ABSTRACT

Title of Dissertation: STRUCTURE AND FUNCTION OF THE GROUP III CHAPERONINS, A UNIQUE CLADE OF PROTEIN FOLDING NANOMACHINES

Sara Elizabeth Rowland, Doctor of Philosophy 2016

Dissertation directed by: Professor Frank T. Robb, Ph.D.
Microbiology and Immunology
University of Maryland School of Medicine

The survival and descent of cells is universally dependent on maintaining their proteins in a properly folded condition. It is widely accepted that the information for the folding of the nascent polypeptide chain into a native protein is encrypted in the amino acid sequence, and the Nobel Laureate Christian Anfinsen was the first to demonstrate that a protein could spontaneously refold after complete unfolding. However, it became clear that the observed folding rates for many proteins were much slower than rates estimated *in vivo*. This led to the recognition of required protein-protein interactions that promote proper folding. A unique group of proteins,

the molecular chaperones, are responsible for maintaining protein homeostasis during normal growth as well as stress conditions.

Chaperonins (CPNs) are ubiquitous and essential chaperones. They form ATP-dependent, hollow complexes that encapsulate polypeptides in two back-to-back stacked multisubunit rings, facilitating protein folding through highly cooperative allosteric articulation. CPNs are usually classified into Group I and Group II. Here, I report the characterization of a novel CPN belonging to a third Group, recently discovered in bacteria. Group III CPNs have close phylogenetic association to the Group II CPNs found in Archaea and Eukarya, and may be a relic of the Last Common Ancestor of the CPN family.

The gene encoding the Group III CPN from *Carboxydothermus hydrogenoformans* and *Candidatus Desulforudis audaxviator* was cloned in *E. coli* and overexpressed in order to both characterize the protein and to demonstrate its ability to function as an ATPase chaperone. The opening and closing cycle of the Chy chaperonin was examined via site-directed mutations affecting the ATP binding site at R155. To relate the mutational analysis to the structure of the CPN, the crystal structure of both the AMP-PNP (an ATP analogue) and ADP bound forms were obtained in collaboration with Sun-Shin Cha in Seoul, South Korea. The ADP and ATP binding site substitutions resulted in frozen forms of the structures in open and closed conformations. From this, mutants were designed to validate hypotheses regarding key ATP interacting sites as well as important stabilizing interactions, and to observe the physical properties of the resulting complexes by calorimetry.

STRUCTURE AND FUNCTION OF THE GROUP III CHAPERONINS, A UNIQUE CLADE OF PROTEIN FOLDING NANOMACHINES

by

Sara Elizabeth Rowland

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

2016

Advisory Committee:
Frank T. Robb, Ph.D., Chair

Feng Chen, Ph.D.

Russell T. Hill, Ph.D.

Steven R. Hutcheson, Ph.D.

George H. Lorimer, Ph.D.

Kevin R. Sowers, Ph.D.

Dedication

This work is dedicated to my family, friends, and dogs for their constant, unending love and support throughout this process. Thank you. Love, SaraYYP.

Acknowledgements

I would like to thank all of my amazing mentors that I have had in the Robb laboratory over the past several years including Dr. Stephen Techtmann, Dr. Suchithra Narayan, Dr. Joel Graham, Dr. Pongpan Laksanalmai and Anchalee Jiemjit. Thank you for everything you have taught me and for being so supportive, especially in the form of cakes, and for all of the dives and happy hours.

I would also like to thank all of our collaborators at the Korea Institute of Ocean Science and Technology without which this work would not be possible. Thank you Dr. Yun Jae Kim, Ae Ran Choi, Jae-Eun Lee, and Drs. Sung Gyun Kang and Hyun Sook Lee for being so incredibly generous during my visit and being so patient in teaching me *Thermococcus* genetics, and for Young Jan An and Dr. Sun-Shin Cha, now at Ewha Womans University in Seoul, Korea, for producing absolutely beautiful crystal structures of the Chy CPN.

Thank you to Dr. Travis Gallagher and Dr. Rob Brinson at the National Institute of Standards and Technology and Institute for Bioscience and Biotechnology Research for assistance with gel filtration/multi-angle light scattering and calorimetry. Thanks to Dario Spigolon for his help with ITC data collection.

To my committee, thank you so much for all of your wisdom, input, and time, and for always being willing to meet and discuss research ideas.

Finally, I would like to thank my advisor, Dr. Frank T. Robb, for providing me with this opportunity, and for all of his positivity, encouragement, and patience. Thank you for always supporting me throughout this journey- your guidance has been invaluable.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

aa	amino acid
bp	base pair
ATP	adenosine 5'-triphosphate
ADP	adenosine diphosphate
AMP	adenosine monophosphate
AMP-PNP	adenylyl-imidodiphosphate
BME	beta-mercaptoethanol
Chy	*Carboxydothermus hydrogenoformans*
CO	carbon monoxide
CODH	carbon monoxide dehydrogenase
CPN	chaperonin
Daud	*Desulforudis audaxviator*
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
HSP	Heat Shock Protein
kb	kilobase
KCl	potassium chloride
NaCl	sodium chloride
PAGE	polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PMSF	phenylmethylsulfonyl fluoride
SDS	Sodium Dodecyl Sulfate

Chapter 1: Introduction

1.1: The Discovery of Chaperonins

1.1.1: Early Discoveries in Protein Folding

Apart from the relatively recent acceptance of the existence of intrinsically disordered or unstructured proteins (Wright and Dyson 1999) it is universally understood that proteins in biological systems must adopt a specific conformation in order to function properly. Anfinsen and his colleagues definitively demonstrated this with his seminal ribonuclease renaturation experiments in the 1950s (Sela *et al.* 1957). In his 1973 publication, he states that the unique amino acid sequence of a protein alone does not give rise to a functional protein until the polypeptide chain reaches a specific folding pattern or native structure in a defined environment (Anfinsen 1973). Levinthal concluded that this folding does not occur by a random search through all possible conformations of a polypeptide chain which would take far too long, but instead is achieved through folding intermediates comprising a pathway to the production of mature, active protein (Levinthal 1968; Kim and Baldwin 1982). Throughout his renaturation experiments, Anfinsen also observed a discrepancy in the rates of refolding specifically due to the reoxidation of RNase *in vitro* versus the predicted much faster *in vivo* rates of active RNase production. It was previously noted that variables including temperature and pH of the reaction were critical considerations for proper, efficient refolding, yet the manipulation of these constraints alone did not yield comparable *in vivo* rates of active RNase production. The addition of compounds including cysteine, B-mercaptoethanol, bovine serum albumin, and native RNase failed to increase the *in vitro* reactivation rates. This led him to deduce that an additional component must be present *in vivo* to facilitate these accelerated folding rates. By adding a rat liver homogenate

to the RNase refolding reaction, reactivation was greatly accelerated from a half time of 20 minutes to 4.5 minutes (Goldberger *et al.* 1963; Venetianer and Straub 1963). Anfinsen acknowledged that a 'disulfide interchange enzyme' must be present *in vivo* to assist the RNase in achieving proper, native folds efficiently (Anfinsen 1973). This work is the basis of our contemporary understanding of proteins and acknowledges that certain proteins in the cell have important roles in assisting other proteins to their native states. While these studies demonstrated the importance of maintaining ideal reaction conditions to achieve the highest amount of enzymatic activity, they did not explore what happens or what systems may be active *in vivo* when cellular conditions deviate from these optima.

Later experiments in the 1970s with strains of *Escherichia coli* containing mutations in the *groE* gene revealed that this gene product was important for bacteriophage λ head morphogenesis at elevated temperatures (Sternberg 1973). *groE*- strains were temperature sensitive, and could not produce viable λ heads at 41°C, nor could they form colonies at 43°C (Georgopoulos 1978). Bacteriophages that were genetically modified to contain a functional *groE*+ bacterial gene and then transduced into *groE- E. coli*, however, were able to form colonies at 43°C and recovered the ability to produce functional bacteriophage λ heads (Georgopoulos and Hohn1978). A second gene was soon discovered that was located next to the previously discovered *groE* locus (Tilly *et al.* 1981). Researchers named these two genes *groEL* and *groES*, and demonstrated their mutual importance in phage λ head formation and growth of *E. coli* at elevated temperatures. These experiments like Anfinsen's, implicated unrelated proteins in the proper maturation of other polypeptide chains, but went further in suggesting that GroEL and GroES were crucial for permitting growth of *E. coli* beyond its optimal growth temperature, and that these proteins may play a larger role *in vivo* in participating in the systemic bacterial response to heat stress. These researchers unknowingly at the time were the first to characterize a chaperonin, a class of molecular chaperone with homologs subsequently identified in nearly all genomes across all domains of life.

1.1.2: Molecular Chaperones

The phrase 'molecular chaperone' was first used to describe a nucleosome assembly protein that facilitated the correct interaction of histones with DNA in *Xenopus laevis* eggs (Laskey *et al.* 1978). Molecular chaperones are primarily responsible for maintaining protein homeostasis, or proteostasis within cells. They operate through a variety of mechanisms to prevent aggregation of unfolded proteins and promote proper protein folding and assembly, and are an integral part of the protein quality control network, active during normal growth as well as stress conditions.

Chaperonins are indispensable components of this proteostasis network (Fayet *et al.* 1989). These molecular chaperones form 1MDa ATP-dependent complexes that directly interact with denatured proteins, and cyclically encapsulate these polypeptides in a chamber to promote refolding and prevent folding intermediates from going off the proper fold pathway. Following the discovery of *groEL* and *groES* involvement in bacterial viability and bacteriophage λ head and T5 tail morphogenesis in the 1970s, a homolog was discovered in higher plant chloroplasts that directly interacted with newly synthesized large subunit of ribulose-1,5-bisphosphate carboxylase-oxygenase, or RuBisCo (Zweig and Cummings 1973; Sternberg 1973; Georgopoulos and Hohn 1978; Barraclough and Ellis 1980; Hemmingsen *et al.*1988). Hemmingsen *et al.* first recognized the ubiquity of this protein complex across domains, and identified its role more broadly as that of a molecular chaperone based on its post-translational assembly of a diverse set of oligomeric protein complexes, thus naming it a chaperonin (Hemmingsen *et al.* 1988).

1.1.3: Structure and Function of GroEL and GroES

Earlier functional and structural work indicated that these *E. coli* groEL chaperonins exhibited ATPase activity and were composed of a single subunit with a molecular weight of approximately 65,000, exhibiting 7-fold cylindrical symmetry with stacked rings totaling 14 subunits 125Å in diameter and 100 Å in height (Georgopoulos and Hohn 1978; Hendrix 1978; Hohn *et al.* 1979; Hendrix 1979). Later studies demonstrated the direct interaction of GroEL with GroES, a ring-like complex with a molecular weight of 80,000 (Chandrasekhar *et al.* 1986; Figure 1.2 (top)). Observations including the co-purification of GroEL with several proteins including RNA polymerase and RuBisCo (Ishihama *et al.* 1976; Barraclough and Ellis 1980), and the requirement of a functional GroEL for *E. coli* growth at elevated temperature (Sternberg 1973; Georgopoulos and Hohn 1978) started to further clarify the biological role of these chaperonins. Herendeen *et al.* noted a marked increase of GroEL abundance in *E. coli* when shifted from 37°C to 46°C growth temperature, going from 1.6% to 15% of total cell protein mass. GroEL homologues in other systems including human cell lines and protozoa exhibited similar heat shock-inducible reactions prompting the classification of this chaperonin as a component of the heat shock response (Herendeen *et al.* 1979; Waldinger *et al.* 1989; McMullin and Hallberg 1987; Neidhardt *et al.* 1988). Another study expanded on GroEL's protein interactions, demonstrating its affinity specifically for unfolded or misfolded protein forms (Bochkareva *et al.* 1988), which supported previous experiments that the presence of intracellular unfolded proteins directly stimulated the induction of heat shock proteins (Parsell and Sauer 1989). These researchers postulated, given the totality of previously collected data, that during normal growth conditions and during stress conditions, the chaperonin's role *in vivo* is to assist nonnative proteins including newly synthesized and translocated polypeptides as well as stress denatured ones in reaching their native states by a transient ATP-driven interaction (Bochkareva *et al.* 1988).

1.1.4: Chaperonin Cycling

Despite being one of the most well studied protein complexes, active controversy surrounds the cycling of chaperonins. Several models have been proposed regarding the interaction of the substrate protein and the chaperonin complex and how it promotes folding. The passive caging model proposes that the chaperonin complex merely provides an environment in which the unfolded substrate is sequestered from the crowded cytosol and given an opportunity to spontaneously refold without any work being done on the substrate by the complex. The inner cavity lining transitions from a hydrophobic surface to a hydrophilic one promoting the exposed hydrophobic regions of the substrate to fold (Fenton and Horwich 2003; Apteri and Horwich 2008; Horwich *et al.* 2009). In the active caging model, it is believed that the chaperonin complex is capable of destabilizing internalized folding intermediates and can thus accelerate substrate folding (Sparrer *et al.* 1997; Gupta *et al.* 2014). Recent work supports yet a third model called the iterative annealing mechanism, whereby the chaperonin complex interacts with the substrate protein for multiple rounds of encapsulation and helps to forcibly unfold and perform work on the substrate to promote refolding attempts. The polypeptide has a half time of 1 second inside the cavity, and can be partially protruding from the chaperonin and may achieve native structure either inside or outside the complex (Shtilerman *et al.* 1999; Motojima and Yoshida 2010; Corsepius and Lorimer 2013; Yang and Lorimer 2013).

In addition to the various models of chaperonin-mediated substrate protein refolding, there are also conflicting reports describing how the rings interact, the timing of their nucleotide binding and hydrolysis, and substrate binding and encapsulation. Currently, chaperonins across all domains are grouped by their structure and phylogeny (Kim *et al.* 1994; Woese *et al.* 1990). Group I chaperonins encompass those present in bacteria and

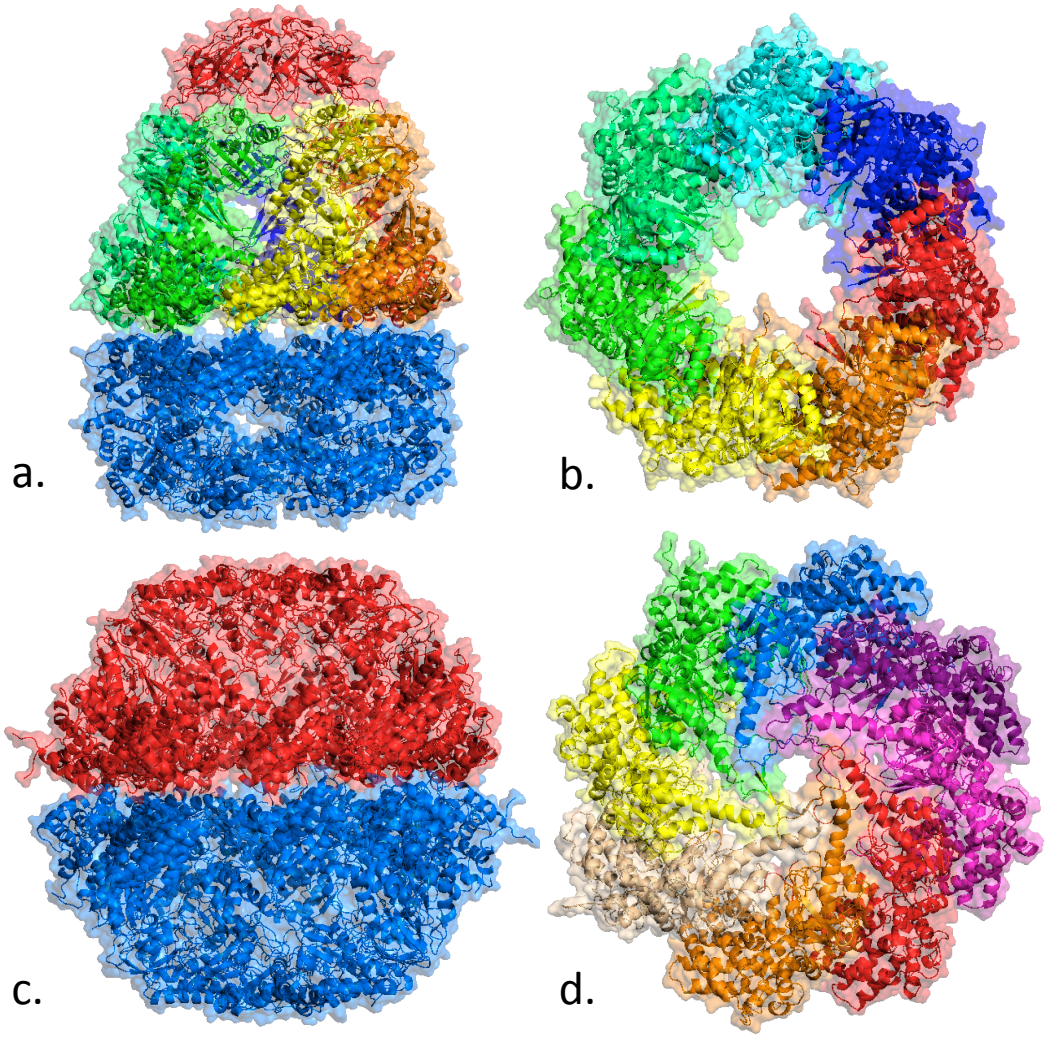

Figure 1.1: Comparison of Group I and Group II chaperonin architecture and their variable cavity closing mechanisms. Top: Crystal structure of *Thermus thermophilus* Group I chaperonin complexed with GroES and 7 ADP molecules. A: GroEL as seen from the side with one ring in blue, and the other ring with each individual shown in a different color interacting with a red GroES heptamer. B: A single monomeric GroEL ring as seen from the apical domain down showing the central cavity of the CPN when complexed with GroES (hidden), (PDB 4V4O; Shimamura *et al*. 2004). Bottom: Crystal structure of *Saccharomyces cerevisiae* Group II chaperonin (CCT). C: Stacked rings shown in red and blue. D: Single ring shown with apical domain on top and with each unique subunit displayed in a unique color. The apical protrusion or built-in-lid is clearly seen in this view (PDB 4V81; Dekker *et al*. 2011). Figures generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC.)

organelles of eukaryotes, and are composed of homomeric 7-membered rings that interact with the co-chaperone known as GroES or Hsp10 for substrate displacement into and closure of the cavitites. Group II chaperoins are present in archaea and the cytosol of eukaryotes, form homo- or heteromeric 8- or 9-membered rings, and function independent of a co-chaperone, instead closing off the central cavities with use of a helical protrusion in the apical domains of subunits known as a built-in lid (Figure 1.1). Both groups form double-stacked ring complexes, with each ring forming a cavity for substrate protein encapsulation and refolding. With Group I chaperonins, a *cis* binding model states that the substrate protein binds at the apical domain of one ring, ATP binds at the same ring, then a conformational change causes the substrate to be released into the chamber upon simultaneous binding of GroES co-chaperone cap. Binding of ATP to the *trans* ring causes displacement of the GroES from the *cis* ring and the cycle repeats (Xu 1997; Horwich 2009). Another model proposes that ATP and GroES can bind both rings simultaneously, forming a symmetric complex that is favored in the presence of substrate and allows for many more iterations of work on a substrate than estimated with the *cis* model (Sparrer *et al.* 1997; Yang and Lorimer 2013).

For Group II chaperonins present in archaea and the cytosol of eukaryotes, ATP hydrolysis is required for complete closure of the built-in lid and release of substrate into the central cavity (Douglas *et al.* 2011). Depending on the size and structure of the substrate, the chaperonin may be able to encapsulate the entire protein, or in the case of larger, multidomain substrates it may partially enclose a segment of the protein leaving the remainder exposed to the cytosol, explaining Group II chaperonins' abilitiy to interact with clients that are larger than the predicted cavity space (Ruessmann *et al.* 2012). All domains of the chaperonin subunit undergo conformational changes throughout the opening and closing cycle (Pereira *et al.* 2010). As with Group I chaperonins, the complete cycle is still unclear. It is suggested that ATP concentrations alter the allosteric function of the rings where low to intermediate concentrations require ADP dissociation in the trans ring for ATP hydrolysis to occur in the

cis ring, while high concentrations of ATP permit simultaneous ring closure (Pereira *et al.* 2010). The built-in lid also appears to play a critical role in the concerted movement of intra-ring subunits as well as in inter-ring communication, and is essential for substrate refolding but not ATP hydrolysis (Reissmann *et al.* 2007).

1.2: The Evolutionary History of Chaperonins

Following the discovery that *E. coli* GroEL and GroES and the chloroplast RuBisCo-binding protein are members of the same chaperonin class of molecular chaperones, more homologues were identified in organisms ranging from pathogenic bacteria, fungi, mammalian mitochondria, and the cytoplasm of mammalian cells (Hutchinson *et al.* 1989; Johnson *et al.* 1989; Picketts *et al.* 1989; Waldinger *et al.* 1989; Gupta 1990). Chaperonins are thought to have originally evolved from a rearrangment of a thioredoxin super-family member protein, specifically one with a peroxidoxin fold (Dekker *et al.* 2011). Peroxiredoxins can form ring-like decamers, and are reported to switch from hyperperoxide detoxification to holdase molecular chaperoning activity via phosphorylation and during stress events such as low pH and increase in temperature (Jang *et al.* 2006; Saccoccia *et al.* 2012; Angelucci *et al.* 2013; Teixeira *et al.* 2015). With the acceptance of the idea of protein moonlighting, it is possible that chaperonins as we classify them today were not even present in the last universal common ancestor, and were instead present in progenitor form such as a peroxiredoxin.

Due to the pervasiveness of this gene across domains, larger evolutionary theories can be supported based on related chaperonin groups in much the same way ribosomal gene sequences had been used to restructure the tree of life (Woese and Fox 1977; Woese *et al.* 1990). For instance, early observations that mitochondrial and chloroplast chaperonin

sequences were most similar to those in purple bacteria and cyanobacteria respectively further supported the endosymbiont origin theory of these organelles in eukaryotic organisms (Gupta *et al.* 1989).

As more homologs are continually being discovered, it has became evident that there are distinct groups within the chaperonin class that adhere to the structuring of the newly described, molecular-based domains of life (Woese *et al.* 1990). Previously characterized polypeptides were found to be homologous to the bacterial-type or groEL chaperonins including the cytosolic eukaryotic TCP-1 (tailless complex polypeptide 1) or CCT (chaperonin containing TCP-1), but these sequences are less similar to GroEL than the chaperonins of eukaryotic organelles. These cytosolic chaperonins are composed of eight different subunits forming heterooligomeric complexes, and are only about 20% identical and 40-60% similar to the chaperonins of bacteria and eukaryotic organelles (Gupta 1990; Rommelaere *et al.* 1993). In archaea, thermosomes or heat shock inducible homo- and heteromeric 8-9 membered rings that form double stacks with ATPase activity and 40% identity to TCP-1 were discovered (Trent *et al.* 1991; Phipps *et al.* 1991; Phipps *et al.* 1993; Marco *et al.* 1994; Guagliardi *et al.* 1994). In the cytosol and in archaea, GroES or hsp10 homologues are notably absent (Gupta 1995).

1.2.1: Group I and Group II Chaperonins

Since the early 1990s, chaperonins have been categorized into 2 distinct groups based on their structure and phylogeny (Kim *et al.* 1994; Woese *et al.* 1990). Structural and mutational studies defined domains and their functions. The crystal structure of *E. coli* GroEL confirmed previous observations of two cylindrical, stacked homohepatmeric rings each with a volume of approximately 85,000 A3, and revealed 3 distinct domains within a single

identical subunit. The equatorial domain forms the interring interface and also houses the ATP-binding pocket. The intermediate domain acts a linker between the equatorial and the substrate binding and GroES interacting apical domain (Braig *et al.* 1994; Figure 1.2). The N-

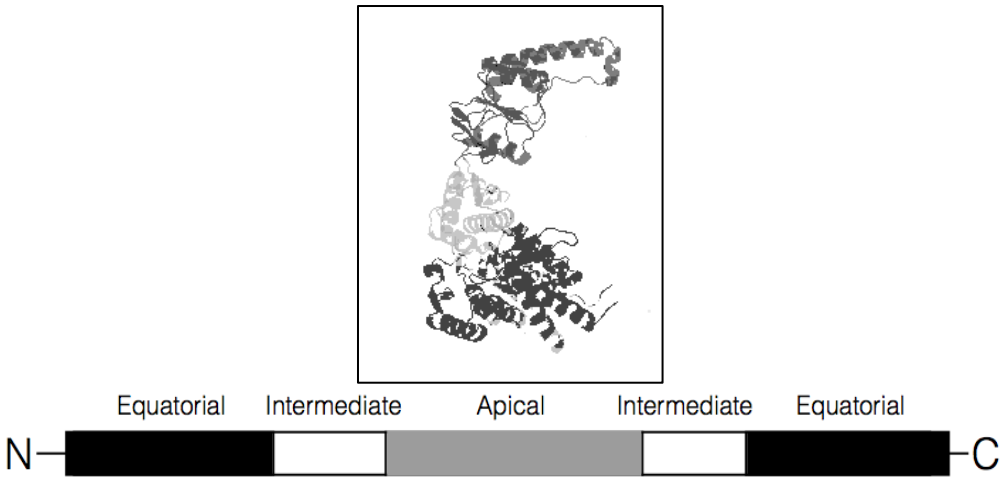

Figure 1.2: Domain architecture of a chaperonin subunit. Top: Side view of a single subunit of Chy CPN-ANP (light grey: apical domain, white: intermediate domain, dark grey: equatorial domain). Bottom: Chaperonin domains as they correspond to the linear polypeptide chain (not drawn to scale). Adapted from Leroux *et al*. 1995, Figure 5.

terminal and C-terminal project into the central cavity at the equatorial domain but are unresolved crystallographically. Hydrophobic residues lining the inside of the apical domain interact with polypeptide substrates as well as the co-chaperone GroES, which displaces substrate into the chamber's cavity upon binding with GroEL (Fenton *et al.* 1994).

groEL and *groES* genes are typically arranged next to each other in an operon; their protein products form the Group I chaperonins, also known as Hsp60 or Cpn60, and Hsp10 or Cpn10, respectively, and are present in all bacteria and eukaryotic organelles except for a few Mycoplasma species (Kim *et al.* 1994; Lund 2009). All known Group I chaperonins form homomeric tetradecamers like the *E. coli* GroEL with the exception of higher plant

chloroplasts which form a heteromeric complex of alpha and beta subunits or beta alone in the presence of MgATP (Martel *et al.* 1990; Dickson *et al.* 2000).

In contrast, the Group II chaperonins are primarily present in archaea and in the cytosol of eukaryotes. These chaperonins can form homo- or heteromeric 8-9 membered double ring structures with similar domain arrangements as the Group I chaperonins, although the apical domain does not rely on interaction with a GroES homolog for complete cycling and instead has a helix-turn-helix protrusion with a large hydrophobic surface referred to as a built-in lid responsible for binding substrate and facilitating folding (Figure [2]; Klumpp *et al.* 1997; Yoshida *et al.* 2002; Reissmann *et al.* 2007). Eukaryotic CCT rings consists of 8 distinct subunits that complex in a specific pattern, while archaeal thermosomes consist of 8-9 membered rings ranging from a single to as many as 5 different subunits (Dekker *et al.* 2011;Leitner *et al.* 2012; Bigotti and Clarke 2008). The heterooligomeric complexes are thought to have arisen via gene duplications. In archaea, paralogs are typically more closely related toeach other than with other archaeal chaperonin genes. These duplications are thought to have occurred independently many times, and also been lost in archaeal genomes (Archibald *et al.*1999). Repeated gene conversions are thought to have slowed differentiation in the substrate-binding domain of these paralogous archaeal genes (Archibald and Roger 2002). In contrast, the CCT subunits in the eukarya are thought to have evolved from a very early duplication event in a protoeukaryote leading to eight distinct subunits, with subunit specific variation in apical domain sequences (Archibald *et al.* 2001). Specific substrates have been identified for eukaryotic chaperonins including the cytoskeletal components, tubulin and actin (Gao *et al.* 1992; Sternlicht *et al.* 1993). Mutagenesis and crosslinking-mass spectrometry studies revealed the residues in each unique subunit that make direct contact with these physiologically relevant substrates. Results suggest that several low-affinity interactions across subunits are required to discern folded versus nonfolded substrate (Joachimiak *et al.* 2014).

1.2.2: Chaperonin Interaction with Co-Chaperones

In addition to GroEL's association with the co-chaperone GroES, chaperonins have been reported to directly interact with other molecular chaperones within the cell. The Group II-containing systems interact with a hexameric holdase chaperone known as prefoldin (PFD or GimC), which offloads nonnative substrates such as eukaryotic actin and tubulin to chaperonins (Vainberg *et al.* 1998). Substrate interaction can occur as the polypeptide chain is being synthesized allowing for efficient, protected delivery to chaperonin cavities for proper folding while minimizing the risk of aggregation in the cytosol (Hansen *et al.* 1999). Prefoldin is a heteromeric complex composed of a double beta barrel body with six coiled coil protrusions each containing hydrophobic patches at the ends for unfolded substrate recognition and interaction (Siegart *et al.* 2000). The termini of the beta subunit are believed to directly interact with chaperonins during substrate delivery (Okochi *et al.* 2004). Other holdases such as small heat shock proteins are thought to function in a similar substrate delivery method as demonstrated with *in vitro* protein folding and salvage assays with chaperonins, although direct contact between the two chaperones has not yet been validated (Laksanalamai *et al.* 2008).

DnaK or Hsp70, a dimeric ATP dependent chaperone that also interacts with nascent polypeptide chains using its C-terminal substrate-binding domain in coordination with co-chaperones DnaJ and GrpE, is a crucial component of the cellular proteostasis network. In proteomic studies using *E. coli*, DnaK demonstrated nonspecific interaction with approximately 700 cytosolic proteins (Calloni *et al.* 2012). DnaK is described as forming productive folding pathways with other chaperones including chaperonins (Beckmann *et al.* 1990). Cytosolic DnaK has been shown to form a stable complex with CCT subunits in eukaryotic systems, although this direct interaction is not demonstrable using the chaperones'

bacterial homologs (Cuellar *et al.* 2008). DnaK is absent in several species of Aquificales and all of the Archaea apart from instances of horizontal gene transfer with bacteria (Warnecke 2012; Macario *et al.* 1991).

1.2.3: Chaperonin Substrates

The foundational understanding of Group I chaperonin substrate recognition and binding is based on the *E. coli* GroEL system. GroEL structures reveal hydrophobic residues in alpha-helices of the apical domain that are crucial for establishing interactions with substrates and the co-chaperone GroES, which upon binding in the presence of MgATP displaces substrate into the central cavity (Fenton *et al.*1994; Fenton and Horwich 2003). Structural studies of bound substrates report findings ranging from loosely organized proteins with variable amounts of secondary structure to random coils devoid of any tertiary structure to disordered polypeptides that associate weakly at a single hydrophobic region but likely become more disordered upon GroEL binding, promoting exposure of more hydrophobic segments of the unfolded substrate, and ultimately permitting a stable substrate interaction across multiple GroEL subunits (Fenton and Horwich 2003; Horst *et al.* 2005; Libich *et al.* 2013). While most substrates are fully encapsulated by GroEL/ES complex during the refolding cycle, there are reports documenting the productive folding of substrates exceeding the size capacity of the internal GroEL cavity (Chaudhuri *et al.* 2001).

There is enough GroEL/ES in an *E. coli* cell during normal growth conditions to assist in the folding of 5% of its proteome (Lorimer 1996). Obligate *in-vivo* GroEL substrates in *E. coli* have been identified (Fujiwara *et al*. 2010). This protein set is responsible for a diverse range of cellular functions, and attempts to generally characterize the requirements for obligate chaperonin interaction include a polypeptide size of 20-60 kDa, and the presence

of maintained, exposed hydrophobic regions mediated either by a slow folding rate, an affinity for a co-chaperone to deliver substrate to the chaperonin, and/or presence of beta-sheets normally buried in the native structure (Ewalt *et al.* 1997; Houry *et al.* 1999; Kerner *et al*. 2005). Substrates with native folding motifs including TIM barrel domains and knotted proteins are also thought to require interaction with the chaperonin system for proper, efficient folding, possibly due to a propensity for proteins with these similar tertiary structures to get stuck more frequently in similar off-pathway intermediates (Kerner *et al.* 2005; Fujiwara *et al.* 2010; Lim and Jackson 2015). A similar *in vivo* GroEL substrate study was completed in the Gram positive bacterium *Bacillus subtilis*, and several substrates absent in *E. coli* including sporulation proteins were identified, indicating that chaperonin interactions may be promiscuous but are still variable across species depending on the host proteome (Endo and Kurusu 1997). Protein folding should be considered in the crowded *in vivo* cellular environment as the polypeptide chain is being synthesized and simultaneously exiting the ribosome. Larger continuous proteins may have domains of variable stabilities, proteins may form different secondary structural elements before others, N and C-termini can not interact until translation is complete, and ribosomal pausing as typical of rare codon insertion into beta-sheet sequences are all factors that can influence protein folding intermediate formation, and potentially their interaction with chaperonins (Tsytlonok and Itzhaki 2013).

