Final Report

Mechanism and function of the chaperonin from Methanococcus maripaludis: implications for archaeal protein homeostasis and energy production

Principal Investigator (PI): Dr. Judith Frydman

Institution: Department of Biology and BioX Program; Stanford University

PI Postal Address:

James H. Clark Center

318 Campus Drive, Room E200A

Stanford University

Stanford, CA 94305-5430

PI Telephone Number: 650-725-7833

PI Email: jfrydman@stanford.edu

Funding Opportunity Announcement Number: *DE-FOA-0001414*

DOE/Office of Science Program Office: Biosciences programs: Physical Biosciences; Office of Basic Energy Sciences, U.S. Department of Energy.

DOE/Office of Science Program Manager Contact: Dr. Robert J. Stack, (301) 903-5652; Robert.Stack@science.doe.gov

Topic Area*: Physical Biosciences

DOE Award Number: DE-SC0008504

Final Report
DOE Award Number: DE-SC0008504

Abstract. Archaea offer a potentially cost effective and renewable source of energy. The methanogen *M. maripaludis,* a fast growing archaea that obtains energy by sequestering H_2 and reducing CO_2 to methane by the methanogenic pathway, is an attractive source for biofuel production. More recently, it has also been suggested that the methanogenesis pathway could be run in reverse, to produce H_2 growing the organism in formate. A multi-level understanding of archaeal protein homeostasis, should be instrumental for improving the functionality and design of the enzyme pathways and complexes involved in energy production and storage. One additional importance consequence of a better understanding of archaeal protein homeostasis will be to increase their stress resistance, since their utilization for the efficient large-scale production of methane (and eventually also of H_2) requires that the organisms are resistance to a range of growth conditions.

This proposal was focused on understanding how archaea achieve protein folding and assembly and maintain protein homeostasis, which are essential for function and viability. We hypothesize that the homo-oligomeric ring shaped chaperonin from *M. maripaludis,* Mm-Cpn, is central to achaeal protein homeostasis and assists folding of a wide spectrum of metabolic, structural and regulatory archaeal proteins. Through a combination of biochemistry, systems biology, computational and structural biology, we have been testing this hypothesis through two complementary efforts: (i) identify the archaeal substrate repertoire of Mm-Cpn, and (ii) define mechanistic and structural principles of Mm-Cpn mediated protein folding. Our achievements in addressing these questions are summarized below:

Description of Results

Specific Aim 1: Identify the archaeal substrate repertoire of Mm-Cpn.

*Proteomic analyses of chaperone interactions. (MS in preparation) T*o obtain a comprehensive description of the Mm-Cpn substrate repertoire, we have carried out immunoprecipitation of Cpn complexes from intact cells using antibodies raised in the lab. Our strategy for the identification of MM-Cpn substrates used a mass spectrometry peptide enrichment identification scheme schematically shown in Fig. 1. Three biological replicates for the Cpn-pull down (PD) and three identical control samples (background; BG) were resolved on an SDS-PAGE gel (Fig 2A). The PD lanes, but not the controls, contained a prominent 60 kDa band corresponding to the Cpn as well as clearly defined associated proteins not present in the BG IPs. We then excised bands from each lane as four fragments (Fig. 2, Frag1-4, note 60 kDa Cpn band in Frag 2) and extracted the peptides using an in-gel proteolysis approach.

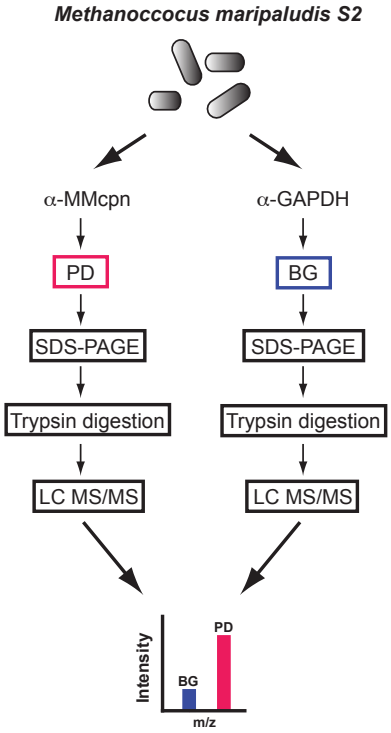

Figure 1. Pulldown-mass spectrometry strategy to identify substrates of the chaperonin from Methonococcus maripaludis, PD: pulldown with anti-Cpn antibodies and BG: background control.

