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*

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 client protein folding, it appears to be invisible or poorly resolved, which suggests that it is disordered. The objective here is to characterize conformational preferences in the C-terminal tails of GroEL, mHsp60 and representative Cpn60s using circular dichroism and nuclear magnetic resonance spectroscopy methods. The tails of GroEL and mHsp60 consist of two segments. The first is rich in residues common in intrinsically disordered proteins, *i.e.* charged, proline, small like Ala, and polar. By contrast, the second segment is consists exclusively (GroEL) or almost entirely (mHsp60) of Gly and Met residues. The spectroscopic results evince that these C-terminal extensions, especially their second segments, are not wholly disordered but adopt high populations of PPII conformations. Regarding the plant cytoplastic chaperonins, whereas the C-terminal segments of Cpn60s are Gly-poor, they are rich in proline and also adopt PPII helix conformations. These results provide insight into the biological activities of the C-terminal tails.

Keywords: Intrinsic disorder, chaperones, nuclear magnetic resonance spectroscopy, circular dichroism dpectroscopy

Conflict of Interests: The authors declare no conflicts of interest.

Introduction

Chaperonins are protein folding machines found in all kingdoms of life [1] [2]. Eubacteria chaperonins, as exemplied by the well characterized protein GroEL from *E. coli* consist of seven subunits, each containing about residues that adopt 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 GroEL is about 60% are present in mitochondria (called Hsp60, sequence identity ca. 60%) with a similar structure [3], and chloroplasts (called Cpn60, sequence identity to GroEL ca. 50%) [4]. This sequence similarity is particularly high in the equitorial domain, which forms most of the interacterions responisble for oligomerization and ATP hydrolysis and is weaker in the apical domain, which concentrates the substrate binding interactions. In addition to the ordered segments, both GroEL, Hsp60 and Cpn60 contain disordered C-terminal segments of about 25 residues in length which appear invisible or diffuse when analyzed by X-ray crystallography and CyroEM and are represented as disordered extensions by AlphaFold (**Figure 1**). Nevertheless, these segments have been reported to contribute to substrate refolding [5] and to be able to form a barrier that blocks substrate from exiting through the bottom of the GroEL barrel[6]. In GroEL and Hsp60, these segments consist of several residues with a residue content typically of intrinsically disordered segments followed by a stretch of about 19 residues which have a remarkably high content of glycine and methothione residues (**Table 1**). Recently a structural class of several proteins containging glycine-rich polyproline II helical bundle domains has emerged [7]. Moreover, the C-terminal Gly/Met-rich stretch of GroEl and mHsp60 is remarkably similar to the sequence of Abductin, a rubber-like protein present at high concentrations in the hinge ligament of bivalves [8], and which is reported to adopt high populations of polyproline II helix [9]. These considerations have led us to hypothesize that the C-terminal extensions of GroEL and mHsp60 might also adopt PPII helical conformations, and testing this hypothesis is a goal of this study.

Chaperonins are also found in Archaebacteria and the cytoplasm of eukaryotics. These proteins share a sequence similarity of about 40% among themselves but have a statistically weak sequence similarity to the chaperonins from eubacteria, mitochondria and chloroplasts [4]. Moreover, they form heterooctamers (each subunit has a slightly different sequence) and have a somewhat larger barrel interior, and seem to lack a disordered C-terminal extension.

The chloroplastic chaperonins, which are key for RuBisCO folding [10], contain two types of subunits, called α and β, have been reported to form heptamers[11]; which then combine to form heterodecatetramers as "football" shaped complexes[12]. It is possible that the α and β subunits of Cpn60 have undergone divergent evolution to

optimize the refolded of distinct client proteins. Like GroEL and mHsp60, these chloroplastic subunits have C-terminal extensions predicted to be disordered by AlphaFold. However, unlike the GroEL and mHsp60 extensions, those of chloroplastic Cpn60 are glycine poor but are often rich in proline residues (**Table 1**). Characterizing these segments is a second goal of this study.

Experimental Procedures

Complete GroEL and mHsp60 C-terminal segments.

Peptides corresponding the complete GroEL and mHsp60 C-terminal segments with sequences: acPKNDAADLGAAGGMGGMGGMGGMMKKK and acPKEEKDPGMGAMGGMGGGMGGGMFKKK, respectively, were obtained from CASLO ApS, Kongens Lyngby, Denmark. These "complete" C-terminal peptides, abbreviated as GroELCtC and mHsp60CtC, respectively. These peptides contain three Lys residues at the C-terminus to promote solubility which are not present in the native sequence. All peptides were N-acetylated as indicated above by the letters "ac" to mimic the lack of charge for this sequence in the context of the full length protein.

Shorter GroEL and mHsp60 C-terminal segments.

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, a C-terminal His residue was added to promote solubility.

A. thaliana and wheat Cpn60 C-terminal segments.

The N-acetylated peptides acPKPKAPAAAPEGLMV and acPKPKPKVAEPAEGQLSV which correspond to the C-terminal segments of *Arabadopsis thaliana* and wheat Cpn60 a subunits, respectively, were purchased from the National Center for Biotechnology (CNB, CSIC) Cantoblanco, Spain.

The sequence and purity of all peptidies was verified by mass spectrometry and NMR spectroscopy. Sample concentrations were estimated by weight and by integration of peptide NMR signals to that of a known concentration of DSS.

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 1**. 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[6] when compared to the $^{13}C\alpha$ predicted for a structureless peptide, which were obtained for each peptide and set of experimental conditions using previously reported parameters (**Figure 1**). $^{3}J_{HNH\alpha}$ couplings also discriminate between α-helical, β-strand, and statistical coil or polyproline II conformations (**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 spectrum of buffer was recorded and substracted.

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 emprical equation of Stellwagen and co-workers : % PPII helix = 100 x ($[\Theta]max + 5560$ $deg \cdot cm^2 \cdot dmol^{-1}$) / 15140 $deg \cdot cm^2 \cdot dmol^{-1}$, where $[\Theta]max$ is the maximum in the spectrum between 210 and 230 nm in $deg \cdot cm^2 \cdot dmol^{-1}$ [13].