Other research has attempted to identify any sequence similarity in *E. coli* chaperonin substrates apart from nonspecific, hydrophobic interactions. The GroES mobile loop sequence that binds to the H and I helices of the GroEL apical domain consists of a hydrophobic region, GGIVLTGAA. Because GroES and unfolded substrate bind the same region of GroEL, researchers looked for a similar sequence pattern in *E. coli* proteins. 60% of *E. coli* proteins contain one hydrophobic patch, and approximately 30% contain multiple hydrophobic patches (Chaudhuri and Gupta 2005; Stan *et al.* 2005). Using this sequence

based approach in combination with previous experimentally derived *in vivo* substrates, it was estimated that 4-5 contact points should be established by the denatured substrate with GroEL across 2-4 subunits. The substrate binding motifs of identified substrate proteins with available structures are buried in their native state, thereby establishing a method for GroEL to recognize denatured versus folded substrates (Stan *et al.* 2005; Stan *et al.* 2006; Houry *et al.* 1999; Kerner *et al.* 2005). Additional studies are required to make more informed predictions regarding GroEL interactors, especially across species (Azia *et al.* 2012). While final folds may have little influence on selection as chaperonin substrate, it is still an interesting trend that the more diverse proteomes in terms of multiple fold families typically contain the most copies of chaperonin genes and subunits as observed in the proteobacteria and eukaryotes (Stan *et al.* 2006; Lund 2009; Kim *et al.* 2012).

1.2.4: Multiple Chaperonins in Bacteria and Archaea

Multiple chaperonin genes are present in approximately 30% of all available sequenced bacterial genomes, either through gene duplication or horizontal gene transfer (Wang *et al.* 2013). These additional chaperonin genes can be found unlinked or in an operon with a *cpn10* gene encoding GroES. In some instances, cross-domain gene transfers are evident. Members of the methanogenic archaeal genus *Methanosarcina* contain both bacterial-type Group I as well as archaeal Group II chaperonins (Deppenmeier *et al.* 2002; Klunker *et al.* 2003). In a number of *Firmicutes*, an archaeal-like chaperonin is annotated in the complete genomes in addition to at least one canonical *groES* and *groEL* operon (Techtmann and Robb 2010; Williams *et al.* 2010), a discovery that forms the basis of this work.

The first reported observation of multiple copies of chaperonins in bacteria was in several *Streptomyces* species containing duplicated *groEL* genes (Guglielmi *et al.* 1991; Mazodier *et al.* 1991). The authors of these studies postulated that these copies are likely retained not for additional chaperoning capacity but for specialized, physiologically distinct functions, such as interaction with unique protein substrates. With certain *alphaproteobacterial* genomes containing as many as 7 chaperonin genes, many subsequent studies have investigated this hypothesis by examining substrate specificities, regulation, and cellular roles of these unique gene products (Lund 2009). Some systems require a specific chaperonin for maturation of key metabolic complexes such as *Bradyrhizobium japonicum*'s nitrogenase, while others are viable when retaining only a single essential housekeeping GroEL gene, indicating that the duplicated copies have overlapping functions or functions not apparent in the study conditions (Fisher *et al.* 1999; Rodriguez-Quinones *et al.* 2005; Bittner *et al.* 2007). Alternatively, these multiple chaperonins may have multiple functions similar to the peroxiredoxins, their postulated antecedents, in which case they would have overlapping functions as well as functions that may not be evident during normal growth conditions (Fares 2014; Dekker *et al.* 2011).

Archaeal genomes that contain multiple copies of chaperonin subunits can form hetero to homomeric complexes of varying sizes depending on the species. Several studies have established differential subunit regulation dependent on temperature. There are two paralogous subunits in *Thermococcus* strain KS-1 with high sequence similarity except for variable C-terminal domain. The alpha subunit contains a GM rich C-terminus, more similar to GroEL subunits, while the beta subunit's C-terminus contains EK repeats. Archaeal chaperonin subunits enriched in EK repeats in their C-termini have been experimentally shown to have increased thermostability (Luo and Robb 2011). Studies examining the ratio of alpha to beta subunits in the protein complex at various temperatures have demonstrated that the beta subunit is in fact more highly expressed in *Thermococcus* KS-1 cells when

grown at an elevated temperature, and that beta subunits demonstrate highest ATPase activity at higher temperatures (Yoshida *et al.* 2002). This indicates that unlike eukaryotic heteromeric chaperonin complexs that have developed distinct substrate specificities with its variable subunits, the archaeal heteromeric complex likely maintains broad specificity and has instead developed as a means to handle fluctuating environmental parameters including temperature (Yoshida *et al.* 2002). A similar observation of temperature-dependent subunit regulation has been made in *Sulfolobus solfataricus* (Chaston *et al.* 2016). This study found that the three subunits formed complexes of different sizes as well as subunit ratios depending on growth temperature- at low temperature, a complex with 9-fold symmetry composed of all three subunits in equal proportion forms, while at normal temperature a 1:1 8-fold symmetrical subunit of beta and alpha subunits complex, and at elevated temperatures a complex composed of entirely beta subunits in 9-membered rings forms. There are subunits with variable thermostabilities, as seen with *Thermococcus* KS-1, as well as possible differential substrate specificites due to the different sizes of complexes formed, though this has yet to be validated experimentally, and a comparative discussion of the sequences of the apical domains of each subunit was not included in the publication (Chaston *et al.* 2016).

Substrate specificities of a dual chaperonin system in the methanoarchaea has also been reported. The methanogenic archaeon *Methanosacrina mazei* evidently acquired a Group I chaperonin via horizontal gene transfer from a bacterium while retaining its Group II chaperonin subunit (Deppenmeier *et al.* 2002). Researchers investigated their substrate specificities via co-immunoprecipitations of the chaperonins followed by LC/MS-MS identification of the interacting proteins. About 13% of the soluble proteome interacted with one or both chaperonin, with an overlapping set of 31% of identified interactors. *Mm* GroEL associated with 183 proteins, while *Mm* thermosome interacted with 252 with no preference for phylogenetic origin of the substrate (Hirtreiter *et al.* 2009). The authors state that the GroEL tended to interact with conserved, more hydrophobic proteins containing alpha/beta

domains while the thermosome favored larger, multidomain substrates of a wider range of domain folds, which may have assisted in the formation of the highly diverse eukaryotic proteomes (Hirtreiter *et al.* 2009). There is experimental evidence for an *E. coli*-derived GroEL/ES system enhancing the laboratory evolution of *Pseudomonas aeruginosa* phosphotriesterase into an arylesterase in an *E. coli* expression system, whereby a 10^4 fold increase in enzymatic activity was achieved when co-expressing the chaperonin system and target protein, suggesting that this chaperone system is important for stabilizing potentially destabilizing mutations and their folding intermediates (Wyganowski *et al.* 2013). The increased diversity of proteome folds in eukaryotes, however, does not appear exclusively dependent on Group II chaperonin-mediated interactions, as bacteria that contain group I chaperonins tend to have more complex proteomes than archaea when comparing diversity of fold families (Kim *et al.* 2012). Other reports suggest that Group II chaperonins may instead have facilitated the shift to an increased number of multidomain proteins due to its ability to partially encapsulate polypeptides at the division of their domains (Ruessmann *et al.* 2012).

1.3: Bacterial and Archaeal-like Chaperonins in Firmicutes

1.3.1: Genomic Context

Firmicutes, specifically the classes *Clostridia*, *Bacilli*, and *Negativicutes*, are currently the only known bacteria that contain an archaeal-like chaperonin gene located within a chaperone gene cluster. This chaperonin complex is structurally more similar to archaeal thermosomes forming homomeric 8-mer rings containing the hallmark protruding apical domain or built-in lid of Group II chaperonins. Phylogenetically, however, these chaperonin sequences cluster in a unique clade separate from both Group I and Group II

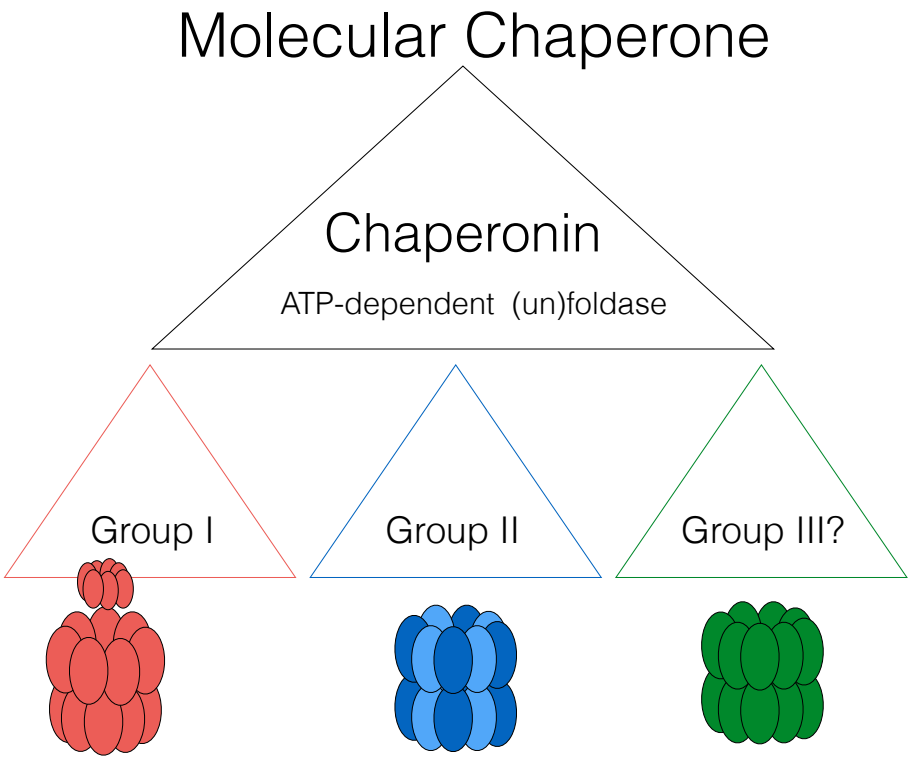

Figure 1.3: Cartoon depicition of the structural and phylogenetic variation of known chaperonins. Group I (red) cpns are found in bacteria and eukaryotic organelles, interact with a co-chaperone lid (Hsp10/GroES), and are monomeric 14-mer complexes. Group II (blue) cpns are found in archaea and the cytosol of eukaryotes, have a built-in lid, and can form homo- or heteromeric complexes of 8 or 9 membered rings. Group III (green) cpns are found in *Firmicutes*, form homomeric complexes of 8-mer rings, and contain a built-in lid.

seqeuneces (Figure 1.3; Techtmann and Robb 2010). All bacteria containing this single archaeal-like chaperonin subunit also contain one or more canonical group I bacterial chaperonin genes, *groEL/ES*. *Carboxydothermus hydrogenoformans*, a CO-oxidizing thermophile, was the first organism in which this archaeal-like Group III chaperonin was observed and characterized (Svetlichny *et al.* 1991; Techtmann and Robb 2010; Williams 2010).

While several lone and less similar archaeal-like chaperonin genes are present in other bacterial species including the cyanobacteria *Gloeobacter violaceus* PCC 7421 and *Gloeobacter kilaueensis* JS1, and the proteobacterium *Oligotropha carboxidovorans* DSM 1227, the Firmicutes' chaperonin gene appears to be under *hrcA* regulation and is situated

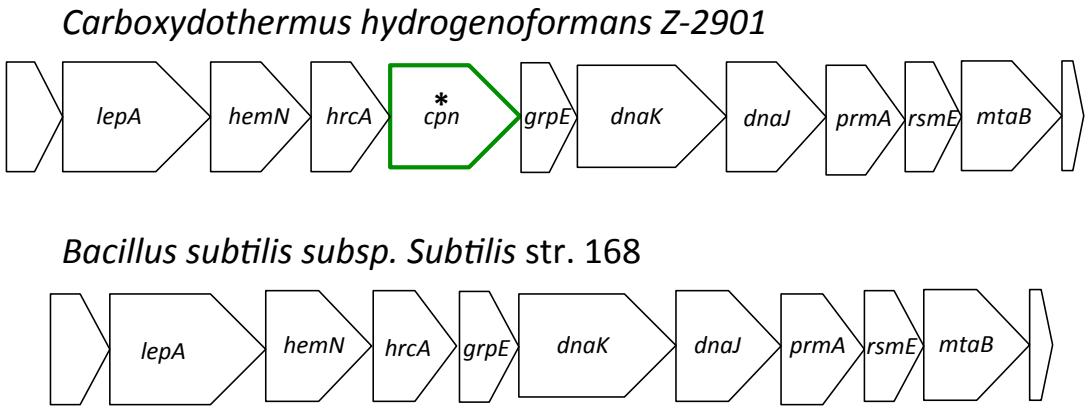

Figure 1.4: Comparision of conserved chaperone operon and surrounding gene ordering in *Firmicutes*. Top: Gene neighborhood of Chy_0413 archaeal-like chaperonin (green) in *Carboxydothermus hydrogenoformans*. Left to right: GTP-binding protein *lepA*; oxygen-independent coproporphyrinogen-3 oxidase *hemN*; heat-inducible transcription repressor *hrcA*; chaperonin *cpn*; molecular chaperone *grpE*; molecular chaperone *dnaK*; molecular chaperone *dnaJ*; LSU ribosomal protein L11 lysine N-methyltransferase *prmA*; 16s rRNA (uracil 1498-N3) methyltransferase *rsmE*; threonylcarbamoyladenosine tRNA methylthiotransferase *mtaB/rimO*. Bottom: Similar gene neighborhood of *dnaK* operon in *Bacillus subtilis subsp. subtilis* str. 168 lacking an archaeal-like chaperonin gene.

downstream of *grpE-dnaK-dnaJ* chaperones. The gene ordering around this apparent chaperone operon is conserved across many Firmicutes, including those that lack the archaeal-like chaperonin gene (Figure 1.4). CIRCE (controlling inverted repeat of chaperone expression) elements have been found upstream of both the *groESL* and *cpn/hsp70* operons in *C. hydrogenoformans*. This conserved sequence element is the site at which the negative repressor HrcA (heat regulation at CIRCE) binds during stress free conditions, inhibiting transcription of downstream chaperones (Narberhaus 1999; Techtmann and Robb 2010). This suggests a coordinated co-regulation and an overlap of function of both the Group I and Group III chaperonins *in vivo.*

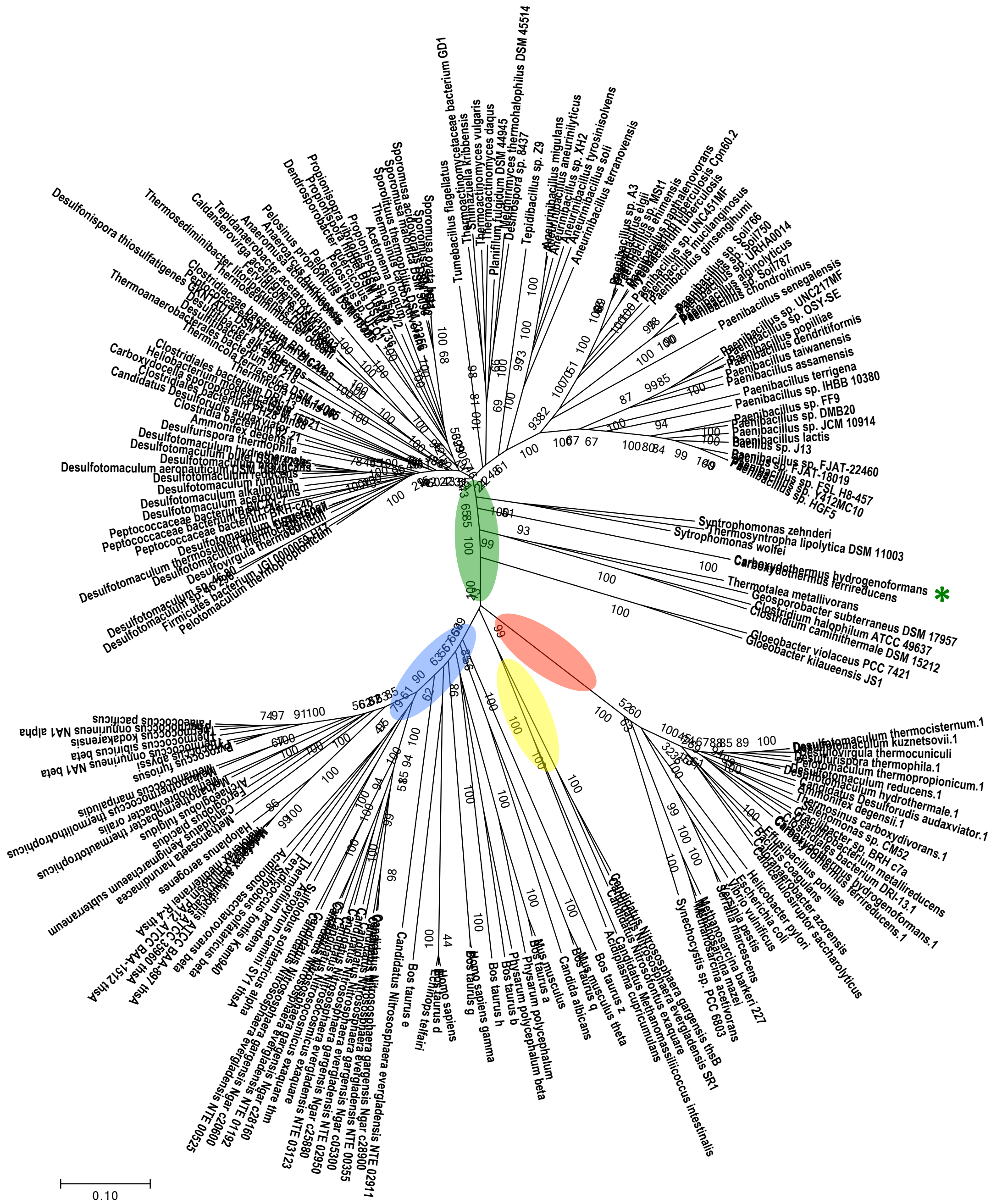

Figure 1.5: Phylogenetic tree depicting chaperonin groups based on their amino acid sequences. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length = 24.90738033 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of amino acid differences per site. The analysis involved 206 full length amino acid chaperonin sequences. All positions containing gaps and missing data were eliminated. There were a total of 346 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.* 2015). The intial alignment was constructed using ClustalW default settings in MacVector. Red: Group I CPNs; Blue: Group II CPNs; Green: Group III CPNs (asterick by Chy CPN sequence); Yellow: new deep branch of archaeal cpn sequences.

1.3.2: Phylogeny of Group III Chaperonins

The phylogenetic origin of Group III chaperonins is intriguing but unclear. Based on the Neighbor-Joining tree of full length amino acid chaperonin sequences, the Group I GroEL subunits cluster together (red branch) while the Group III (green branch) cluster in a distinct phylogenetic group (Figure 1.5; see Appendices for BLASTp table). All currently identified Group III cpn-containing organisms also have at least one copy of the canonical bacterial GroEL/ES operon which raises the question of how these *Firmicutes* acquired an archaeal-like chaperonin and why it has been retained along with the Group I chaperonin when it is absent in closely related bacteria (Figure 1.4). While the majority of Group II chaperonins cluster together including the eight distinct eukaryotic subunits (blue branch), a new chaperonin branch (yellow) has recently emerged consisting of *Thaumarchaeota* and *Euryarchaeota* species in part due to the microbial dark matter sequencing initiative and other culture-independent methods (Rinke *et al.* 2013). The organisms in this group with complete genomes contain multiple archaeal chaperonin, or thermosome, subunits including *Candidatus Nitrosphaera gargensis* Ga9-2 and *Candidatus Nitrosphaera evergladensis* SR1 which encode for 5 and 7 annotated thermosome subunits respectively, in addition to the novel annotated chaperonin sequence that clusters more closely to the Group III sequences. These archaea also contain the bacterial chaperone genes *dnaK-dnaJ-grpE* as well as many heavy metal, in particular copper, resistance related genes (Spang *et al.* 2012; Zhalnina *et al.* 2014). NTE_03148 (*C. N. everglafensis* SR1 *dnaK*) shares 58% amino acid identity to *Tepidanaerobacter sp.* Re1, 59% *Ammonifex degensii* KC4, 56% *Caldanaerovirga acetigigens* DSM 18802, and 57% *Carboxydothermus hydrogenoformans* along with many other Group III-containing bacterial *dnaK* genes after top hits to other *Thaumarchaea* and methanogens. Another member of this novel yellow branch includes *Acidiplasma*

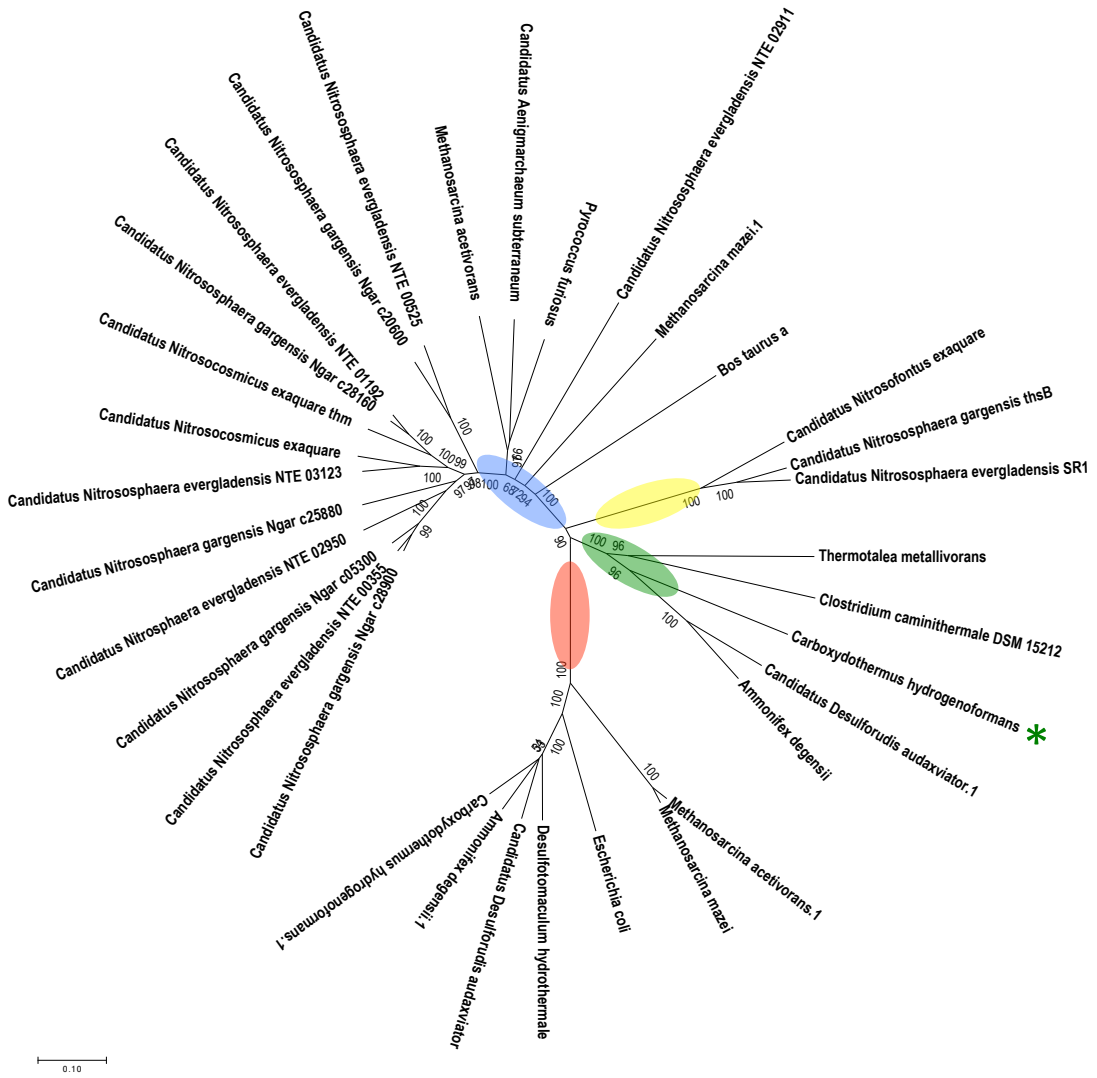

Figure 1.6: Simplified chaperonin phylogenetic tree depiciting novel groups. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length = 5.94216536 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of amino acid differences per site. The analysis involved 33 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 468 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.* 2015). The intial alignment was constructed using ClustalW default settings in MacVector. Red: Group I cpns; Blue: Group II cpns; Green: Group III cpns (asterick by Chy CPN sequence); Yellow: new deep branch of archaeal CPN sequences.

cupricumulans (Hawkes *et al.* 2006; Golyshina *et al.* 2009), an organism isolated from an environment containg 0.4% copper. A proteomic study of a closely related species, *Ferroplasma acidarmanus*, revealed a significant up-regulation of DnaK and thermosome in copper exposed cultures (Baker-Austin *et al.* 2005), indicating that these proteins are important in mitigating cellular damage during metal stress events. It is not yet known if any of the five genes represented in the yellow branch are actively transcribed.

Carboxydothermus hydrogenoformans' genome encodes for other genes that share high similarity to archaeal genes including an annotated copper-translocating P-type ATPase (Chy_0940). A recent study demonstrated this gene product's crucial role in preventing copper toxicity during anaerobic conditions caused by the destruction of [4Fe-4S] clusters in enzymes such has HemN (Azzouzi *et al.* 2013). The BLASTp results of this 90kDa protein sequence against the Joint Genome Institute's Integrated Microbial Genomes system display hits the methanogenic archaeon *Methanohalobium evestigatum* DSM 3721 with e-values of zero within the first 40 sequences, 47% identity, and scores of 1967/4374. For the Chy_0413 archaeal-like chaperonin, however, the first archaeal hit occurs with the halophilic archaeon *Haloferax denitrificans* S1 with an e-value of 3E-67, 35% identity, and a score of 608/2697 as hit 145. *M. evestigatum* is an extremely halophilic, anaerobic methanogen isolated from a microbial mat in Arabat Lagoon in Crimea, Russia (Zhilina and Svetlichnaya 1989). *H. denitrificans* was initially isolated from mud and brine of salterns in San Francisco Bay grown in anaerobic conditions in the presence of nitrate and high salt 30°C-55°C (Tomlinson *et al.* 1986). Several other Group III chaperonin-containing microbes also contain a highly similar copper-translocating P-type ATPase including *Carboxydothermus ferrireducens*, Clostridiales bacterium PH28_bin88, *Desulfotomaculum thermosubterraneum*, *Thermincola ferriacetica*, *Thermincola potens*, *Desulfotomaculum thermocisternum*, *Desulfotomaculum australicum*, *Desulotomaculum kuznetsovii*, *Desulfovirgula thermocuniculi*, *Sporomusa*

malonica, *Sporomusa acidovorans*, *Dendrosporobacter quercicolus*, *Thermosediminibacter litoriperuensis*, *Carboxydocella sporoproducens*, *Thermosediminibacter oceani*, *Desulfitibacter alkalitolerans*, and *Candidatus Desulforudis audaxviator* among others. Additionally, many of these organisms including *Desulfotomaculum kuznetsovii* 17, DSM 6115 (55%), *Carboxydocella sporoproducens* DSM 16521, *Desulfovirgula thermocuniculi* DSM 16036, and *Pelotomaculum thermopropionicum* SI (53%) also contain an annotated coenzyme F420 hydrogenase beta subunit with 34% amino acid identity to *Methanobacterium* sp. SMA-27 and 36% to *Candidatus Methanoperedens nitroreducens* ANME-2d. *C. hydrogenoformans* encodes for a selenocysteine-containing heterodisulfide alpha subunit with 0.0 e value to many polyextremophilic bacteria, and a heterodisulfide reductase subunit beta with 45% identity to *Methanoperedens nitroreducens* ANME-2d. This *hdrABC* cluster is found in many Group III chaperonin-containing species next to a flavin oxidoreducatse (*FlxABCD*). These complexes are though to operate in a similar fashion as in methanogens whereby hdrABC-MvhADG hydrogenase or FdhAB formate dehydrogenase (in place of *FlxABCD* in bacteria) complex reduces the heterodisulfide and oxidized ferredoxin via flavin-based electron bifurcation (FBEB) with hydrogen and formate being the electron donor respectively (Ramos *et al.* 2015; Kaster *et al.* 2011). As the bacterial Hdr complexes seem to lack association with the coenzymes M and B, dsrC is a proposed final electron acceptor or donor depending on metabolism along with ferredoxin (Ramos 2015). Both *Carboxydothermus hydrogenoformans* and *Candidatus Desulforudis audaxviator* possess all the genes necessary for dissimilatory sulfate reduction including dsrC. FBEB appears to be a shared strategy present in fermenting clostridia and thermophilic bacteria, acetogenic bacteria, and methanogens, in which unfavorable reactions including the reduction of ferredoxin are coupled to energetically favorable ones so that reduced ferredoxin can contribute to cellular anabolic processes or chemiosmtic energy conservation. *Candidatus Desulforudis audaxviator* contains an archaeal-type acetyl-CoA decarbonylase/synthase alpha

subunit (Daud_0105) with 53% amino acid similarity and e-value of 0.0 to *Methanothermobacter thermautotrophicus* CaT2, and many other Group III chaperonin organisms also contain CO-utilizing genes although of the bacterial type of carbon monoxide dehydrogenases. Many of these organisms are also hydrogenogens and coexist syntrophically with H_2-consuming methanogens to keep H_2 partial pressues low and maintain favorable conditions for other metabolic reactions including volatile fatty acid oxidation in *Pelotomaculum thermopropionicum* strain SI (Ishii *et al.* 2005).

Several Group III-containing bacteria contain a high number of genes with postulated archaeal origin, with 23 genes retaining 30% or higher identity to 22 *Euryarchaeota* genes (22.73% *Methanomicrobiales*, 13.64% *Archaeoglobales*, *Methanococcales*, and *Methanosarcinales* of the total archaeal hits) and 1 Crenarchaeota hit in *Candidatus Desulforudis audaxviator*, and 14 *Euryarchaeota* and 1 *Thaumarchaeota* hit in *Thermincola ferriacetica* DSM 14005, while *Carboxydothermus hydrogenoformans* has 3 Euryarchaeota hits. Alternately, *C. Nitrosphaera gargensis* Ga9-2 contains 21 genes with bacterial Clostridia origin, specifically 16 from *Clostridiales*, 1 *Halanaerobiales*, and 4 *Thermoanaerobacterales* (11.29% of 186 bacterial genes total hitting Clostridia, 16.67% hittting *Cyanobacteria*, 10.22% hitting *Alphaproteobacteria* and 9.68% hitting *Bacilli*), while *Methanobrevibacter oralis* JMR01 hits 39 *Clostridiales* genes and 1 *Thermoanaerobacterales* (37.61% of 109 bacterial genes total hitting *Clostridia* followed by 20.18% hitting *Bacilli*), and *Methanocella arvoryzae* MRE50 hits 41 *Clostridiales* and 6 *Thermoanaerobacterales* (18.50% of 254 bacterial genes total hitting *Clostridia* followed by 12.99% from *Cyanobacteria* and 10.63% from *Bacilli*) (Markowitz *et al.* 2012).

These findings support the shared ecological niche hypothesis whereby horizontal gene transfer events likely occur between exclusive groups of predominantly polyextremophilic Firmicutes and methanogenic archaea due to cohabitation in energy

limited, heavy metal, H_2, and CO-rich environments (Williams 2010). These gene transfers likely confer advantages to the recipients, and may potentially aid in colonizing new environments as theorized with the adaptation of archaea to more mesophilic growth termperatures via the acquisition of mesophilic bacterial Hsp70 genes (Lopez-Garcia *et al.* 2004). It is unclear whether or not the Group III chaperonin represents a relic from a horizontal transfer event with an archaeon, or if they represent an ancient chaperonin that existed in the last universal common ancestor prior to the splitting of archaeal and bacterial lineages. Relatively recent cross-domain horizontal gene transfer events including the acquisition of Group I chaperonins by methanogens are clearly discernible considering the tight grouping of their GroEL/ES with other Group I CPN sequences (Figure 1.6; Deppenmeier *et al.* 2002; Klunker *et al.* 2003). The green and yellow branches of these trees represent sequences from many organisms that are difficult to sample and/or culture in the laboratory. Techtmann argued that the absence of the conserved gene ordering of the Group III chaperonin operon in archaea suggests that this sequence was not obtained from an archaeal genome (Techtmann 2009; Techtmann and Robb 2010). The conserved gene ordering of the Group III chaperonin in related bacteria lacking the CPN gene instead suggests a possible gene loss event (Figure 1.3). The continued sequencing of the global microbiome will hopefully provide more evidence regarding the origin of this important protein complex.