Subsequent LC MS/MS of the samples was carried out on an Orbitrap XL elite mass spectrometer and the spectra were assigned to proteins in the *MM* proteome using the software MSQuant. This experimental setup allows us to compare the median intensity (derived from triplicates) of each protein identified as a ratiometric parameter between the PD and BG values. This analysis has resulted in the identification of 154 proteins which are enriched in PD as defined by a log ratio PD/BG >0.3 (Fig. 2B). Of these, 56 proteins were only observed in the PD samples, and represent high confidence chaperonin-interacting proteins. Our triplicate ratiometric approach to the analysis enhances our confidence in the identification of bona fide substrates. Preliminary analysis of these putative substrates reveals enzymes involved in many important metabolic pathways, including methanogenesis.

Substrate confirmation and validation: In future rounds of substrate identification we will repeat the immunoprecipitations (IPs) comparing different growth conditions and temperatures. We will also examine the effect of ATP and ADP on the Cpn interactions: we predict that substrates may be released in the presence of hydrolysable ATP, while cofactors may be unaffected. As a complementary strategy, we plan to implement a SILAC based approach, by growing MM on labeled (^{15}N) or unlabeled ammonia as a sole nitrogen source and then carrying IPs in parallel for Cpn and BG and analyzing the samples together by MS to detect the ratio between labeled and unlabeled peptides. This type of an approach may further improve the confidence level of our identified substrates. Another isolation strategy is to carry out crosslinking mass spectrometry (XL MS) to identify in more detail the different contact points between the chaperonin and its endogenous substrates on a sequence level. This information will further help discriminate between substrates and cofactors and facilitate the structural interpretation of how substrates interact the structural rearrangements of the complex that drive productive substrate folding.

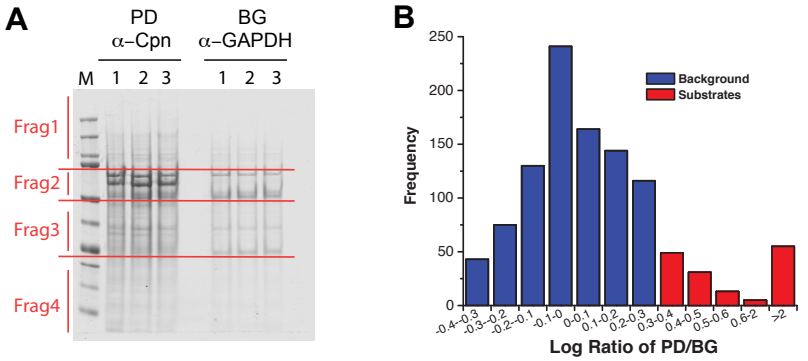

Figure 2. Mass Spectrometry identification of Methanococcus maripaludis chaperonin substrates. (A) SDS-PAGE colloidal blue stained gel of triplicate PD and BG fractions. Each lane was excised as four separate fragments and proteolyzed in-gel for peptide isolation. (B) Bioinformatic analysis of LC MS/MS identified 154 significantly enriched proteins as substrates using ratiometric analysis.

Future Direction: Exploring the MM chaperone network. We have created a series of plasmids in the MM expression plasmid for the *in vivo* expression of tagged and untagged versions of other archaeal chaperones (pfdA and pfdB) and components of the quality control machinery (cdc48). These will allow us to assess the interplay of substrate utilization between chaperones in the archaeal protein homeostasis network. These plasmids have been shipped to our collaborators in the Leigh Lab at the University of Washington for transformation and initial biochemical analysis. Pending this preliminary analysis, large scale pellets will be harvested and shipped to the Frydman Lab where the affinity purification will be carried out. Samples will be processed and analyzed by MS onsite.

