNH\alpha}$ coupling constants and small or slightly negative ^{13}Cα conformational chemical shifts are indicative of statistical coil and extended (β-strand or PPII helical) conformations

(**Figure 3A,B**). Similar results are seen for the C-terminal segments of *Arabadopsis thaliana* and wheat Cpn60β subunits (**Sup. Figure 6A**). The CD spectra of peptides corresponding to the C-terminal segments of *Arabadopsis thaliana* and wheat Cpn60α recorded at 5, 25, 37 and 65 ºC are shown in **Figure 3C,D**. These spectra show a weak maximum near 224 nm at 5ºC which is suggestive of PPII helical conformations. Except for minor differences in the position of the peak maxima, which could be ascribed to a difference in the content of proline and aromatic residues, the results for the C-terminal segments of *Arabadopsis thaliana* and wheat Cpn60β subunits are similar (**Sup. Figure 6B,C**). Upon heating, the spectra show significant changes including a decreased signal around 220 nm. Like the difference spectra of GroELCtC and mHsp60CtC, the chloroplastic difference spectra closely resemble typical PPII helical spectra. Based on the CD data, the populations of PPII helix in the C-terminal segments of *A. thaliana* and wheat Cpn60α are about 30% and are slightly higher (32% and 34%) for *A. thaliana* and wheat Cpn60β, respectively, at 25ºC.

Figure 3. NMR spectral parameters and CD spectra of *Arabadopsis thaliana* and Wheat Cpn60 C-terminal segments.



^{13}Cα conformational chemistry shifts (black bars, left y-axis) and $^{3}J_{HNH\alpha}$ coupling constants (blue diamonds, right y-axis) for the *A. thaliana* (panel **A**) and wheat (panel **B**) Cpn60 C-terminal segments. Blue dashed horizontal lines mark the range of constants (between 5.5 & 8.0 Hz) expected for statistical coil or polyproline II conformations.

Far UV-CD spectra recorded at 5ºC(blue), 25ºC(green) 37ºC(orange) and 65ºC(red) and after recooling to 5ºC (purple) for the *A. thaliana* (panel **C**) and wheat (panel **D**) Cpn60α C-terminal segments. Insets show the difference CD spectra for 5ºC-25ºC (green), 5ºC-37ºC(orange) and 5ºC-65ºC(red).

Discussion

The main finding of this study is that peptides corresponding to the C-terminal extensions of eubacterial, mitochondrial and chloroplastic chaperonins, which appear disordered in cryoEM and X-ray crystallographic analyses, adopt high populations of polyproline II helix in aqueous solution. This is consistent with the high intrinsic propensity of glycine residues to adopt isolated PPII helices[31] as well as PPII helical bundles in synthetic peptides[32], natural protein domains[33–35] and quasi infinite honeycomb networks[36].

Despite the sequence similarity to the "snow flea" antifreeze protein[23] and abductin[9], which contain glycine-rich polyproline II helices assembled into bundles, the C-terminal chaperonin peptides appear adopt isolated polyproline II helices under the conditions employed here. Nevertheless, since the seven C-terminal tails would be present at concentrations of dozens of mM and subject to crowding effects[37] within chaperonin barrel, the possibility that they may associate under physiological conditions can not be ruled out. The associations between PPII helices of the C-terminal tails could plug the bottom of the chaperonin barrel and prevent client proteins from escaping[6]. Incidentally, the association of the PPII helices formed by Gly-rich C-terminal segments of GroEL into putative PPII helical bundle structures could explain why they are much more resistant to protease K degradation than the C-terminal segments of chloroplastic Cpn60[38]. The latter, being Pro-rich but Gly-poor, can form isolated PPII helices but cannot associate[7].

The results show significant populations of PPII helix in both the α and β subunits of Cpn60. Sequence comparison of Cpn60 α and β subunits (**Sup. Table 1**) suggests that the C-terminal extensions of their α subunits often start with three cationic residues and end with two hydrophobic residues. By contrast, β-subunits frequently begin with two anionic residues and end in a SGYGY motif that is reminiscent of the GY segments of TDP-43 and FUS proteins, which is known to promote π/π and cation/π interactions[39,40]. These differences suggest that the C-terminal tails of the α and β subunits may have specialized through evolution to fold client proteins rich in anionic and aliphatic (Cpn60α subunit) or cationic and aromatic (Cpn60β subunit).

Experimental Procedures

Peptides

Peptides corresponding the complete GroEL and mHsp60 C-terminal segments with sequences: acPKNDAADLGAAGGMGGMGGMGGMMKKK and acPKEEKDPGMGAMGGMGGGMGGGMFKKK, respectively, were obtained from CASLO ApS, Denmark. These "complete" C-terminal peptides, are abbreviated as GroELCtC and mHsp60CtC, respectively. These peptides contain three extra Lys residues at the C-terminus to promote solubility. All peptides were N-acetylated as indicated by "ac" to mimic the lack of charge in the context of the full length protein.

Shorter peptides with just the final Gly/Met rich stretch of these peptides were also obtained from Calso ApS. The peptides, named GroELCtS and mHsp60CtS (the last "S" standing for short) have the sequences: acAGGMGGMGGMGGMMH and acAMGGMGGGMGGGMFGH, respectively. For these peptides, one C-terminal His residue was added to promote solubility.

The N-acetylated peptides acPKPKAPAAAPEGLMV, acPKPKPKVAEPAEGQLSV, acEPEPVPVGNPMDNSGYGY and acEPEAAPLANPMDNSGFGY which correspond to the C-terminal segments of *Arabadopsis thaliana* Cpn60α, wheat Cpn60α, *Arabadopsis thaliana* Cpn60β, and wheat Cpn60β subunits respectively, were purchased from the National Center for Biotechnology (CNB, CSIC) Cantoblanco, Spain.

The sequence and purity of all peptides was verified by mass spectrometry and NMR spectroscopy. Sample concentrations were measured by integration of 1D ^{1}H peptide NMR signals to that of a known concentration of sodium trimethyl-silylpropanesulfonate, which was also used as the internal chemical shift standard.

NMR Spectroscopy

A series of NMR spectra; namely: 2D ^{1}H-^{1}H COSY, TOCSY and NOESY and 2D ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC were recorded on 1 – 3 mM peptide samples in 10 mM K_2HPO_4 buffer (pH 6) and 5ºC on the 600 MHz (^{1}H) Bruker AvanceNeo spectrometer belonging to the "Manuel Rico" NMR laboratory, IQF/CSIC. The spectral parameters are summarized in **Sup. Table 2**. The spectra were assigned and analyzed to obtain ^{13}Cα chemical shifts and $^{3}J_{HNH\alpha}$ coupling constants. The ^{13}Cα chemical shifts reveal conformational tendencies[41] when compared to reference ^{13}Cα chemical shifts predicted for a structureless peptide, which were obtained for each peptide and set of experimental conditions using previously reported parameters[42] (**Figure 1**). $^{3}J_{HNH\alpha}$

couplings also discriminate between α-helical, β-strand, and statistical coil or polyproline II conformations[43] (**Figure 1**).