Chapter 2: The Crystal Structure and Functional Mutations of *Carboxydothermus hydrogenoformans* Archaeal-like Chaperonin

2.1: Abstract

Carboxydothermus hydrogenoformans is the first observed bacterial genome containing an archaeal-like chaperonin gene that clusters independent of all previously characterized chaperonins, forming a new third group of chaperonins (Wu 2005; Techtmann 2009; Techtmann and Robb 2010). Preliminary Chy CPN studies indicate that this double-ringed, homomeric hexadecamer is an active ATPase capable of protein folding (Techtmann 2009; Techtmann and Robb 2010). No studies, however, have been completed investigating Group III structure and function. This body of work aims to investigate the function of key residues in Group III chaperonin cycling based on the novel high resolution crystal structure of Chy CPN in comparison with previous Group I and Group II structure-function studies (An *et al.* 2016 (accepted)). The Chy_0413 gene was subcloned into a new vector without a terminal tag for validation of previous work and further characterization. Site-directed mutagenesis of residue R155 followed by ATPase assays and isothermal titration calorimetry with ATP indicate that at a basic residue at this site is critical for proper ATP binding and subsequent hydrolysis, and therefore overall cpn cycling and substrate refolding. Chy CPN D373 is a crucial catalytic residue for ATP hydrolysis, with D373A resulting in 95.3% diminished ATPase activity after one hour at 65°C, while a deletion of the last 16 reisudes of the C terminus results in a down shift in optimum temperature from 65°C to 50°C, and only about a third of the ATPase acitivty. Chy CPN though phylogenetically distinct is overall structurally and functionally similar to previously characterized Group II chaperonins with comparable key residues performing analogous functions in ATP sensing and hydrolysis, and

providing stabilizing interactions of the complex. The physiological role of this group of chaperonins remains to be investigated.

2.2: Introduction

Carboxydothermus hydrogenoformans is an anaerobic, carbon monoxide utilizing thermophilic bacterium originally isolated from Kunashir Island. It is capable of growing in 100% CO, and can use this C1 compound as its sole carbon and enery source (Svetlichny e*t al.* 1991). Its genome was fully sequenced and analyzed in 2005, and resulted in the discovery of the first instance of an archaeal-like chaperonin gene in a bacterial genome (*Chy_0413*; Wu *et al.* 2005; Techtmann 2009; Techtmann and Robb 2010). This chaperonin gene is structurally more similar to Group II chaperonins but phylogenetically distinct from both known CPN groups (Techtmann 2009; Techtmann and Robb 2010). It has subsequently been identified in approximately 80 other bacterial genomes, the majority of which are closely related Firmicutes that share similar gene ordering with the Group III CPN gene residing in a chaperone operon (Table 3; Figure 1.3). In all instances the occurrence of a Group III chaperonin gene coccurs with at least one copy of a Group I chaperonin gene (Techtmann 2009; Techtmann and Robb 2010; Table 1.3). Any distinct physiological roles of both types of chaperonins *in vivo* are as of yet undetermined.

A basic characterization of Chy CPN (Chy_0413) protein has been completed (Techtmann 2009; Techtmann and Robb 2010). Chy_0413 quaternary structure was investigated by native PAGE, GF/MALS, and cryoEM, revealing that this single subunit chaperonin forms double stacked eight membered rings with an optium ATPase activity at 65°C. Chy_0413 is capable of refolding denatured substrates, and hydrolyzing other nucleotides including TTP and GTP at 60% the rate of ATP (Techtmann 2009). The crystal

structures for Chy CPN bound with ADP and a non-hydrolyzable ATP analog were obtained to confirm previous structural observations as well as provide a more detailed model of Group III chaperonins (An *et al.* 2016). Mutations at putative catalytic residues, the ATP binding site, and other complex stabilizing interactions based on structurally and alignment observations were constructed to experimentally validate.

Each individual chaperonin subunit regardless of Group affiliation has a single nucleotide-binding site (Figure 2.6). All Group II chaperonins, including the eight distinct subunits of the eukaryotic complex, have a conserved lysine residue in the putative nucleotide-sensing region of the complex (top sequence; Figure 2.1). This lysine is thought to interact specifically with the gamma phosphate of ATP. Studies with *Methanococcus maripaludis* K161A revealed a significant decrease in ATP binding and only ~25% of the ATPase activity of the WT *Mm* cpn (Pereira *et al.* 2012). Amino acid alignments of Group II and Group III chaperonin sequences reveal a completely conserved arginine

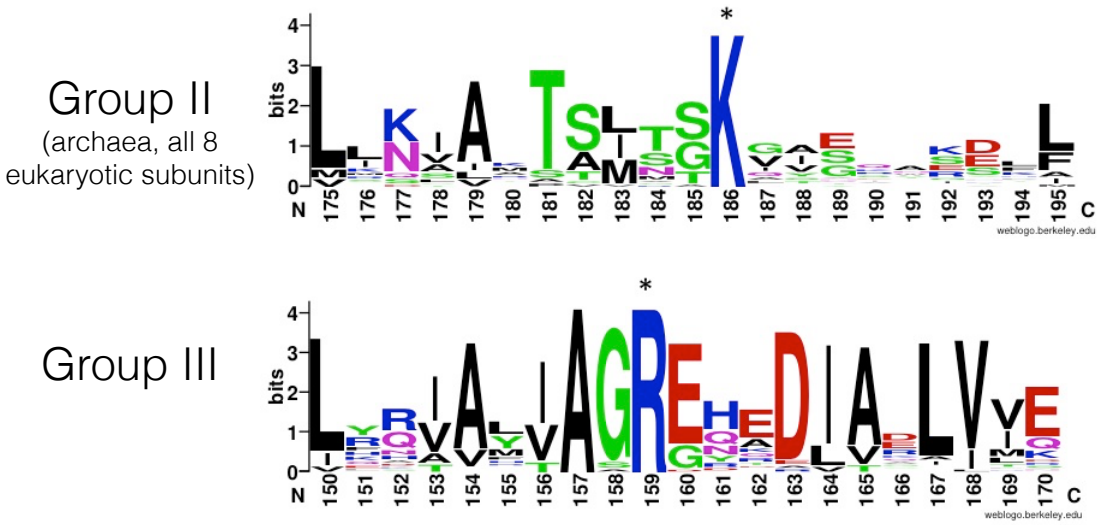

Figure 2.1: Comparison of highly conserved putative ATP gamma-phosphate interacting residue (denoted by asterisks) in Group II (top) and Group III (bottom) chaperonins. Top: Logo created using WebLogo (Schneider *et al.* 1990; Crooks *et al.* 2004) and ClustalW amino acid alignment of 26 Group II chaperonins including representatives from all 8 eukaryotic subunits and archaeal sequences. Bottom: Group III chaperoning alignment of 79 sequences pulled from NCBI database by BLAST against Ch cpn amino acid sequence (Altschul *et al.*1997).

residue at this location (Figure 2.1). A Chy_0413 R155 mutant series was constructed to determine whether this residue is important for ATP binding.

The flexible N and C-termini of chaperonins are highly variable in sequence, and have yet to be fully resolved in structures. Several studies have provided evidence for their roles in chaperonin cycling. In Group II chaperonins, the termini are the primary determinant in the thermostability of the complex. An increased number of charged residues in the C-termini tightly correlate with an increased temperature optimum of its native organism (Luo and Robb 2011). CryoEM studies examining the termini locations of archaeal chaperonins in open and closed states revealed that electrostatic interactions mediate a connected network of the termini in the central portion of the equatorial domain of the complex. This network is crucial for assembly and thermostability, and may transition from mass overlapping interring connections to exclusive intraring associations depending on the opening and closing of the complex, although the timing of this shift is disputed (Zhang *et al.* 2010; Zhang *et al.* 2013).

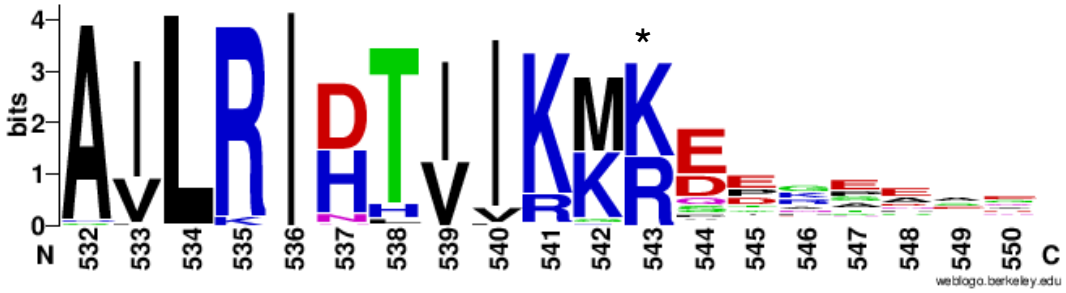

Figure 2.2: C-terminal amino acid alignment of all Group III chaperonin sequences available in JGI IMG database using ClustalW and WebLogo. The asterick indicates the terminal residue of the Chy_0413 C-terminal deletion mutant (Crooks *et al.* 2004; Schneider *et al.*1990).

With Group I chaperonins, however, there is evidence that indicates that the termini play a more direct role in substrate interaction. Eubacterial chaperonin C-termini typically contain a repeated GGM motif. Some reports suggest that the last 27 residues of the C-termini are dispensable for normal growth in *E. coli* (Burnett *et al.* 1994), while *in vitro* studies with a variant lacking the last 23 residues result in significantly decreased substrate

folding rates (Weaver and Rye 2014). The observed slowed refolding cycling of the C-termini deletion mutant is explained by proposing a direct interaction by the C-termini with unfolded substrate, which plays a vital role in bringing the substrate down into the cavity where it promotes protein unfolding for proper refolding to occur (Weaver and Rye 2014). A Chy_0413 C-terminal deletion mutant omitting the last 16 residues of the subunit was constructed in order to investigate the role of this flexible region in the cycling and assembly of Group III chaperonins based on an alignment of known Group III chaperonin amino acid sequences (Figure 2.2).

Intra-ring contacts are crucial for assembly and coordinated chaperonin opening and closing. In the Group II chaperonin from *Methanococcus maripaludis*, residues D45 and R511 form ion pairs across subunits in the opened and closed conformation, and are thought to be a critical hinge mediating intraring movements (Zhang *et al.* 2010). Mutants were constructed to determine whether comparable residues in the Chy_0413 Group III chaperonin, D41 and K497, are important for intra-ring communication (Figure 2.18).

2.3: Materials & Methods

2.3.1: Cloning, Expression, and Purification

The Ch cpn gene was cloned downstream of the tac promoter of the expression plasmid pTac24b(+) using *NdeI* and *XhoI* (NEB Inc.; pTac24b(+) courtesy of Dr. Zeev Pancer; constructed using pET24b(+) vector (Novagen) with 5'- CCATGGTTGACAATT AATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTC cloned into the BglII and XbaI sites, replacing the T7 promoter with tac promoter (de Boer 1983)). The initial leucine codon was changed to a methionine using forward primer 5'-

AAAAAACATATGAAAAAAGAAAAGGAAA and reverse primer 5'-TTTTTTCTCG AGTTATCTTTTTCCTCCATC (IDT Inc.). FideliTaq PCR Master Mix was used for gene amplification using 12.5ng of *C. hydrogenoformans* DSM 6008 genomic DNA as template in a 50µl reaction (Affymetrix). Static cultures overpressured with 95-100 kPa of CO were prepared in accordance with the DSM 6008 medium 507 protocol. Genomic DNA was extracted from DSM 6008 cell pellets following DOE JGI's Bacterial DNA Isolation using CTAB Protocol (Feil *et al.* 2012). A portion of this gDNA was sent to JGI for resequencing of DSM 6008 culture collection strain to compare with the available genome for strain Z-2901 and validate suspected frameshifts (Wu *et al.* 2005). All PCR was performed using a PTC-200 Peltier Thermal Cycler Sequence (MJ Research). Verified plasmid DNA was transformed into the *E. coli* BL21 (Novagen) and expressed via autoinduction at 25 C for 72 hours with 50µg/ml kanamycin (see A.2 for full protocol; Studier 2005). Cells were harvested by centrifugation at 5,000xg for 10 minutes at room temperature, then resuspended (5% wt/vol) in 50mM NaH2PO4, pH 8.0, 300mM NaCl or 50mM HEPES, pH 8.0, 100mM NaCl in the presence of 1mM PMSF followed by lysis via French Press. Lysate was spun for 20 minutes at 20,000xg at room temperature. The supernatant was heated at 50°C for 30 minutes followed by another 20 minute 20,000xg centrifugation. Heat treated and clarified supernatant was then loaded onto anion exchange High Q column (Bio-Rad) and eluted over a gradient up to 1M NaCl. Protein concentrations were determined by Coomassie Plus (Bradford) reagent (Pierce) at 595nm. All proteins used in ATPase assays were initially resuspended and lysed in HEPES buffer to avoid high phosphate background.

2.3.2: Site-Directed Mutagenesis

The QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) was used to construct all mutants. Primers were provided by Integrated DNA Technologies (see A.1 for sequences). Reverse primers for mutagenesis were the reverse and complement of forward sequences. 5µl of 10x reaction buffer, 20ng of dsDNA template (WT Chy_0413 in pTac24b(+)), 1 µl dNTP mix, and 125ng of each primer were included in a thin-walled PCR tube containing a final reaction volume up to 50µl using dH_2O. 2.5U of *PfuUltra* HF DNA polymerase was then added to the reaction. After 30 seconds at 95°C, the PCR cycled from 95°C/30 seconds, 55°C/1 minute, to 68°C/6.5 minutes for 12 cycles for point mutations, 16 cycles for single amino acid changes, and 18 cycles for multiple amino acid changes as per the kit's protocol. After PCR, 10U of DpnI was added directly to the reaction, mixed by pipetting several times, followed by a 30 second centrifugation and incubation for 1 hour at 37°C. The reaction was then chilled on ice, 1µl of which was transformed into thawed chemically competent XL1-Blue cells. After 30 minutes on ice, the transformation reaction was heat shocked for 45 seconds at 42°C and chilled on ice for 2 minutes after. 0.5ml of NZY+ broth pre-heated to 37°C (A.2) was added to the reaction, and then placed in incubator-shaker for 1 hour at 37°C, 200rpm. Transformations were plated on LB agar plates containing 50µl/mg of kanamycin and 0.5% glucose. Sequence verified mutants were subsequently transformed into *E. coli* BL21 (Novagen). Expression and purification of mutant Ch cpns were identical to the Ch cpn wild type protocol.

2.3.3: Native PAGE

Continuous native polyacrylamide gels were cast using 1:1 mix of 3 and 15% polyacrylamide gel formulations drawn into a serological pipet and mixed using an air

bubble. A 3% formulation consisted of 3.225ml dH_2O, 0.375ml 40% bis-acrylamide (Bio-Rad), 1.3ml 1.5M Tris, pH 8.8, 50µl 10% ammonium persulfate (Sigma), and 5.5µl TEMED (Sigma). 15% formulation consisted of 1.0ml sterile glycerol, 0.7ml dH2O, 1.9ml 40% bis-acrylamide (Bio-Rad), 1.3ml 1.5M Tris, pH 8.8, 50µl ammonium persulfate (Sigma), and 2µl TEMED (Affymetrix USB products). NativeMark Unstained Protein Standard (LC0725; Novex by Life Technologies) was used for all native PAGE. All gels were run using Mini-PROTEAN Tetra Cell and system (Bio-Rad) and Tris-Glycine running buffer. Samples were incubated at 65°C for 10 minutes prior to loading with or without the presence of 2mM ATP and 5mM $MgCl_2$.

2.3.4: Gel Filtration/Multi-angle Light Scattering

The gel filtration/multi-angle light scattering studies were performed using a Shodex Protein KW-804 column, miniDAWN TREOS multi-angle light scattering instrument (Wyatt Technology, laser wavelength of 658nm, cell type K5), variable-wavelength UV absorbance at 280nm (Agilent Technologies 1200 series), and refractive index instrument OptilabrEX (Wyatt Technology). ASTRA software was used for analysis of results. Prior to loading, the chaperonin samples were buffer exchanged into column buffer, 100mM ammonium sulfate, 10mM sodium phosphate, 0.01% NaN_3 at pH 6.4-NaOH, and incubated at 60°C for 10 minutes in the absence and presence of 1mM AMP-PNP and 1mM $MgSO_4$. Any aggregates were centrifuged out of supernatant at 11,000xg for 5 minutes at room temperature followed by supernatant filtration using 0.22µM cellulose acetate filters. 20µl of 7.852 mg/ml purified Chy_0413 was injected into the column. Samples were run at a flow rate of 0.5ml/min. I received the assistance of Dr. Travis R. Gallagher of the National Institute of Standards and Technology/Institute for Bioscience & Biotechnology Research during these experiments.

2.3.5: Open and Closing Analysis with Proteinase K Digestion and SDS-PAGE

Chy_0413 was exposed to proteinase K, a protease with broad specificity, in the presence of several different nucleotides to determine how accessible the chaperonin is to digestion in different nucleotide-induced conformations. Proteinase K (fungal; Ambion) was resuspended in 10mM Tris, pH 7.5, 20mM CaCl2, 50% glycerol. Chy_0413 (30 nM complex) were preincubated at 65°C for 10 minutes in 50mM HEPES, pH 7.5-KOH, 50mM KCl, 20mM MgCl2, with 2mM freshly prepared nucelotide or analogue (adenosine-5'-monophosphate (Sigma), adenosine-5'-diphosphate (Boehringer Mannheim GmbH), adenosine-5'-triphosphate (Sigma), adenylyl imidodiphosphate (Sigma), or 10mM EDTA to chelate Ca2+ and interfere with proteinase K actvitiy. Proteinase K (0.5µg) was added to each reaction at room temperature and incubated for 10 minutes, then stopped with 5mM PMSF and place on ice. Samples were then loaded into a discontinuous 15% SDS-PAGE and run in Tris-Glycine buffer system using PageRuler Unstained Broad Range Protein Ladder (26630; ThermoFisher Scientific) and the Mini-PROTEAN Tetra cell (Bio-Rad). Gels were stained using 0.2% Coomassie brilliant blue R-250 (Amresco), 40% methanol, 10% acetic acid solution, and destained using a 40% methanol, 10% acetic acid solution.

2.3.6: ATPase Activity of Chy CPN and Chy CPN Mutants

Malachite Green, a colormetric reagent that reports the amount of free inorganic phosphate in a system, was used to determine all chaperonin ATPase activities. Malachite Green forms a complex with phosphomolybdate at low pH and absorbs wavelengths ranging from 620-660nm (Itaya and Ui 1966). All reactions were carried out in 50mM HEPES pH 7.5-KOH, 50mM KCl, 20mM $MgCl_2$ with ATP. The dye regeant was made by combining

36ml of 0.045% Malachite Green in H_2O (Fluka), 12ml of 4.2% ammonium molybdate in 4N HCl (Sigma), and 1ml of 1% Triton X-100 (Bio-Rad). The dye solution was clarified with a 0.22μM filter. Chaperonin samples were preincubated at the indicated temperatures for 10minutes before the addition of ATP (0.1-2mM final concentrations). Reactions were sampled at indicated time points by mixing 8μl of samples with 128μl dye solution and 6μl of buffer. The reaction was stopped after 1-minute with 16μl of 34% citric acid. All samples were read at 640-660nm using a Beckman DU640 spectrophotometer, and assayed in triplicate.

2.3.7: Differential Scanning Calorimetry

All differential scanning calorimetry was performed using a Nano DSC in manual mode with cells pressurized to 3atm (TA Instruments). Samples were scanned from 25-100°C, 1°C per minute with 600s equilibration. Four rounds of heating and cooling were performed for buffer baselines, and two rounds were performed for samples. The protein samples' buffer was exchanged into 20mM NaH_2PO_4-NaOH, pH 7.4, 100mM NaCl using nanopure H_2O and centrifugal concentrators with a 30 kDa cut-off (Amicon). Both buffer blanks and samples were degassed for 10 minutes at 426mmHg, 20°C, prior to loading at a concentration of 1.03 mg/ml into the Nano DSC (TA Instruments). DSCRun and NanoAnalyze software were used for analysis of results including subtraction of buffer baselines from samples (TA Instruments). All cloning, protein expression, and purification procedures are described in section 2.1.1 and 2.6.1, with primer sequences in A.1.

2.3.8: SV2

Dr. C. M. Santosh Kumar of the National Center courteously provided SV2 for Cell Science, Pune, India. Only pTac24b(+) constructs were used for this experiment as SV2 lacks an integrated T7 polymerase gene. Chy_0413, a polycistronic Chy_0806_0807 construct, *E. coli* GroEL, and empty pTac24b(+) were all transformed into previously prepared SV2 chemically competent cells (see A.3). Cultures containing 0.5% glucose and 50 µg/ml of kanamycin were grown at 30°C, 200rpm to an OD600 of 0.6, and then induced with 1mM IPTG for 2 hours. The density of cultures were then again compared and adjusted to comparable levels using sterile LB/Kan broth. 10µl samples of each culture were then drawn and mixed throughly by pipetting with 90µl of sterile 0.85% NaCl. Serial dilutions were made for each sample using the sterile NaCl solution. 3µl of each sample was then spotted onto duplicate LB/Kan agar plates. Both were incubated overnight, one at 30°C and the other at 42-44°C Colony counts were performed the following day. Initial and end point samples (1ml) were taken from cultures to assess expression of chaperonins via SDS-PAGE.

2.3.9: Isothermal Titration Calorimetry with ATP

All ITC experiments were conducted by Dario Spigolon. Measurements of the ATP binding were made in a Nano ITC 1 ml volume (New Castle, DE, TA Instruments) using the Chy_0413 D373A/R155 mutant series. ATP concentrations were measured spectrophotometrically at 259 nm with an extinction coefficient of 15,400 M. All experiments were carried out at 55°C in 20mM Tris-HCl, 20 mM $MgCl_2$ and 10 mM KCl (pH 7.6). Prior to use, all solutions were degassed under vacuum to eliminate air bubbles. Titration experiments were performed by successive 5 µL injections of 200 µM ATP prepared in the same buffer as the samples into the proteins (0.43 µM for the hexadecamer). The interval

between injections was 300 s, and stirring at 300 rpm ensured a good mixing. Binding isotherms were corrected by subtracting the ligand dilution isotherms, determined by titrating ATP solution into buffer alone. The observed binding isotherms were fitted to the "one set of sites" model. In this model, the putative chaperonin has several identical nucleotide binding sites, which are independent of each other and have a uniform binding constant, K_a, and enthalpy change, ΔH. Data analysis was carried out using MicroCal Origin 7.0.

2.3.10: X-ray Crystallography of Chy CPN by the Cha Lab at KIOST

Our collaborators at the Korea Institute of Ocean Science and Technology performed crystallization and X-ray crystallographic analysis of Chy_0413 as described in An *et al.* 2016. I provided purified protein and provided insight on the effects of ADP and nucleotide analog binding to the Chy CPN complex. High-resolution structures for both the open and the closed conformations were obtained by incubation of protein with 3mM ADP and AMP-PNP respectively. All figures were generated using PyMOL unless contributed by the Cha laboratory (The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC).

2.4: Results

2.4.1: Complex Assembly in the Presence of Nucleotides

Native PAGE of Chy_0413 revealed major bands between 720 and 1048 kDa. Two bands are visible in the lane 1 without ATP, while a single band is present in the sample incubated with ATP prior to loading (Figure 2.3a).

Gel filtration/multi-angle light scattering analysis of Chy_0413 alone resulted in a major peak of approximately 925.4 kDa around the 15-minute elution time (Figure 2.4a). Preincubation with 1mM AMP-PNP and 1mM $MgSO_4$ resulted in a major peak 952.8 kDa (Figure 2.4b). Proteinase K digestion of Chy_0413 is decreased when the chaperonin is incubated with ATP and AMP-PNP. No nucloetide, AMP, and ADP incubated with Chy_0413 result in increased digestion (Figure 6). EDTA (10mM) also offers protection to Chy_0413 from digestion by chelating Ca^{2+} in solution, impairing Proteinase K's catalytic activity (Figure 2.3.b).

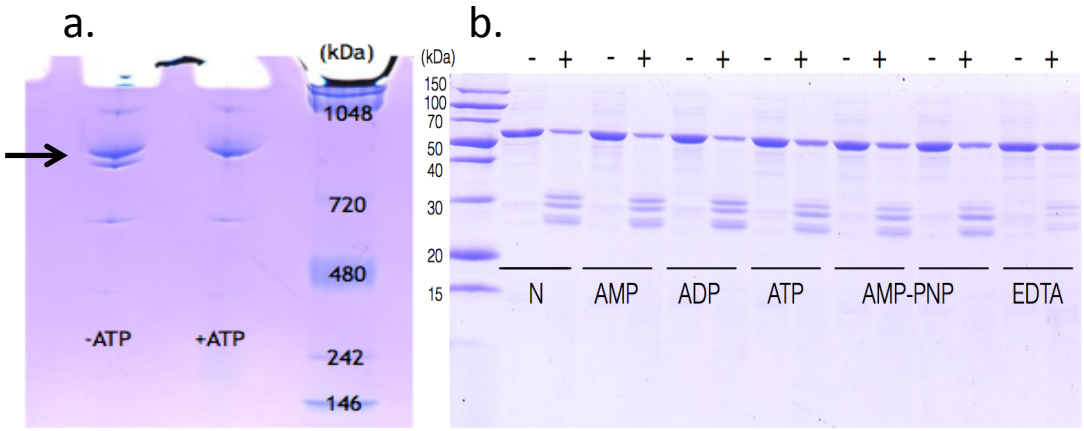

Figure 2.3: The response of Chy CPN to various nucelotides. Left (a.): Native PAGE of Chy_0413 WT without (-; lane 1) and with (+; lane 2) 1mM ATP/5mM $MgCl_2$. 10µl NativeMark Unstained Protein Standard (Novex) loaded in lane 3. Right (b.): 15% SDS-PAGE of Chy_0413 (0.3µM complex) incubated with (+) and without (-) proteinase K at 65°C for 10 minutes. Proteinase K buffer alone (10mM Tris, pH 7.5, 20mM $CaCl_2$, 50% glycerol) was added to (-) samples. 10µl of PageRuler Unstained Broad Range Protein Ladder (ThermoFisher Scientific) in lane 1. (N): Chy_0413 alone. AMP-PNP loaded twice from two separate batches of analogue (Sigma).

The crystal structures obtained by the Cha laboratory revealed an open complex in the presence of ADP and a partially or fully closed structure in the presence of the non-hydrolyzable ATP analogue, AMP-PNP or ANP at ~3Å resolution (Figure 2.5; An *et al.* 2016).

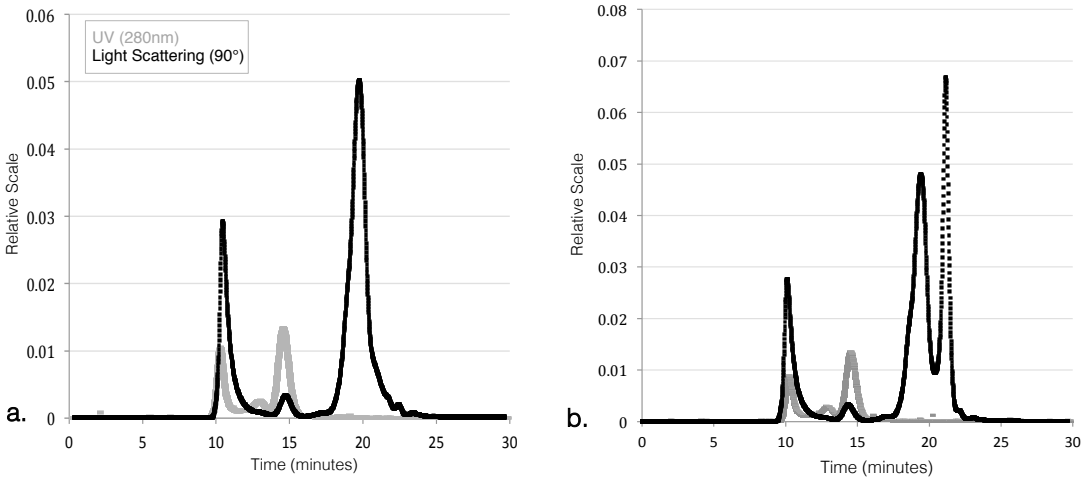

Figure 2.4: Size determination of Chy CPN using Gel Filtration/Multi-Angle Light Scattering. a. Chy cpn alone incubated at 60°C for 10 minutes prior to loading. b. Chy cpn inucbated with 1mM AMP-PNP, 1mM $MgSO_4$ at 60°C for 10 minutes prior to loading. Both runs consisted of a 20µl injection of 7.9mg/ml of Chy CPN in 100mM ammonium sulfate, 10mM sodium phosphate, 0.01% NaN_3 at pH 6.4-NaOH.

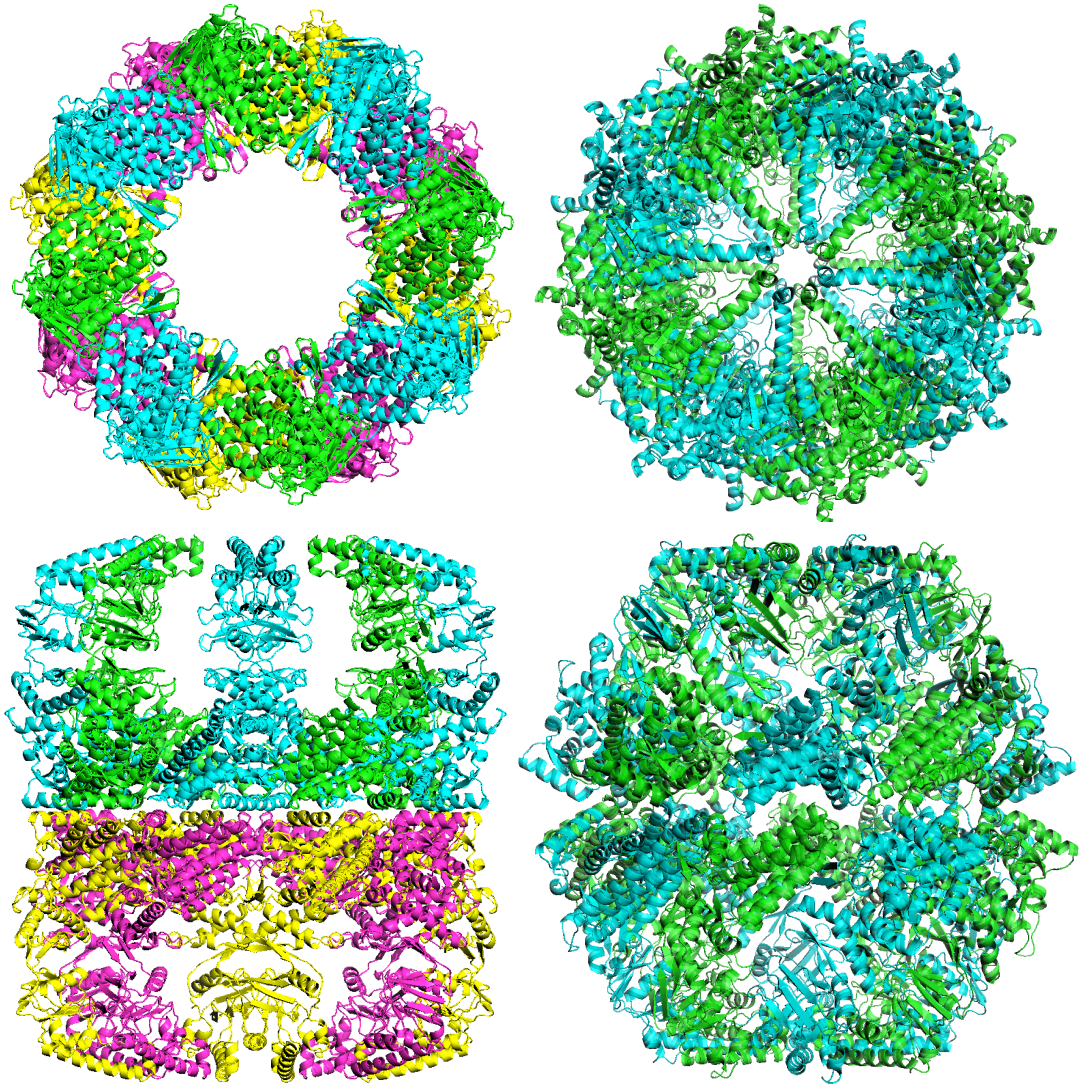

Figure 2.5: Comparison of ADP (left) and AMP-PNP (right) bound Chy CPN crystal structures (An *et al.* 2016; The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC)). Top: top down view of open/ADP and closed/AMP-PNP complexes. Bottom: side view of open/ADP and closed/AMP-PNP stacked rings. Individual subunits are shown with random alternating colors.