Results

Short peptides with just the Gly-rich stretch contain 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 show poor ^{1}HN and $^{13}C\alpha$ chemical shift dispersion and degenerate $^{1}H\alpha$ glycine signals (**Sup. Fig. 1**). Moreover the $^{13}C\alpha$ conformational chemical shifts are very small and indicate an absence of α-helix and β-strand conformations (**Figure 2A,B**). Due to signal overlap, only a few $^{3}J_{HNH\alpha}$ coupling constants could be measured. These are consistent with extended conformational like β-strands, statistical coil or PPII helix.

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 maxiumum near 212 nm which is characteristic of glycine-rich PPII helices [14]. Upon heating, this maximum weakens in intensity. The CD spectral bands are stronger for Hsp60CtS than GroELCtS. This is consistent with a greater number of glycine residues in the former, as the intrinstic PPII forming propensity of Gly is higher than Met [15, 16]. Based on their maximum CD signals of 1600 $deg \cdot cm^2 \cdot dmol^{-1}$ and 1000 $deg \cdot cm^2 \cdot dmol^{-1}$, values for the % populations of PPII helix at 0 ºC would be 47% and 43% for Hsp60CtS and GroELCtS, respectively. 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 sigmodial transition typically observed for folded protein, GroELCtS and mHsp60CtS show a linear loss of signal upon heating (**Fig. 1C,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 ([10] or partly populated PPII helices in intrinsically disordered proteins [11]). Afterwards, the sample was recooled to 0ºC and once again it recovered its original CD signal (data not shown).

Full length peptides contain a substantial portion of polyproline II helix.

To further corroborate and assess the position of the PPII helix, we also characterized peptides corresponding to the complete segments of GroEL and mHsp60 that are invisible to X-ray crystalographic 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 an absence of α-helical and β-strand structure [17],[18]. Whereas two different ^{1}Hα signals were seen for H-bonded glycine residues in sfAFP, which consists of a well folded bundle of six PPII helices [19], no distinct chemical shifts are observed for the two ^{1}Hα of the glycine residues in GroELCtC or mHsp60CtC or the short versions of these peptides (**Sup. Figure 1**). This relevant because the seven C-terminal segments extending from the equatorial domains towards the interior of GroEL or mHsp60 might associate to form a well 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 [20], [21], [22], rule out the presence of α-helices or β-strands and evince a conformational ensemble dominated by statistical coil and/or polyproline II helices (**Sup. Figure 4A&B**).

Circular dichroism (CD) spectroscopy can distinguish between statistical coil and polyproline II conformations [23]. The presence of weak maxima in the CD spectra of GroELCtC and mHsp60CtC (**Sup Fig. 4C&D**) indicate the presence of polyproline II helix. These peaks become weaker upon heating; this is consistent with a PPII helix to coil transition and has been observed previously for other PPII helices [13][24]. The difference spectra, which were calculated by subtracting a spectra recorded at high temperature from one recorded at low temperature, are very similar to the CD spectra of peptides [25] [26], collagen [27] and proteins [14] known to adopt PPII helices (**Sup. Fig 4C&D insets**). Based on the [Θ]max values, the % of polyproline II helix present at 37ºC is estimated to be 31% and 29% for GroELCtC and mHsp60CtC, respectively.

Chloroplast homologs of GroEL, which are Pro rich and Gly-poor, 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**). The CD spectra of peptides corresponding to the C-terminal segments of *Arabadopsis thaliana* and wheat Cpn60 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. Upon heating, the spectra show significant changes including a decreased signal around 220 nm. By subtracting the spectra measured at 25, 37 and 65ºC from the one recorded at 5ºC, difference spectra were calculated. Like the difference spectra of GroELCtC and mHsp60CtC, these difference spectra closely resemble standard PPII helical spectra and strongly suggest that these peptides adopt sizable populations of PPII helical conformations of approximately 37% and 34% at 0ºC and 32% and 28% at 25ºC, for *A. thaliana* and wheat Cpn60 C-terminal segments, respectively, which unravel partially at higher temperatures.

Discussion

The main finding of this study is that peptides corresonding to the C-terminal extensions of eubacterial, mitochondrial and chloroplastic chaperonins, which appear disordered in cryoEM and X-ray crystallographic analyses, adopt a high population of polyproline II helix in aqueous solution. The ability of these glycine rich sequences to adopt large populations of PPII helix is consistent with the high intrinsic ability of glycine residues to adopt isolated polyproline II helices [28] as well as polyproline II helical bundles in synthetic peptides [29], natural protein domains [30][31][32] and quasi infinite honeycomb networks [33].

Despite the sequence similarty to sfAFP and abductin, 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 peptide segments would be present at a much higher concentration of dozens of mM and subject to tethering and crowding effects [34] within chaperonin barrel, the possiblity that they may associate under physiological conditions can not be ruled out by our results. Incidently, the hypothetical association of the polyproline II helices formed by Gly-rich C-terminal segments of GroEL into putative polyproline II helical bundle structures could explain why they are much more resistant to protease K degradation than the C-terminal segments of chloroplastic Cpn60[35]. The latter, being Pro-rich but Gly-poor, can form isolated PPII helices but can not associate to form bundles[7]. Client proteins do not escape out of the bottom of chaperonin barrels despite a lack of ordered structure [6]. The possible associations between PPII helices of the C-terminal extensions, which might also be driven by hydrophobic interactions between the numerous Met residues in eubacterial and mitochondrial chaperonins, could form a barrier at the bottom of the chaperonin barrel and might help explain why client proteins do not escape.

Unlike eubacterial and mitochondrial chaperonins, in their chloroplastic homologs the two barrels are assembled from two different subunits, called α and β. Sequence comparison of Cpn60 α and β subunits from a dozen representative species (**Table 1**) suggests that the α subunits often end with two or three hydophobic residues

right at the C-terminus. By contrast, β-subunits frequently end in a SGYGY motif that is reminisent of the GY segments of TDP-43 and FUS proteins, which is known to promote π/π and cation/π interactions [36],[37]. These differences suggest that the α and β subunits may have undergo divergent evolution to specialize in the refolding of aliphatic (α subunit) or aromatic (β subunit) rich proteins.

Acknowledgements

This study is part of the project PID2022-137806OB-I00, funded by the Spanish Ministry of Science, Innovation and Universities: MICIN/AEI/10.13039/501100011033/FEDER, UE. The funding agency had no role in the study design or the interpretation of the data. 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).