Circular Dichroism Spectroscopy

Far UV circular dichroism (CD) spectra were recorded on a Jasco 810 spectropolarimeter equipped with a Peltier temperature control unit at 190 – 260 nm or 198 – 260 nm with a 50 nm/min scan speed, 1.67 nm bandwidth, peptide concentrations ranging from 130 to 300 μM in 10 mM K_2HPO_4 buffer prepared in milliQ water at pH 6. Ten scans were recorded and averaged per spectrum. A reference buffer spectrum was subtracted.

Thermal unfolding experiments were monitored by CD at 217 nm from 0ºC to 65ºC employing a 1ºC/min heating rate and a 1.0 nm bandwidth. Spectra were recorded at 0ºC before and after the thermal denaturation experiment to test reversibility. The empirical equation of Stellwagen and co-workers[44] :

Eqn. 1: $\% \text{ PPII helix} = 100 \cdot ([\Theta]max + 5560 \text{ deg} \cdot cm^2 \cdot dmol^{-1}) / 15140 \text{ deg} \cdot cm^2 \cdot dmol^{-1}$,

where [Θ]max is the maximum in the spectrum between 210 and 230 nm in $deg \cdot cm^2 \cdot dmol^{-1}$, was used to estimate the population of PPII helix.

Supporting Information Available: Tables with sequence information, NMR spectral parameters, Figures with NMR & CD spectra and spectral results are included in the Supporting Information.

Acknowledgements: We are grateful to Dr. I. Tascón, Prof. I. Ubarretxena, and Prof. J. M. Valpuesta for helpful conversation and encouragement.

Funding and Additional Information: This study is part of projects PID2019 109306RB-I00 and PID2022-137806OB-I00, funded by the Spanish Ministry of Science, Innovation and Universities: MICIN/AEI/10.13039/501100011033/FEDER, UE. NMR experiments were performed in the "Manuel Rico" NMR laboratory (LMR) of the Spanish National Research Council (CSIC), a node of the Spanish Large-Scale National Facility (ICTS R-LRB).

REFERENCES

(1) Horwich, A. L.; Neupert, W.; Hartl, F. U. Protein-Catalysed Protein Folding. *Trends Biotechnol* **1990**, *8* (5), 126–131. https://doi.org/10.1016/0167-7799(90)90153-o.

(2) Horovitz, A.; Reingewertz, T. H.; Cuéllar, J.; Valpuesta, J. M. Chaperonin Mechanisms: Multiple and (Mis)Understood? *Annu Rev Biophys* **2022**, *51*, 115–133. https://doi.org/10.1146/annurev-biophys-082521-113418.

(3) Gomez-Llorente, Y.; Jebara, F.; Patra, M.; Malik, R.; Nisemblat, S.; Chomsky-Hecht, O.; Parnas, A.; Azem, A.; Hirsch, J. A.; Ubarretxena-Belandia, I. Structural Basis for Active Single and Double Ring Complexes in Human Mitochondrial Hsp60-Hsp10 Chaperonin. *Nat Commun* **2020**, *11* (1), 1916. https://doi.org/10.1038/s41467-020-15698-8.

(4) Kim, S.; Willison, K. R.; Horwich, A. L. Cystosolic Chaperonin Subunits Have a Conserved ATPase Domain but Diverged Polypeptide-Binding Domains. *Trends Biochem Sci* **1994**, *19* (12), 543–548. https://doi.org/10.1016/0968-0004(94)90058-2.

(5) Weaver, J.; Rye, H. S. The C-Terminal Tails of the Bacterial Chaperonin GroEL Stimulate Protein Folding by Directly Altering the Conformation of a Substrate Protein. *J Biol Chem* **2014**, *289* (33), 23219–23232. https://doi.org/10.1074/jbc.M114.577205.

(6) Ishino, S.; Kawata, Y.; Taguchi, H.; Kajimura, N.; Matsuzaki, K.; Hoshino, M. Effects of C-Terminal Truncation of Chaperonin GroEL on the Yield of In-Cage Folding of the Green Fluorescent Protein. *J Biol Chem* **2015**, *290* (24), 15042–15051. https://doi.org/10.1074/jbc.M114.633636.

(7) Rodríguez, C. S.; Laurents, D. V. Architectonic Principles of Polyproline II Helix Bundle Protein Domains. *Arch Biochem Biophys* **2024**, *756*, 109981. https://doi.org/10.1016/j.abb.2024.109981.

(8) Kelly, R. E.; Rice, R. V. Abductin: A Rubber-like Protein from the Internal Triangular Hinge Ligament of Pecten. *Science* **1967**, *155* (3759), 208–210. https://doi.org/10.1126/science.155.3759.208.

(9) Bochicchio, B.; Jimenez-Oronoz, F.; Pepe, A.; Blanco, M.; Sandberg, L. B.; Tamburro, A. M. Synthesis of and Structural Studies on Repeating Sequences of Abductin. *Macromol Biosci* **2005**, *5* (6), 502–511. https://doi.org/10.1002/mabi.200500007.

(10) Ma, K.; Kan, L.; Wang, K. Polyproline II Helix Is a Key Structural Motif of the

Elastic PEVK Segment of Titin. *Biochemistry* **2001**, *40* (12), 3427–3438. https://doi.org/10.1021/bi0022792.

(11) Wang, Y.; Jardetzky, O. Probability-Based Protein Secondary Structure Identification Using Combined NMR Chemical-Shift Data. *Protein Sci* **2002**, *11* (4), 852–861. https://doi.org/10.1110/ps.3180102.

(12) Ellis, R. J. Molecular Chaperones: The Plant Connection. *Science* **1990**, *250* (4983), 954–959. https://doi.org/10.1126/science.250.4983.954.

(13) Nishio, K.; Hirohashi, T.; Nakai, M. Chloroplast Chaperonins: Evidence for Heterogeneous Assembly of Alpha and Beta Cpn60 Polypeptides into a Chaperonin Oligomer. *Biochem Biophys Res Commun* **1999**, *266* (2), 584–587. https://doi.org/10.1006/bbrc.1999.1868.

(14) Zhao, Q.; Zhang, X.; Sommer, F.; Ta, N.; Wang, N.; Schroda, M.; Cong, Y.; Liu, C. Hetero-Oligomeric CPN60 Resembles Highly Symmetric Group-I Chaperonin Structure Revealed by Cryo-EM. *Plant J* **2019**, *98* (5), 798–812. https://doi.org/10.1111/tpj.14273.

(15) Zhao, Q.; Liu, C. Chloroplast Chaperonin: An Intricate Protein Folding Machine for Photosynthesis. *Front Mol Biosci* **2017**, *4*, 98. https://doi.org/10.3389/fmolb.2017.00098.

(16) Woody, R. W. Circular Dichroism Spectrum of Peptides in the Poly(Pro)II Conformation. *J Am Chem Soc* **2009**, *131* (23), 8234–8245. https://doi.org/10.1021/ja901218m.