2.4.2: ATP Binding and Hydrolysis: Chy CPN R155 and D373

The crystal structures bound with ADP and ANP revealed a single nucleotide-binding site per subunit with a total of 16 ANP molecules binding per complete Chy CPN complex (Figure 2.6; An *et al.* 2016).

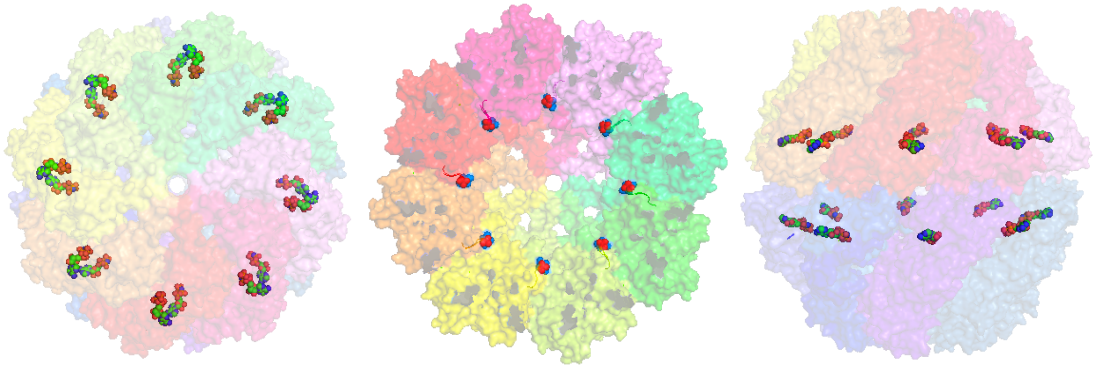

Figure 2.6: Illustration of a single molecule of ANP (bold spheres) bound to each individual Chy CPN subunit represented in a unique color with transparent surfacing. Left: Top down view of complex. Middle: Single ring bottom (equatorial region) up view. Right: Side view of double-ringed complex (An *et al.* 2016; The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC).

Incubation of purified R155 proteins and WT Chy_0413 at 65°C with 0.1mM ATP resulted in 106 (+/-9.8) µmol/min/mg for WT, 99.6 (+/-3.2) µmol/min/mg for R155, 8.9 (+/-9.7) µmol/min/mg for R155E, and 0.22 (+/-2.7) µmol/min/mg for R155L. To investigate the ATP binding affinities of WT Chy_0413 versus R155, mutants with minimal rates of ATP hydrolysis were constructed (Figure 2.7).

Chy_0413 D373A series was made to interfere with the coordination of the water molecule required for cleavage of ATP gamma phosphate using WT and all previously described Chy_0413 R155 mutants as templates (see A.1 for primer sequences). The malachite green assay as described in section 2.3.6 was performed using the R155/D373A

series to determine ATPase activities. D373A displayed decreased activity at the experimentally relevant temperature of 65°C (72.0 (+/- 4.3) for WT; 3.4 (+/- 9.0) for D373A; units in µmol/min/mg). WT Chy CPN showed approximately 17% of ATPase activity at room temperature versus 65°C after 60 minutes. D373A cpn had comparable, minimal activity under both conditions (Figure 2.8a).

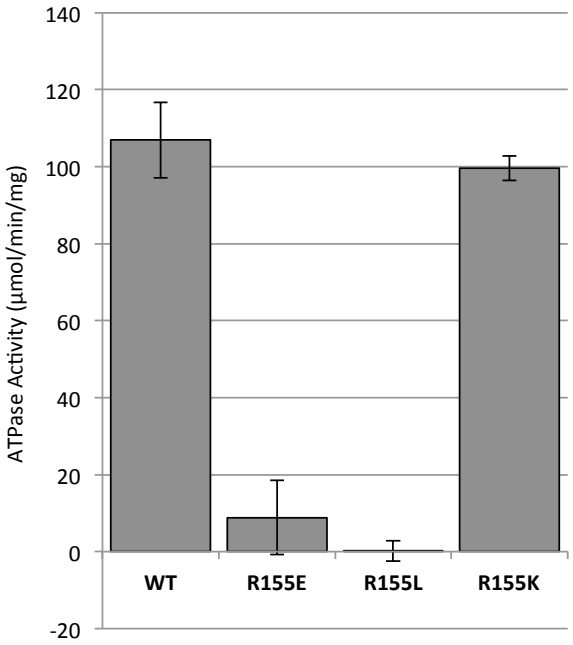

Figure 2.7: ATPase activity of nucleotide-binding site Chy CPN R155 mutant series compared with wild type (WT) Chy CPN. Samples (~0.1 mg/ml protein) were incubated at 65°C for 60 minutes and assayed using the Malachite Green assay as described in section 2.3.6.

All nucleotide binding mutants were assayed together at 65°C. WT and R155 Chy CPN liberated 31.2 (+/-8.9) and 34.3 (+/-5.4) µmol inorganic phosphate respectively, while R155E, R155L, and all D373A variants liberated -3.2 (+/- 6.7) (D373A/R155L) to 3.9 (+/- 10.4) (D373A/R155K) µmol phosphate (Figure 2.8b).

Table 1. Binding affinities and number of binding sites (n) of Chy CPN D373A/R155 series with ATP as determined via Isothermal Titration Calorimetry at 55°C (2.3.9).

Proteins (All D373A)	n	Kd (M)	Ka (M^{-1})
Chy413	16	$2.5*10^{-8}$	$4*10^{7}$
Chy413 R155K	16	$3*10^{-8}$	$3*10^{7}$
Chy413 R155E	13	$5*10^{-7}$	$2*10^{6}$
Chy413 R155L	8	$3.5*10^{-7}$	$2.8*10^{6}$

All ITC was performed using Chy CPN D373A/R155 mutants with ATP as the titrant at 55°C to avoid unnecessarily long equilibration times required at higher temperature. D373A cpn and D373A/R155K both had a calculated n of 16, while D373A/R155E and D373A/R155L had n values below the maximum 16 at 13 and 8 respectively. D373A had Ka (M-1) of $4*10^{7}$ and a Kd (M) of $2.5*10^{-8}$, comparable to D373A/R155K, while D373A/R155E and D373A/R155L had lower Ka and higher Kd values (Table 1).

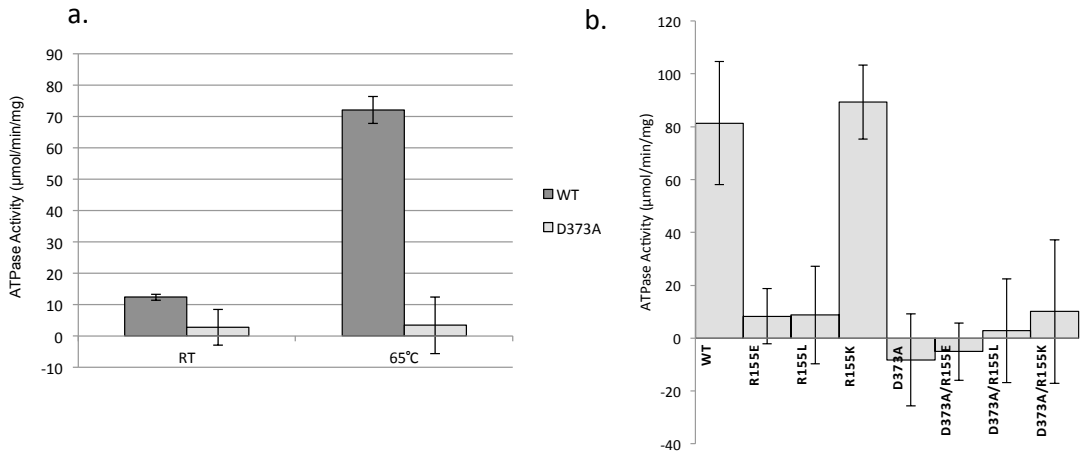

Figure 2.8: ATPase activity of WT Chy CPN and all nucleotide-binding mutants. Left (a.): Chy CPN WT and D373A at room temperature (RT) and 65°C for 60 minutes (final concentration of both 0.063mg/ml). Right (b.): ATPase activity of WT Chy CPN, R155 mutant series, D373A, and D373A/R155 series at 65°C for 30 minutes (final concentration of all protein samples 0.063mg/ml). All ATPase assays performed as described in section 2.3.6.

2.4.3: Thermostability and Stabilizing Interactions of Chy CPNs

Differential scanning calorimetry of Chy_0413 (cpn) and Chy_0807 (GroEL) revealed that the Chy CPN is more thermostable than the GroEL subunit alone. The archaeal-like chaperonin has two major thermal transitions at approximately 83.9°C and 91.5°C, although the raw data suggests that these major peaks may each represent two superimposed

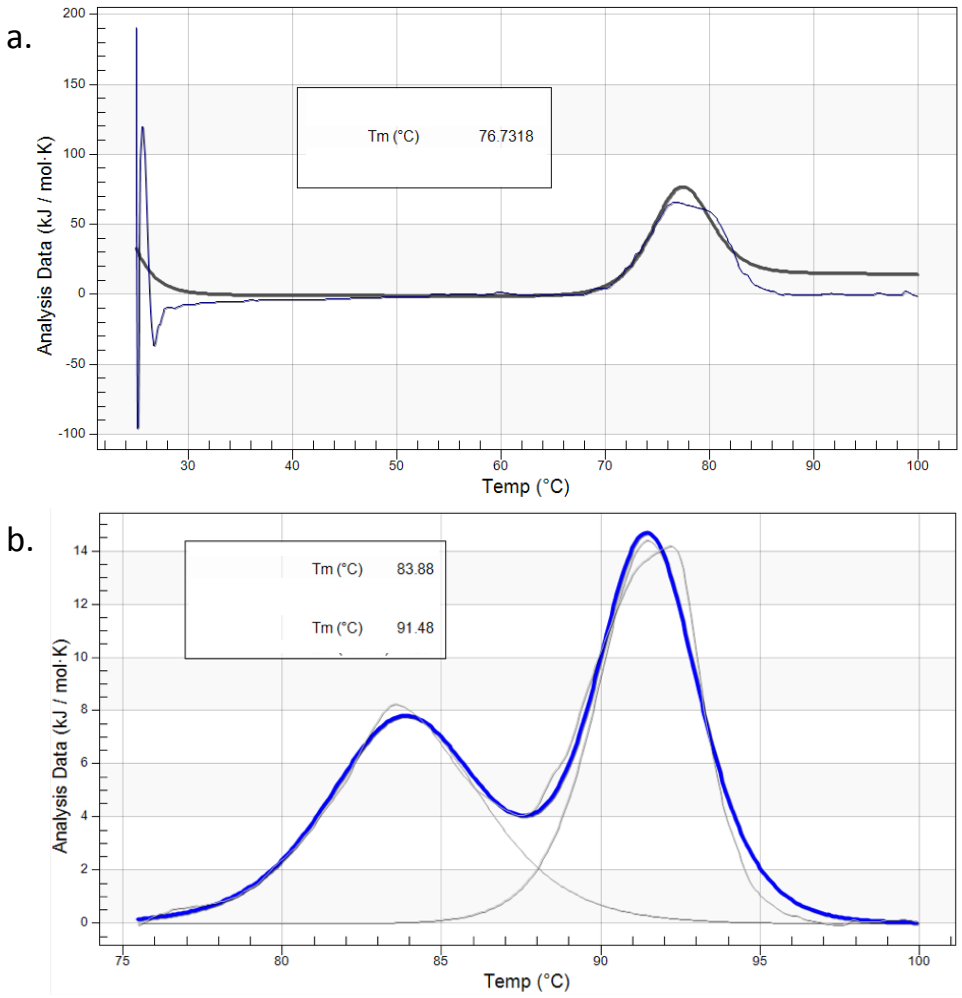

Figure 2.9: The thermostability of Group I and Group III Chy CPNs as determined via DSC. Top (a.): Chy_0807 (GroEL) alone. Bottom (b.): Chy_0413 (archaeal-like CPN). The archaeal-like chaperonin has at least two distinct thermal transistions at approximately 83.9°C and 91.5°C while the GroEL alone has two closely associated transistions at around 76°C and 81°C (the thin lines represent raw data visible under the bold models).

Transitions (Figure 2.9b; raw data in light grey). The GroEL alone has two closely associated transitions at around 76°C and 81°C (Figure 2.9a). A repeated run of Chy_0413 (20µM monomer) in 50mM HEPES-KOH, pH 7.5, 100mM KCl produced two major transitions at 83°C and 90.5°C (Figure 2.13).

Continuous native PAGE (3-15%) was run to determine any deficiencies in assembly of Chy_0413 506* compared with wild type Chy_0413 as previously described in section 2.3.3. Native PAGE of Chy CPN WT and 506* display similar patterns- both can form 16-mer complexes that shift to a single conformation with the addition of ATP (Figure 2.10). There is also an additional faint band directly above the top double band in the 506* lanes, along with another band across all lanes slightly below 1048 kDa.

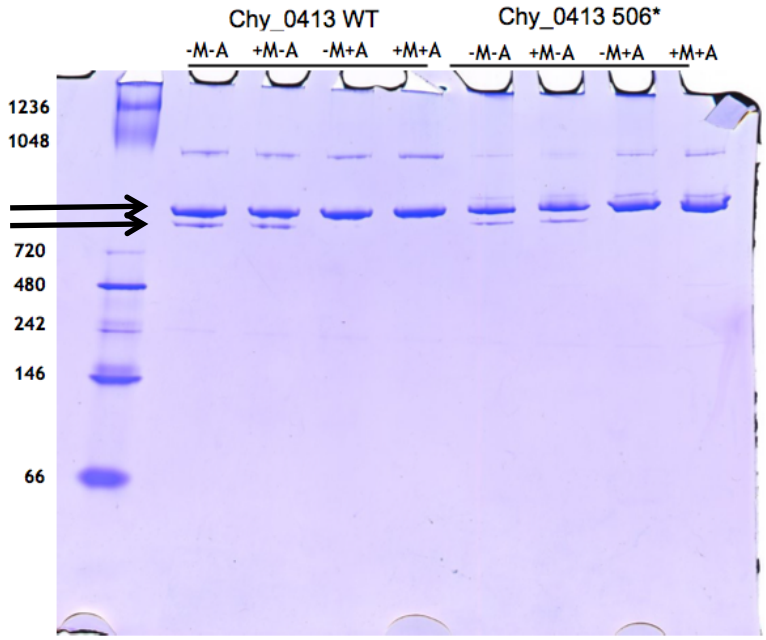

Figure 2.11: Assessment of complex assembly via 3-15% Native PAGE of Chy CPN WT and C-terminal deletion (506*). Samples were incubated at 65°C for 10 minutes prior to loading +/- 1mM ATP (A), 5mM MgCl2 (M) in 10mM HEPES, 20mM NaCl pH 8.0. 15µl of ~0.78µM complex were loaded then run for 5 hours at 100 V. NativeMark Unstained Protein Standard (Novex) is loaded in the first lane.

Malachite green ATPase assays were conducted as previously described in section 2.3.6 to detect any change in thermostability of the 506* mutant versus wild type Chy_0413.

The optimum temperature for the recombinant WT Chy CPN is 65°C as previously reported (Techtmann 2009; Techtmann and Robb 2010). The C-terminal deletion, however, displays virtually no activity at 65°C and only 27% of the ATPase activity of the WT at it's decreased optimum temperature of 49.9°C, liberating only 6.8μM compared with 24.9μM of phosphate produced by the WT cpn during a 45 minute incubation period (Figure 2.11).

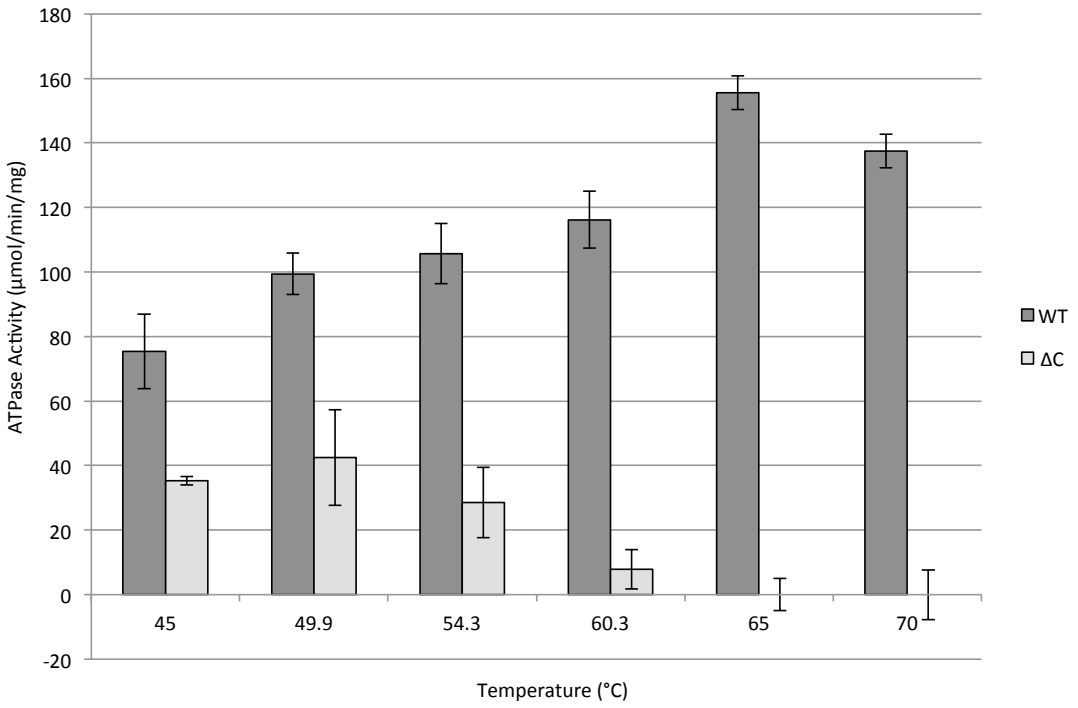

Figure 2.11: Analysis of the thermostability of Chy CPN WT and C-terminal deletion (506*; ΔC) via ATPase assays across temperature. Samples were incubated for 45 minutes with 0.1mM ATP across temperatures 45-70°C and assayed as described in section 2.3.6. All final protein concentrations were 0.06 mg/ml.

Native PAGE was run as described in section 2.3.3 to determine if disruption of this ion pair would yield assembly deficient mutants. Samples were incubated at 60°C for 10 minutes in 50mM HEPES, pH 8.0 with and without 0.1 or 1mM ATP and 5mM Mg2+ prior to loading. Both Chy cpn D41K and D41K/K497A appear to form 16-mer complexes

between 720 and 1048 kDa that respond to ATP in a similar fashion to WT (Figure 2.12). The ATP-induced shift to a single, higher MW complex does not appear to be fully complete with the Chy D41K/K497A CPN. Differential scanning calorimetry of Chy D41K CPN sample in

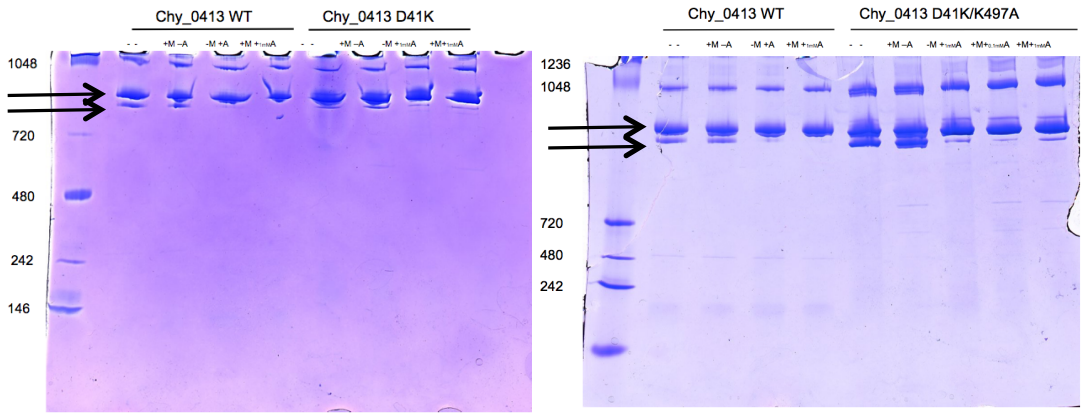

Figure 2.12: Visulatiztion of CPN complex with 3-15% Native PAGE of Chy CPN WT, D41K, and D41K/K497A. Samples were incubated at 60°C for 10 minutes with (+) or without (-) 0.1mM ATP, 1mM ATP (A), and 5mM Mg^{2+} (M) prior to loading. Ladder in lane 1 (unstained native mark; Novex).

50mM HEPES, 100mM KCl pH 7.5 (6.2 mg/ml) was run from 30-100°C. Transitions are apparent at 73.9, 83.7, and 91.0°C (Figure 2.14b). Chy CPN WT at 1/5 the concentration but the same buffer background displays a transition at 83°C as previously observed, and also may have a smaller transition at around 78.0°C (Figure 2.13a).

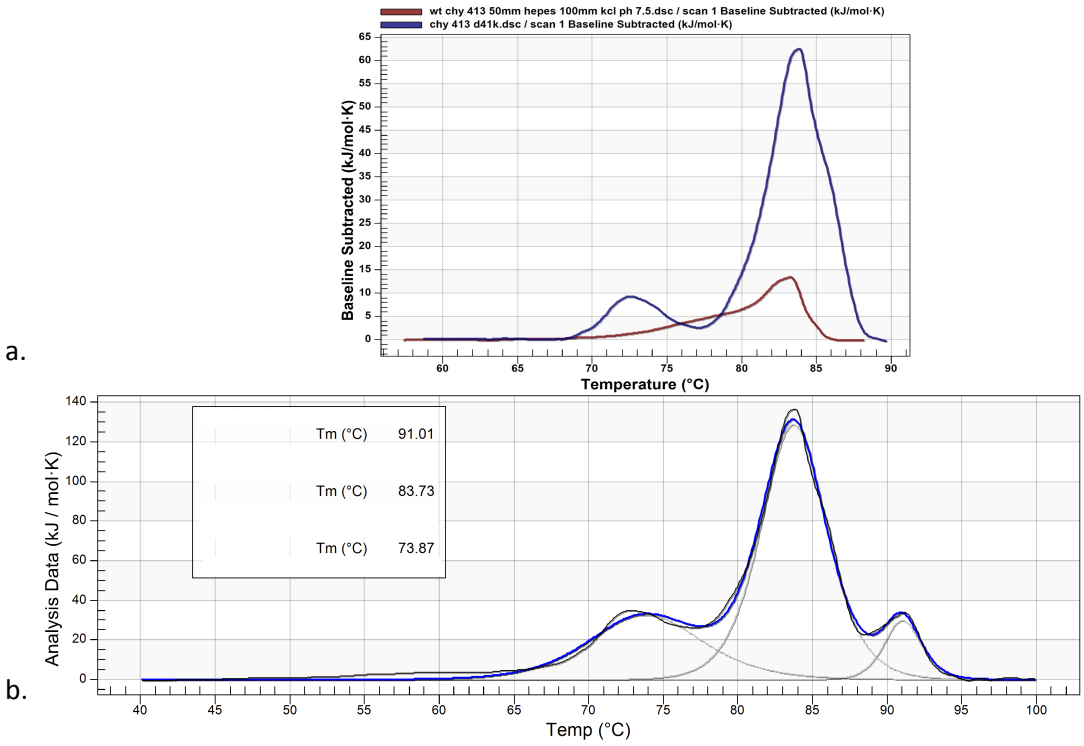

Figure 2.13: The thermostability of Chy CPN D41K as determined by DSC across 30-100°C. Top (a.): Raw data overlay of Chy CPN D41K (blue; 6.2mg/ml) and Chy CPN WT (red; 1mg/ml) in 50mM HEPES, 100mM KCl pH 7.5 shown from 60-90°C. Bottom (b.): Full scan with model (blue) over raw buffer subtracted data (light grey) and transitions indicated at 73.9, 83.7, and 91.0°C.

2.4.4: Complementation of Chy CPNs in SV2

GroEL44 (SV2) is a temperature sensitive strain of *Escherichia coli* K-12 with a mutation in the GroEL gene resulting in E191G at the hinge region of the intermedite and apical domains, inhibiting colony formation above 42°C (Klein and Georgopoulos 2001). As such, this strain is used for GroEL complementation experiments in *E. coli.* This experiment was attempted multiple times- nonpermissive growth temperatures adjusted from 42-44°C, different types of plates were used (YT versus LB; plates with and without IPTG included),

etc. and all results were the same- Chy CPN and Chy GroEL/ES did not complement GroEL E191G in strain SV2. The *E. coli* g*roEL*/pTac24b(+) positive control was the only culture that grew at nonpermissive temperatures, successfully complemented SV2 (Figure 2.14). Expression of chaperonins was visable via SDS-PAGE (not included).

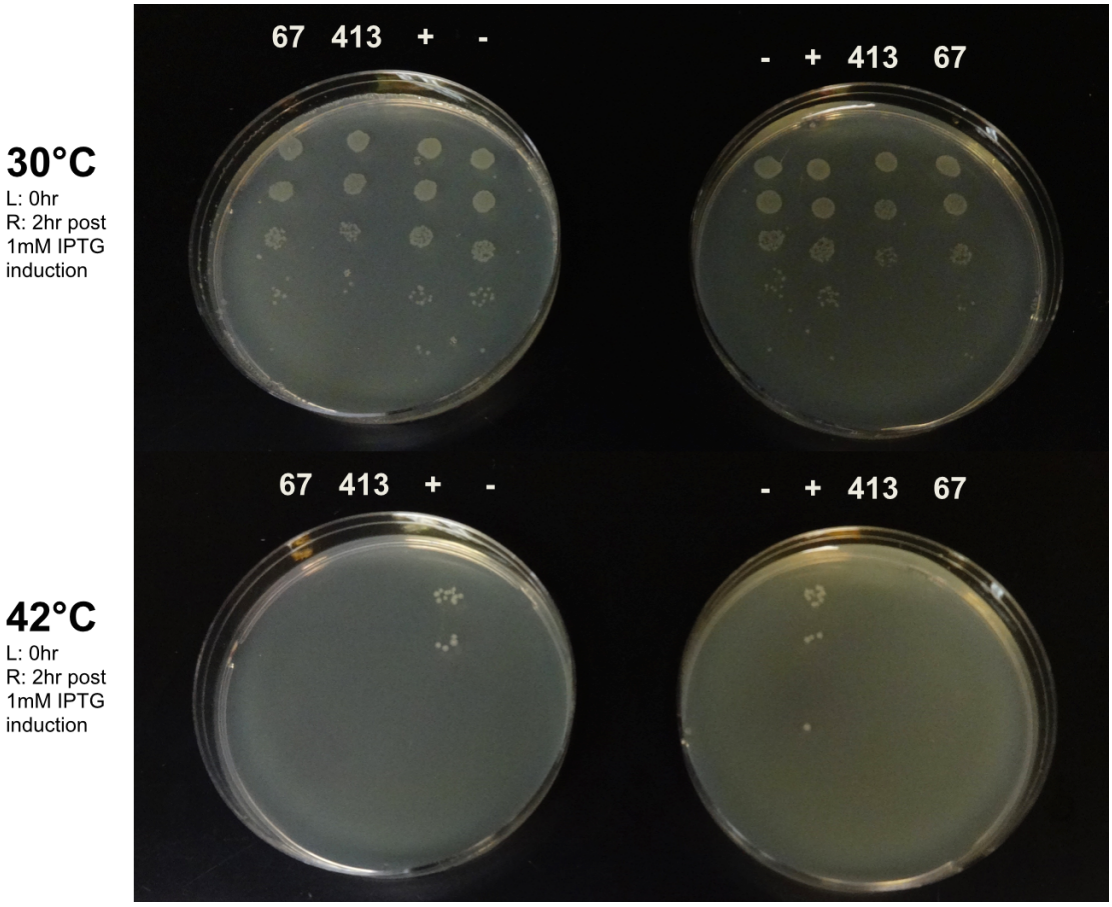

Figure 2.14: Complementation of Chy Group I and Group III CPN in GroEL deficient *E. **coli.*** *E. coli* SV2 transformed with Chy CPN (413), Chy groESL (67), *E. coli* groEL (+), and empty pTac24b(+) (-). Plates on the left were sampled pre-IPTG addition; plates on the right were sampled after 2hour induction at 30°C with 1mM IPTG. Plates on the top were incuabted overnight at 30°C while plates on the bottom were incubated at the retrictive temperature of 42°C.

2.5: Discussion

The open, ADP bound and closed, AMP-PNP associated crystal structures of Chy CPN were obtained at a resolution of 4.0 and 3.0 Å respectively (Figure 2.5; An *et al.* 2016). The subunits of different rings interact with each other in a direct subunit to subunit (one to one) fashion at their equatorial domains similar to group II chaperonins as opposed to the staggered interaction of subunits in separate rings seen with group I chaperonins (Figure 2.15). Overall, there appear to be several short insertions in Group II amino acid alignments compared with Group III chaperonin sequences resulting in no drastically different structural arrangements (Figure 2.15).

The crystal structure complexed with AMP-PNP and the increased resistance of the oligomer to proteinase K digestion when incubated with ATP and AMP-PNP indicate that these nucleotides facilitate at least partial closure of the complex (Figure 2.3b). While there is appreciable protection of the chaperonin in the presence of ATP and AMP-PNP, there appears to be less protection than observed with *Methanococcus maripaludis* Group II chaperonin (Reismann *et al*. 2007; Zhang *et al*. 2010). This may be attributed to differences in exposure of Group II and Group III cpn apical domains or indicative of incomplete closure of the complex. The crystal structure of closed yeast CCT solved with ADP-BeF at 3.8 Å and *M. maripaludis* complexed with AMP-PNP at 3.2 Å (top right inset) versus the *C. hydrogenoformans* cpn-AMP-PNP (3.0 Å) shows a marked difference (Dekker *et al.* 2011; Pereira *et al.* 2010; Figure 2.16). The protruding helices of the yeast CCT apical domains point in towards each other when closed, whereas the group III chaperonin helices point directly in towards the center of the chamber. If in fact the Chy CPN-AMP-PNP structure is fully closed, this may facilitate the interaction and folding of different size range of substrates compared with other chaperonins, and should be investigated in the future.

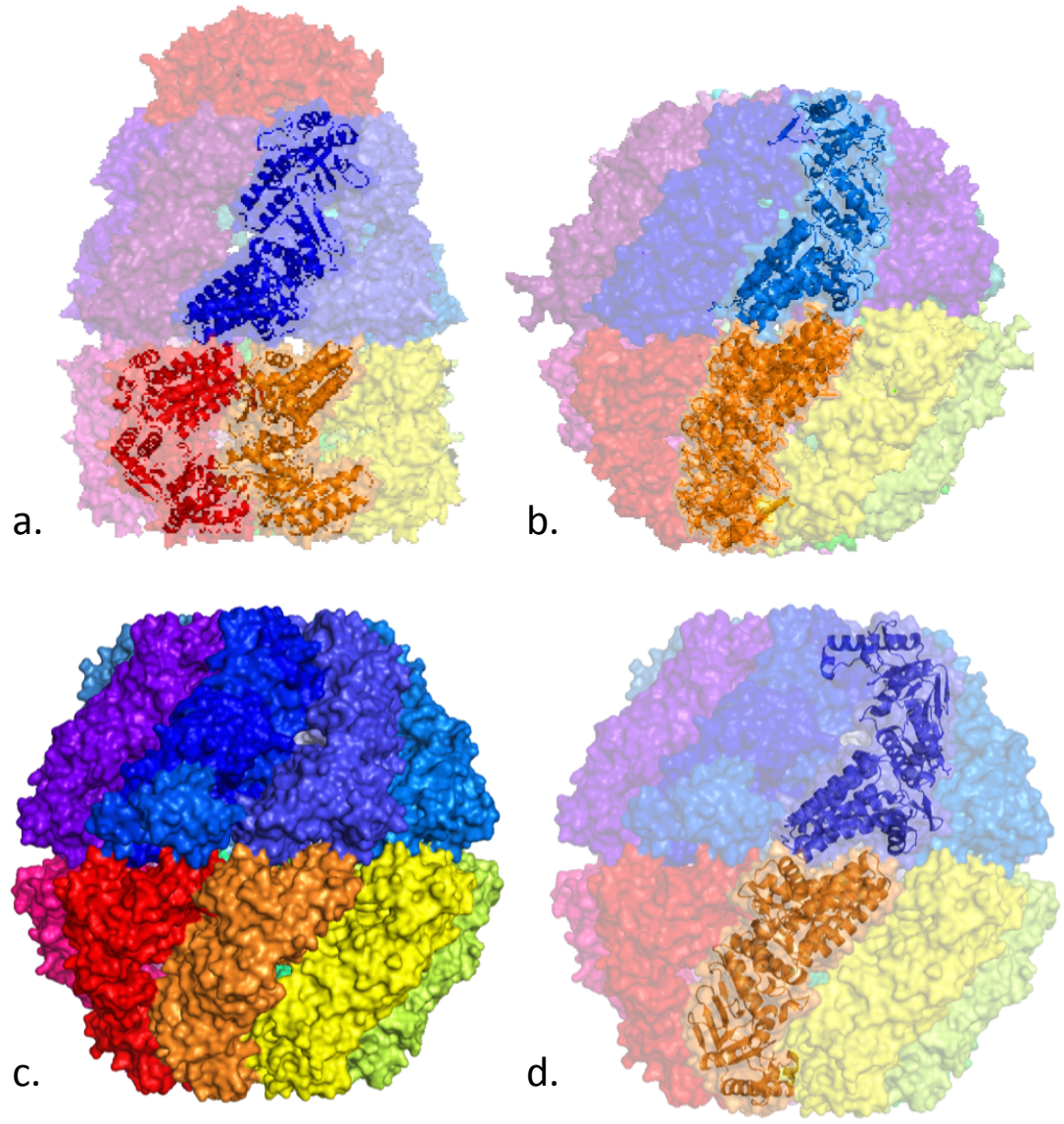

Figure 2.15: Comparison of closed CPNs across groups depicting inter-ring subunit orientation. Top left (a.): *Thermus thermophilus* GroEL/Group I CPN (2.8 Å; PDB 4V4O: Shimamura *et al*. 2004). Top right (b.): *Saccharomyces cerevisiae* CCT/Group II CPN (3.8 Å; PDB 4V81: Dekker *et al.* 2011). Bottom (c., d.): Closed *C. hydrogenoformans* CPN/Group III CPN (3 Å; An et al., 2016). Images generated with PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC).