Network analyses of chaperone interactions. Our mass spectrometry analysis of Mm-Cpn substrates identified 154 proteins that associate preferentially with the endogenous chaperonin, as expected from bona fide substrates. During our analyses of these substrates we found that Cpn interactome is highly enriched for highly connected proteins, with many interactions (not shown). While a number of substrates are involved in the ribosome, tRNA charging, DNA replication and repair (similar to the eukaryotic chaperonin); approximately a third are enzymes involved in various metabolic pathways. Fig. 3 shows the KEGG pathway analysis for two important pathways mediated in part by chaperonin-substrates, namely, methanogenesis and nucleotide and amino acid metabolism. The reactions catalyzed by a Cpn substrate are highlighted in red (Fig. 3).

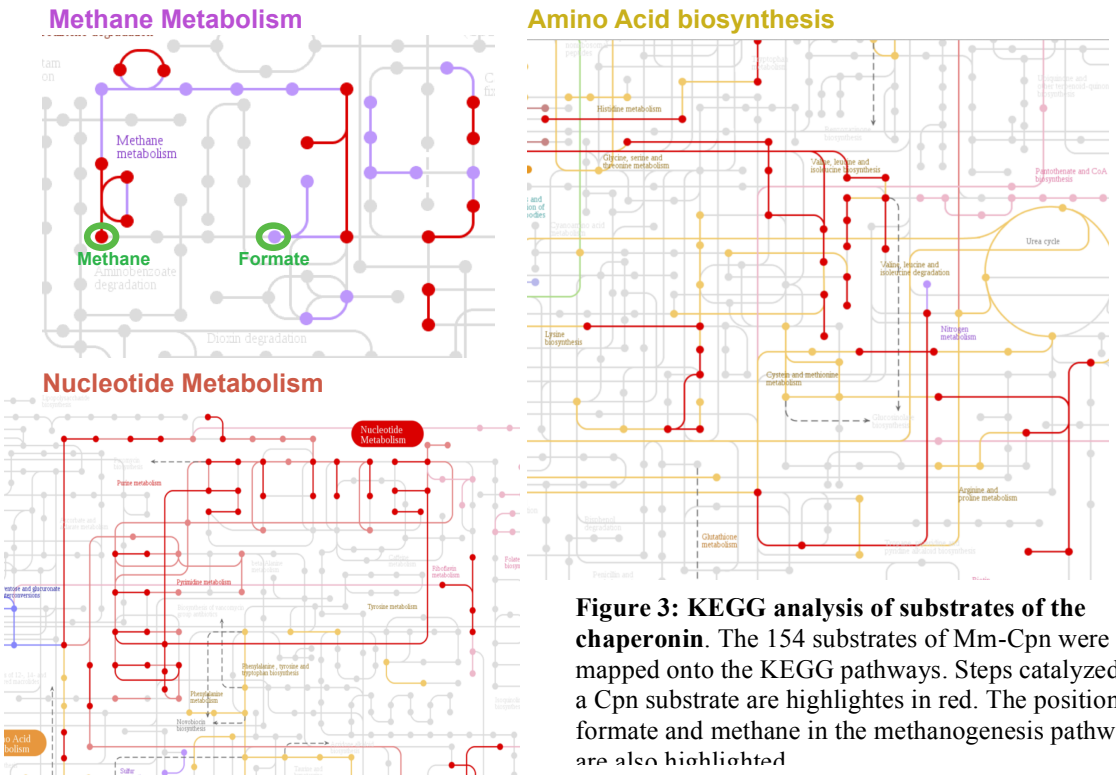

Figure 3: KEGG analysis of substrates of the chaperonin. The 154 substrates of Mm-Cpn were mapped onto the KEGG pathways. Steps catalyzed by a Cpn substrate are highlightes in red. The position of formate and methane in the methanogenesis pathway are also highlighted.