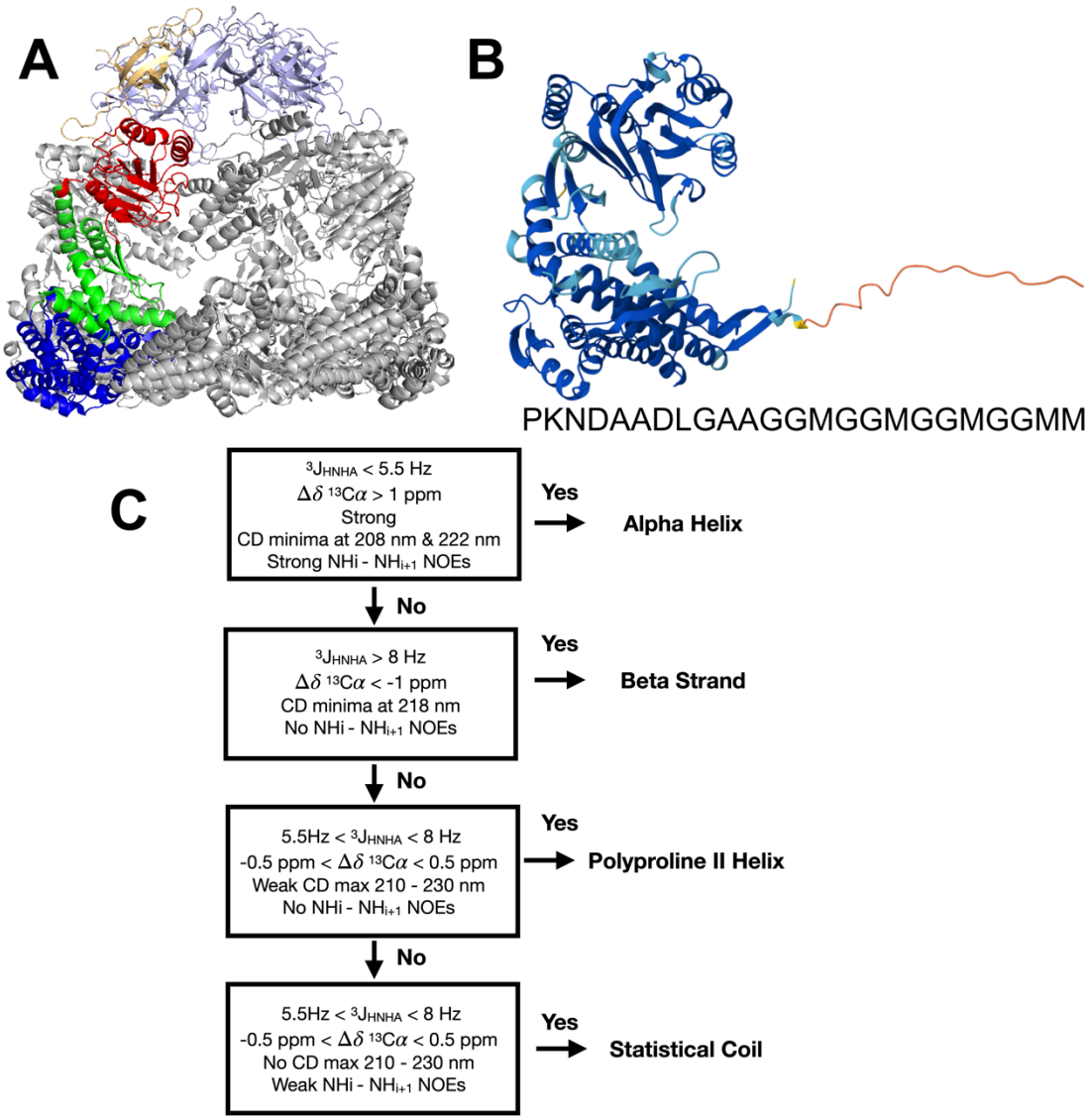
Table 1

Amino-acid residue 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**D**P**TK**V**SCC**I**A**I**L$_{571}$
Malus domesticus (Apple tree) β	A0A498K869	C$_{589}$**VVV**E**I**KE**P**E**P**AV**P**A**G**N**P**MDNS**G**Y**G**Y$_{614}$
Eucalyptus grandis β subunit	A0A498K869	A$_{591}$**I**P**W**TTQ**V**TD**I**E**G**F**F**EDREQ**I**NLD**LL**SEE**V**E$_{620}$
Theobromo cacao β subunit	UPI000848E310	K$_{583}$E**P**E**P**AAA**G**N**P**MDNS**G**Y**G**Y$_{608}$
Lactuca sativa (Lettuce) β subunit	A0A2J6LIA5	K$_{585}$E**P**E**P**M**V**A**G**N**P**MDNS**G**Y**G**Y$_{603}$
Oryza sativa (Rice) β subunit	A0A0E0HK09	K$_{583}$E**P**E**P**A**P**V**T**N**P**MDNS**G**Y**G**Y$_{601}$
Helianthus annuus (Sunflower) β subunit	A0A251SG14	K$_{587}$E**P**E**P**V**T**V**G**N**P**MDNS**G**Y**G**Y$_{605}$
Maize β subunit	A0A3L6DCR4	K$_{561}$E**P**EAA**P**V**A**N**P**MDNS**G**KR**H**C**H**$_{581}$
Cicer Arietinum (Chickpea) β subunit	A0A1S3DXB5	K$_{575}$E**P**E**P**V**V**A**G**N**P**MDNS**G**Y**G**M$_{593}$