Purified untagged Chy_0413 forms a 16-mer complex with an estimated oligomeric molecular weight of 905,000 Da. Native PAGE and GF/MALS of Chy_0413 revealed large protein complexes around this molecular weight that appear to change conformation in

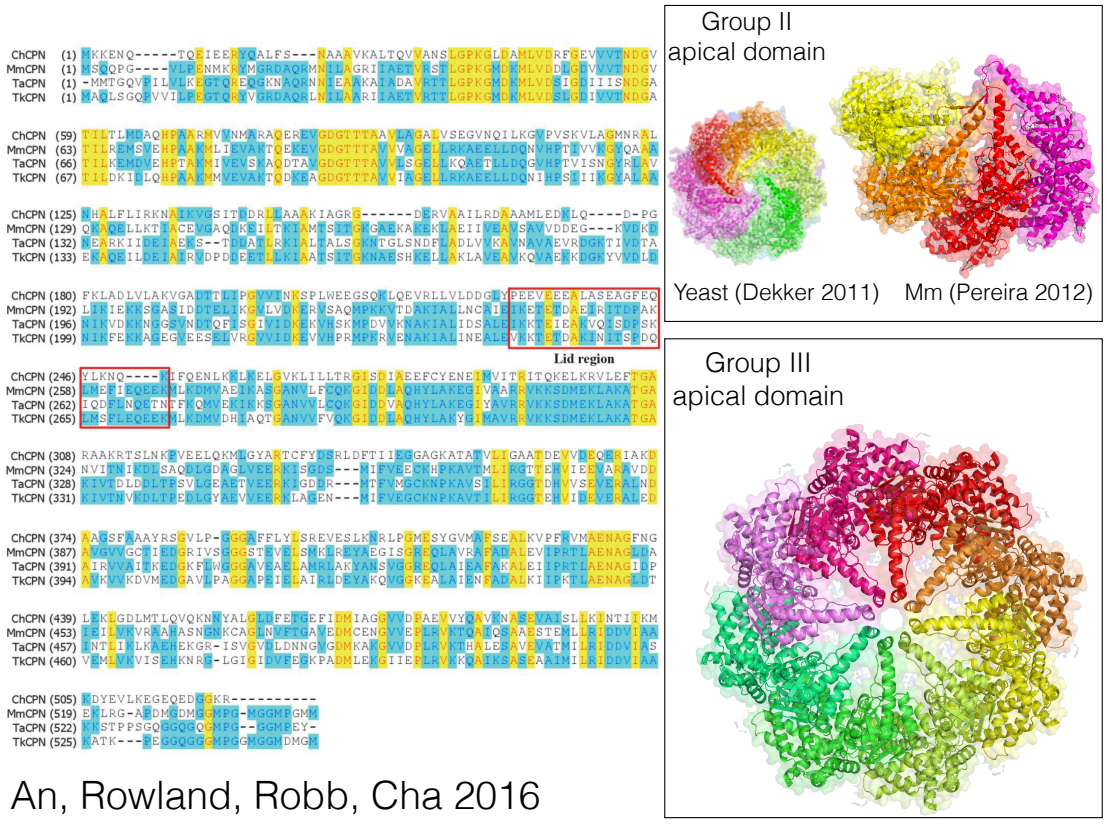

Figure 2.16: Comparison of lid region of Group III Chy CPN with Group II CPNs. Bottom right: Top view of the apical ('built-in lid') domain of *C. hydrogenoformans* chaperonin (Chy_0413-AMP-PNP) crystal structure (An 2016; The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC). Top right: Similar view of *S. cerevisiae* CCT-ADP-BeF and *M. maripaludis* CPN-AMP-PNP (PDB 4V81: Dekker *et al*. 2011; PDB 3KFB: Pereira *et al.* 2010). Left: Amino acid alignment of *Ch* CPN versus sequences of archaeal group II CPNs with available structures including *M. maripaludis* (Mm), *Thermococcus* strain KS-1 (Tk), and *Thermoplasma acidophilum* (Ta) (generated by Cha group; published An *et al*. 2016). Yellow: residues conserved across groups; Green: residues conserved across 2 or more sequences.

response to nucleotides including ATP and the nonhydrolyzable ATP analogue AMP-PNP as previously reported with the tagged protein (Techtmann 2009). Previous GF/MALS run using the N-terminal polyhistadine tagged construct resulted in an estimated MW of 937.3 kDa (Techtmann 2009). The untagged Chy_0413 protein forms a slightly smaller complex at 925.4 kDa. There is an apparent increase in the molecular weight of the chaperonin complex in the presence of ATP and Mg^{2+} (952.8 kDa; Figure 2.4). This shift in weight is also seen on the native PAGE and is likely due to a change in conformation of the complex to a partially

or fully closed conformation. This result is supported by previous proteinase K protection assays in the presence of ATP and AMP-PNP, and by the structure obtained from crystals formed with Chy_0413 and AMP-PNP (Figure 2.3.b; Figure 2.6). Other researchers, however, speculate that this close double banding pattern observed on native PAGE may in fact be a variation in ring symmetry of thermosomes, vacillating from 8 and 9 membered rings of single subunits. The banding observed in the Zhang *et al.* 2013 study, however, maintains two bands in the presence of ATP unlike the response observed wtih Chy_0413, and no analysis of the Chy CPN complex in any condition exceeds the estimated MW of an 18-mer at 1,018.8 kDa. The ATP cycle of the thermosome is still being actively investigated. Many studies indicate that ATP binding facilitates a partial closure of the complex with a 45° counterclockwise rotation of the apical domains. Full closure and releases substrate into the cavity is only achieved after hydrolysis of ATP (Zhang *et al.* 2011; Douglas *et al.* 2011; Nakagawa *et al.* 2014).

Interestingly, the Chy_0413 demonstrates increased thermostability compared with the Chy_0807 (GroEL). At least two major thermal transistions of 83.9°C and 91.5°C were observed for the Chy CPN while the GroEL alone has two closely associated transistions at around 76°C and 81°C (Figure 2.9). Additional DSC studies of archaeal thermosomes demonstrate comparable double and triple transition patterns (personal communication; Spigolon and Robb 2016, unpublished). The highest temperature transition is likely the melt of individual subunits, while the lower temperature transition may represent the melt of the complex. Therefore, Chy_0413 may be able to fold Chy_0807 at temperatures between 76 and 83°C. This could potentially explain the physiological role of the Group III chaperonin in bacteria. It has been previously theorized that the transfer of mesophilic bacterial chaperones to archaea assisted in their transistion to growth at lower tempereatures (Lopez-Garcia *et al.* 2004). It is possible that the group of Firmicutes containing an archaeal-like chaperonin, many of which were isolated from polyextremophilic habitats, then acquired a thermosome

gene from a more thermophilic archaeon to permit maintenance of proteostasis and growth at increased temperatures. This DSC experiment should be repeated using Chy GroEL-GroES, and additional ATPase assays should be performed to compare temperature optimas of the two CPNs in similar conditions.

Despite the putative gamma phosphate interacting site in Group II and Group III chaperonin being completely conserved as a K and R respectively (Figure 2.1), these residues appears to be functioncally interchangeable in regards to ATP cycling (Figure 2.7, 2.8). *M. maripaludis* K161A mutation results in 74% decrease in ATPase activity as well as inhibited ATP binding by ^{32}P-ATP scintillation counts (Pereira *et al.* 2012).

D373A does appear to be a critical residue required for ATP hydrolysis of the Chy chaperonin. Comparable ATPase deficient mutants of archaeal thermosomes include *Mm* D386A and TKS1 D64/D393A (Douglas *et al.* 2011; Nakagawa *et al.* 2014). These residues are responsible for coordinating a water molecule that acts as a nucleophile during hydrolysis of ATP (Douglas *et al.* 2011). D373A, however, does not appear to inhibit ATP binding. Non-similar variations at R155, however, do result in decreased affinity for ATP (Table 1). This result supports the proposed role of this completely conserved residue in Group III chaperonins as being analogous to the conserved *Mm* K161 ATP gamma phosphate interacting site (Figure 2.17; An *et al.* 2016; Pereira *et al.* 2012). Why this residue is a conserved K rather than an R when these basic side chains appear to be functionally interchangeable is unknown. It may have to do with the absence of the nucleotide sensing-loop (NSL) directly after this residue and a potential difference in structural rigidity of this region. The significance of the absence of this structural element in Group III chaperonins has yet to be determined. Arginine residues are slightly bulkier than lysine and may encounter less steric hindrance with the absence of the NSL. If this is the case, it is expected that

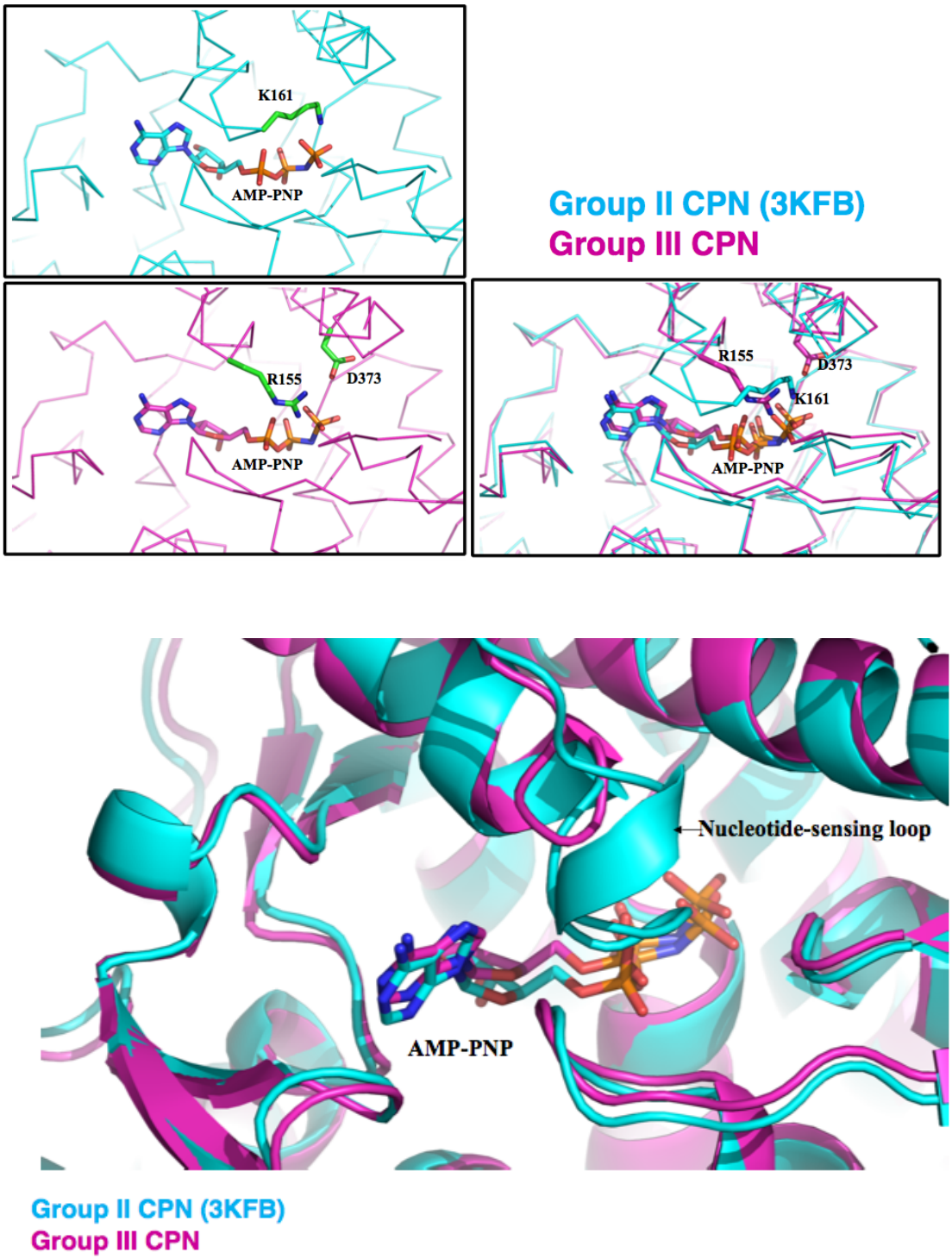

Figure 2.17: Superposition of Group II and Group III CPN nucleotide binding site. Top panels: Comparison of Chy CPN (purple; An *et al.* 2016) and *M. maripaludis* (blue; Pereira *et al.* 2010) nucleotide binding regions with the putative gamme phosphate interacting residue highlighted for each. D373A in Chy CPN is also shown as this residue appears to be key for hydrolytic activity of ATP. Bottom: Superimposed Chy cpn (purple; An *et al.* 2016) and *M. maripaludis* (blue; Pereira *et al.* 2010) structures focused on the nucleotide binding regions. The nucleotide-sensing loop characterized in Mm is absent in Chy CPN (Pereira *et al.* 2012). Figures created by collaborators at KIOST/Cha lab 2016 for upcoming publication.

Mm K161R would likely display decreased ATPase activity and binding affinity. Further, these side chains have a pKa of 12.5 and 10.5 respectively. This variation might be important in maintaining a positive charge depending on the microenvironments in which they operate.

The estimated number of ATP molecules binding per complex indicates that all sites are accessible in a resting state. Thus, the negative allosteric inter-ring model of ATP binding to a single ring while decreasing ATP affinity in an opposite ring of Group I and II chaperonins may not apply to Group III chaperonin cycling (Yifrach and Horovitz 1995; Horovitz *et al.* 2001; Reismann *et al.* 2007).

The last 19 residues of the Chy CPN-AMP-PNP crystal structure are not resolved. While the Chy CPN C-terminus mutant with the last 16 residues deleted does appear to form complexes, they are not as active or as thermostable as WT cpn. This indicates that the last 16 residues of this cpn are involved in forming stabilizing interactions with other subunits of the chaperonin complex as observed with other thermosomes (Zhang *et al.* 2013). It is unclear as to how many residues can de deleted from the C-termius before resulting in assembly deficient subunits, and should be investigated in the future. It is also not known whether or not the termini interact with substrate protein during the refolding cycle as with GroELs, although it highly likely that Chy 506* CPN would exhibit decreased refolding cycling due to its decreased basal ATPase activity (Weaver and Rye 2014). ATPase assays should be repeated with denatured protein added to the reaction to determine if activity exceeds the observed basal activity. *In vitro* refolding assays may be performed, ideally with a thermostable substrate so that refolding rates can be calculated at each protein's optimum temperature for ATPase activity, 65°C for WT CPN and 49.9°C for 506* CPN, for a more meaningful comparison. DSC can also help to determine if the 506* CPN complex disassociates at a lower temperature versus WT Chy CPN.

Analysis of Chy CPN D41 and K497 mutants suggests that the interaction and therefore communication between Chy D41K/K497A subunits in a ring may be impaired

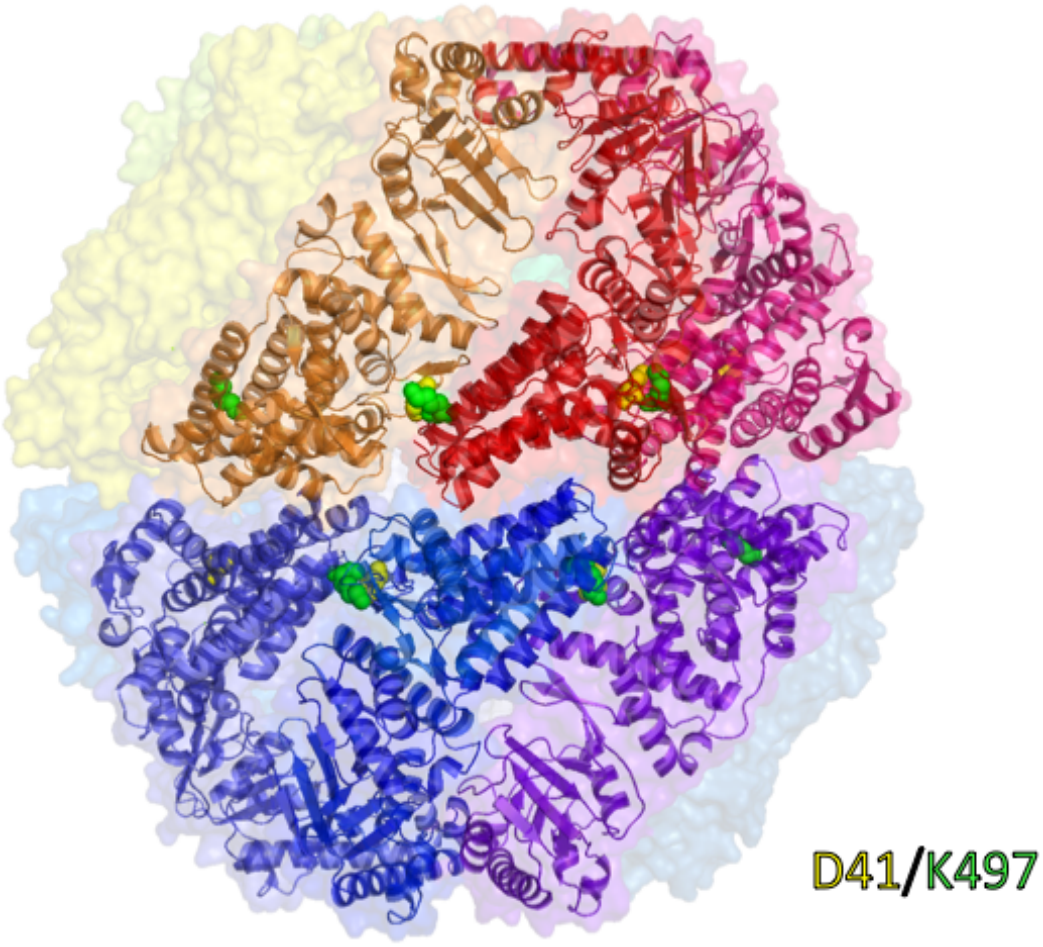

Figure 2.18: Potential intra-ring ion pair interaction D41/K497 in Chy CPN-AMP-PNP. The amino acids are highlighted in the equatorial domain: D41 (yellow sphere) and K497 (green sphere). Figure generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC).

(Figure 2.18). As a result, when ATP binds, subunits are not able to fully close in a coordinated fashion as seen with the persistance of double bands on the native PAGE in the presence of ATP (Figure 2.12). DSC data of Chy CPN D41K revealed a prominent transition at 73.87°C in addition to the two seen with Chy CPN WT (83.73, and 91.01°C) (Figure 2.13, 2.9). The raw data overlay of Chy CPN WT and D41K, both in the same buffer, hints at a third small transition at ~78°C for the WT (Figure 2.13). Chy CPN WT repeated at the same

concentration will help to reveal whether or not this is a transition. Regardless, there does appear to be destabilization of Chy CPN D41K resulting in a transition at a lower temperature of 74°C (Figure 2.13). This prominent third transition of a lower temperature has also been observed with mutants of a homooligomeric Group II CPN (Spigolon and Robb 2016; not published). Additional analysis of the Chy CPN will help to reveal how the subunits of this homooligomeric complex communicate with each other and perform such dynamic rearrangements in the presence of nucleotides and substrate protein.

Given the estimated optimum temperature of Chy CPN and Chy GroESL at 65 and ~60°C as determined by ATPase activity, it is not surprising that there was no observed complementation by these constructs in a GroEL-impaired strain of the mesophilic lab model, *E. coli* (not included; Figure 2.14). It is also possible that the thermophilic chaperonins did not recognize and interact with the *E. coli* proteome, although given the promiscuous characterization of chaperonins and their documented ability to refold nonnative denatured substrate, this seems unlikely. Other studies using SV2 have also only documented complementation with the positive control, including those investigating the function of multiple chaperonins from the mesophile, *Mycobacterium tuberculosis*, likely due to its inability to form complexes when derived from the *E. coli* system (Kumar *et al.* 2009). Given that recombinant Chy CPN has a well documented ability to assemble (Techtmann 2009; Techtmann and Robb 2010; An *et al.* 2016), it seems most likely that the variation in operating temperatures is in fact the main cause of lack of complementation in SV2.

2.6: Future Directions

C. hydrogenoformans groES (Chy_0806), groEL (Chy_0807), grpE (Chy_0414), dnaK (Chy_0415), and *dnaJ (Chy_0416)* were cloned into pTac24b(+)-KanR after being

initially inserted into pSC-A-AmpR/KanR using a StrataClone kit (Agilent Technologies). *Chy_0806* and *Chy_0807* were also cloned into pET33b(+) using NdeI and XhoI restriction sites (see A.1 for primer sequences). Due to the genomic organization of the group III chaperonin within this chaperone gene cluster, it is postulated that these chaperones may form an interactive network *in vivo* (Figure 2.19). Further *in vitro* assays are required to determine if the combination of these chaperones stimulates and increases the refolding rates of denatured substrates.

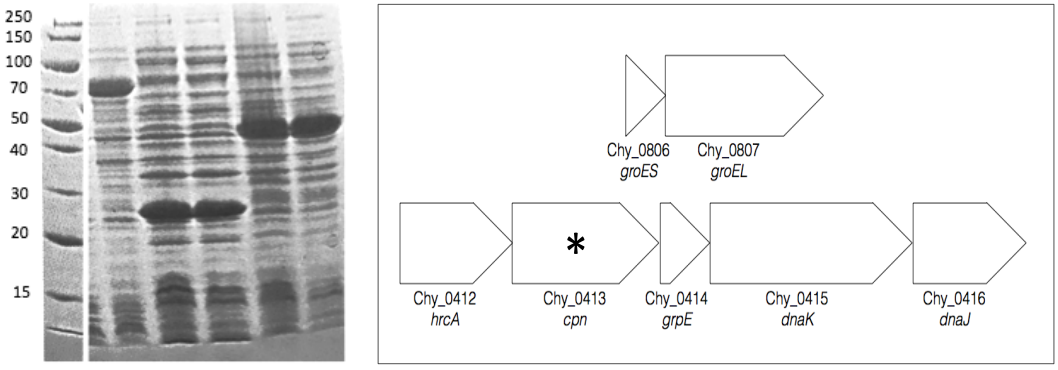

Figure 2.19: The organization and cloning and expression of the chaperones downstream of Group III Chy CPN. Left: Chaperonin gene arrangements in the genome of *C. hydrogenoformans.* Right: SDS-PAGE of Chy chaperone expression trials (image brightness has been adjusted from scanned gel image in laboratory notebook and nonrelevant lanes excised). Lane 1: 8µl PageRuler (kDa; Thermo Fisher Scientific); 2. 5µl Chy_0415 (DnaK)/pTac24b(+), BL21 4hr post-1mM IPTG induction at 37°C; 3. 5µl Chy_0414 (GrpE)/pTac24b(+), BL21 24hr 25°C autoinducation (A.2); 3. 5µl 48hr; 4. 5µl Chy_0416 (DnaJ)/pTac24b(+), BL21 24hr 25°C autoinducation (A.2); 5. 5µl 48hr. All cultures included 50µg/ml of kanamycin and were shaken 180-200rpm. Estimated MW (kDa) of Chy_0414 (GrpE) is 23.24, Chy_0415 (DnaK) is 66.11, and Chy_0416 (DnaJ) is 41.8.

Carboxydothermus hydrogenoformans lacks a gene for prefoldin within its genome. Previous studies have demonstrated direct interaction of prefoldin and archaeal chaperonins (Okochi 2004). In eukaryotic systems, *Bos taurus* Hsc70 and CCT interactions have been previously reported via electron microscopy, while direct contact with bacterial homologues Hsp70 and GroEL has not been observed (Cuellar *et al.* 2008). Mutants based on the results

of this study were produced to investigate the ability of *Chy* group III chaperonin to directly interact with the upstream DnaK/Hsp70.

B. taurus Hsc70 was observed interacting with CCT from V189GAERNV195 (Cuellar *et al.* 2008). A ClustalW amino acid alignment was constructed to determine the analogous position in *C. hydrogenoformans* (Chy_0415), *C. D. audaxviator, B. taurus,* and *E. coli* for putative CCT binding site swaps in *Chy_0415* (Figure 2.20).

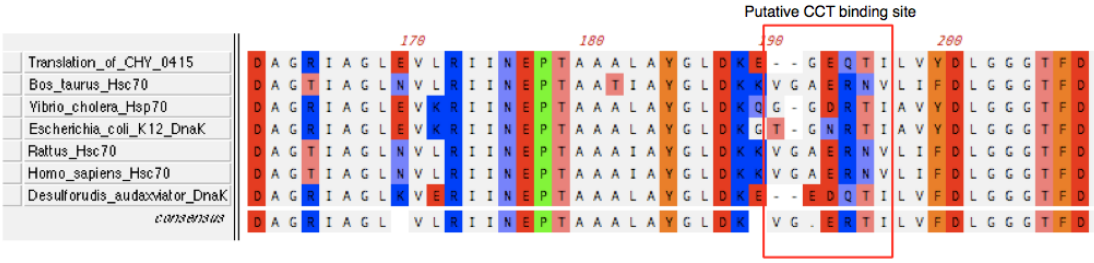

Figure 2.20: Full length DnaK amino acid alignment of Chy DnaK with the organisms featured in Cuellar *et al.* 2008. The red box indicates the putative CCT interaction site across sequences.

The QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) was used to construct all mutants with the primers listed in Table 2. All reverse primers were the reverse and complement the forward primer sequences. The template DNA used in reactions was previously cloned N-terminal polyhistadine tagged WT *Chy_0415* into pET33b(+)-KanR using NheI and XhoI restriction sites. Sequence verified constructs were transformed into *E. coli* BL21 and expressed via autoinduction at 25°C for 72hrs shaking 180rpm in 1L cultures in the presence of 50µg/ml kanamycin. Pellets were spun down at 5,000xg for 5 minutes at room temperature, and then resuspended (5% wt/vol) in 50mM NaH_2PO_4, 300mM NaCl, pH 8.0 buffer containing 2mM EDTA, 1mM PMSF, followed by French Press lysis and centrifugation at 20,000xg for 20 minutes at room temperature. Proteins were purified from the clarified lysate via Protino Ni-TED 1000 kit (Macherey-Nagel). Preliminary Biacore analyses were performed but resulted in considerable inconsistencies. Additional experiments

including CD to confirm proper folding of proteins followed by ITC to determine binding constants of DnaK variants with the group III chaperonin versus the GroEL (Chy_0807) are required to assess any specific interactions with DnaK and CPN III.

Table 2. Sequences of DnaK Putative CCT binding site swaps in *Chy DnaK*.

ORGANISM	FORWARD PRIMER SEQUENCE (5'-3')	RESULTING AMINO ACID SEQUENCE
C. hydrogenoformans	N/A (WT)	$G_{161}EQTI_{165}$
C. D. audaxviator	GCTTACGGCCTTGATAAAGAA**GAGGAC**CAAACT A-TCTTAGTTTACGACCTG	$E_{161D}QTI_{165}$
E. coli	GCTTACGGCCTTGATAAAGAA**ACGGGAAACCGA -**ACTATCTTAGTTTACGACCTG	$T_{185}GNRTI_{191}$
B. taurus	GCTTACGGCCTTGATAAAGAA**GTAGGGGCAGAA -AGGAACGTA**ACTATCTTAGTTTACGACCTG	$V_{189}GAERNV_{195}$

Chapter 3: Characterization of Proteins from an Uncultured Deep Subsurface Bacterium *Candidatus Desulforudis audaxviator* MP104C

3.1: Abstract

Candidatus Desulforudis audaxviator MP104C and its close relatives are important members of the deep subsurface microbial biosphere having been sampled in terrestrial as well as marine environments, and as deep as 2.8km subsurface where it represented over 99.9% of recovered microbes in what is currently the least diverse known ecosystem (Chivian *et al.* 2008; Jungbluth *et al*. 2014). Despite its dominance and ability to thrive independently throughout the deep, *C. D. audaxviator* cannot be continually grown and maintained in the laboratory. Therefore, the majority of insights into its metabolism and growth requirements have been inferred from its genome (Chivian *et al.* 2008). There continues to be increased sequencing of microbial dark matter through single cell genomics, and metagenomic and metatranscriptomic studies which can tell us what organisms are present and what they are doing at a given time point, but provides us with very little biochemical parameters that can be used to inform culturing attempts (Rinke *et al.* 2013). Here we describe the initial characterization of synthesized *Ca. D. audaxviator* (Daud) genes as a proof of concept study using recombinantly produced proteins from uncultivable organisms with known genomes to provide information that can assist in stable, continuous growth in the laboratory. The Group III Daud CPN has a temperature optimum of 50°C and functions at a near neutral pH, supporting previous environmental sampling (Chivian *et al.* 2008). Continued characterization of the Daud carbon monoxide dehydrogenases will reveal insight into how this microbe can thrive alone in the deep.

3.2: Introduction

The majority of microbes evade laboratory-based cultivation. *Candidatus Desulforudis audaxviator* MP104C is a deep subsurface sulfate reducing bacterium that cannot be cultured. Its genome was sequenced and closed while attempting a microbial environmental survey of ~5,600 liters (~60°C) of filtered fracture water from 2.8km deep in the Mponeng South African Gold Mine. Upon sequencing, it was discovered that greater than 99.9% of the genomic material collected originated from a single species, making this the least diverse ecosystem to date (Chivian *et al.* 2008). Relatives of this organism are prevalent in a variety of subsurface environments. Clones in the genus *Desulforudis* accounted for 17% of the Firmicute-related 16S rRNA sequences sampled from a 147m deep subseafloor borehole (39°C) into a 1.24 million year old aquifer (Jungbluth *et al.* 2014). In a gold mine neighboring Mponeng in Johannesburg, Evander Au Mine, the microbial composition of flowing subterranean boreholes ranging from 1.88-2.28km and a temperature of ~45°C was monitored for 3 months (Davidson *et al.* 2011). *Firmicutes*, *Methanosarcinales*, and *Thermoproteacaea* dominated these samples, and the *dsrAB* sequences isolated by metagenomic analysis had greater than 99% similarity to *C. D. auxadviator*. Abiogenic methane composed 50% or greater of the methane analyzed by stable isotope MS and the aquifer was at high pressure favoring sulfate reduction over methanogenesis in this environment by organisms related to the genera *Desulfotomaculum* and *Desulforudis*. Thermodynamic analysis revealed that sulfate reduction coupled with methane oxidation, and acetogenesis and methanogenesis from CO were generally the most favorable predicted reactions (Davidson *et al.* 2011). In another deep fracture of the Witwatersrand Basin, South Africa, metagenomic analysis from fracture fluids collected from 3km depth revealed that the Wood-Ljungdahl/ reductive acetyl-CoA pathway is the most abundant carbon fixation

pathway of the six known pathways, with 57.4% and 22.3% of single copy protein encoding genes from *Firmicutes* and *Euryarchaeota* respectively (Magnabosco *et al.* 2016; Berg 2011).

The genome of *C. D. audaxviator* contains genes thought to be of archaeal origin including a sulfate adenylytransferase (*sat*) and an additional CODH (*Daud_0105*; *cdhA*) as well as the Group III chaperonin (*Daud_2059*). A recent single cell genomics study of 5 *C. D. audaxviator* cells suggests that despite low cell abundance and a presumed long generation time, this organism undergoes horizontal gene transfer and is infected with viruses at an unusually high frequency (Labonte *et al.* 2015). Further, models based on biological racemization rates of planktonic isolates from the same borehole from which *C. D. audaxviator* genomic DNA was sampled in 2008 predict an amino acid turnover at 3km (54°C) of only 1-2 years, suggesting that subsurface communities are much more active than previously predicted (Onstott *et al.* 2014).