Flux through the chaperonin under different growth conditions: Hydrogenotrophic methanogenic Archaea are defined by an H_2 requirement for growth. Despite this requirement, M. maripaludis can also grow with formate as an electron donor for methanogenesis. Because several enzymes in the methanogenesis pathway were substrates of the chaperonin, we decided to

examine the substrate flux through the chaperonin when the cells are grown in the presence of either electron donor. To this end, we established a system to grow M. maripaludis in the lab under conditions were we can carry out stable isotope labeling by amino acids in cell culture (SILAC) (Fig. 4). Briefly, *M. maripaludis* strain S2 was grown in pressurized flasks in ^{14}N- or ^{15}N- containing media that provided either H_2 or formate as electron donors for methanogenesis (Fig. 4A). Lysates were generated from 100 mL cultures grown to an approximate OD_{600} 0.25. Lysates were precipitated with antibodies against the chaperonin or GAPDH as a control. After washing, the resulting pull downs from opposing antibodies and SILAC labels were pooled. Samples were separated by SDS-PAGE prior to in gel digestion and mass spectrometry (Figure 4A).

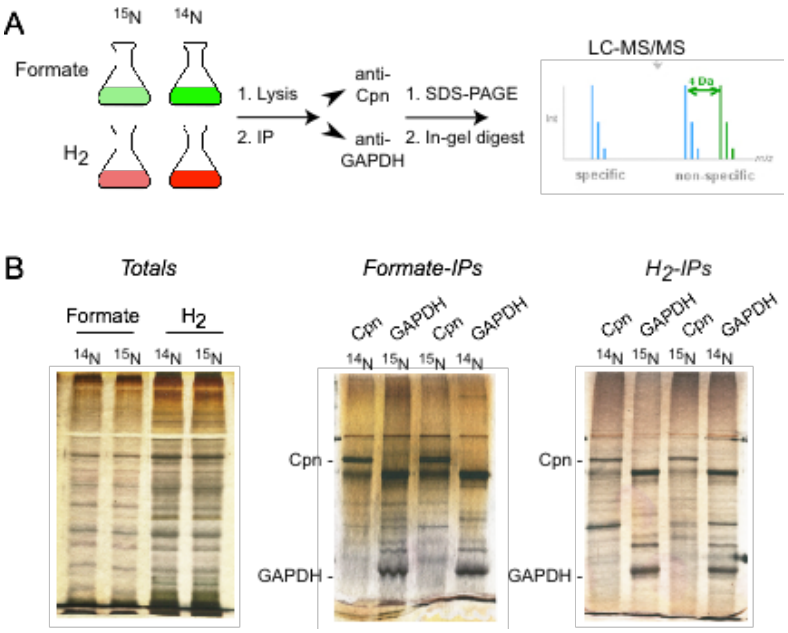

Figure 4. SILAC Approach to examine in vivo interactions with Cpn . (A) Scheme of experimental approach (B) Left: SILAC labeling does not affect protein content if cells grown with H2 or Formate. Middle and Right panels: SDS-PAGE analysis of immunoprecipitations with Cpn or GAPDH antibodies. Reactions were analyzed in parallel by mass spectrometry.

We observed differences in the total protein composition between H_2 or formate, but little change caused by the SILAC label (Figure 2B). Isolation of Chaperonin-containing complexes by immunoprecipitation showed clear differences from the GAPDH control (Figure 4B). The mass spectrometry analysis of chaperonin substrates has already been carried out and we are awaiting analysis of the SILAC data.

Specific Aim 2: Define mechanistic and structural principles of Mm-Cpn mediated folding.

Mm-Cpn uses ATP cycling to drive a conformational cycle that promotes polypeptide folding. We are combining biochemical, biophysical and structural approaches to reveal how Mm-Cpn interacts with its substrates and what are the mechanics of Mm-Cpn folding. A number of exciting findings and approaches have been developed in this first year, including a novel observation hinting that the Chaperonin C-terminal tails play a central role in the chaperonin folding cycle. Our major efforts are briefly summarized below:

Mapping the conformational dynamics of Mm-Cpn along the folding cycle (MS in preparation):
We started using HD-exchange in combination with mass spectrometry to map the conformational dynamics of the chaperonin in the different states of the folding cycle, in collaboration with Matthias Mayer in the University of Heidelberg. Preliminary experiments show that the unliganded state is extremely dynamic, whereas the ATP-bound closed state in

significantly more rigid. This approach will be applied to chaperonin-substrate complexes to gain insights into the nature of folding intermediates.