(17) Gates, Z. P.; Baxa, M. C.; Yu, W.; Riback, J. A.; Li, H.; Roux, B.; Kent, S. B. H.; Sosnick, T. R. Perplexing Cooperative Folding and Stability of a Low-Sequence Complexity, Polyproline 2 Protein Lacking a Hydrophobic Core. *Proc Natl Acad Sci U S A* **2017**, *114* (9), 2241–2246. https://doi.org/10.1073/pnas.1609579114.

(18) Andrews, B.; Zhang, S.; Schweitzer-Stenner, R.; Urbanc, B. Glycine in Water Favors the Polyproline II State. *Biomolecules* **2020**, *10* (8). https://doi.org/10.3390/biom10081121.

(19) Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A. F.; Creamer, T. P. Host-Guest Study of Left-Handed Polyproline II Helix Formation. *Biochemistry* **2001**, *40* (48), 14376–14383. https://doi.org/10.1021/bi011043a.

(20) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II Structure in a Sequence of Seven Alanine Residues. *Proc Natl Acad Sci*

U S A **2002**, *99* (14), 9190–9195. https://doi.org/10.1073/pnas.112193999.
(21) Kjaergaard, M.; Nørholm, A.-B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. Temperature-Dependent Structural Changes in Intrinsically Disordered Proteins: Formation of Alpha-Helices or Loss of Polyproline II? *Protein Sci* **2010**, *19* (8), 1555–1564. https://doi.org/10.1002/pro.435.
(22) Wishart, D. S.; Sykes, B. D. The 13C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using 13C Chemical-Shift Data. *J Biomol NMR* **1994**, *4* (2), 171–180. https://doi.org/10.1007/BF00175245.
(23) Treviño, M. Á.; Pantoja-Uceda, D.; Menéndez, M.; Gomez, M. V.; Mompeán, M.; Laurents, D. V. The Singular NMR Fingerprint of a Polyproline II Helical Bundle. *J Am Chem Soc* **2018**, *140* (49), 16988–17000. https://doi.org/10.1021/jacs.8b05261.
(24) Pardi, A.; Billeter, M.; Wüthrich, K. Calibration of the Angular Dependence of the Amide Proton-C Alpha Proton Coupling Constants, 3JHN Alpha, in a Globular Protein. Use of 3JHN Alpha for Identification of Helical Secondary Structure. *J Mol Biol* **1984**, *180* (3), 741–751. https://doi.org/10.1016/0022-2836(84)90035-4.
(25) Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. Intrinsic Propensities of Amino Acid Residues in GxG Peptides Inferred from Amide I' Band Profiles and NMR Scalar Coupling Constants. *J Am Chem Soc* **2010**, *132* (2), 540–551. https://doi.org/10.1021/ja9058052.
(26) Tiffany, M. L.; Krimm, S. Effect of Temperature on the Circular Dichroism Spectra of Polypeptides in the Extended State. *Biopolymers* **1972**, *11* (11), 2309–2316. https://doi.org/10.1002/bip.1972.360111109.
(27) Park, S. H.; Shalongo, W.; Stellwagen, E. The Role of PII Conformations in the Calculation of Peptide Fractional Helix Content. *Protein Sci* **1997**, *6* (8), 1694–1700. https://doi.org/10.1002/pro.5560060809.
(28) Tiffany, M. L.; Krimm, S. Circular Dichroism of Poly-L-Proline in an Unordered Conformation. *Biopolymers* **1968**, *6* (12), 1767–1770. https://doi.org/10.1002/bip.1968.360061212.
(29) Sreerama, N.; Woody, R. W. Poly(pro)II Helices in Globular Proteins: Identification and Circular Dichroic Analysis. *Biochemistry* **1994**, *33* (33), 10022–10025. https://doi.org/10.1021/bi00199a028.
(30) Drzewiecki, K. E.; Grisham, D. R.; Parmar, A. S.; Nanda, V.; Shreiber, D. I. Circular Dichroism Spectroscopy of Collagen Fibrillogenesis: A New Use for an Old Technique. *Biophys J* **2016**, *111* (11), 2377–2386.

https://doi.org/10.1016/j.bpj.2016.10.023.
(31) Ohnishi, S.; Kamikubo, H.; Onitsuka, M.; Kataoka, M.; Shortle, D. Conformational Preference of Polyglycine in Solution to Elongated Structure. *J Am Chem Soc* **2006**, *128* (50), 16338–16344. https://doi.org/10.1021/ja066008b.
(32) Lotz, B.; Keith, H. D. The Crystal Structures of Poly(LAla-Gly-Gly-Gly)II and Poly(LAla-Gly-Gly)II. *J Mol Biol* **1971**, *61* (1), 195–200. https://doi.org/10.1016/0022-2836(71)90216-6.
(33) De Munck, S.; Provost, M.; Kurikawa, M.; Omori, I.; Mukohyama, J.; Felix, J.; Bloch, Y.; Abdel-Wahab, O.; Bazan, J. F.; Yoshimi, A.; Savvides, S. N. Structural Basis of Cytokine-Mediated Activation of ALK Family Receptors. *Nature* **2021**, *600* (7887), 143–147. https://doi.org/10.1038/s41586-021-03959-5.
(34) Pentelute, B. L.; Gates, Z. P.; Tereshko, V.; Dashnau, J. L.; Vanderkooi, J. M.; Kossiakoff, A. A.; Kent, S. B. H. X-Ray Structure of Snow Flea Antifreeze Protein Determined by Racemic Crystallization of Synthetic Protein Enantiomers. *J Am Chem Soc* **2008**, *130* (30), 9695–9701. https://doi.org/10.1021/ja8013538.
(35) Dunne, M.; Denyes, J. M.; Arndt, H.; Loessner, M. J.; Leiman, P. G.; Klumpp, J. Salmonella Phage S16 Tail Fiber Adhesin Features a Rare Polyglycine Rich Domain for Host Recognition. *Structure* **2018**, *26* (12), 1573-1582.e4. https://doi.org/10.1016/j.str.2018.07.017.
(36) Crick, F. H.; Rich, A. Structure of Polyglycine II. *Nature* **1955**, *176* (4486), 780–781. https://doi.org/10.1038/176780a0.
(37) Rivas, G.; Minton, A. P. Macromolecular Crowding In Vitro, In Vivo, and In Between. *Trends Biochem Sci* **2016**, *41* (11), 970–981. https://doi.org/10.1016/j.tibs.2016.08.013.
(38) Bai, C.; Guo, P.; Zhao, Q.; Lv, Z.; Zhang, S.; Gao, F.; Gao, L.; Wang, Y.; Tian, Z.; Wang, J.; Yang, F.; Liu, C. Protomer Roles in Chloroplast Chaperonin Assembly and Function. *Mol Plant* **2015**, *8* (10), 1478–1492. https://doi.org/10.1016/j.molp.2015.06.002.
(39) Pantoja-Uceda, D.; Stuani, C.; Laurents, D. V.; McDermott, A. E.; Buratti, E.; Mompeán, M. Phe-Gly Motifs Drive Fibrillization of TDP-43's Prion-like Domain Condensates. *PLoS Biol* **2021**, *19* (4), e3001198. https://doi.org/10.1371/journal.pbio.3001198.
(40) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.;

Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174* (3), 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006.