The cohabitation of *C. D. audaxviator* or a related sulfate reducing bacterium with methanogens likely explains its acquisition and retention of several archaeal-like genes. While there is a healthy initiative to sequence microbial dark matter from environmental samples, biochemical and biophysical studies supporting these sequence results are lacking (Rinke *et al.* 2013). To learn more about the putative growth physiology of this unculturable organism and inform future attempts at continued laboratory growth, the Group III chaperonin of *C. D. audaxviator* was synthesized and characterized.

Many Group III chaperonin-containing organisms are also CO-utilizers including *C. hydrogenoformans* and *C. D. auxdaviator*, with roughly 6% of all sequenced bacterial and archaeal genomes containing one or more copies of a carbon monoxide dehydrogenase (Techtmann *et al.* 2012). Evidence of horizontal gene transfer of CODHs indicates that the utilization of CO confers a strong selective advantage in many environments including the energy limited deep subsurface where these enzymes facilitate energy conserving schemes and carbon fixation via the Wood-Ljungdahl pathway (Chivian *et al.* 2008; Techtmann *et al.*

2012; Sant'Anna *et al.* 2015; Magnabosco *et al.* 2016). *C. D. auxdaviator* is unique in possessing a bacterial (*Daud_0870*) as well as an archaeal type (*Daud_0105*) CODH catalytic subunit most similar to a *Dehalococcoidia* bacterium (59% identity) followed by the anaerobic, thermophilic methanogen *Methanothermobacter thermautotrophicus* (53% identity; Chivian 2008; Zeikus 1972; Wasserfallen 2000). This methanogen and related species have close association with bacteria, in particular *Firmicutes*, as evidenced by their acquisition of the bacterial DnaK/Hsp70 system and experimentally validated growth as syntrophic co-cultures (Popp 2009; Ishii 2005; JGI IMG BLASTp results include scores of 762/1198 *C. Nitrosofontus exaquare*, 774/1198 *C. Nitrososphaera evergladensis Ga9-2*, and 759/1198 *T. acetatoxydans Re1*, 726/1198 *C. D. audaxviator* all with e values of 0.0). Further, *Firmicutes* and methanogenic *Euryarchaeota* dominate metagenomic samples from deep subsurface fracture fluids, unequivocally validating mutually productive cohabitation (Magnabosco *et al.* 2016).

Anaerobic bacterial (CooS) and archaeal (Cdh) CODHs both contain [Ni, Fe] active sites, although their sequences share low homology (Techtmann *et al.* 2012; Svetlichnyi *et al.* 2004). Structurally, Cdh contain two [4Fe-4S] clusters absent in CooS subunits (Gencic *et al.* 2010). These enzymes catalyze the water-gas shift reaction, or the reversible oxidation of CO to CO_2 ($CO + H_2O = CO_2 + 2e^- + 2H^+$), which can be coupled to many different metabolic functions. In *C. hydrogenoformans*, an organism with five distinct CODHs, linked functions range from reducing equivalents generating NADPH via a ferredoxin: NADP+ oxidoreducatse, and H_2 via a potentially H^+ translocating hydrogenase (Svetlichnyi *et al.* 2001). Other microbes can use these reducing equivalents in autotrophic and heterotrophic metabolic strategies for metal and sulfate reduction, as well as CO_2 reduction to CO for acetate production and use in the Wood-Ljungdahl pathway (Ragsdale 2008).

Both the *cooS* and *cdhA* genes from *C. D. auxdaviator* were synthesized for *in vitro* characterization to learn more about the CO metabolism of this unculturable microbe, and by

extension carbon transformations in deep subsurface commmunities. Accessory proteins including CooC, a Ni-inserting ATPase, play a critical role in the proper maturation of these anaerobic CODH active sites both in the native organism and during recombinant CODH production (Jeoung *et al.* 2009; Inoue *et al.* 2014). Two different *cooC* genes were cloned from *C. hydrogenoformans* for co-expression trials aimed at generating active Daud CODHs.

3.3: Materials & Methods

3.3.1: Gene Synthesis

The *C. D. audaxviator* group III chaperonin gene (*Daud_2059*) was synthesized and cloned into pUC57 with SacI at 5' end and XbaI site at the 3' end (GenScript). *Daud_2059* was amplified off of the cloning vector with primers introducing restriction sites of NdeI at the 5' end and Bpu1102 I at the 3' end for introduction into pTac24b(+) for expression. *Daud_2059* was also cloned into pET33b(+) at NdeI 5' and HindIII 3' to introduce a N-terminal His-tag Daud_2059 product.

3.3.2: Expression & Purification

Sequence verified *Daud_2059* in pTac24b(+) was transformed into chemically competent BL21 (Novagen). A 1L autoinduction (A.2) containing 50µg/ml of kanamycin was inoculated with a single colony, and shaken at 25°C, 200rpm for 72 hours. Cells were harvested at 5,000xg for 5 minutes at 4°C, and then resuspended in 20mM HEPES pH 8.0-KOH, 100mM KCl, 1mM PMSF (5% wt/vol) and lysed via French Press at 1,000 PSIG. Lysate was centrifuged down at 20,000xg for 20 minutes. Daud_2059 was salted out of the supernatant with 30% ammonium sulfate (Sigma), centrifuged at 10,000xg for 10 minutes,

solubilized in lysis buffer, and dialyzed against fresh lysis buffer overnight at 4°C on a stir plate. The supernatant was then loaded onto an anion exchange High Q column (Bio-Rad) and eluted over a gradient up to 1M KCl. Pure fractions were pooled and concentrated via Ultra-15 30K cutoff centrifuge concentrators (Amicon). *Daud_2059* in pET33b(+) was induced as described above and lysed (10% wt/vol) via French Press at 1,000 PSIG in 20mM NaH_2PO_4 150mM NaCl, pH 7.2-NaOH, 0.1mM PMSF and 2mM DTT. Lysate was clarified at 20,000xg for 20 minutes and then loaded onto an anion exchange High Q column (Bio-Rad) and eluted over a gradient up to 1M NaCl at a flow rate of 5ml/min and a 1-minute fraction collection. Pure fractions were pooled and concentrated as described. Protein concentrations of samples were determined using the Coomassie Plus Bradford Reagent read at 595nm (Pierce).

3.3.3: Native PAGE

Native PAGE was prepared as described in section 2.3.3. Purified preparation of 1.1µM Daud_2059 (complex; untagged) in 50mM HEPES-KOH pH 7.5, 50mM KCl was incubated for 30 minutes at temperatures ranging from 4-100°C, then clarified at 11,000xg for 2 minutes. Supernatants (17.5µl) of with 2X native sample buffer were loaded into the wells with NativeMark Unstained Protein Standard (Novex) in the first lane.

3.3.4: Differential Scanning Calorimetry

DSC was performed using the same equipment, software, and general methods as described in section 2.2.4. 0.7µM of Daud_2059 (complex) in 20mM NaH2PO4-NaOH, pH 7.4, 100mM NaCl was heated at 45-50°C for 15 minutes, then centrifuged at 11,000xg for 5

minutes. The supernatant was degassed, loaded, and run as previously described. 0.7μM Daud_2059 (complex) in 50mM acetate, 100mM NaCl pH 5.0 at 45°C was also prepared and run to determine effects of pH on complex stability.

3.3.5: MDH Refolding Assays

The activity of porcine heart malate dehydrogenase was measured as an indirect assay of chaperonin refolding capabilities. 7μM malate dehydrogenase was incubated for 20 minutes at 37°C in 25mM HEPES-KOH pH 8.0, 300mM KCl, 1mM MgCl2, 5mM DTT, and 2M guanidine-HCl. This reaction was then diluted 1/100 into 0.5M ammonium sulfate, 5mM ATP (final concentration). 74.5nM of Daud_2059 complex was incubated with denatured MDH at 37°C. 10μl of the reaction was pulled at various time points and mixed with 315μl of 90mM HEPES-KOH, pH 8.0, 0.22mM B-NADH, and 0.55mM oxaloacetate in a temperature controlled cuvette at as described in Luo 2011. MDH activity was monitored at 340nm over 1 minute in Beckman DU-640 spectrophotometer. Results were analyzed by comparing the amount of activity recovered in the chaperonin-dMDH reaction compared to dMDH and native MDH alone. Reagents were purchased from Sigma unless otherwise noted.

3.3.6: ATPase Assays

The hydrolysis of ATP by Daud_2059 complex was assayed using the Malachite Green Assay as described in section 2.3.6. Heat-treated supernatant (untagged; incubated clarified lysate at 55°C for 30 minutes and then spun at 20,000xg for 20 minutes) with 2mM ATP and 0.08mg/ml of purified N-polyhistadine tagged Daud_2059 with 0.1mM ATP were assayed across temperatures ranging from 30-70°C after 30minute incubations.

3.3.7: Gel Filtration/Multi-angle Light Scattering

Gel filtration/multi-angle light scattering was performed using the same instrument, software, and general methodology as previously described in section 2.3.4 to verify 16-mer complex formation of Daud_2059. 20µl of 2µM Daud_2059 (untagged complex; 8.98 mg/ml) was injected into the column after 30 minute incubation at 50°C with 2mM ATP and 1mM $MgSO_4$ followed by centrifugation. 20µl of N-histadine tagged Daud_2059 (9.9mg/ml) was injected after 10 minutes at 45°C with no nucleotides followed by centrifugation for 2 minutes at 11,000xg. All samples were initially buffer exchanged using 10kDa cutoff centrifugation concentrators into column equilibrated buffer consisting of 100mM $NaSO_4$, 80mM $NaPO_4$ pH 6.7, 0.05% sodium azide. I received the assistance of Dr. Travis R. Gallagher of the National Institute of Standards and Technology/Institute for Bioscience & Biotechnology Research throughout these experiments.

3.3.8: Daud Protein Modeling

Daud_2059 (CPN), Daud_0105 (CdhA), and Daud_0870 (CooS) wild type amino acid sequence was input into the Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) using normal modeling mode (Kelley *et al.* 2015). The models generated of Daud proteins were visualized using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC.)

3.3.9: Cloning, Expression, and Purification of Daud CODHs: Daud_0105 and Daud_0870

Daud_0105 (*cdhA*) was synthesized and inserted into pBSK(+) (*ampR*) at NdeI/XhoI restriction sites (Biomatik), and Daud_0870 (*cooS*) was synthesized and inserted into pUC57 (*kanR*) at NdeI/XhoI as well (GenScript). Both constructs were subcloned into pTac24b(+) (*kanR*) and pET33b(+) (*kanR*). Chy_1221 (*cooS*), Chy_1220 (*cooC*), and Chy_1225 (*cooC*) were amplified from *C. hydrogenoformans* genomic DNA as described in section 2.1.1. Chy_1221 was inserted into pTac24b(+), while Chy_1220 and Chy_1225 were inserted into pCDF-1b-SmR (Novagen) at NcoI/XhoI restriction sites for CODH co-expression trials. CODH and CooC constructs were co-transformed into BL21 (Novagen). Colonies were grown in LB or YT media supplemented with 50µg/ml of kanamycin and spectinomycin and 0.5% glucose. Aerobic cultures were grown at 37°C to an OD600 of 0.4-0.6 and induced with 1mM IPTG (American Biomedicals), then shifted to 30°C for 4 hours shaking at 190rpm before harvesting cells. Anaerobic cultures were initially grown aerobically at 37°C to an OD600 of 0.6-0.7, then switched over to a 100% N_2 atmosphere. After 30 minutes shaking 190rpm at 30°C, cultures were induced with 0.5mM IPTG, 0.5mM $NiCl_2$, 1mM $FeSO_4$, 50mM KNO_3, 2mM L-cysteine (Inoue *et al.* 2014). Mature cultures were spun down at 5,000xg for 5 minutes. Cultures of Daud_0105 and Daud_0870 in pET33b(+) were lysed in 50mM NaH2PO4 pH 8.0-NaOH, 300mM NaCl, clarified, and then loaded onto a NiTED 1000 column from the Protino kit and eluted along a gradient up to 250mM imidazole (Macherey-Nagel). CODH pTac24b(+) cultures were lysed and clarified to assess solubility and expression. No attempts at purification were made. Daud_0105 was solubilized by suspending cell pellets with 100mM NaH_2PO_4 pH 8.0-NaOH, 150mM NaCl, 10% glycerol, 5mM CHAPS, 2mM DTT, 1mM EDTA, 0.1mM PMSF (5%wt/vol) and lysing via French Press at 1,000 PSIG. 1% sarkosyl was added to the lysate that was then placed on a rocker for

14hours at room temperature. After the sarkosyl rocking incubation, lysate was centrifuged for 20 minutes at 20,000xg.

3.3.10: Phylogenetic Analysis of Daud CODHs

The 106 intact carbon monoxide dehydrogenase amino acid sequences currently in the databases including representatives from archaeal (Cdh) and bacterial (CooS) subunits, extracted from DOE JGI and NCBI databases, were aligned using ClustalW in MacVector (default settings; MacVector, Inc.). The alignment was imported into MEGA7 for molecular phylogenetic analysis by the maximum likelihood method (Jones *et al.* 1992; Kumar *et al.* 2015).

3.4: Results

3.4.1: Daud CPN (Daud_2059) Results

Daud_2059 overexpressed readily in the heterologous expression host, *E. coli* BL21 (Novagen; Figure 3.1). Daud_2059 has an estimated molecular weight 55.1 kDa of and an isoelectric point of 5.08 as calculated by ExPASy pI/MW Tool (Gasteiger *et al.* 2005). Purified Daud_2059 forms a complex between 720 and 1,048 kDa as was observed with Chy CPN (Figure 3.2; Figure 2.3a). The estimated molecular weight of the 16-mer is 881,600 Da. GF/MALS using purified untagged Daud_2059 in the presence of 2mM ATP and 1mM $MgSO_4$ revealed a major peak with a predicted MW of 848.8 kDa (Figure 3.3a), while N-His tagged Daud_2059 revealed two major peaks around 448 kDa and 784 kDa (+/-10%; Figure

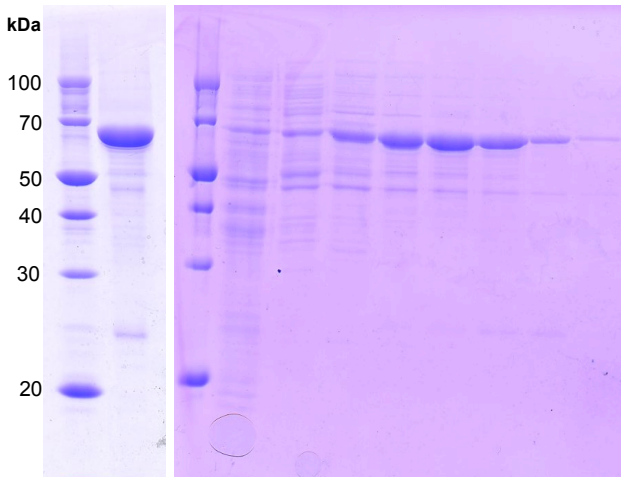

Figure 3.1: Expression and purification of Daud CPN from *E. coli*. 15% SDS-PAGE of Daud_2059/pET33b(+) purification process after expression as described in 3.3.2. Left: Pooled clean Q fractions resulting in 9.9mg/ml solution of N-His tagged Daud_2059 used in GF/MALS experiment (2µl loaded of 2x sample buffer preparation). Right: (left to right): Flow through and wash of Daud_2059 N-H Q run, fractions 6, 8, 10, 12, 14, 16, and 18 over a 20 minute gradient of 150mM to 1M NaCl with 1-minute fraction collections. Protein markers in both gels PageRuler Unstained Broad Range Protein Ladder (ThermoFisher Scientific).

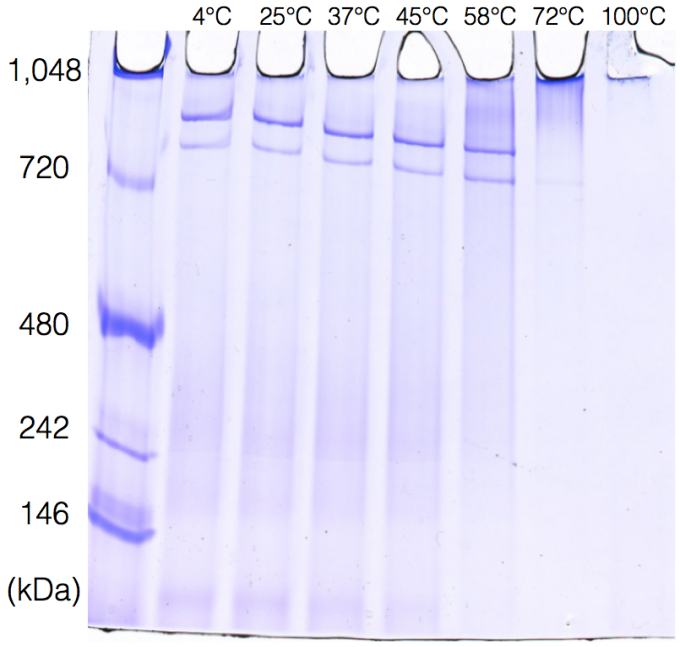

Figure 3.2: Visualization of complex assembly and thermostability of Daud CPN with 3-15% Gradient Native PAGE. 1.1µM of Daud_2059 complex was incubated for 30 minutes at temperatures ranging from 4-100°C, centrifuged for 2 minutes at 11,000xg (room temperature) and loaded with 2x native loading dye. NativeMark Unstained Protein Standard (life technologies) is loaded in the first lane.

3.3b). Precipitate appears to form at 58°C and complexes start to disappear from solution at 72°C. At 100°C no bands are observed on the gel (see Figure 3.2).

Major thermal transitions are observed for purified Daud_2059 (6µM monomer) at 62.2°C and 74.0°C in 20mM NaH_2PO_4-NaOH, pH 7.4, 100mM NaCl. In 50mM acetate, 100mM NaCl pH 5.0 at 45°C, Daud_2059 (6µM monomer) appears to have one or two transitions at 70.7°C with a shoulder 74°C (Figure 3.4). These double transitions are consistent with observed DSC patterns for Chy_0413 (inset; Figure 3.4). ATPase data across

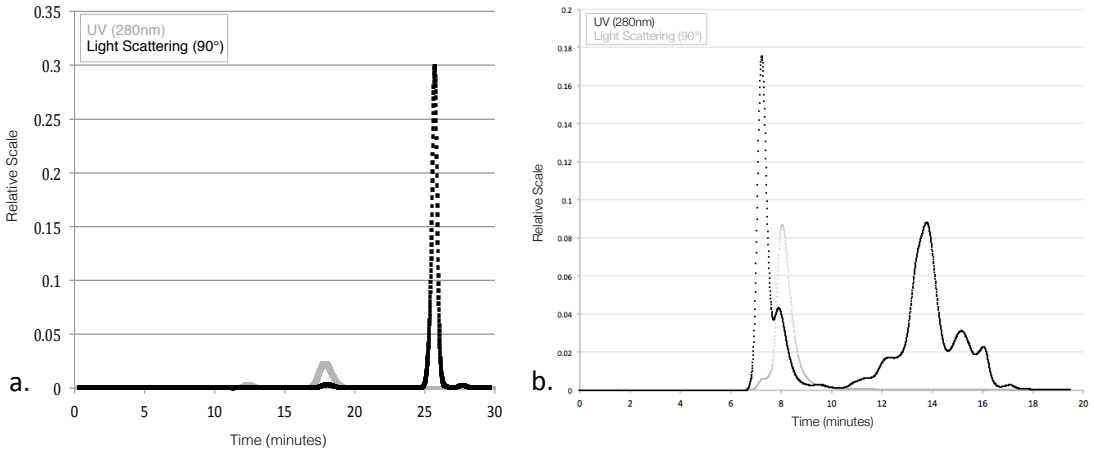

Figure 3.3: Size determination of Daud CPN using GF/MALS. Two different preparations of purified Daud_2059 were analyzed. Left (a.): Untagged Daud_2059 after ammonium sulfate purification preparation as described in section 3.3.2. The sample was incubated with 2mM ATP and 1mM $MgSO_4$ at 50°C for 30minutes followed by centrifugation of 11,000xg for 2minute. An injection of 20µl of this 9mg/ml protein solution resulted in a major peak of 848.9kDa. Right (b.): N-terminal His-tagged Daud_2059 (Figure 3.1) was heated for 10 minutes at 45°C, clarified as described in (a.) and injected at 10mg/ml resulting in major peaks at 448kDa and 784 kDa (+/-10%). Both runs used samples buffer exchanged with the column buffer 100mM $NaSO_4$, 80mM $NaPO_4$ pH 6.7, 0.05% sodium azide.

temperatures ranging from 35°C to 70°C using purified protein (open circles, Figure 3.5 sample at a final reaction concentration of 0.08mg/ml) and heat-treated Daud_2059

supernatant (closed circles, Figure 3.5) prepared as described in section 3.3.6 results in a temperature optimum of 50°C for both Daud_2059 protein preparations.

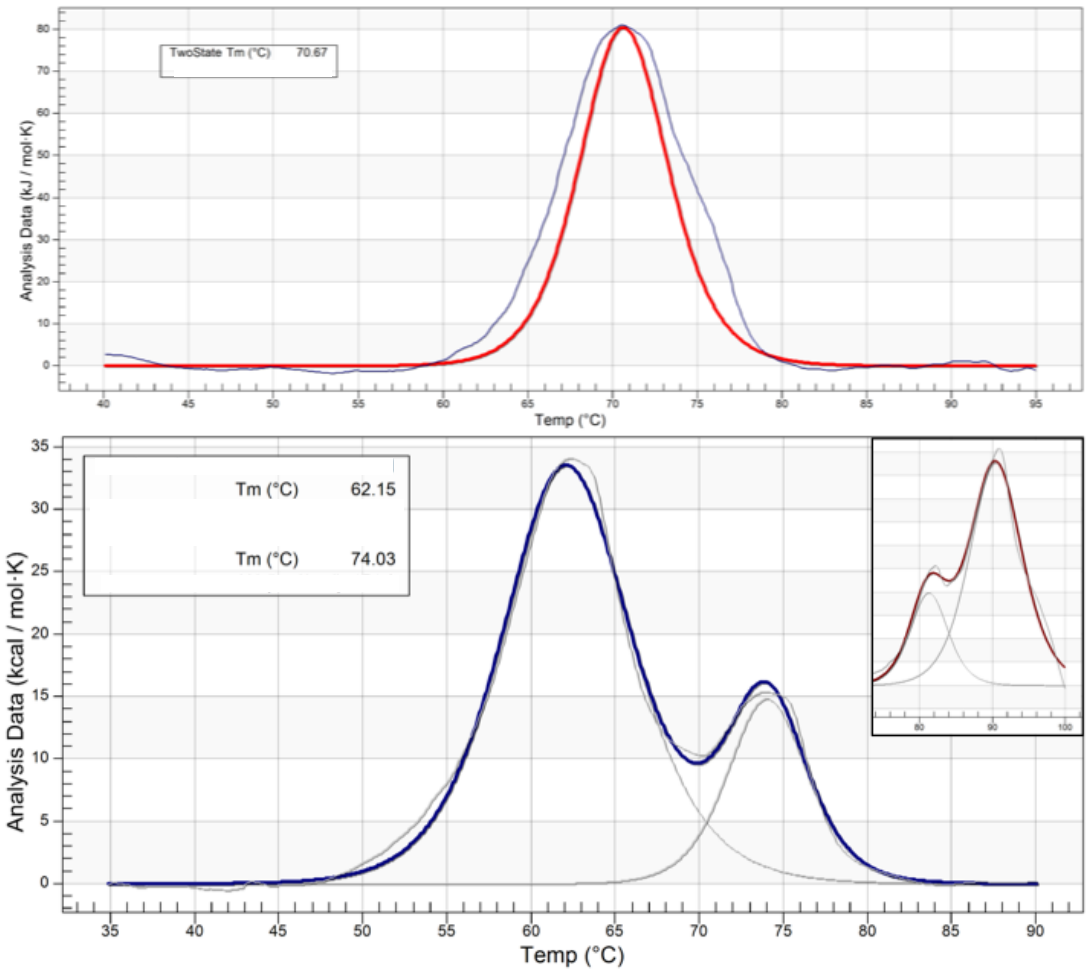

Figure 3.4: Thermostability of Daud CPN as determined by DSC. Top (red model): 6µM (monomer) of Daud_2059 run up to 100°C in 50mM acetate, 100mM NaCl pH 5.0 at 45°C. Bottom (blue model): 6µM (monomer) of Daud_2059 heated to 100°C in 20mM NaH_2PO_4-NaOH, pH 7.4, 100mM NaCl. Bottom inset (dark red): 20µM (monomer) of Chy_0413 in 50mM HEPES-KOH, pH 7.5, 100mM KCl.

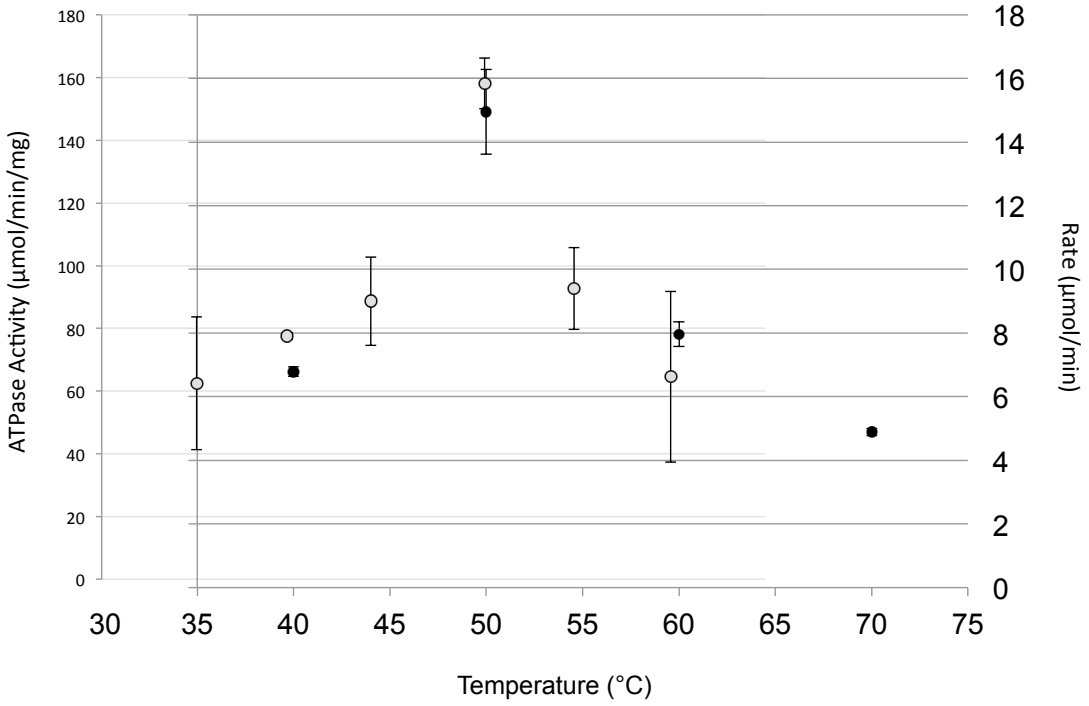

Figure 3.5: Temperature profile of ATPase activity of Daud CPN. ATPase activity of heat-treated Daud_2059 supernatant (untagged) and purified N-terminal His-tagged Daud_2059 across temperatures. ATP hydrolysis was assayed using the Malachite Green Assay as described in section 2.3.6. Heat-treated supernatant with 2mM ATP (closed circles; untagged; incubated clarified lysate at 55°C for 30 minutes and then spun at 20,000xg for 20 minutes) and 0.08mg/ml of purified N-polyhistadine tagged Daud_2059 with 0.1mM ATP (open circles) were incubated for 30 minutes at the indicated temperatures ranging from 35-70°C. ATPase activity (µmol/min/mg) corresponds to open circles while Rate (µmol/min) of inorganic phosphate production corresponds to data in closed circles.

Denatured MDH when incubated with purified Daud CPN for 10 minutes as described in section 3.3.5 at 37°C results in recovery of 37.7% (+/- 3.6) MDH activity as monitored by NADH depletion at 340nm (Figure 3.6).

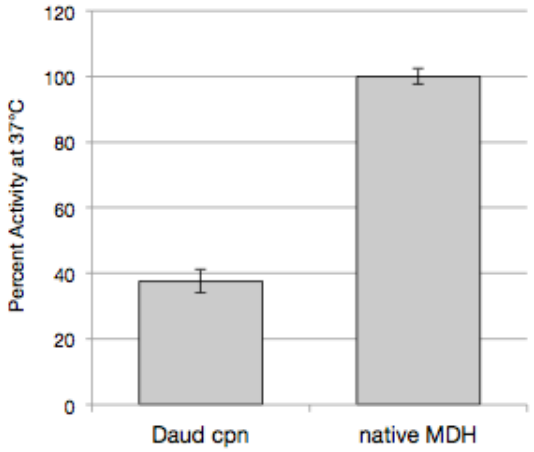

Figure 3.6: Refolding activity of Daud CPN on denatured MDH. Active dimeric MDH (70nM) activity resulting from refolding by Daud_2059 (74.5nM) at 37°C for 10 minutes. Spontaneous refolding of prepared denatured MDH alone was subtracted as background from the Daud+dMDH sample. The amount of recovered activity is determined by simultaneously assaying native MDH (70nM).

Daud CPN subunit model was generated using Phyre2 as described in section 3.3.8 and visualized in PyMOL (Figure 3.7; Kelley *et al.* 2015; The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC). The input WT Daud_2059 amino acid sequence aligned with 96% of the highest scoring template sequence (all available PDB structures are used during this analysis) with 100% confidence. The template used was the t-complex protein 1 subunit/bovine TRiC/CCT Ca model Chain E derived from 4.0Å cryo-EM map (c3iygE/PDB 3IYG; Cong *et al.* 2010). Daud CPN and Chy CPN sequence share 48% amino acid identity and 20% similarity, while Daud CPN and Bt TRiC/CCT chain E share 26% amino acid identity (Kelley *et al.* 2015).

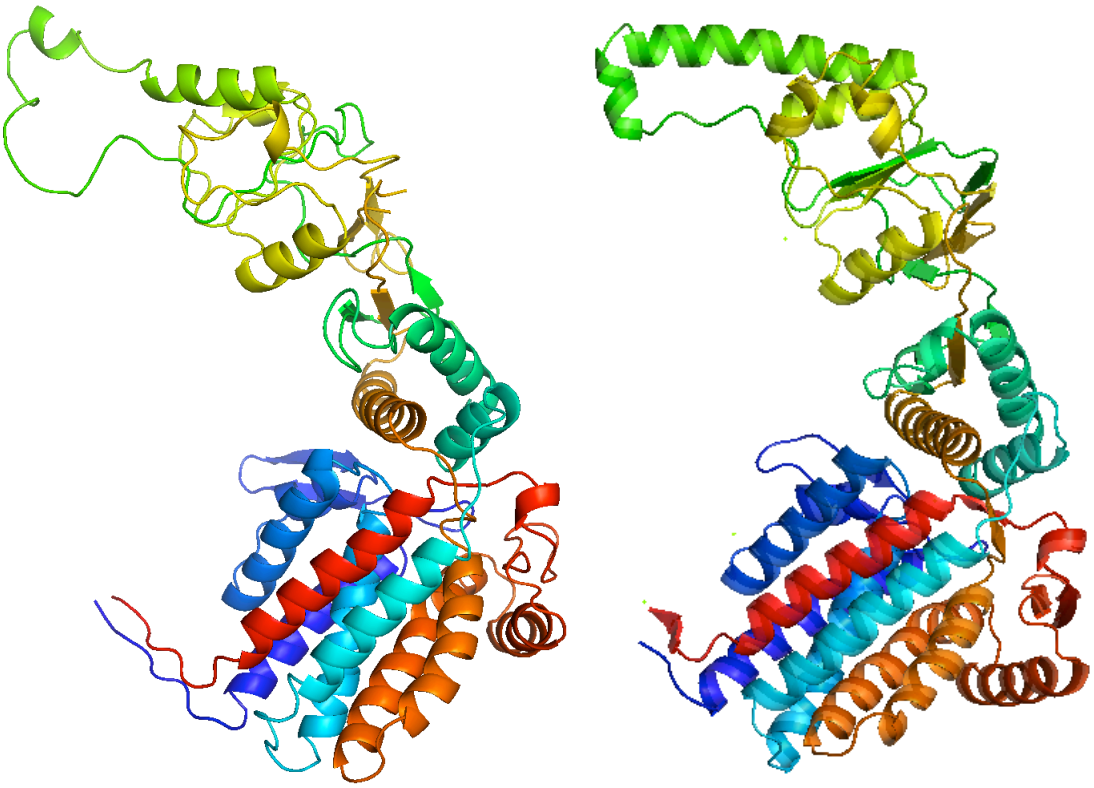

Figure 3.7: Monomers of Daud CPN model and Chy CPN crystal structure. Side view of model Daud CPN subunit (left) produced using Phyre2 compared with Chy CPN-ANP subunit (right; An *et al.* 2016) (Kelley et al. 2015). N-termini are blue and C-termini are blue. The first 8 residues of the N-terminus and the last 13 residues of the C-terminus are absent in this Daud CPN representation. 96% of the input WT Daud_2059 amino acid sequence with 26% identity modeled with 100% confidence by the single highest scoring template, the t-complex protein 1 subunit/bovine TRiC/CCT Ca model derived from 4.0Å cryo-EM map (c3iygE/PDB 3IYG; Cong *et al.* 2010). Image generated using PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Shroedinger, LLC).