Structure of the open, substrate binding state (MS in preparation): In collaboration with Paul Adams (LBL-Berkeley) we are examining the structure of the open state, which is the substrate acceptor state. *The preliminary data obtained so far has permitted us to formulate a hypothesis that the C-terminal tails, a hitherto uncharacterized region of the chaperonin, plays an important role in the substrate folding cycle* (as discussed below).

Role of the chaperonin tails in the folding cycle: Chaperonins fold substrates within their central chamber. chaperonins possess flexible C-termini that extend from the base of their equatorial domains into the central cavity (Fig. 5, orange). In previous experiments, carried out in year 1, we obtained biochemical evidence that these tails are intimately involved in the function of the chaperonin. For instance, deletion of the C-terminal 16 amino acids of MM-cpn does not affect its assembly and overall structure, but impairs its ATPase activity and ability to fold substrates. Despite the importance of this domain of the chaperonin, its relation to the structural mechanism remains rather opaque. We also used an *in silico* approach to model the tails in the two mmCpn crystal structures (Fig. 5A). The conformational ensembles of the C-termini were markedly different for these two different points in the nucleotide cycle of mmCpn. Particularly, in the

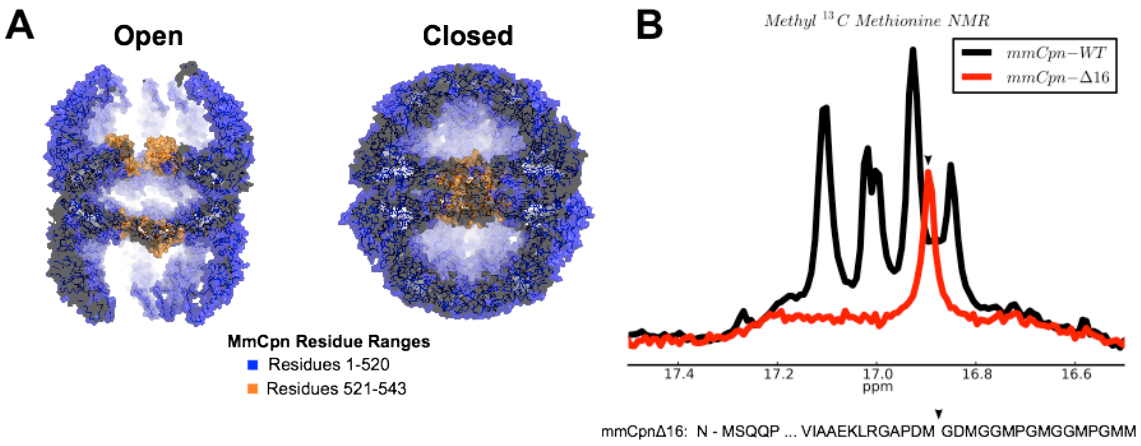

Figure 5. Structural characterization of the C-terminal tails during the chaperonin cycle. (A) Spacefill representation of the MM-Cpn in the open (apo) and closed (ATP) states. Modeling of the C-terminal tails (in orange) predicts different conformations in the open and closed state. (B) 13C-Methyl-methionine NMR allows direct experimental determination of the conformation and interactions of the C-terminal tails during the folding cycle. 1-D carbon NMR spectra acquired on an 800Mhz Varian spectrometer in 1 hour. Black trace: WT CPN; Red trace: Cpn lacking the C-terminal methionine rich tail, as indicated.

closed state conformational ensemble, the C-termini pack into the bottom of the cavity with the tails from the cis and trans rings interacting to form a septum which presumably prevents the migration of a substrate from one ring to another during folding. By contrast, in the open state, the termini from one ring do not appreciably interact with the termini of the other. The open state termini also appear to interact with residues further up the cavity walls providing a tantalizing hint that they may be poised to interact with substrates during this portion of the nucleotide cycle.