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



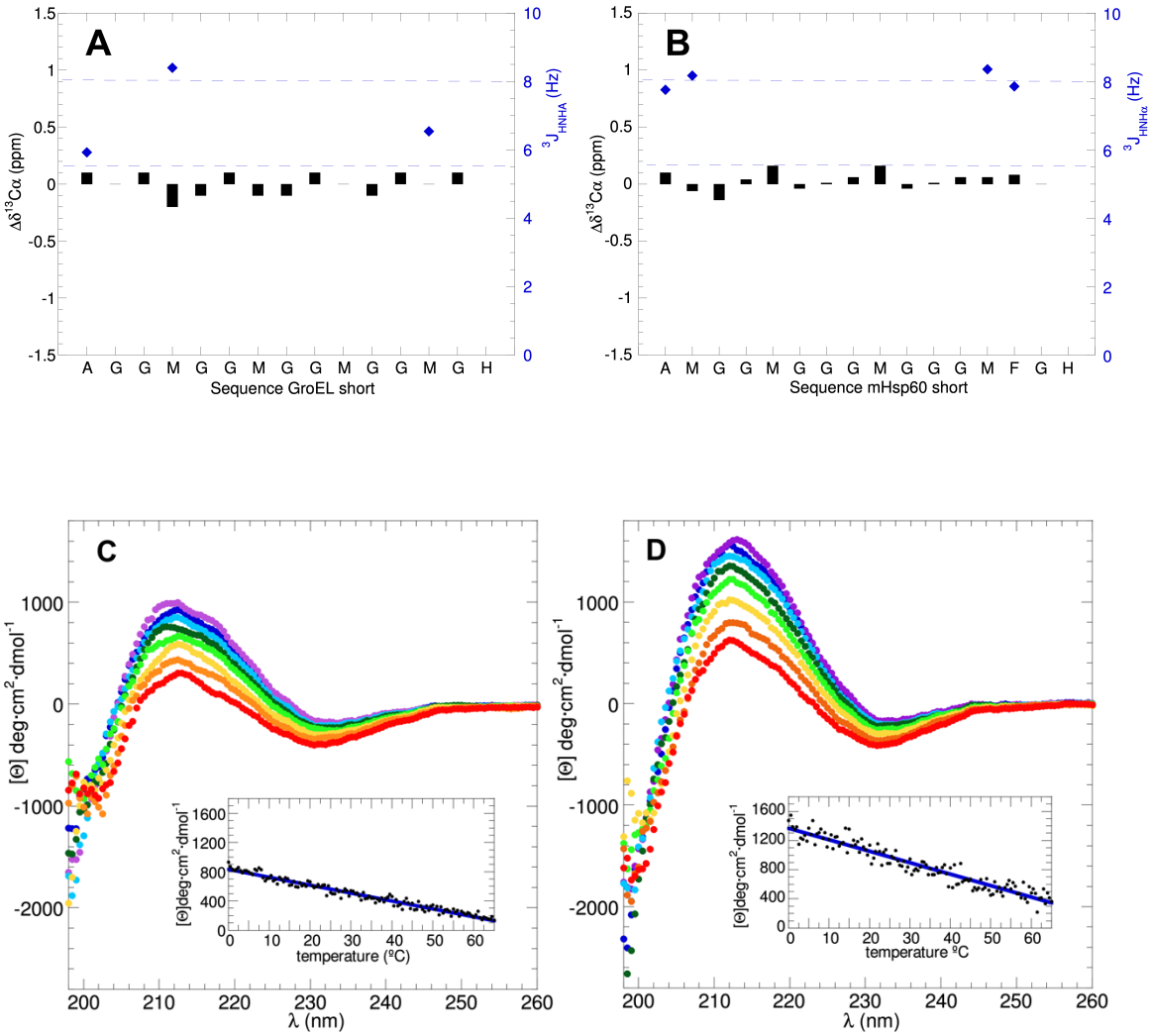
A. Structure of mHsp60/mHsp10 complex (mitochondrial GroEL/ PDB 6MRD). Six mHsp10 subunits are tinted blue/white, the seventh is colored gold/ Six mHsp60 subunits are colored gray, the seventh is colored red, green and blue for its apical, intermediate and equatorial domains, respectively.

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 or very low confidence are colored orange and red; namely the C-terminal segment, whose sequence is written below the figure.

C. Summary of criteria from NMR chemical shift deviations (Δδ), coupling constants, NOE crosspeaks 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.* [38] but substituting the ^{1}Hα Δδ for ^{13}Cα Δδ, as the latter are more reliable [17].

Figure 2

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



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

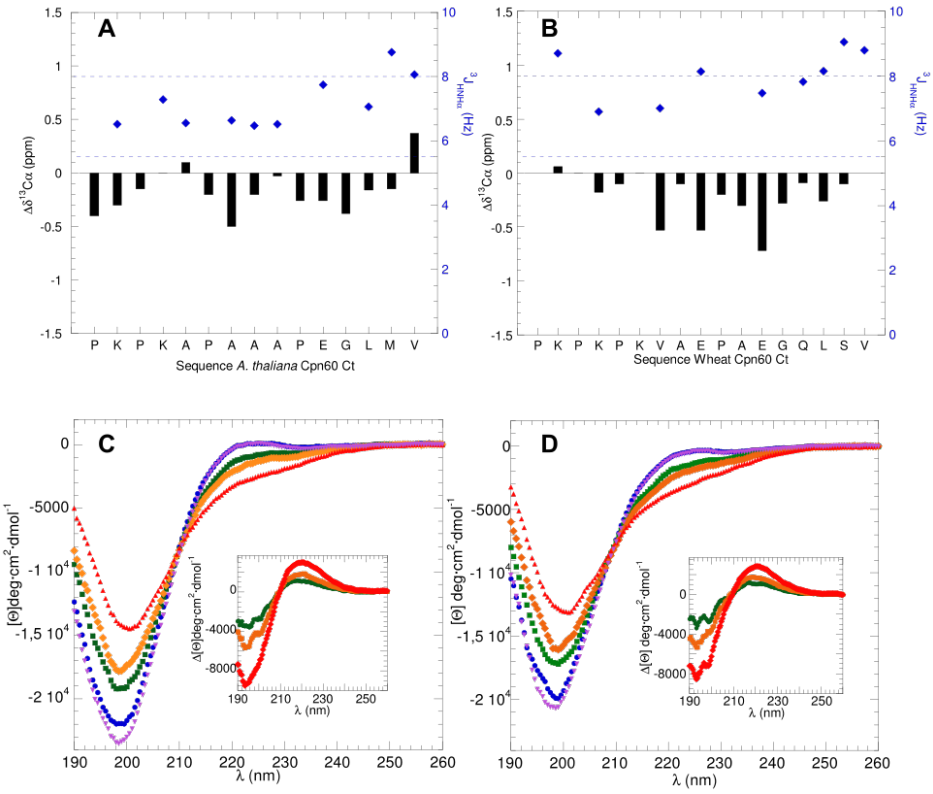
B. mHsp60CtC ^{13}Cα conformation chemical shifts (black bars, left y-axis) and $^{3}J_{HNH\alpha}$ (blue diamonds, right y-axis). The latter are measured for non-overlapping, non-Glycine residues.

C. Far UV-CD spectra of GroELCtC 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 C-terminal segment of GroEL recorded at 217 nm and a heating rate of 1ºC per minute. The fit of a linear equation (blue line) to the data (black points) is $[\Theta]_{217nm} = 835 - 10.8$ (TºC) R= 0.98.