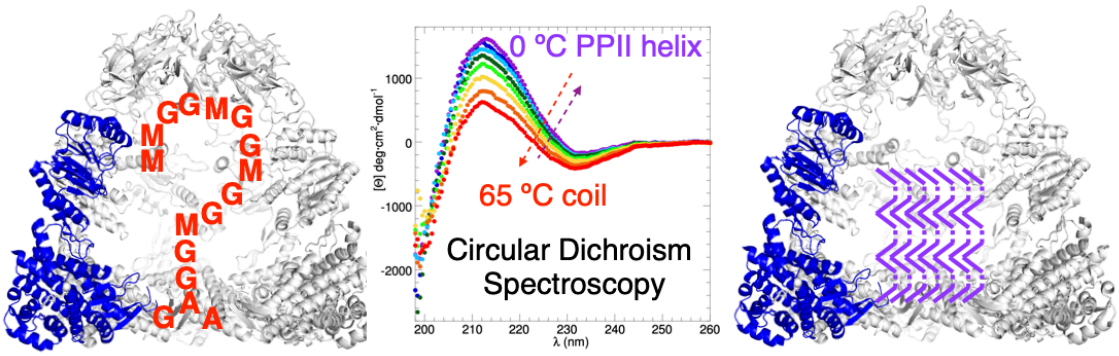
(41) Wishart, D.; Skyes B. D. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR*. **4**: 171-180.

(42) Kjaergaard, M.; Brander, S.; Poulsen, F. M. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J. Biomol. NMR*. **49**: 139-149.

(43) Pardi, A.; Billeter, M.; Wüthrich, K. Calibration of the angular dependence of the amide proton-C^{α} proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein: Use of $^{3}J_{HN\alpha}$ for identification of helical secondary structure. *J. Mol Biol.* **180**: 741-751.

(44) Park, S-H.; Shalongo, W.; Stellwagen, E. The role of PII conformations in the calculation of peptide fractional helix content. *Prot. Sci.* **6**: 1694-1700.

Table of Contents Graphic



Supporting Information for

The C-terminal tails of GroEL and its mitochondrial and chloroplastic homologs adopt polyproline II helices.

Cristian Segura Rodríguez[1], Rubén López-Sánchez [1] & Douglas Vinson Laurents[1*],

Institute for Physical Chemistry "Blas Cabrera", CSIC

Serrano 119, 28006 Madrid, Spain

**e-mail for correspondence: dlaurents@iqf.csic.es*

Contents

I. Supporting Table 1: Sequences of Representative Chaperonin C-terminal Tails

II. Supporting Table 2: NMR spectral parameters.

III. Supporting Figure 1: NMR spectra of GroELCtS and mHsp60CtS

IV. Supporting Figure 2: NMR Spectra of GroELCtC

V. Supporting Figure 3: NMR Spectra of mHsp60CtC

VI. Supporting Figure 4: NMR Parameters and CD Spectra of GroELCtC and mHsp60CtC Segments

VII. Supporting Figure 5: NMR Spectra of *A. thaliana* and wheat Cpn60α

VIII. Supporting Figure 6: NMR Parameters and CD Spectra of *A. thaliana* and wheat Cpn60β

Supporting Table 1

Aminoacid sequences of C-terminal disordered segments in *E. coli* GroEL, human mitochondrial mHsp60 and representative chloroplastic Cpn60 α and β subunits. Glycine, proline, aliphatic and aromatic residues are colored green, blue, bold black and purple, respectively.

Chaperonin	Uniprot Code	Sequence
E. coli GroEl	P0A6F5	P_{525}KNDAADLGAAGGMGGMGGMGGMGGMM$_{548}$
H. sapiens mitochondrial mHsp60	P10809	P_{550}KEEKDPGMGAMGGMGGGMGGGMF$_{573}$
Arabidopsis thaliana (Mouse ear cress) α subunit	P21238	P_{571}KPKAPAAAPEGLMV$_{586}$
Triticum aestivum (Wheat) Cpn60 α subunit	P08823	P_{527}KPKPKVAEPAEGQLSV$_{543}$
Fagus sylvatica (Beech tree) Cpn60 α subunit	A0A2N9GXR7	K_{570}PKPRSPVAAPPQGLTV$_{586}$
Citrus sinensis (Sweet orange) α	A0A067F287	K_{565}PKPKTPVAAPPQGLMV$_{582}$
Malus domesticus (Apple tree) α	A0A498INK5	K_{632}AKPKSAAAAAPQGMTV$_{648}$
Eucalyptus grandis α subunit	A0A059D558	K_{642}PKAPAAASPQGLTV$_{656}$
Theobromo cacao α subunit, isoform 1	A0A061GX48	K_{570}PKPKAPAAAAAQAMEIKDDPRKWLTYSGEARS$_{602}$
Lactuca sativa (Lettuce) α subunit	A0A2J6K5Y6	K_{572}PKPKAPMAAAPQGMSI$_{588}$
Oryza sativa (Rice) α subunit	Q0INX9	K_{561}PKPKAPVAEPAEGTLTV$_{578}$
Helianthus annuus (Sunflower) α subunit	A0A251VPF5	K_{632}PKPRAPRAAEPQGLTV$_{648}$
Maize α subunit	B7ZZZ2	K_{461}PKKAPAAAAAAAP$_{474}$
Cicer Arietinum (Chickpea) α subunit	A0A1S2YKM3	K_{572}PKPRAPIAGSPQGLTV$_{588}$
Arabidopsis thaliana (Mouse ear cress) β subunit	P21240	E_{583}PEPVPVGNPMDNSGYGY$_{600}$
Triticum aestivum (Wheat) Cpn60 β subunit	A0A3B6I2A6	I_{565}KEPEAAPLANPMDNSGFGY$_{684}$
Fagus sylvatica (Beech tree) Cpn60 β subunit	A0A2N9F1H9	K_{590}EPESAVPAGNPMDNSGYGY$_{609}$

Citrus sinensis (Sweet orange) β	A0A067EKN6	**I**$_{555}$**I**DPTK**V**SCC**I**A**I**L$_{571}$
Malus domesticus (Apple tree) β	A0A498K869	C$_{589}$**VVV**E**I**KEPEPAVPAGNPMDNSGYGY$_{614}$
Eucalyptus grandis β subunit	A0A498K869	A$_{591}$**I**PWTTQ**V**TD**I**EGFEDREQ**I**NLD**LL**SEE**V**E$_{620}$
Theobromo cacao β subunit	UPI000848E310	K$_{583}$EPEPAAAGNPMDNSGYGY$_{608}$
Lactuca sativa (Lettuce) β subunit	A0A2J6LIA5	K$_{585}$EPEP**MV**AGNPMDNSGYGY$_{603}$
Oryza sativa (Rice) β subunit	A0A0E0HK09	K$_{583}$EPEPA**P**V**T**NPMDNSGYGY$_{601}$
Helianthus annuus (Sunflower) β subunit	A0A251SG14	K$_{587}$EPEP**V**T**V**GNPMDNSGYGY$_{605}$
Maize β subunit	A0A3L6DCR4	K$_{561}$EPEAA**P**VANPMDNSGKRHCH$_{581}$
Cicer Arietinum (Chickpea) β subunit	A0A1S3DXB5	K$_{575}$EPEP**VV**AGNPMDNSGYG**M**$_{593}$