In order to assess the validity of these computer simulations, and to examine the function and conformation of the termini during the folding cycle, we have developed an NMR labeling scheme using ^{13}C-methyl methionine which will allow the spectroscopic characterization of the termini. Owing to the rapid conformational dynamics of the chaperonin C-termini, Methionines in this region present as sharp peaks in carbon spectra clearly visible above background signal from the balance of the complex (Fig 5). 1D-NMR spectra exhibit six well dispersed peaks in the methyl region (Fig. 5B). Of these six peaks, five disappear upon the deletion of the C-terminal 16 residues of MM-Cpn (Fig. 5B). This system will also allow us to probe the C-termini spectroscopically during the ATPase cycle and in the presence and absence of substrates. We plan to assign each of the six peaks in this C-terminal region via single point mutations. To study interactions between the chaperonin tails, the cavity and the folding substrates we will introduce nitroxide spin labels at defined locations within the chaperonin or the substrates and measure paramagnetic relaxation enhancements.

We are now in the process of leveraging these resonances to localize the C-terminus. We designed a cysteine-less Mm-Cpn and showed it to be active. By attaching 4-maleimido-TEMPO spin labels to distinct engineered unique cysteines and quantitating the relaxation enhancement felt by the termini, we will triangulate the equilibrium position of the tails in the cavity. Once this has been ascertained, we will utilize this workflow to interrogate the relationship between the C-terminal position and nucleotide cycle. At present, we have demonstrated a measurable relaxation enhancement for spin labels incorporated at position 237 (Fig 6).

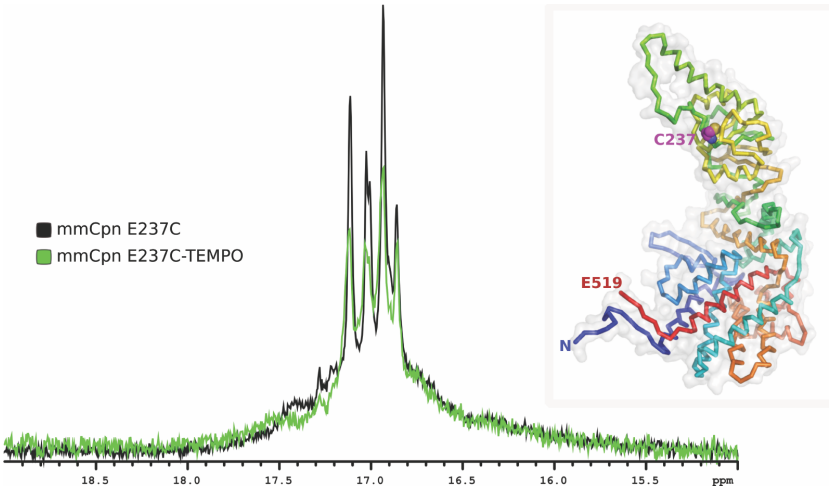

Figure 6. E237-TEMPO enhances the relaxation of C-terminal Methionines. 1-D Carbon NMR spectra of 13C-methyl-methionine labeled cysteine-less mmCpn E237C. Spectra of unmodified (black) show higher, sharper peaks than those of the TEMPO labeled sample (green) indicating that the C-termini are proximal to the spin label.

Elucidating the Functional Role of archaeal chaperonin Residues via Statistical Coupling Analysis (Lopez, Dalton et al, Nature Struct and Mol Biology 2017): Statistical coupling analysis (SCA) can identify clusters of co-evolving residues within a protein family which form spatially contiguous functional networks. Coupling analyses have heretofore largely been restricted to small domains and monomeric proteins. The archaeal group II chaperonins pose a particular problem for established coupling workflows inasmuch as they form a complex oligomer composed of highly conserved multidomain monomers for which few (less than one thousand) sequences are available. To meet these challenges, we have a devised and implemented a novel statistical coupling metric which renders analysis of very conserved, sparsely sequenced protein families. The metric constitutes an agnostic information theoretical measure which