D. Far UV-CD spectra of mHsp60CtC 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 mHsp60 recorded at 217 nm and a heating

rate of 1ºC per minute. The fit of a linear equation (blue line) to the data (black points) is $[\Theta]_{217nm} = 13200 - 15.8\ (T^{o}C)$ R= 0.96.

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 $^{3}J_{HNHα}$ 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**) Cpn60 C-terminal segments. Insets show the difference CD spectra for 5ºC-25ºC (green), 5ºC-37ºC (orange) and 5ºC-65ºC (red).

References

1. Horwich AL, Neupert W, Hartl FU. Protein-catalysed protein folding. Trends Biotechnol. 1990;8:126–31.
2. Horovitz A, Reingewertz TH, Cuéllar J, Valpuesta JM. Chaperonin Mechanisms: Multiple and (Mis)Understood? Annu Rev Biophys. 2022;51:115–33.
3. Gomez-Llorente Y, Jebara F, Patra M, Malik R, Nisemblat S, Chomsky-Hecht O, et al. Structural basis for active single and double ring complexes in human mitochondrial Hsp60-Hsp10 chaperonin. Nat Commun. 2020;11:1916.
4. Kim S, Willison KR, Horwich AL. Cystosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. Trends Biochem Sci. 1994;19:543–8.
5. Weaver J, Rye HS. 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;289:23219–32.
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;290:15042–51.
7. Rodríguez CS, Laurents DV. Architectonic principles of polyproline II helix bundle protein domains. Arch Biochem Biophys. 2024;756:109981.
8. Kelly RE, Rice RV. Abductin: a rubber-like protein from the internal triangular hinge ligament of pecten. Science. 1967;155:208–10.
9. Bochicchio B, Jimenez-Oronoz F, Pepe A, Blanco M, Sandberg LB, Tamburro AM. Synthesis of and structural studies on repeating sequences of abductin. Macromol Biosci. 2005;5:502–11.
10. Ellis RJ. Molecular chaperones: the plant connection. Science. 1990;250:954–9.
11. 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;266:584–7.
12. Zhao Q, Zhang X, Sommer F, Ta N, Wang N, Schroda M, et al. Hetero-oligomeric CPN60 resembles highly symmetric group-I chaperonin structure revealed by Cryo-EM. Plant J. 2019;98:798–812.
13. Park SH, Shalongo W, Stellwagen E. The role of PII conformations in the calculation of peptide fractional helix content. Protein Sci. 1997;6:1694–700.
14. Gates ZP, Baxa MC, Yu W, Riback JA, Li H, Roux B, et al. 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;114:2241–6.
15. Andrews B, Zhang S, Schweitzer-Stenner R, Urbanc B. Glycine in Water Favors the Polyproline II State. Biomolecules. 2020;10.
16. Kelly MA, Chellgren BW, Rucker AL, Troutman JM, Fried MG, Miller AF, et al. Host-guest study of left-handed polyproline II helix formation. Biochemistry. 2001;40:14376–83.
17. Wang Y, Jardetzky O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci. 2002;11:852–61.
18. Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR. 1994;4:171–80.
19. Treviño MÁ, Pantoja-Uceda D, Menéndez M, Gomez MV, Mompeán M, Laurents DV. The Singular NMR Fingerprint of a Polyproline II Helical Bundle. J Am Chem Soc. 2018;140:16988–7000.

20. 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;180:741–51.
21. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci U S A. 2002;99:9190–5.
22. Hagarman A, Measey TJ, 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;132:540–51.
23. TIFFANY ML, KRIMM S. Circular dichroism of a random polypeptide chain. Biopolymers. 1969;4:347–59.
24. Tiffany ML, Krimm S. Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. Biopolymers. 1972;11:2309–16.
25. Tiffany ML, Krimm S. Circular dichroism of poly-L-proline in an unordered conformation. Biopolymers. 1968;6:1767–70.
26. Sreerama N, Woody RW. Poly(pro)II helices in globular proteins: identification and circular dichroic analysis. Biochemistry. 1994;33:10022–5.
27. Drzewiecki KE, Grisham DR, Parmar AS, Nanda V, Shreiber DI. Circular Dichroism Spectroscopy of Collagen Fibrillogenesis: A New Use for an Old Technique. Biophys J. 2016;111:2377–86.
28. Ohnishi S, Kamikubo H, Onitsuka M, Kataoka M, Shortle D. Conformational preference of polyglycine in solution to elongated structure. J Am Chem Soc. 2006;128:16338–44.
29. Lotz B, Keith HD. The crystal structures of poly(LAla-Gly-Gly-Gly)II and Poly(LAla-Gly-Gly)II. J Mol Biol. 1971;61:195–200.
30. Pentelute BL, Gates ZP, Tereshko V, Dashnau JL, Vanderkooi JM, Kossiakoff AA, et al. X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. J Am Chem Soc. 2008;130:9695–701.
31. Dunne M, Denyes JM, Arndt H, Loessner MJ, Leiman PG, Klumpp J. Salmonella Phage S16 Tail Fiber Adhesin Features a Rare Polyglycine Rich Domain for Host Recognition. Structure. 2018;26:1573-1582.e4.
32. De Munck S, Provost M, Kurikawa M, Omori I, Mukohyama J, Felix J, et al. Structural basis of cytokine-mediated activation of ALK family receptors. Nature. 2021;600:143–7.
33. Crick FH, Rich A. Structure of polyglycine II. Nature. 1955;176:780–1.
34. Rivas G, Minton AP. Macromolecular Crowding In Vitro, In Vivo, and In Between. Trends Biochem Sci. 2016;41:970–81.
35. Bai C, Guo P, Zhao Q, Lv Z, Zhang S, Gao F, et al. Protomer Roles in Chloroplast Chaperonin Assembly and Function. Mol Plant. 2015;8:1478–92.
36. Pantoja-Uceda D, Stuani C, Laurents DV, McDermott AE, Buratti E, Mompeán M. Phe-Gly motifs drive fibrillization of TDP-43's prion-like domain condensates. PLoS Biol. 2021;19:e3001198.
37. Wang J, Choi J-M, Holehouse AS, Lee HO, Zhang X, Jahnel M, et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell. 2018;174:688-699.e16.
38. Ma K, Kan L, Wang K. Polyproline II helix is a key structural motif of the elastic PEVK segment of titin. Biochemistry. 2001;40:3427–38.