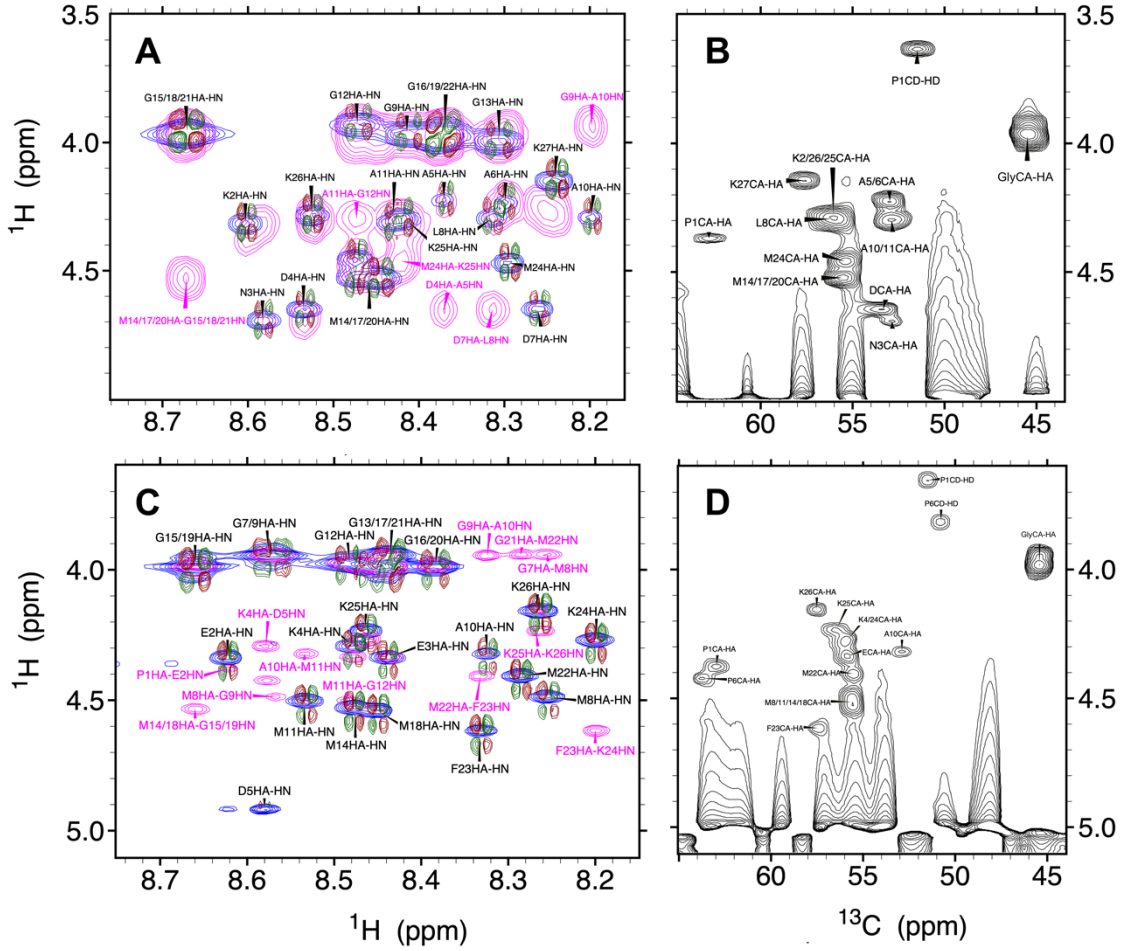
Supporting Table 2: NMR Spectral Parameters

Type of Spectrum	Number of Scans	Sweep Width (ppm)	Matrix Size	Mixing Time (ms)
1D ^{1}H	64	12	-	-
2D ^{1}H, ^{1}H COSY	16	11 x 11	8192 x 336	-
2D ^{1}H, ^{1}H TOCSY	16	11 x 11	2048 x 256	40, 60 or 100*
2D ^{1}H, ^{1}H NOESY	40	11 x 11	2048 x 288	100-120
2D ^{1}H-^{13}C HSQC	144	10.4 x 50 (center 40 ppm)	1024 x 80	-
2D ^{1}H-^{15}N HSQC†	336	10 x 23	2048 x 128	-

* The default TOCSY mixing time was 60 ms, but some additional experiments were run using shorter or longer mixing times to distinguish and assign longer side chains.

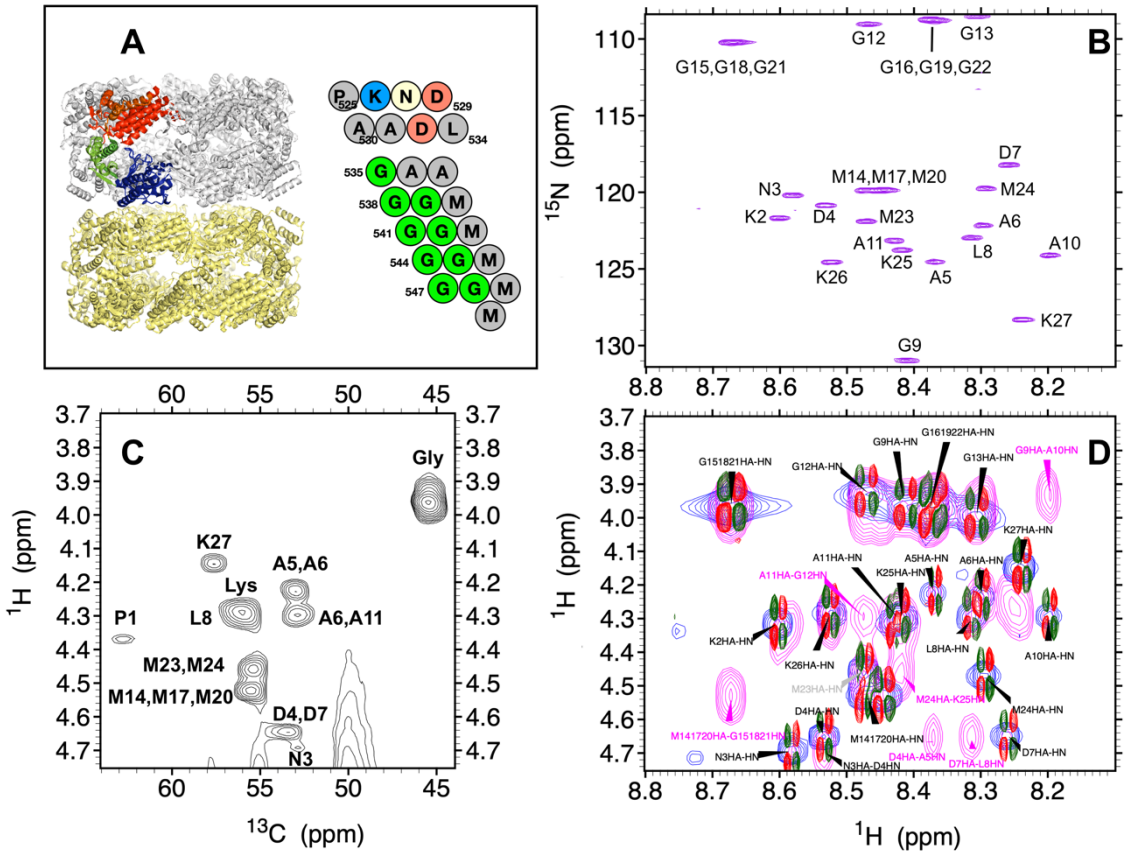
† Recorded only for the longer peptides, namely: GroELCtC and mHsp60CtC.