3.4.2: Daud CODH Preliminary Data

Two distinct clades present themselves in the CODH phylogenetic tree representing archaeal and bacterial subunits. The archaeal-like CODH in *C. D. audaxviator* (red; Figure 3.8) is currently most similar to another archaeal-like bacterial CODH found in Dehalococcoidia bacterium SCGC AB-539-J10, an organism isolated from 10 cm subseafloor in Aarhus Bay, Denmark (Wasmund *et al.* 2014). This acetyl-CoA decarbonylase/synthase alpha gene shares 59% identity with *Daud_0105* and has an e-value of 0.0. These bacterial gene sequences are clustered in archaeal sequences that are overwhelmingly methanogens. In the bacterial CODH clade, there are also several genes present from methanogens including *Methanosarcina barkeri* 227, *Methanotorris igneus*, and *Methanoregula boonei* 6A8. These organisms were isolated from varied environments including a mesophilic acidophilic bog and a hyperthermophilic submarine vent system in Iceland (Bruggraf *et al.* 1990; Braeuer *et al.* 2011).

Of the structures available in the protein database (RCSB PDB), Daud_0105 has 95% sequence coverage and 44% identity with *Methanosarcina barkeri* acetyl-CoA decarboxylase/synthase alpha subunit of the carbon monoxide dehydrogenase, and Daud_0870 has 97% sequence coverage and 39% identity with the alpha subunit of the carbon monoxide dehydrogenase/acetyl-CoA synthase from *Moorella thermoacetica* followed closely by 96% input sequence coverage and 41% identity with *Carboxydothermus*

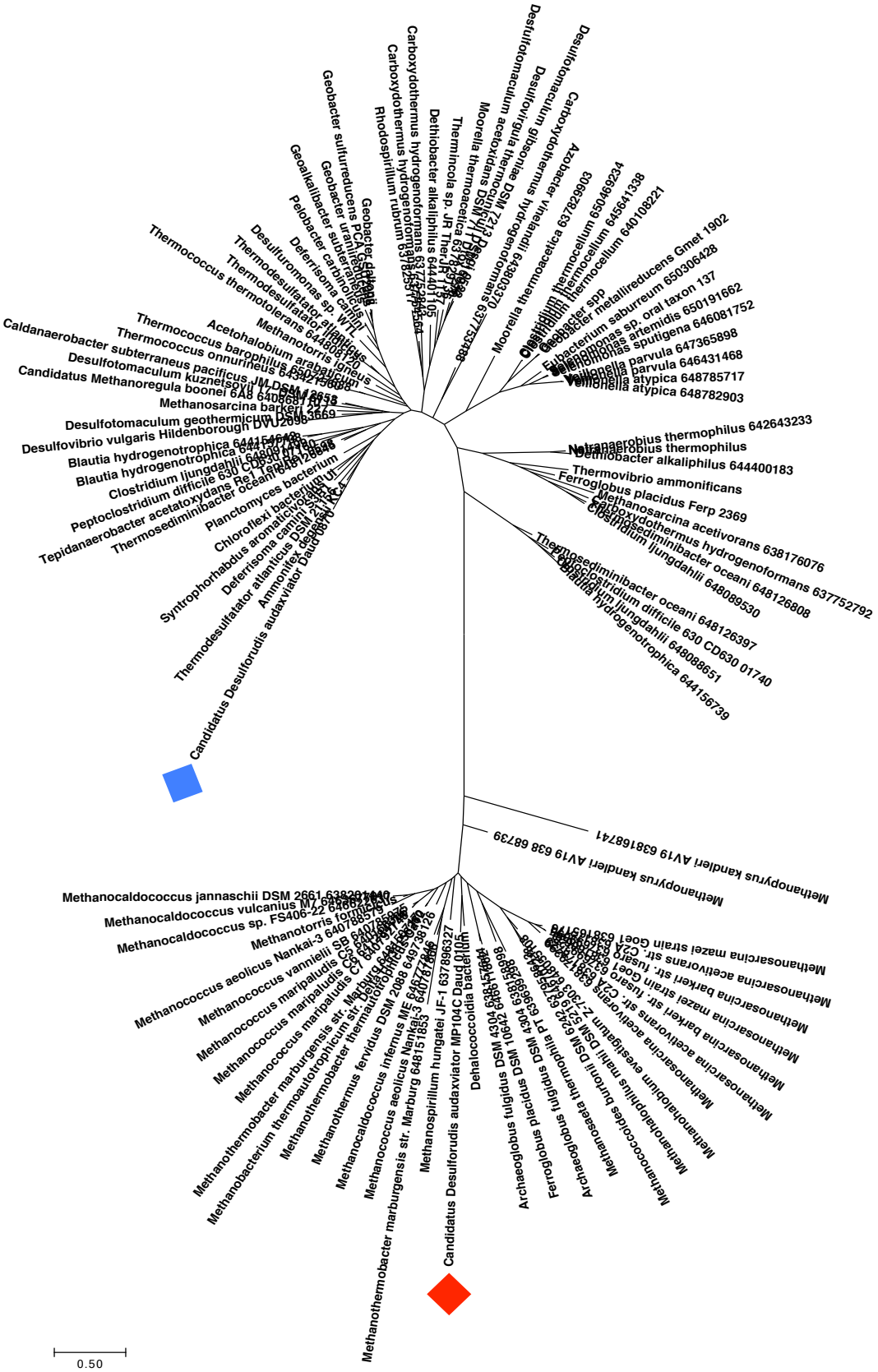

Figure 3.8: Phylogeny of CODH full amino acid sequences. Daud_0870 (CooS; blue), Daud_0105 (CdhA; red). The evolutionary history was inferred by using the Maximum Likelihood method based

on the JTT matrix-based model (Jones *et al.* 1992). The tree with the highest log likelihood (-30681.9824) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 106 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 253 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.* 2015).

hydrogenoformans carbon monoxide dehydrogenase (Figure 3.10; 3CF4, Gong *et al.* 2008; 1OAO, Darnault *et al.* 2003; 1SU7, Dobbek et al. 2004 Kelley *et al.* 2015).

Daud_0105 overexpressed well in *E. coli* (estimated MW 83.9 kDa; Figure 3.9). No discernable strong band was seen for Daud_0870 during expression trials (not included). Assays to evaluate activity with and without co-expression of CooC need to be completed.

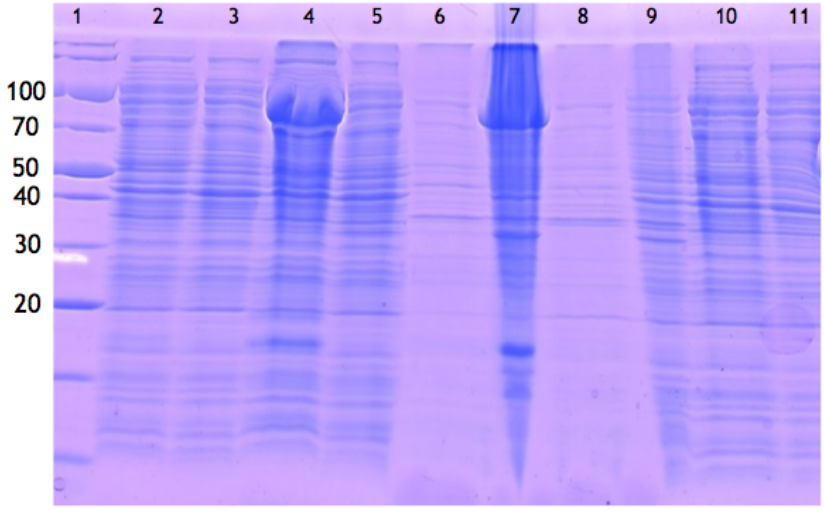

Figure 3.9: Solublization experiment with overexpressed Daud CdhA in *E. coli*. 1. Marker lane, 2. Daud_0105 aerobic supernatant (L), 3. Daud_0105 anaerobic supernatant (L), 4. Daud_0105 aerobic supernatant (1%), 5. Daud_0105 anaerobic supernatant (1%), 6. Daud_0105 aerobic supernatant (10%), 7. Daud_0105 aerobic pellet (10%), 8. Daud_0870 supernatant (10%), 9. Daud_0870 pellet (10%), 10. Daud_0870 aerobic supernatant (L), 11. Daud_0870 anaerobic supernatant (L). KEY:(L)-lysate in 100mM NaH_2PO_4 pH 8.0-NaOH, 150mM NaCl, 10% glycerol, 5mM CHAPS, 2mM DTT, 1mM EDTA, 0.1mM PMSF(1%)- lysate (L) + 1% sarkosyl, incubated at room temperature with rocking ~14hr (10%)- resuspended pellet (lysate in 1XLEW (50mM NaH_2PO_4 pH 8.0, 300mM NaCl), centrifugation 20,000xg, 20min) in 10% sarkosyl, incubated at room temperature with rocking ~17hr.

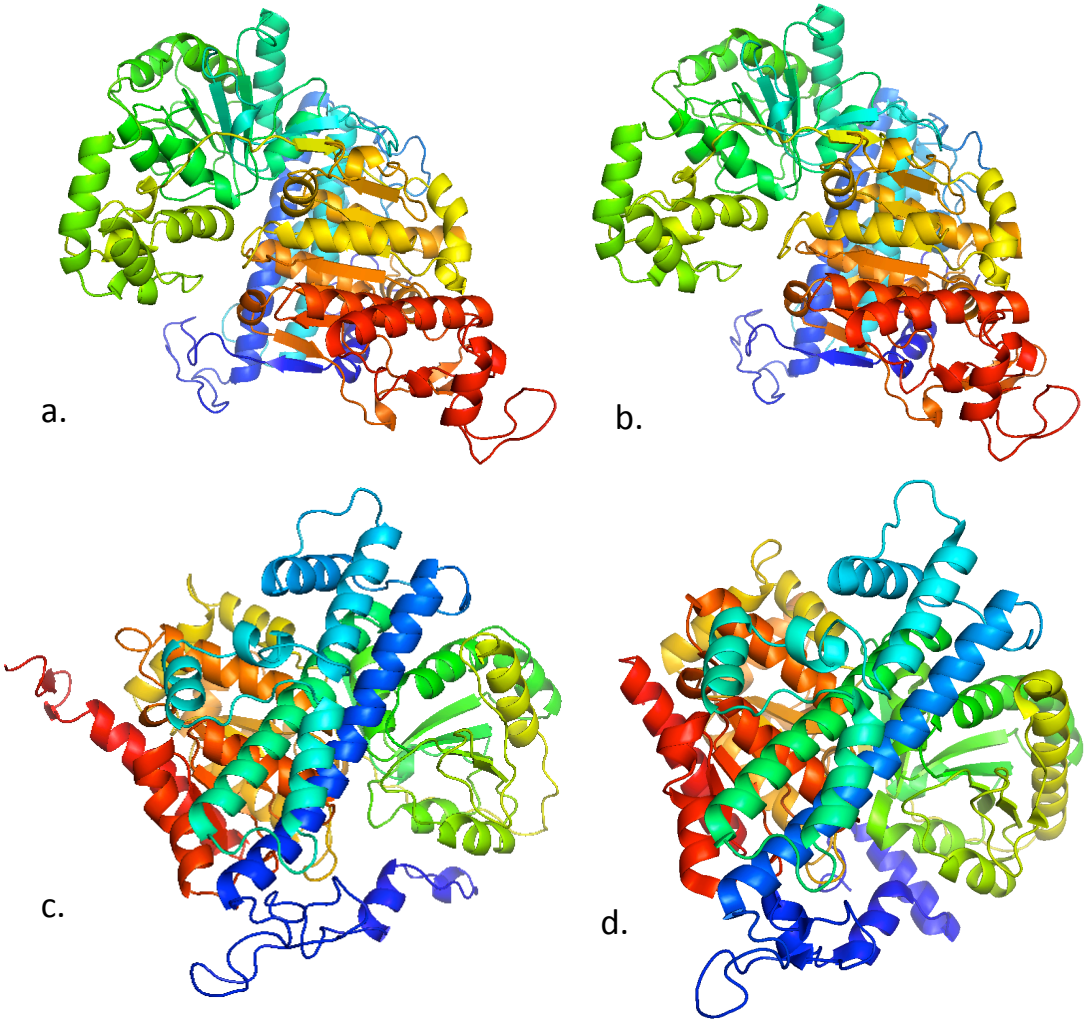

Figure 3.10: Models of Daud CODHs with their highest scoring template structures. Daud_0105 (CdhA (a.)) modeled using the template with 95% coverage of the input sequence and 44% identity *Methanosarcina barkeri* chain A ((b.) 3CF4; Gong *et al.* 2008). Daud_0870 (CooS (c.)) modeled using the highest scoring template with 97% input sequence coverage and 39% identity with alpha chain of *Moorella thermoacetica* Acetyl-CoA Synthase/Carbon Monoxide Dehydrogenase (1OAO; Darnault *et al.* 2003) and 96% input sequence coverage and 41% identity with *Carboxydothermus hydrogenoformans* carbon monoxide dehydrogenase ((d.) 1SU7; Dobbek *et al.* 2004).

3.5: Discussion

Reported sampling temperatures from gold mines in the Witwatersand Basin, South Africa range from 27-56°C depending on location and depth of collection, with 2.8km fracture fluids in the Mponeng mine maintaining an estimated temperature closer to 60°C and pH 9.3 (Onstott *et al.* 2014; Chivian *et al.* 2008). To determine if these sampling site temperatures are an accurate reflection of the optimal growth temperature of this planktonic organism, the thermostability of Daud_2059 was analyzed by native PAGE, DSC, and *in vitro* assays. Based on the observed data, it appears that the Daud_2059 complex has a thermostability upper limit at 70°C. Daud_2059 displays a first melt between 62.15-70.67°C, although complex dissociation is likely closer to the 62.15°C transition as observed under neutral, and assumed physiologically relevant pH conditions (Figure 3.4). This dissociation is also observed in the 70°C lane of the native PAGE (Figure 3.2). Recombinant Chy CPN has an initial transition between 81-83°C, and an observed optimum ATPase activity at 65°C (Inset Figure 3.4). *C. hydrogenoformans* is typically grown at temperatures ranging from 65-70°C (Svetlichny *et al.* 1991). Chy_0413 and Daud_2059 have major transitions about 8°C and 12°C apart, exhibiting a similar double melt pattern at neutral pH (Figure 3.4). Further, Daud_2059 has transitions about 20°C below Chy_0413. Therefore, from these results an optimum functional temperature may extrapolated of about 45°C. ATPase assays of purified Daud_2059 and heat-treated, clarified supernatant display an optimal ATPase activity at 50°C (Figure 3.5). Purified Daud_2059 using ammonium sulfate prior to anion exchange chromatography resulted in loss of ATPase activity (see section 3.3.2). Daud CPN is a presumed housekeeping gene with demonstrated transcription of both Group III and Group I CPNs in *C. hydrogenoformans* cultures during normal growth conditions (Techtmann 2009). Recombinant Daud CPN functions optimally at neutral pH at 50°C which is likely slightly

below the optimal growth temperature of the organism. This data is consistent with the noted temperature of ~60°C at the *C. D. audaxviator* sampling site (Chivian *et al.* 2008).

The predicted size of Daud CPN 16-mer is 881.6 kDa and a single ring is 440.8 kDa (Gasteiger *et al.* 2005). The major peak upon incubation of untagged Daud CPN with 2mM ATP and 1mM $MgSO_4$ yields a major complex of 848.8 kDa. (Figure 3.3a). In the absence of nucleotide, purified N-His-tagged Daud CPN produces major complexes of 448 kDa and 784 kDa +/- 10% (Figure 3.3b), likely representing single and double ringed complexes respectively. Native PAGE supports these results with bands present between 720 and 1,048 kDa for purified Daud CPN samples, indicating that as with Chy CPN, Daud CPN also forms a homomeric hexadecamer (Figure 3.2). The Daud and Chy CPN amino acid sequences share 48% amino acid identity and 20% similarity, and are likely very similar structurally. The modeled Daud CPN subunit using the Bt CCT chain E as a template which are only 26% identical produces a structure with similar overall architecture compared with the closed Chy CPN monomer (Figure 3.7).

There are typically 2 distinct, larger groupings of CODHs based on their phylogenetic origins (Techtmann *et al.* 2012). *C. D. audaxviator*, however, contains two annotated CODHs- a bacterial type (CooS; Daud_0870) as well as an archaeal subunit (CdhA; Daud_0105). The genome of *C. D. audaxviator* contains several archaeal-like gene annotations including the Group III CPN and CdhA (Chivian *et al.* 2008). While the origins of the Group III chaperonins are not very clear due to their clustering in a clade independent of known archaeal sequences (Figure 1.5, 1.6), the CdhA represents a gene that was more likely acquired via recent horizontal gene transfer with an archaeon or a bacterium with a recently acquired archaeal CdhA due to Daud_0105's deep association with known archaeal CODH subunits (Figure 3.8). Both CODHs have structural representatives in PDB with ~40% identity compared with 26% identity of the highest scoring template for the Daud

Group III CPN (Figure 3.7, 3.10). Additionally, several methanogens contain annotated bacterial-type CooS CODHs that cluster with the known bacterial sequences, similar to the Group I CPNs also found in methanogens (Figure 3.8; Figure 1.5). These gene transfer events are probably due to cohabitation and metabolic crossfeeding of *C. D. audaxviator* with methanogens (Figure 3.8). Hydrogenogens like *C. hydrogenoformans* can supply H_2 to hydrogen-utilizers including sulfate reducers like *C. D. audaxviator* and methanogens (Techtmann 2009). Certain *Dehalococcoides* with incomplete Wood-Ljungdahl pathways lacking CODHs have demonstrated biogenic CO production (Zhuang *et al.* 2014), which can in turn be used by methanogens and carboxydotrophic bacteria. Acquiring genes from potential competitors in this energy limited and resource sparse space can help microbes utilize new compounds, and thrive in challenging conditions.

3.6: Future Directions

A more thorough characterization of Daud CPN and CODHs remains to be accomplished. Further assays determining parameters including optimal salt concentrations and pH need to be completed to provide more meaningful data for use during *C. D. audaxviator* culturing attempts. This study provides preliminary biochemical data that supports environmental sampling information, and should be repeated for sequence data lacking such supporting data to provide useful information of *in situ* environment from which the sequence was sampled as well as potential guidance during in laboratory culturing attempts.

Chapter 4: Discussion and Future Directions

This work provides the first structural study of a Group III CPN with supporting functional and cross-Group comparative analysis. We have demonstrated the close structural similarities to known archaeal as well as eukaryotic Group II structures with high-resolution ADP and ANP bound structures of Group III Chy CPN (An *et al.* 2016; Pereira *et al.* 2010; Cong *et al.* 2010; Dekker *et al.* 2011). Currently, the highest scoring template structure available in RSCB PDB is the alpha CPN subunit from *Thermococcus* strain KS-1 with 29% identity to 95% sequence coverage (1Q2V, Shomura *et al.* 2004; Kelley *et al.* 2015). Chy CPN is a compact Group II-like variant and may be representative of an ancient CPN, and can therefore provide more complete insight into the development of this ubiquitous protein-folding complex.

The amino acid alignment of Chy CPN compared with archaeal subunits displays the most divergence in the apical lid region (Figure 2.17; An *et al.* 2016). Structurally, a lack of strict cavity closure facilitated by the beta strands in the apical domain of Group II CPNs is observable in the Chy CPN-ANP complex (Figure 2.17; Pereira *et al.* 2010 (3KFB); Shomura *et al.* 2004 (1Q3Q)). The apical domains of the Chy CPN-ANP do not overlap as with Group II CPNs, which may be due to the sequence disparity between the Groups, or due to incomplete closure with ANP as reported by other groups (Douglas *et al.* 2011). Eukaryotic ANP-bound structures vary from available archaeal structures, with *Bos taurus* heteromeric CCT remaining in an asymmetrical and open conformation (4A0V, Cong *et al.* 2012). Obtaining a Chy CPN structure bound with AlF or BeF, molecules that can lock complexes immediately post-ATP hydrolysis, will help determine whether there is any incomplete closure in the ANP bound complex. Additionally, the substrate displacement mechanisms and

the size limit of substrates that are able to be accommodated within the Group III CPN cavity should be investigated.

The overall distribution of hydrophobic residues in this region including the helical extension and the preceding loop protrusion, however, is very similar between archaeal Group II and Group III CPNs (Chy CPN Y227, V232, L237, F243, Y246, L247, I252, F253; Ta CPN alpha subunit I249, I254, V258, I266, F269, L270, F277). These residues may be important in mediating unfolded protein interactions (Klumpp *et al.* 1997). The distribution of charged residues, however, does vary slightly between the two types of CPNs. Chy CPN contains a cluster of negatively charged residues in the closed conformation located in the interior of the apical loop protrusion (E230, E231, E233, E234, E235), while the archaeal Group II subunits typically contain 2-3 positively charged residues in this region. Eukaryotic CCT gamma subunits also contain a predominance of positively charged residues in their apical regions (Pappenberger *et al.* 2002). The variation in charges in the apical domain may mediate different interactions with different proteomes and substrates of the Group II versus Group III organisms, or it may be due to varied direct interactions with co-chaperones. Prefoldin, a holdase chaperone complex found in archaea, is thought to directly interact with the charged residues located in the apical domain in Group II CPNs (Vainberg *et al.* 1998; Sahlan *et al.* 2010). This is a mode of substrate delivery to the chaperonin complex for refolding. The absence of this chaperone in bacterial genomes may indicate that another chaperone such as the neighboring DnaK is functioning as a direct delivery mechanism, which has been observed with eukaryotic homologs, or that the Group III CPNs can function efficiently without the assistance of any co-chaperones (Cuellar *et al.* 2008). Work detailed in section 2.6 should also be completed to understand the significance of the conserved Group III chaperonin-containing chaperone gene cluster.

The physiological significance of Group III chaperonins in bacteria and the novel chaperonin branch including the *Thaumarchaeota*, however, remains an enigma (Figure 1.5). Both types of newly identified chaperonins are present in organisms whose genomes encode at least one additional chaperonin encoding gene- Group I CPNs in the bacteria (green) and multiple Group II-clustering CPNs in the archaea (yellow) (Figure 1.5, 1.6). Quantitative RT-PCR targeting the Group I and Group III chaperonin of *C. hydrogenoformans* during different growth conditions including high heat, and growth using CO versus pyruvate as carbon sources, will help to determine if there is condition-dependent variation in transcriptional regulation between these chaperonins, and if these chaperonins harbor any specialized roles *in vivo* or are functionally redundant. Supporting proteomic determination of CPN Group III protein expression correlated to transcription should also be performed, along with proteomic analysis of native interacting partners analogous to the work performed with the co-immunoprecipitations of Group I and Group II chaperonins in *M. mazei* (Hirtreiter *et al.* 2009). Multiple copies of chaperonins in bacteria are thought to allow for the development of specilialized protein interactions (Lund 2009), and homologs of DnaK in *E. coli* have demonstrable specific protein interactions and functions such as HscA's involvement in the maturation of iron-sulfur proteins (Hoff *et al.* 2000; Hoff *et al.* 2002; Genevaux *et al.* 2007). Therefore, it is possible that these Group III CPNs may have specialized interactions *in vivo* either with DnaK itself as a co-chaperone or with substrate protein.

The isothermal titration calorimetry data obtained in section 2.4.2 reveals ATP binding at all available nucleotide-binding sites for both the D373A Chy CPN and the D373A/R155K Chy CPN with comparable binding affinities (Table 1). This implies a lack of negative allosteric regulation between the double-stacked rings as observed with Group I and II CPNs (Yifrach and Horovitz 1995; Horovitz *et al.* 2001). To confirm this observation, this experiment should be repeated using the non-hydrolyzable ATP analog, AMP-PNP, to further

reduce any possibility of ATP hydrolysis during measurements. Mutations probing the allosteric mechanisms of intra-ring and inter-ring interactions need to be designed. An inter-ring mutation at Chy CPN L439 was attempted multiple times as described in section 2.3.2 but never successfully generated. This residue appears to interact across rings based on the structures and may be important in forming the hexadecamer complex as well as communicating ATP and substrate cycling events (An *et al.* 2016).

To determine if Group III chaperonins are functionally equivalent to Group II chaperonins *in vivo*, the genes from *C. hydrogenoformans* could be introduced into a genetically tractable archaeon such as *Thermococcus onnurineus* NA1 as previously attempted (Sara Rowland, NSF EAPSI Fellowship in Seoul, 2013). Further, a genetic system should be developed in a Group III chaperonin-containing organism so that separate knock-outs of both the Group I and Group III chaperonins may be constructed. The resulting genetically modified organisms may then be cultured to determine any deficiencies in growth across a variety of conditions. This work may help to determine if there is overlap in function, or essential independent roles in the cell.

Comparative genomic analyses of Group III CPN-containing organisms can shed light on the evolutionary history of this novel protein-folding complex and help to inform its functional significance through any discovered similarities (Table I). It was previously observed that the first several Group III CPN sequences were derived primarily from CO utilizing bacteria including *C. hydrogenoformans*, resulting in the hypothesis that this archaeal-like CPN may be involved in the maturation of the large, multi-subunit metalloenzymes involved in the metabolism of CO (Techtmann 2009; Techtmann and Robb 2010).

The Daud CPN work provides a proof of concept study whereby an unculturable organism can be partially characterized through the biochemical analysis of well studied, pervasive, and constitutively expressed proteins. The derived temperature optimum of Daud

Group III CPN through *in vitro* experimentation (50°C; Figure 3.2, 3.4, 3.5) supports the environmental site sampling of *C. D. audaxviator* (Chivian *et al.* 2008). This demonstrates that this type of study can be translated to sequence with no corresponding environmental analysis at the sample site to inform culture growth parameters. Additionally, the basic characterization of Daud Group III CPN validates the initial Chy CPN studies as a homomeric hexadecamer capable of utilizing ATP to fold denatured substrate (Techtmann 2009; Techtmann and Robb 2010).

The Robb laboratory is continuing the characterization of the Daud and Chy CODHs. The CODHs will be co-expressed with Chy CooCs, Nickel-inserting chaperones, as well as the Group III CPNs in an attempt to produce the most amount of active enzyme possible (Inoue *et al.* 2014). These assays will help determine the roles of the dual Daud CODHs *in vivo*, namely whether they favor the forward or reverse gas-water shift reaction or if they perform redundant metabolic functions. Of the Chy CODHs, Chy CODH V is still not assigned a specific role in the cell like NADPH generation or energy conservation functions of other Chy CODHs and should be investigated (Svetlitchnyi *et al.* 2001; Wu *et al.* 2005).

Extant cells rely on multiple types of chaperones to maintain proteostasis and survive throughout a variety of conditions. The ATP-dependent molecular chaperones known as chaperonins are present in organisms across all kingdoms, with complexes forming with single subunits in archaea and bacteria, and up to eight distinct polypeptide chains in the cytosol of eukaryotes. The homomeric Group III CPNs may represent a mechanistically primitive version of these protein-folding nanomachines from which more complex CPNs including CCTs are derived.

Appendices

A.1: Primers

Chy_0413 FWD
5’ AAAAAACATATGAAAAAAGAAAACCAAA
Chy_0413 REV
5’ TTTTTTCTCGAGTTATCTTTTTCCTCCATC

Chy_0413 R155E FWD
5’-CGGCCAAAATTGCCGGAGAGGGTGATGAAAGGGTTGC
Chy_0413 R155 E REV
5’- GCAACCCTTTCATCACCCTCTCCGGCAATTTTGGCCG

Chy_0413 R155L FWD
5’-GAGGCCAAAATTGCCGGACTGGGTGATGAAAGGGTTGC
Chy_0413 R155L REV
5- GCAACCCTTTCATCACCCAGTCCGGCAATTTTGGCCGC

Chy_0413 R155K FWD
5’-CGGCCAAAATTGCCGGAAAGGGTGATGAAAGGGTTGC
Chy_0413 R155K REV
5’- GCAACCCTTTCATCACCCTTTCCGGCAATTTTGGCCG

Chy_0413 D373A FWD
5’ CAGGAAAGAATTGCTAAAGCTGCGGCCGGAAGTTTTGCCC
Chy_0413 D373A REV
5’ GGCAAAACTTCCGGCCGCAGCTTTAGCAATTCTTTCCTG

Chy_0413 N55A FWD
5’ GATAGATTTGGGGAAGTGGTGGTTACCGCCGACGGAGTAACGATTTTG
Chy_0413 N55A REV
5’ CAAAATCGTTACTCCGTCGGCGGTAACCACCACTTCCCCAAATCTATC

Chy_0413 D506* FWD
5’ ACCATAATTAAAATGAAATAATAGGAAGTTTTAAAGGAAGGGGAGCAG
Chy_0413 D506* REV
5’ CTGCTCCCCTTCCTTTAAAACTTCCTATTATTTCATTTTAATTATGGT

Chy_0413 G156E FWD
5’- GCCAAAATTGCCGGACGGGAGGATGAAAGGGTTGCTGCTATCC
Chy_0413 G156E REV
5’- GGATAGCAGCAACCCTTTCATCCTCCCGTCCGGCAATTTTGGC

Chy_0806 FWD
5’-AAACATATGAAGCGGGATTATTACGAAA
Chy_0806 REV
5’-AAACTCGAGTTATAACGCAAATTTGTTTT

Chy_0807 FWD
5’-CATATGGCTGGCAAACAAATATTATTT

Chy_0807 REV
5’-CTCGAGTTATTACATCATGTCCATGTT

Chy_0414 FWD
5’-CATATGGAGGAAAAAGATAAAGAGGA
Chy_0414 REV
5’-CTCGAGTTATTGTTTTTTCGAAA

Chy_0415 FWD
5’-AAGCTAGCATGGGTAGAATTGTTGGTATT
Chy_0415 REV
5’-TTCTCGAGTTATTTATCGTCATTTACCT

Chy_0416 FWD
5’-CATATGTTGAAGCGGGATTATTACGA
Chy_0416 REV
5’-GGATCCTTATAAGCGAAATTTGTTTTT

Daud_2059 FWD
5’-AAAAAACATATGTTGAGTCTCAAAAAGCA
Daud_2059 REV
5’-TTTTTTGCTGAGCTCACTCATCCTTCCT

Daud_2059 N-His Tag FWD
5’-AAAAAACATATGAGTCTCAAAAAGCA
Daud_2059 N-His Tag REV
5’-TTTTTTAAGCTTTCATCACTCATCCTTCCT

Daud_2059 E159G FWD
5’- GTGGCCGGCCGGGGGCATGCGGACATCGCCG
Daud_2059 E159G REV
5’- CGGCGATGTCCGCATGCCCCCGGCCGGCCAC

Chy_1220 FWD
5’-AAAAAACCATGGGGTAGAATGAAGTTAG
Chy_1220 REV
5’- TTTTTTCTCGAGTTATCCCACCTCCAAA

Chy_1221 FWD
5’-AAAAAACATATGCCAAGGTTTAGAGATTT
Chy_1221REV
5’-TTTTTTCTCGAGTTAATAGTTTTGGCAGA

Chy_1225 FWD
5’- AAAAAAACCATGGCCTTTAAAATTGCGGT
Chy_1225 REV
5’- TTTTTTCTGGAGTCAGATAATACCGCGAT

T7 FWD
5’-TAATACGACTCACTATAGG
T7 REV
5’-AAGGGGTTATGCTAGTTATTGCTCAGC
pTac FWD

5'-TAATACGACTCACTA
M13R
5'-CAGGAAACAGCTATGAC

A.2: Autoinduction Protocol and Various Broths

Zym-5052 Autoinduction (10mls)
9.58ml ZY
20ul 1M $MgSO_4$
2ul 1000x metals
200ul 50X5052
200ul 50xM
+antibiotic(s)

ZY
10g tryptone
5g yeast extract

50XM (100ml)
80ml H2O
17.75g Na_2HPO_4
17.0g KH_2PO_4
13.4g NH_4Cl
3.55g Na_2SO_4
(pH of 50-fold dilution should be 6.7)

50X5052 (100ml)
25g glycerol
73ml H_2O
2.5g glucose
10g alpha-lactose monohydrate

1000x metals (100ml)
36ml sterile H_2O
50ml 0.1M $FeCl_3$ in 0.12M HCl
2ml 1M $CaCl_2$
1ml 1M $MnCl_2$-$4H_2O$
1ml 1M $ZnSO_4$-$7H_2O$
1ml 0.2M $CoCl_2$-$6H_2O$
2ml 0.1M $CuCl_2$-$2H_2O$
1ml 0.2M $NiCl_2$-$6H_2O$
2ml 0.1M Na_2MoO_4-$2H_2O$
2ml 0.1M Na_2SeO_3-$5H_2O$
2ml 0.1M H_3BO_3

NZY+ Broth (1L) for XL1-Blue Competent Cell Transformations (QuikChange II Site-Directed Mutagenesis Kit)
10g NZ amine (casein hydrolysate)

5g yeast extract
5g NaCl
dH2O up to 1L
Adjust to pH 7.5 with NaOH; autoclave, then add the following filter sterilized solutions:
12.5ml 1M $MgCl_2$
12.5ml 1M $MgSO_4$
20ml of 20% (w/v) glucose

A.3: Chemically Competent Cells Protocol (Adapted from Grose Lab/BYU protocol)

1. Grow from single colony overnight at 37 C, 200 rpm in 10ml SOB and any required antibiotic.
2. Dilute 1/100 in fresh SOB at 37 C, 200 rpm to OD550 0.4.
3. Chill on ice for 5-10minutes.
4. Spin 3,000-4,000k for 10 minutes, 4C.
5. Carefully decant and discard supernatant. Gently resuspend pellet in 150ml TFB I on ice. Leave on ice for 5 minutes.
6. Spin 3,000-4,000k for 10 minutes, 4C.
7. Carefully decant and discard supernatant. Gently resuspend pellet in 15ml TFB II. Leave on ice for 15 minutes.
8. Gently aliquot cells into sterile microcentrifuge tubes and snap freeze (ethanol, dry ice). Immediately move cells to -80 C for storage.