Supporting Material

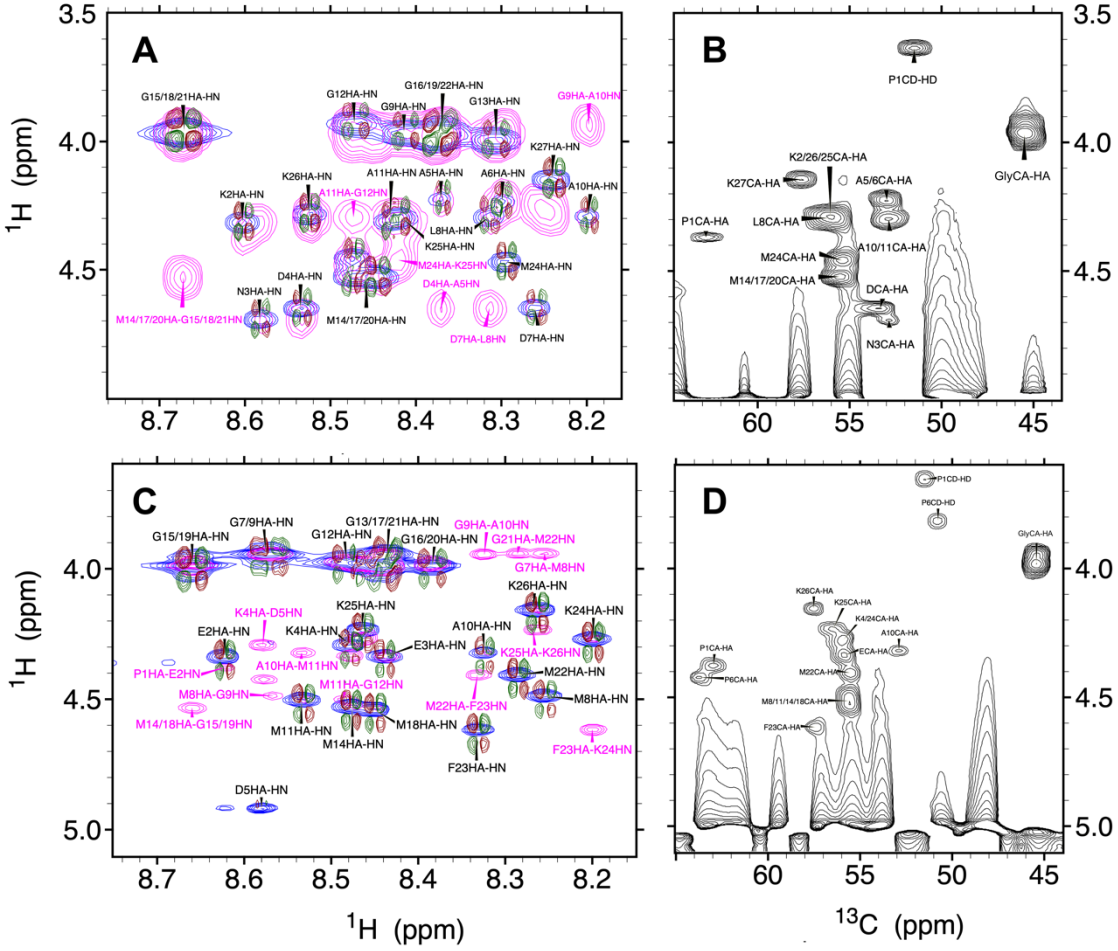
Supporting Table 1: NMR Spectral Parameters

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.

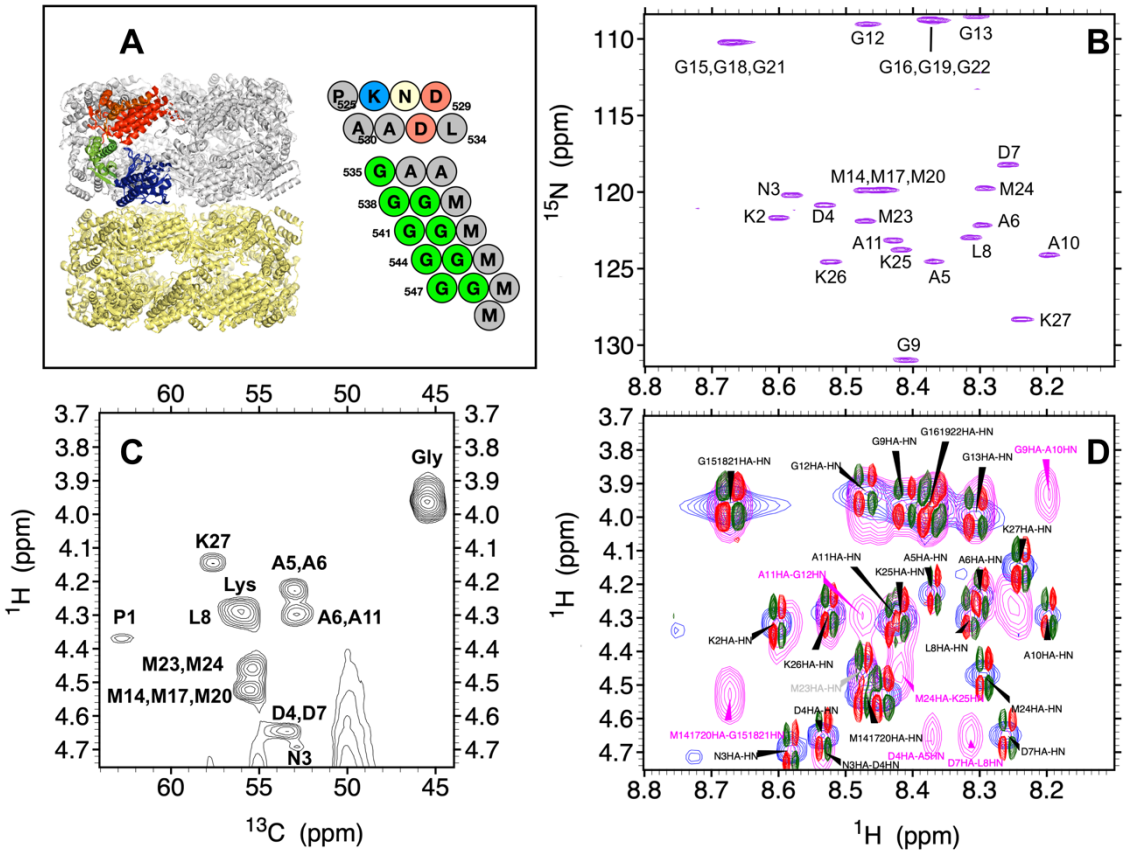
† Recorded only for the longer peptides, namely: GroELCtC and mHsp60CtC.