Supporting Figure 1:
NMR spectra of GroELCtS and mHsp60CtS



A. GroELCtC 2D ^{1}H-^{1}H COSY (maroon/green), TOCSY (blue) and NOESY (magenta) NMR signals of the HN/Hα region are shown. Intra-residual crosspeaks are labeled in black and inter-residual signals are labeled in magenta.
B. 2D ^{1}H-^{13}C HSQC signals from the ^{13}Cα/^{1}Hα region. Overlapped glycine signals are labeled "Gly" and overlapped Asp are labeled with a "D".
C. mHsp60 CtC 2D ^{1}H-^{1}H COSY (maroon/green), TOCSY (blue) and NOESY (magenta) signals of the HN/Hα region are shown. Intra-residual crosspeaks are labeled in black and inter-residual signals are labeled in magenta.
D. mHsp60 CtC 2D ^{1}H-^{13}C HSQC signals from the ^{13}Cα/^{1}Hα region. Overlapped glycine signals are labeled "Gly".

Supporting Figure 2: NMR spectra of GroEL C-terminal Segment



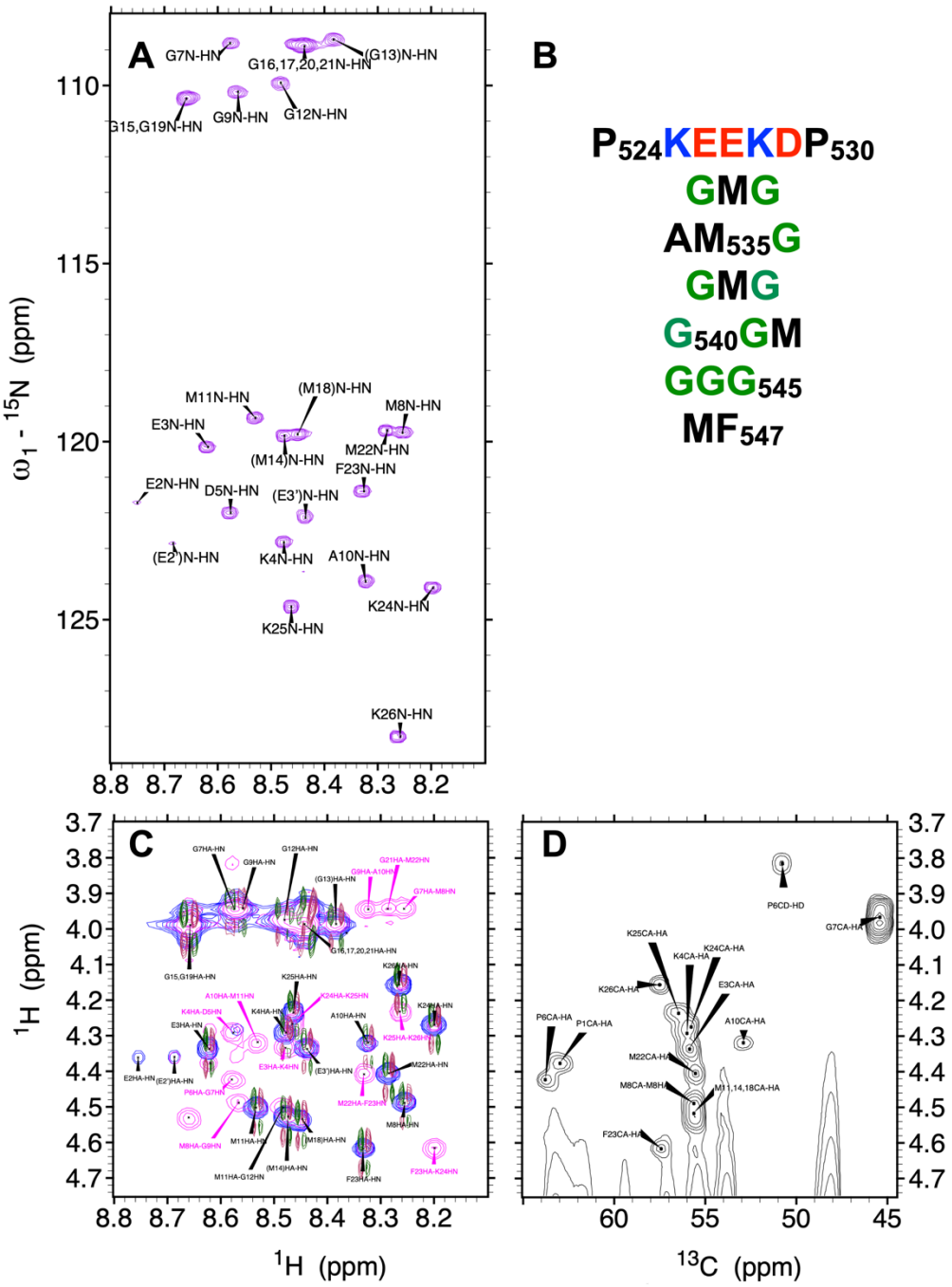
A. (*left*) Structure of GroEL (PDB 2EU1) showing seven subunits of the lower barrel in gold and six subunits of the upper barrel in silver. The apical, intermediate and equatorial domains of one subunit are shown in red, green and blue, respectively. (*right*) The residues of the GroEL C-terminal tail, which are invisible to CryoEM or X-ray diffraction.

B. ^{1}H-^{15}N HSQC spectrum of the C-terminal tail of GroEL. The signal of G9 is folded.

C. ^{1}HN-^{1}Hα region of 2D ^{1}H-^{1}H COSY (maroon/dark green), TOCSY (blue) and NOESY (magenta) spectra. Inter-residue crosspeaks are colored maroon and intra-residues crosspeaks are colored black.

D. ^{1}HN-^{1}Hα region of 2D ^{1}H-^{13}C HSQC spectrum. Peaks with overlapped signals from multiple residues are labeled with the three-letter code (*e.g.* Gly, Lys).