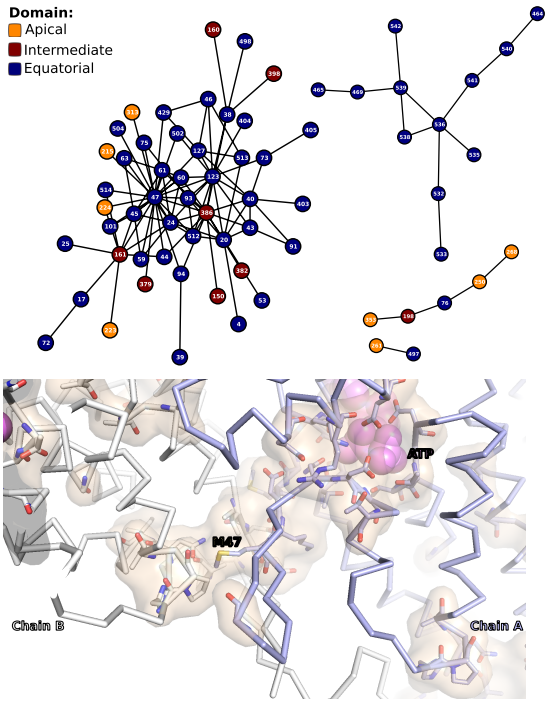

Figure 7. MISC predicts an intersubunit allosteric network in archaeal chaperonins. Top: pairwise residue interaction network. Nodes in the network correspond to mmCpn residues and are colored in correspondence with their domain localization within the mmCpn crystal structure. The graph shows one predominant network. Bottom: The coupling network residues rendered on the crystal structure of mmCpn bound to ADP (PDBID: 3RUW). The

quantifies the degree of coupling between two columns in a multiple sequence alignment. Formally, it consists of the well-known variation of information which has been normalized by the joint entropy.

Application of this new metric to an alignment of 1186 archaeal group II sequences successfully predicted a network of coevolving residues in MmCpn (Fig 7 top panel). The network contains the catalytic aspartate-386 responsible for nucleotide hydrolysis as well as lysine-161, which is implicated in sensing the departure of the γ-phosphate of ATP. This network is spatially compact and spans the lateral interface between MmCpn monomers in the chaperonin complex. Using this algorithm, we identified a cluster of co-evolving residues that make up a portion of the nucleotide binding pocket in the equatorial domain of the chaperonin subunit and extends laterally from the nucleotide to the neighboring subunit within the same ring of the complex. The clusters from adjacent subunits appear to "talk" primarily through a single methionine residue at position 47 (Fig. 7 in yellow in bottom panel). Mutation of this residue to an alanine, however, generated stable soluble monomers of the chaperonin Analyses of this network revealed the allosteric regulation of group II chaperonins.

Publications supported by the grant:

1) Lopez T; Dalton K; Thomlinson A; Pande V and **Frydman J** (2017). An information theoretic framework reveals a tunable allosteric network in group II chaperonins. **Nature Structural & Molecular Biology**, *24(9):726-733. doi: 10.1038/nsmb.3440.*
2) Lopez T, Dalton K, **Frydman J**. (2015) The Mechanism and Function of Group II Chaperonins. **J Mol Biol**.;427(18):2919-30. PMID: 25936650
3) Walzthoeni T, Joachimiak LA, Rosenberger G, Röst HL, Malmström L, Leitner A, **Frydman J,** Aebersold R. (2015) xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. **Nat Methods.**. doi:10.1038/nmeth.3631. PMID: 26501516
4) Dalton KM, **Frydman J**, Pande VS. (2015) The dynamic conformational cycle of the group I chaperonin C-termini revealed via molecular dynamics simulation. **PLoS One**. 30;10(3):e0117724. PMID: 25822285

5) Joachimiak LA, Walzthoeni T, Liu C, Aebersold R, **Frydman J**. (2014) The Structural Basis of Substrate Recognition by the Eukaryotic Chaperonin TRiC/CCT. **Cell**, (5):1042-55, PMID: 25416944
• *Perspective Cell 2014*

6) Leitner A, Joachimiak LA, Unverdorben P, Walzthoeni T, **Frydman J,** Förster F, Aebersold R. (2014) Chemical Cross-linking/Mass Spectrometry Targeting Acidic Residues in Proteins and Protein Complexes. **Proc Natl Acad Sci U S A**, 111(26):9455-60, PMID: 24938783, PMC4084482

7) Pereira JH, Ralston CY, Douglas NR, Kumar R, Lopez T, McAndrew RP, Knee KM, King JA, **Frydman J**, Adams PD. (2012) Mechanism of Nucleotide Sensing in Group II Chaperonins. **The EMBO Journal**, 31(3):731-40, PMID: 22193720, PMC3273386