Supplementary 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



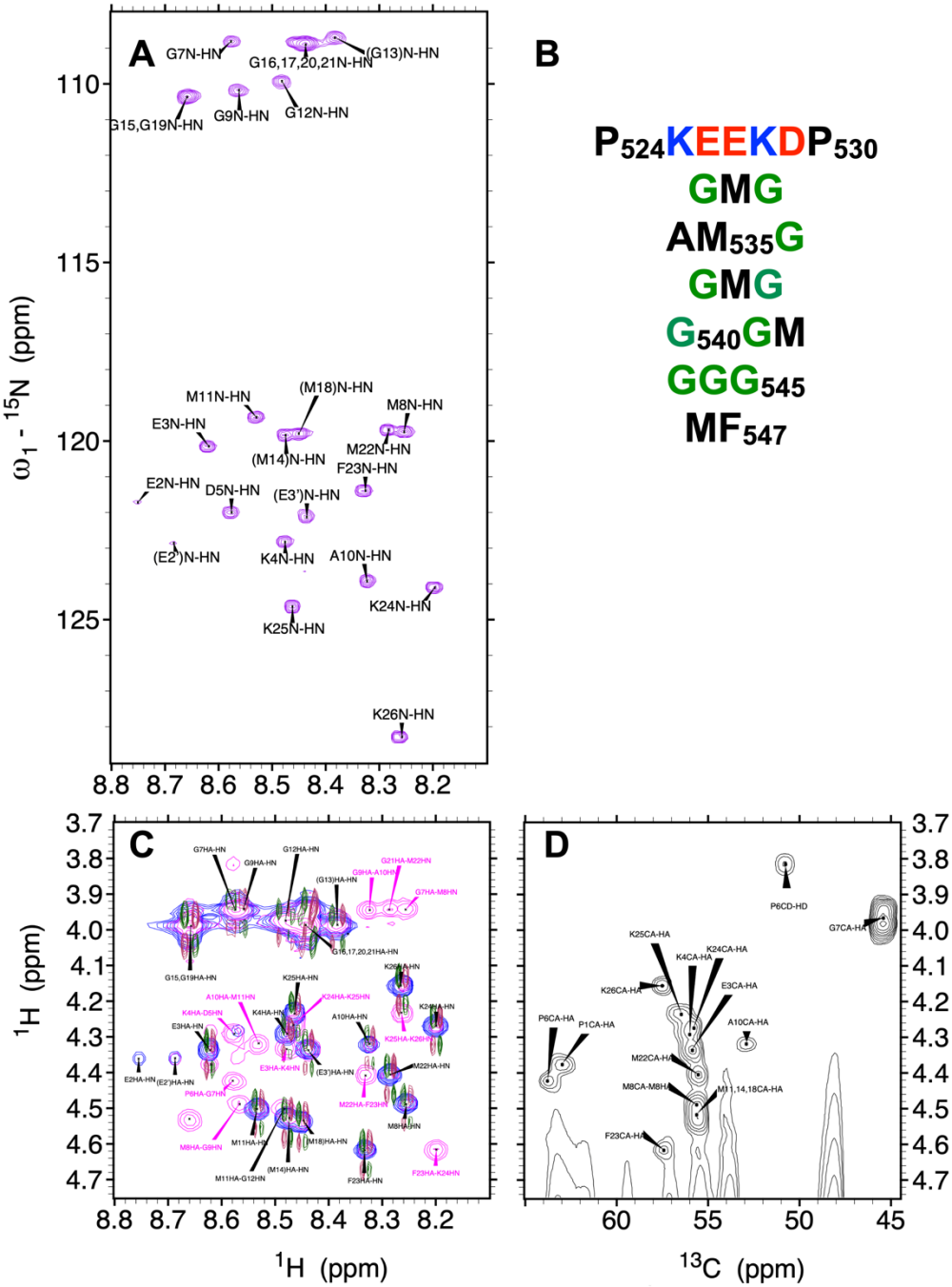
A. (*left*) Structure of GroEL (PDB 2EU1) showing seven subunits of the lower barrel in gold and six subunits of the upper barrel in silver. The apical, intermediate and equatorial domains of one subunit are shown in red, green and blue, respectively. (*right*) The residues of the GroEL C-terminal tail,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