Supporting Figure 3: NMR spectra of mHsp60 C-terminal Segment



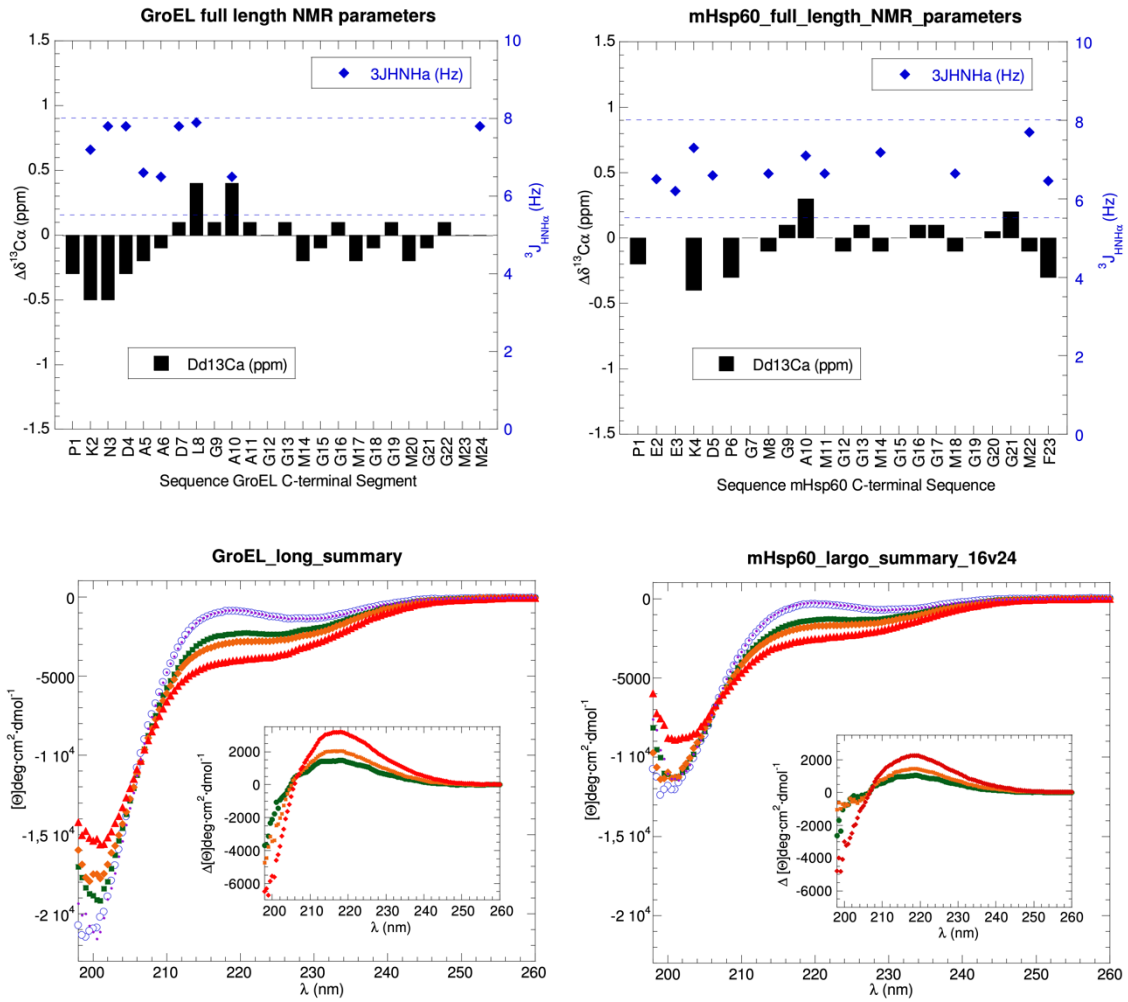
A. ^{1}H-^{15}N HSQC spectrum of the C-terminal tail of mHsp60. Ambiguous assignments are shown in parentheses.

B. The residues of the mHsp60 C-terminal tail, which are invisible to CryoEM or X-ray diffraction. Cationic, anionic, nonpolar and glycine residues are colored blue, red, black and green, respectively.

C. ^{1}HN-^{1}Hα region of 2D ^{1}H-^{1}H COSY (maroon/dark green), TOCSY (blue) and NOESY (magenta) spectra. Inter-residue crosspeaks are colored maroon and intra-residues crosspeaks are colored black.

D. ^{1}HN-^{1}Hα region of 2D ^{1}H-^{13}C HSQC spectrum.

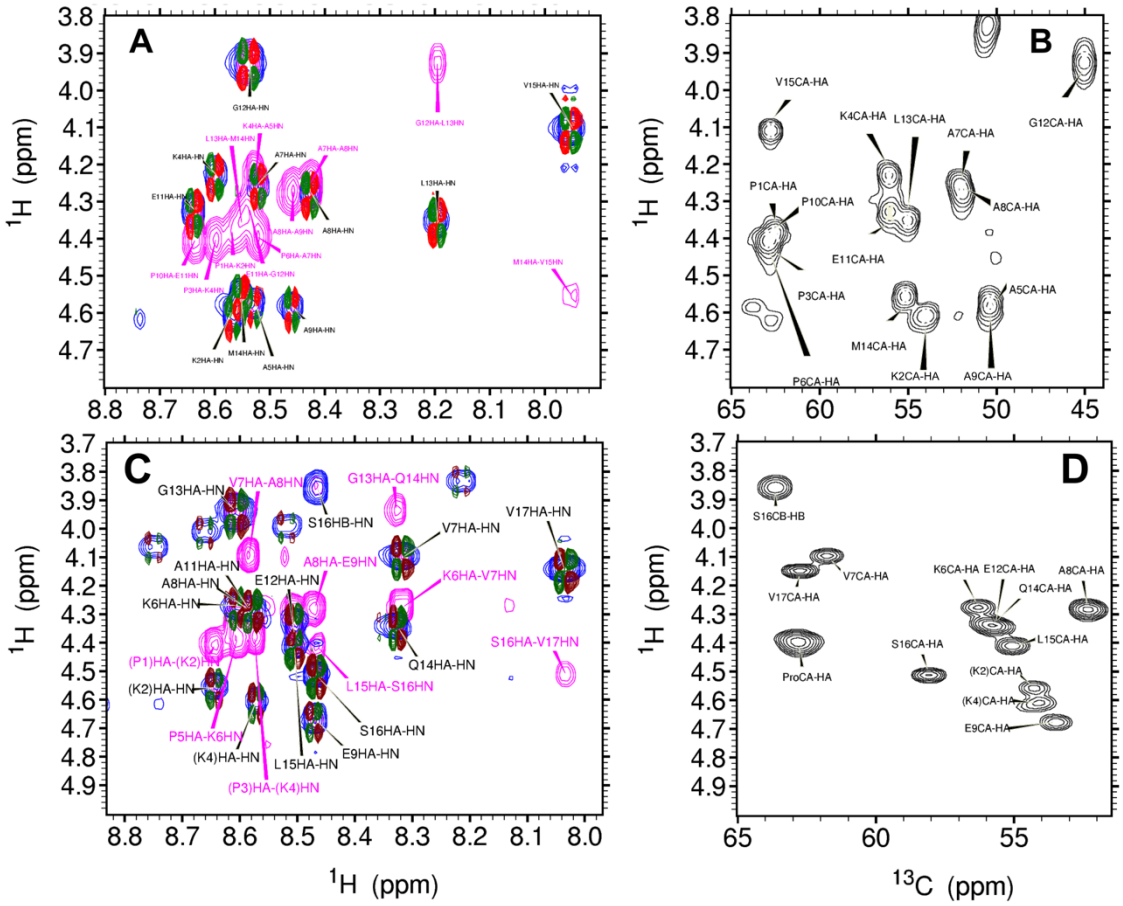
Supporting Figure 4: NMR Parameters and CD Spectra of GroEL and mHsp60 C-terminal Segment (complete)



^{13}Cα conformational chemistry shifts (black bars, left y-axis) and $^{3}J_{HNH\alpha}$ coupling constants (blue diamonds, right y-axis) for the GroEL (top left panel) and mHsp60 (top right panel) C-terminal segments. Blue dashed horizontal lines mark the range of constants (between 5.5 & 8.0 Hz) expected for statistical coil or polyproline II conformations.

Far UV-CD spectra recorded at 5ºC (blue), 25ºC (green) 37ºC (orange) and 65ºC (red) and after recooling to 5ºC (purple) for the C-terminal tails of GroEL (bottom left panel) and mHsp60 (bottom right panel). Insets show the difference CD spectra for 5ºC-25ºC (green), 5ºC-37ºC (orange) and 5ºC-65ºC (red).

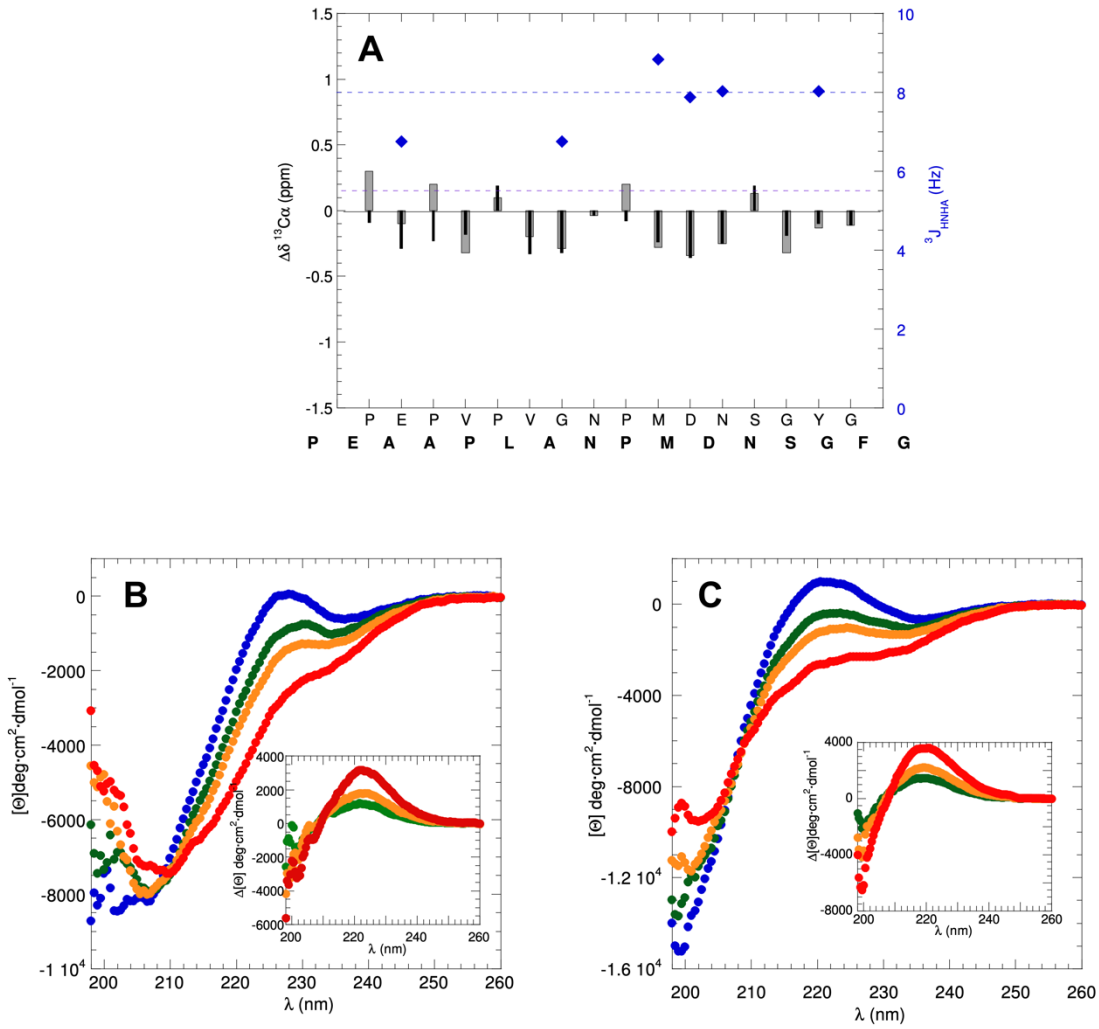
Supporting Figure 5:
NMR Spectra of *Arabadopsis thaliana* and wheat Cpn60α C-terminal segments



A. ^{1}HN-^{1}Hα region of 2D ^{1}H-^{1}H TOCSY (blue), NOESY (magenta), and COSY (maroon/dark green) spectra of *A. thaliana* Cpn60α Ct. Intraresidue crosspeaks are labeled in black and sequential NOESY crosspeaks are labeled in magenta.
B. ^{1}Hα-^{13}Cα region of the 2D ^{1}H-^{13}C HSQC spectrum of *A. thaliana* Cpn60α Ct. The signals arising from the proline residues are overlapped. Contours are plotted with a 1.4 multiplication factor.
C. ^{1}HN-^{1}Hα region of 2D ^{1}H-^{1}H TOCSY (blue), NOESY (magenta), and COSY (maroon/dark green) spectra of Wheat Cpn60α Ct. Intraresidue crosspeaks are labeled in black and sequential NOESY crosspeaks are labeled in magenta. Assignments for K2 and K4 as well as P1 and P3 (in parentheses) are ambiguous.
D. ^{1}Hα-^{13}Cα region of the 2D ^{1}H-^{13}C HSQC spectrum of Wheat Cpn60α Ct. Assignments for K2 and K4 (in parenthesis) are ambiguous and the signals arising from the proline residues (labeled "Pro") are overlapped. Contours are plotted with a 1.4 multiplication factor.

Supporting Figure 6

NMR Parameters and CD Spectra of *A. thaliana* and wheat Cpn60β



A. ^{13}Cα conformational chemistry shifts of *A. thaliana* (black narrow bars, left y-axis) and wheat (gray wide bars, left y-axis) and wheat $^{3}J_{HNHα}$ coupling constants (blue diamonds, right y-axis) Cpn60β C-terminal segments. Blue dashed horizontal lines mark the range of constants (between 5.5 & 8.0 Hz) expected for statistical coil or polyproline II conformations.

Far UV-CD spectra recorded at 5ºC (blue), 25ºC (green) 37ºC (orange) and 65ºC (red) and after recooling to 5ºC (purple) for the C-terminal tails of *A. thaliana* Cpn60β (panel **B**) and wheat Cpn60β (panel **C**). Insets show the difference CD spectra for 5ºC-25ºC (green), 5ºC-37ºC (orange) and 5ºC-65ºC (red).