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



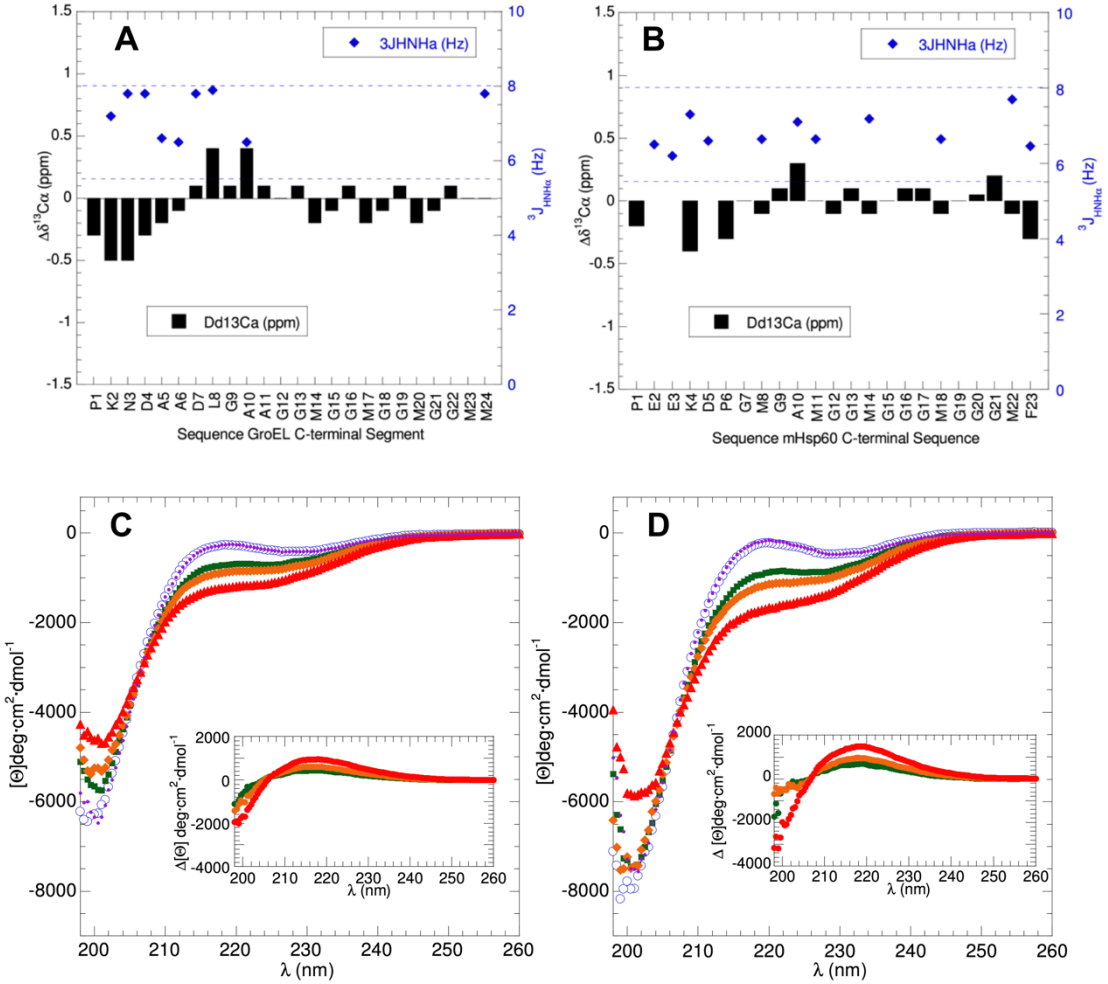
A. ^{1}H-^{15}N HSQC spectrum of the C-terminal tail of mHsp60. Ambiguous assignments are shown in parentheses.

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. Peaks with overlapped signals from multiple residues are labeled with the three-letter code (*e.g.* Gly, Lys).

Supplementary Figure 4: NMR Parameters and CD Spectra of GroEL and mHsp60 C-terminal Segment (full length)

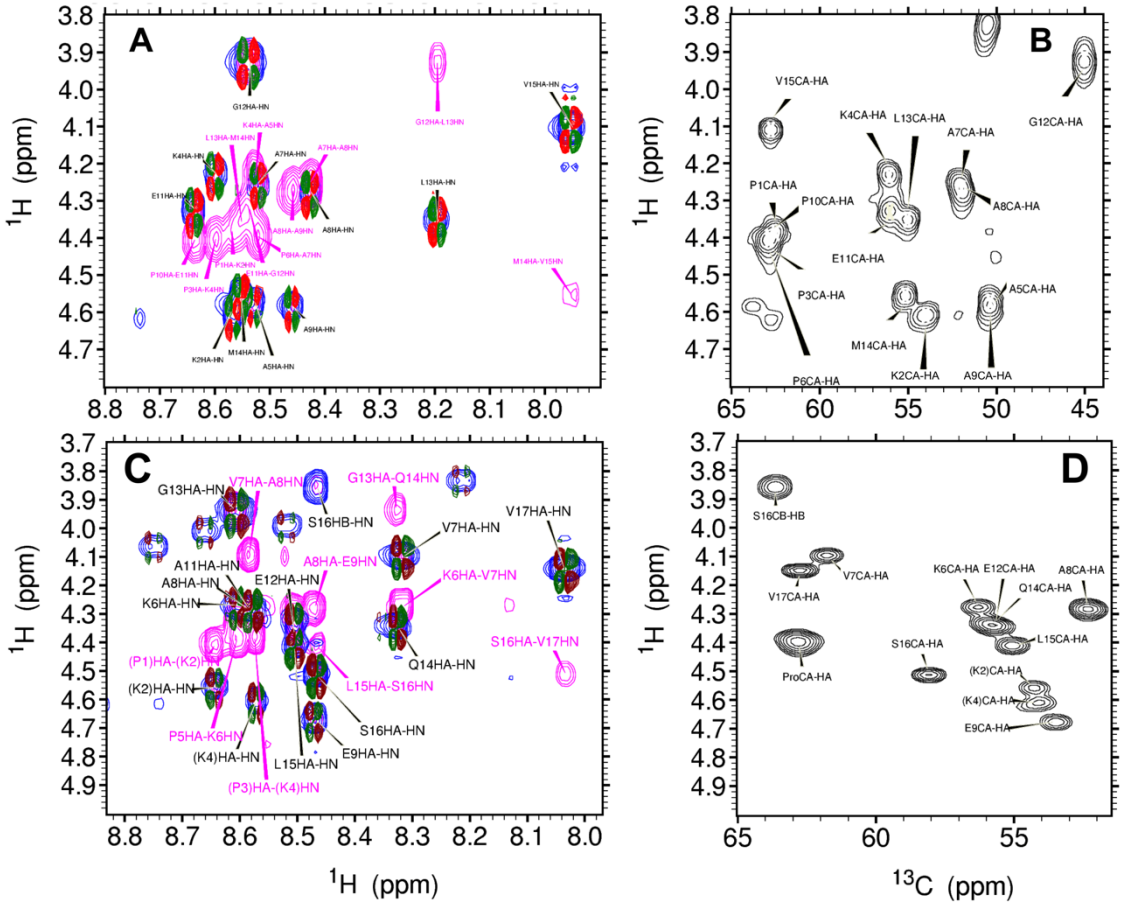


^{13}Cα conformational chemistry shifts (black bars, left y-axis) and $^{3}J_{HNH\alpha}$ coupling constants (blue diamonds, right y-axis) for GroEL (panel **A**) and mHsp60 (panel **B**) complete C-terminal segments. Blue dashed horizontal lines mark the range of constants (between 5.5 & 8.0 Hz) expected 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 GroEL (panel **C**) and mHsp60 (panel **D**) complete C-terminal segments. 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 Cpn60 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* Cpn60 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* Cpn60 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 Cpn60 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 paraentheses) are ambiguous.

D. ^{1}Hα-^{13}Cα region of the 2D ^{1}H-^{13}C HSQC spectrum of Wheat Cpn60 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.