SOB (400ml)
8g tryptone
2g yeast
0.2g NaCl
4ml 0.25M KCl
pH 7.0 using 5N NaOH, autoclave
add 2ml filter sterilized 2M $MgCl_2$

TFB I
30mM potassium acetate
100mM KCl
10mM $CaCl_2$
50mM $MnCl_2$
15% glycerol
pH 5.8 using acetic acid; Filter sterilize

TFB II
10mM MOPS
75mM $CaCl_2$
10mM KCl
15% glycerol
pH 6.5 using KOH; filter sterilize

Table 3. Chy_0413 BLASTp results in JGI IMG (Altschul *et al.* 1997).

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes	*Thermoanaerobacteraceae*	*Carboxydothermus hydrogenoformans Z-2901*	0		
Firmicutes	*Thermoanaerobacteraceae*	*Carboxydothermus hydrogenoformans DSM 6008*	0	hot spring, Kunashir Island, Russia; anaerobic 65C	Svetlichny et al. 1991
Firmicutes	*Thermoanaerobacteraceae*	*Carboxydothermus ferrireducens DSM 11255*	0	hot spring, USA, Yellowstone, Calcite Spring Area; anaerobic 65C	Slobodkin et al. 1997; Slobodkin et al. 2006
Firmicutes		*Clostridiales bacterium DRI-13*	2E-167		
Firmicutes	*Peptococcaceae*	*Desulfotomaculum nigrificans DSM 574*	2E-166	Netherlands, fresh water, anaerobic, 55C	Werkman and Weaver 1927; Campbell and Postgate 1965 emend. Visser et al. 2014; synonyms: Desulfotomaculum carboxydivorans
Firmicutes	*Peptococcaceae*	*Desulfotomaculum carboxydivorans CO-1 SRB DSM14880*	2E-166	sludge from an anaerobic bioreactor treating paper mill wastewater; anaerobic; 55C	Parshina et al. 2005
Firmicutes	*Peptococcaceae*	*Thermincola potens JR*	9E-162		
Firmicutes	*Peptococcaceae*	*Thermincola ferriacetica Z-0001*	9E-162		
Firmicutes	*Peptococcaceae*	*Thermincola ferriacetica DSM 14005*	9E-162	ferric deposits of a terrestrial hydrothermal spring; Kunashir Island; anaerobic 60C	Zavarzina et al. 2007
Firmicutes	*Peptococcaceae*	*Desulfotomaculum reducens MI-1*	1E-159		
Firmicutes	*Peptococcaceae*	*Desulfotomaculum ruminis DL, DSM 2154*	3E-159	rumen of hay-fed sheep; anaerobic; 37C	Campbell and Postaget 1965
Firmicutes	*Peptococcaceae*	*Candidatus Desulforudis audaxviator*	2E-158		

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes	*Veillonellaceae*	*Thermosinus carboxydivorans*	2E-156		
Firmicutes	*Veillonellaceae*	*Sporolituus thermophilus DSM 23256*	4E-156	microbial mat sample from the outflow of a GAB bore, Great Artesian Basin, Australia, Queensland, New Lorne Bore; anaerobic 55C	Ogg and Patel 2009
Firmicutes	*Peptococcaceae*	*Desulfotomaculum putei DSM 12395*	9E-156	2.7km deep terrestrial subsurface; USA, VA, Taylorsville, Triassic Basin; anaerobic 55C	Liu et al. 1997
Firmicutes		*Thermosediminibacter oceani JW/IW-1228P, DSM 16646*	1E-153	core samples of deep sediment, Peru; anaerobic, 68C	Lee et al. 2006
Firmicutes	*Veillonellaceae*	*Sporomusa sp. KB1*	6E-153		
Firmicutes	*Veillonellaceae*	*Dendrosporobacter quercicolus DSM 1736*	1E-152	discolored tissue in a living oak, USA; anaerobic 28C	Stankewich et al. 1971; Stroempl et al. 2000
Firmicutes		*Firmicutes bacterium JGI 0000059-H17*	2E-152		
Firmicutes	*Veillonellaceae*	*Sporomusa ovata DSM 2662*	2E-152	sugar beet leaf silage, Germany; anaerobic 30C	Moeller et al. 1985
Firmicutes	*Heliobacteriaceae*	*Heliobacterium modesticaldum Ice1 DSM 9504*	2E-152	soil from near hot springs, Iceland, Reykjanes; anaerobic in light, 50C	Kimble et al. 1996
Firmicutes	*Veillonellaceae*	*Sporomusa ovata DSM 2662*	4E-152		
Firmicutes		*Clostridiales bacterium PH28_bin88*	6E-152		
Firmicutes	*Veillonellaceae*	*Sporomusa acidovorans DSM 3132*	9E-152	effluent from alcohol-distillery plant, France; anaerobic 35C	Ollivier et al. 1990
Firmicutes		*Carboxydocella sporoproducens DSM 16521*	3E-151	hot spring of Karymskoe Lake, Russia; anaerobic 60C	Slepova et al. 2006
Firmicutes	*Peptococcaceae*	*Pelotomaculum thermopropionicum SI (reannotation)*	3E-151		
Firmicutes	*Peptococcaceae*	*Pelotomaculum thermopropionicum SI*	3E-151		
Firmicutes	*Veillonellaceae*	*Sporomusa malonica DSM 5090*	4E-151	freshwater mud, Germany; anaerobic 30C	Dehning et al. 1990

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes	*Peptococcaceae*	*Desulfotomaculum aeronauticum DSM 10349*	4E-151	corroded aluminum alloy in an aircraft; anaerobic 37C	Hagenauer et al. 1997
Firmicutes	*Peptococcaceae*	*Desulfitibacter alkalitolerans DSM 16504*	4E-151	metal coupon retrieved from a corrosion-monitoring reactor of a Danish district heating plant, Denmark, Skanderborg; anaerobic, 37C	Nielsen et al. 2006
Firmicutes	*Peptococcaceae*	*Desulfotomaculum acetoxidans 5575, DSM 771*	1E-150	piggery waste, Germany; anaerobic 37C	Widdel and Pfenning 1977
Firmicutes	*Peptococcaceae*	*Desulfotomaculum hydrothermal Lam5 DSM 18033*	2E-150	terrestrial hot spring at 280m elevation, North East Tunisia; anaerobic 55C	Haouari et al. 2008
Firmicutes	*Peptococcaceae*	*Desulfotomaculum hydrothermal Lam5*	2E-150		
Firmicutes	*Thermoanaerobacteraceae*	*Tepidanaerobacter sp. Re1*	2E-149		
Firmicutes	*Thermoanaerobacterales*	*Tepidanaerobacter acetoxydans Re1 DSM 21804*	2E-149	silage of alfalfa in a stirred laboratory reactor, Sweden, Uppsala; anaerobic, 37C	Schnuerer
Firmicutes	*Peptococcaceae*	*Desulfurispora thermophila DSM 16022*	4E-149	sulfidogenic FBR treating acidic metal-containing wastewater, Finland; anaerobic 50C	Kaksonen et al. 2007
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus migulans Nagano E1*	1E-148		
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus migulans Nagano*	1E-148		
Firmicutes		*Thermosediminibacter litoriperuensis DSM 16647*	1E-148	core sample of deep sediment, Peru Margin; 64C, anaerobic	Lee et al. 2006
Firmicutes	*Veillonellaceae*	*Pelosinus propionicus DSM 13327*	1E-147	gut of termite Thoracotermes macro thorax, P1 section, Congo, Brazzaville, Mayombe Forest; 30C anaerobic	Boga et al. 2007, Moe et al. 2012
Firmicutes	*Syntrophomonadaceae*	*Fervidicola ferrireducens*	1E-147		
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus aneurinilyticus ATCC 12856*	3E-147	human feces; 30C	Shida et al.
Firmicutes	*Veillonellaceae*	*Pelosinus sp. UFO1*	5E-147		

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes	*Veillonellaceae*	*Acetonema longum APO-1, DSM 6540*	2E-146	gut termite, Pterotermes occidentis; anaerobic 30C	Kane and Breznak 1992
Firmicutes	*Veillonellaceae*	*Pelosinus fermentans B4*	9E-146		
Firmicutes	*Veillonellaceae*	*Pelosinus fermentans B3*	9E-146		
Firmicutes	*Veillonellaceae*	*Pelosinus fermentans R7, DSM 17108*	9E-146	kaolin deposit Zhuravlinii Log, Russia; anaerobic 25C	Shelobolina et al. 2007
Firmicutes	*Veillonellaceae*	*Pelosinus fermentans A11*	9E-146		
Firmicutes	*Veillonellaceae*	*Pelosinus sp. HCF1 DSM 26827*	9E-146	aquifer sediment, USA, Washington State, Handford 100H site; anaerobic 30C	Beller
Firmicutes	*Veillonellaceae*	*Pelosinus fermentans JBW45*	1E-145		
Firmicutes	*Thermoanaerobacteraceae*	*Desulfovirgula thermocuniculi DSM 16036*	2E-145	Toyoha mine, Japan; anaerobic 70C	Kaksonen et al. 2007
Firmicutes		*Caldanaerovirga acetigignens DSM 18802*	2E-144	soil sample, USA, Nevada, Trego Hot Spring; anaerobic 62C	Wagner et al. 2009
Firmicutes	*Peptococcaceae*	*Desulfotomaculum kuznetsovii 17, DSM 6115*	2E-143	underground thermal mineral water, USSR; anaerobic 60C	Nazina et al. 1990
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus migulans DSM 2895*	3E-143		Takagi et al. 1993, Shida et al. 1996 emend. Heyndrickx et al. 1997
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus migulans NCTC 7096*	3E-143		
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus migulans DSM 2895*	3E-143		
Firmicutes	*Veillonellaceae*	*Anaeromusa acidaminophila DSM 3853*	2E-142	anaerobic purification plant, Netherlands; anaerobic 37C	Baena et al. 1999; synonyms: Selenomonas acidaminophil a
Firmicutes	*Syntrophomonadaceae*	*Syntrophomonas zehnderi*	3E-142		

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes		*Thermovenabulum gondwanense*	3E-142	Fe(III)-reducing bacterium; microbial mats from Great Artesian Basin bore runoff channel	Ogg et al. 2010
Firmicutes	*Peptococcaceae*	*Desulfotomaculum thermosubterraneum DSM 16057*	3E-142	geothermally active underground mine, Japan; anaerobic 60C	Kaksonen et al. 2006
Firmicutes	*Veillonellaceae*	*Anaeroarcus burkinensis DSM 6283*	4E-142	rice field soil, West Africa, Burkina Faso; anaerobic 35C	Ouattara et al. 1992, Stroempl et al. 1999; synonyms: Anaerovibrio burkinabensis Ouattara et al. 1992
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus terranovensis DSM 18919*	1E-141	soil, Antarctica; 40C	Allan et al. 2005
Firmicutes	*Peptococcaceae*	*Desulfotomaculum australicum DSM 11792*	6E-141	runoff channel of bore well, Australia, Great Artesian basin; 65C anaerobic	Love et al. 1993
Firmicutes	*Peptococcaceae*	*Desulfotomaculum thermocisternum DSM 10259*	7E-141	formation water, oil reservoir, Norway, North sea; anaerobic 60C	Nilsen et al. 1996
Firmicutes	*Veillonellaceae*	*Propionispora hippei DSM 15287*	3E-140	sewage treatment plant; anaerobic 37C	Abou-Zeid et al. 2004
Firmicutes	*Veillonellaceae*	*Propionispora vibrioides DSM 13305*	3E-140	compost, Germany; anaerobic 30C	Biebl et al. 2001
Firmicutes	*Peptococcaceae*	*Desulfonispora thiosulfatigenes GKNTAUT, DSM 11270*	4E-137	sewage plant, Germany; anaerobic 30C	Denger et al. 1999
Firmicutes	*Paenibacillaceae*	*Aneurinibacillus thermoaerophilus DSM 10154*	1E-136	sugar beet extraction juice, Austria, Leopoldsorf; 55C	Meier-Stauffer et al. 1996, Heyndrickx et al. 1997; synonyms: Bacillus thermoaerophilus
Firmicutes	*Peptococcaceae*	*Desulfotomaculum alkaliphilum DSM 12257*	1E-136	mixed sample of cow and pig manure, Russia, near Moscow; anaerobic 50C	Pikuta et al. 2000
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. URHA0014*	3E-135		

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. Soil724D2*	4E-135		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. Root444D2*	4E-135		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. Soil766*	4E-135		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. Soil750*	1E-133		
Firmicutes	*Paenibacillaceae*	*Paenibacillus mucilaginosus KNP414*	4E-132		
Firmicutes	*Paenibacillaceae*	*Paenibacillus mucilaginosus K02*	4E-132		
Firmicutes	*Paenibacillaceae*	*Paenibacillus mucilaginosus 3016*	4E-132		
Firmicutes	*Syntrophomonadaceae*	*Syntrophomonas wolfei Goettingen, DSM 2245B*	6E-132	anaerobic digester sludge; Germany, Goettingen; anaerobic 35C	McInerney et a. 1982 emend. Lorowitz et al. 1989
Firmicutes	*Paenibacillaceae*	*Paenibacillus tianmuensis CGMCC 1.8946*	9E-132		
Firmicutes	*Paenibacillaceae*	*Paenibacillus elgii B69*	3E-131		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. YR247*	3E-131		
Firmicutes	*Paenibacillaceae*	*Paenibacillus lautus Y412MC10*	3E-131		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. CL141A*	3E-131		
Firmicutes	*Paenibacillaceae*	*Paenibacillus ehimensis A2*	4E-131		
Firmicutes	*Paenibacillaceae*	*Paenibacillus naphthalenovorans 320-Y*	6E-131		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. MSt1*	2E-130		
Firmicutes	*Thermoanaerobacteraceae*	*Ammonifex degensii KC4*	2E-130		Hubber and Stetter 1996
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. HGF5*	7E-130		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. FSL H8-457*	8E-130		
Firmicutes	*Paenibacillaceae*	*Paenibacillus terrigena DSM 21567*	2E-129	soil; Japan, Chiba seacoast; 30C	Xie and Yokota 2007
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. Soil787*	3E-129		
Firmicutes	*Paenibacillaceae*	*Paenibacillus chondroitinus OK414*	5E-129		

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes	*Paenibacillaceae*	*Paenibacillus alginolyticus DSM 5050*	1E-128	soil; 30C	Nakamura 1987, Shida et al. 1997; synonyms: Bacillus alginolyticus
Firmicutes	*Paenibacillaceae*	*Paenibacillus vortex V453*	1E-128		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. FSL R5-808*	1E-128		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. UNC451MF*	2E-128		
Firmicutes	*Thermoactinomycetaceae*	*Planifilum fulgidum DSM 44945*	1E-127	hyperthermal compost, Japan; anaerobic 60C	Hatayama et al. 2005
Firmicutes	*Paenibacillaceae*	*Paenibacillus alvei DSM 29 (PAV)*	2E-127		
Firmicutes		*Bacillus sp. FJAT-18019*	7E-127		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. UNC217MF*	8E-127		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. CL6Col*	9E-127		
Firmicutes	*Paenibacillaceae*	*Paenibacillus alvei A6-6i-x*	9E-127		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. FJAT-22460*	9E-127		
Firmicutes	*Syntrophomonadaceae*	*Thermosyntropha lipolytica DSM 11003*	2E-126	sediment, Kenya, Lake Bogoria; anaerobic 60C	Svetlitshnyi et al. 1996
Firmicutes	*Paenibacillaceae*	*Paenibacillus uliginis N3/975*	2E-126		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. DMB20*	3E-126		
Firmicutes		*Bacillus sp. J13*	3E-126		
Firmicutes	*Paenibacillaceae*	*Paenibacillus lactis 154*	3E-126		
Firmicutes	*Paenibacillaceae*	*Paenibacillus alvei TS—15*	4E-126		
Firmicutes	*Paenibacillaceae*	*Paenibacillus assamensis DSM 18201*	7E-126	warm spring water, India, Assam; 30C	Saha et al. 2005
Firmicutes	*Alicyclobacillaceae*	*Tumebacillus flagellatus GST4*	1E-124		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. 1_12*	1E-124		
Firmicutes	*Paenibacillaceae*	*Paenibacillus senegalensis JC66*	6E-124		

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Actinobacteria	*Mycobacteriaceae*	*Mycobacterium tuberculosis (Cpn 60 2)*	6E-123		
Actinobacteria	*Mycobacteriaceae*	*Mycobacterium tuberculosis*	8E-123		
Firmicutes	*Thermoactinomycetaceae*	*Thermoactinomyces sp. DSM 45891*	4E-123		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. IHBB 10380 null replaces 81672*	1E-122		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. IHBB 10380*	1E-122		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. OSY-SE*	4E-122		
Firmicutes	*Paenibacillaceae*	*Paenibacillus dendritiformis C454*	4E-122		
Firmicutes	*Thermoactinomycetaceae*	*Shimazuella kribbensis DSM 45090*	1E-121	forest soil, South Korea; 28C	Park et al. 2007
Firmicutes	*Paenibacillaceae*	*Paenibacillus ginsengihumi DSM 21568*	2E-121	soil of a ginseng field, South Korea; 30C	Kim et al. 2008
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. JCM 10914*	2E-121		
Firmicutes	*Paenibacillaceae*	*Paenibacillus taiwanensis DSM 18679*	5E-121	farmland soil, Taiwan, Wu-Feng; 30C	Lee et al. 2007
Firmicutes	*Paenibacillaceae*	*Paenibacillus popilliae ATCC 14706*	2E-119	26C, aerobic	Haynes et al. 1961, Roberts et al. 1996; Pettersson et al. 1999
Firmicutes	*Thermoactinomycetaceae*	*Thermoactinomyces daqus H-18*	2E-117		
Firmicutes	*Thermoactinomycetaceae*	*Thermoactinomycetaceae bacterium GD1*	3E-117		
Firmicutes	*Thermoactinomycetaceae*	*Thermoactinomyces sp. CDF*	3E-115		
Firmicutes	*Thermoactinomycetaceae*	*Thermoactinomyces sp. Gus2-1*	3E-115		
Firmicutes	*Thermoactinomycetaceae*	*Thermoactinomyces vulgaris DSM 43016*	3E-115	compost; 50C	Tsilinsky 1899, Kurup et al. 1975
Firmicutes	*Veillonellaceae*	*Pelosinus fermentans A12*	1E-110		
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. CL123*	2E-108		

Phylum	*Family*	*Species*	E Value	Isolation Site	References
Firmicutes	*Paenibacillaceae*	*Paenibacillus sp. UNCCL117*	2E-108		
Firmicutes	*Clostridiaceae*	*Clostridium caminithermale DSM 15212*	2E-101	deep-sea hydrothermal vent, Atlantic Ocean Ridge; Anaerobic 42C	Brisbarre et al. 2003
Firmicutes	*Clostridiaceae*	*Clostridium halophilum ATCC 49637; DSM 5387*	1E-100	pasteurized hypersaline sediment, Maldive Islands; anaerobic 37C	Fendrich et al. 1991
Firmicutes	*Clostridiaceae*	*Geosporobacter subterraneus DSM 17957*	1E-100	water-producing deep artesian well, France, Paris Basin; anaerobic 40C	Klouche et al. 2007
Firmicutes	*Thermoanaerobacteraceae*	*Melghirimyces thermohalophilus DSM 45514*	8E-100	soil from a salt lake, Algeria, Chott Melghir; 55C	Addou et al. 2013
Firmicutes		*Composite genome from Lake Mendota Epilimnion*	3E-98		
Firmicutes	*Thermoactinomycetaceae*	*Desmospora sp. 8437*	7E-96		
Cyanobacteria		*Gloeobacter violaceus PCC 7421*	2E-93		
Cyanobacteria		*Gloeobacter violaceus PCC 7421 (re-annotation)*	3E-93		
Cyanobacteria		*Gloeobacter kilaueensis JS1*	3E-89		
Firmicutes	*Paenibacillaceae*	*Paenibacillus alvei E194 (re-annotation, Oct 2015)*	8E-83		
Firmicutes	*Paenibacillaceae*	*Gorillibacterium massiliense G5; PARTIAL*	6E-73	stool sample, wild Gorilla gorilla subsp. gorilla, Cameroon; facultative anaerobe 37C	Keita et al. 2014
Euryarchaeota	*Halobacteriaceae*	*Haloferax denitrificans S1*	3E-67	SCORE= 238 bits (608) vs. Chy 1043 bits (2697)	

Bibliography

Altschul, S. F., Madden, T. L., Schaeffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25, 3389-3402 (1997).

An, Y. J., Rowland, S. E., Robb, F. T., and Cha, S. S. Purification, crystallization and preliminary X-ray crystallographic analysis of Group III chaperonin from *Carboxydothermus hydrogenoformans*. Journal of Microbiology, Accepted for publication (2016).

Anfinsen, C. B. Principles that govern the folding of protein chains. Science 181, 223-230 (1973).

Angelucci, F., Saccoccia, F., Ardini, M., Boumis, G., Brunori, M., Di Leandro, L., Ippoliti, R., Miele, A. E., Natoli, G., Scotti, S., and Bellelli, A. Switching between the alternative structures and functions of a 2-Cys peroxiredoxin, by site-directed mutagenesis. Journal of Molecular Biology 425, 4556-4568 (2013).

Apetri, A. C. and Horwich, A. L. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. Proceedings of the National Academy of Sciences 105, 17351-17355 (2008).

Archibald, J. M., Logsdon Jr, J. M., and Doolittle, W. F. Recurrent paralogy in the evolution of archaeal chaperonins. Current Biology 9, 1053-1056 (1999).

Archibald, J. M., Blouin, C., and Doolittle, W. F. Gene Duplication and the evolution of group II chaperonins: Implications for structure and function. Journal of Structural Biology 135, 157-169 (2001).

Archibald, J. M. and Roger, A. J. Gene duplication and gene conversion shape the evolution of archaeal chaperonins. Journal of Molecular Biology 316, 1041-1050 (2002).

Azia, A., Unger, R., and Horovitz, A. What distinguishes GroEL substrates from other *Escherichia coli* proteins? The Federation of European Biochemical Societies Journal 279, 543-550 (2012).

Azzouzi, A., Steunou, A. S., Durand, A., Khalfaoui-Hassani, B., Bourbon, M., Astier, C., Bollivar, D. W., and Ouchane, S. Coproporphyrin III excretion identifies the anaerobic coproporphyrinogen III oxidase HemN as a copper target in the Cu+-ATPase mutant *copA*- of *Rubrivivax gelatinosus*. Molecular Microbiology 88, 339-351 (2013).

Baker-Austin, C., Dopson, M., Wexler, M., Sawers, R. G., and Bond, P. L. Molecular insight into extreme copper resistance in the extremophilic archaeon 'Ferroplasma acidarmanus' Fer1. Microbiology 151, 2637-2646 (2005).

Barraclough, R. and Ellis, R. J. Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. Biochimica et Biophysica Acta 608, 19-31 (1980).

Baykov, A. A., Kasho, V. N., and Avaeva, S. M. Inorganic pyrophosphate as a label in heterogenous enzyme immunoassay. Analytical Biochemistry Journal 2, 271-276 (1988).

Beckmann, R. P., Mizzen, L. E., and Welch, W. J. Interaction of Hsp 70 with newly synthesized proteins: implications for protein folding and assembly. Science 18, 850-854 (1990).

Berg, I. A. Ecological aspects of the distribution of different autotrophic CO2 fixation pathways. Applied and Environmental Microbiology 77, 1925-1936 (2011).

Bigotti, M. G and Clarke, A. R. Chaperonins: The hunt for the group II mechanism. Archives of Biochemistry and Biophysics 474, 331-339 (2008).

Bittner, A. N., Foltz, A., and Oke, V. Only One of five groEL genes is required for viability and successful symbioses in *Sinorhizobium meliloti*. Journal of Bacteriology 189, 1884-1889 (2007).

Bochkareva, E. S., Lissin, N. M., and Girshovich, A. S. Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. Nature 336, 254-257 (1988).

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature 371, 578-586 (1994).

Braeuer, S. L., Cadillo-Quiroz, H., Ward, R. J., Yavitt, J. B., and Zinder, S. H. *Methanoregula boonei* gen. nov., sp. nov., and acidiphilic methanogen isolated from an acidic peat bog. International Journal of Systematic and Evolutionary Microbiology 61, 45-52 (2011).

Burggraf, S., Fricke, H., Neuner, A., Kristjansson, J., Rouvier, P., Mandelco, L., Woese, C. R., and Stetter, K. O. *Methanococcus igneus* sp. nov., a novel hyperthermophilic methanogen from a shallow submarine hydrothermal system. Systematic and Applied Microbiology 13, 263-269 (1990).

Burnett, B. P., Horwich, A. L., and Low, K. B. A carboxy-terminal deletion impairs the assembly of GroEL and confers a pleiotropic phenotype in *Escherichia coli* K-12. Journal of Bacteriology 176, 6980-6985 (1994).

Calloni, G., Chen, T., Schermann, S. M., Chang, H. C., Genevaux, P., Agostini, F., Tartaglia, G. G., Hayer-Hartl, M., and Hartl, F. U. DnaK functions as a central hub in the *E. coli* *c*haperone network. Cell Reports 1, 251-264 (2012).

Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R., and Georgopoulos, C. Purification and properties of the groES Morphogenetic protein of *Escherichia coli*. The Journal of Biological Chemistry 261, 12414-12419 (1986).

Chaston, J. J., Smits, C., Aragao, D., Wong, A. S. W., Ahsan, B., Sandin, S., Molugu, S. K., Molugu, S. K., Bernal, R. A., Stock, D., and Stewart, A. G. Structural and functional insights into the evolution and stress adaptation of type II chaperonins. Structure 24, 364-374 (2016).

Chaudhuri, T. K., Farr, G. W., Fenton, W. A., Rospert, S., and Horwich, A. L. GroEL/ES-mediated folding of a protein too large to be encapsulated. Cell 107, 235-246 (2001).

Chaudhuri, T. K., and Gupta, P. Factors governing the substrate recognition by GroEL chaperone: a sequence correlation approach. Cell Stress & Chaperones 10, 24-36 (2005).

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E, Dehal, P. S., DeSantis, T. Z., Gihring, T. M., Lapidus, A., Lin, L. H., Lowry, S. R., Moser, D. P., Richardson, P. M., Southam, G., Wanger, G., Pratt, L. M., Andersen, G. L., Hazen, T. C., Brockman, F. J., Arkin, A. P., and Onstott, T. C. Environmental genomics reveals a single-species ecosystem deep within Earth. Science 322:275-278 (2008).

Cong, Y., Baker, M. L., Jakana, J., Woolford, D., Miller, E. J., Reissmann, S., Kumar, R. N., Redding-Johanson, A. M., Batth, T. S., Mukhopadhyay, A., Ludtke, S. J., Frydman, J., and Chiu, W. 4.0-A resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. Proceedings of the National Academy of Sciences 107, 4967-4972 (2010).

Cong, Y., Schroeder, G. F., Meyer, A. S., Jakana, J., Ma, B., Dougherty, M. T., Schmid, M. F., Reissmann, S., Levitt, M., Ludtke, S. L., Frydman, J., and Chiu, W. Symmetry-free cryo-EM structures of the chaperonin TRiC along its ATPase-driven conformational cycle. The EMBO Journal 31, 720-730 (2012).

Corspeius, N. C. and Lorimer, G. H. Measuring how much work the chaperone GroEL can do. Proceedings of the National Academy of Sciences E2451-E2459, doi:10.1073/pnas.1307837110 (2013).

Crooks G. E., Hon G., Chandonia J. M., Brenner S. E. WebLogo: A sequence logo generator. Genome Research 14, 1188-1190 (2004).

Cuellar, J., Martin-Benito, J., Scheres, S. H. W., Sousa, R., Moro, F., Lopez-Vinas, E., Gomez-Puertas, P., Muga, A., Carrascosa, J. L., and Valpuesta, J. M. The structure of CCT-Hsc70 NBD suggests a mechanisms for Hsp70 delivery of substrates to the chaperonin. Nature Structural & Molecular Biology 15, 858-864 (2008).

Darnault, C., Volbeda, A., Kim, E. J., Legrand, P., Vernede, X., Lindahl, P. A., Fontecilla-Camps, J. C. Ni-Zn-[Fe4-S4] and Ni-Ni-[Fe4-S4] clusters in closed and open alpha subunits of acetyl-CoA synthase/carbon monoxide dehydrogenase. Nature Structural & Molecular Biology 10, 271 (2003).

Davidson, M. M., Silver, B. J., Onstott, T. C., Moser, D. P., Gihring, T. M., Pratt, L. M., Boice, E. A., Sherwood Lollar, B., Lippmann-Pipke, J., Pfiffner, S. M., Kieft, T. L., Seymore, W., and Ralston, C. Capture of planktonic microbial diversity in fractures by long-term monitoring of flowing boreholes, Evander Basin, South Africa. Geomicrobiology Journal 28, 275-300 (2011).

de Boer, H. A., Comstock, L. J., and Vasser, M. The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. Proceedings of the National Academy of Sciences 80, 21-25 (1983).

Dekker, C., Willison, K. R., and Taylor, W. R. On the evolutionary origin of the chaperonins. Proteins: Structure, function, and bioinformatics 79, 1172-1192 (2011).

Dekker, C., Roe, S. M., McCormack, E. A., Beuron, F., and Willison, K. R. The crystal structure of yeast CCT reveals intrinsic asymmetry of eukaryotic cytosolic chaperonins. The EMBO Journal 30, 3078-3090 (2011).

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Baeumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemeke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H. P., Gunsalus, R. P., Fritz, H. J., and Gottschalk, G. The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. The Journal of Moleular Microbiology and Biotechnology 4, 453-461 (2002).

Dickson, R., Weiss, C., Howard, R. J., Alldrick, S. P., Ellis, R. J., Lorimer, G., Azem, A., and Vitanen, P. V. Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. The Journal of Biological Chemistry 275, 11829-11835 (2000).

Dobbek, H., Svetlitchnyi, V., Liss, J., and Meyer, O. Carbon monoxide induced decomposition of the active site [Ni-4Fe-5S] cluster of CO dehydrogenase. The Journal of American Chemical Society 126, 5382-5387 (2004).

Douglas, N. R., Reissmann, S., Zhang, J., Chen, B., Jakana, J., Kumar, R., Chiu, W., and Frydman, J. Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. Cell 144, 240-252 (2011).

Endo, A., and Kurusu, Y. Identification of *in vivo* Substrates of the chaperonin GroEL from *Bacillus subtilis*. Bioscience, Biotechnology, and Biochemistry 71, 1073-1077 (2007).

Ewalt, K. L., Hendrick, J. P., Houry, W. A., and Hartl, F. U. *In vivo* observation of polypeptide flux through the bacterial chaperonin system. Cell 90, 491-500 (1997).

Fares, M. A. The evolution of protein moonlighting: adaptive traps and promiscuity in the chaperonins. Biochemical Society Transactions 42, 1709-1714 (2014).

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. Journal of Bacteriology 171, 1379-1385 (1989).

Feil, W. S., Feil, H., and Copeland, A. Bacterial genomic DNA isolation using CTAB. http://1ofdmq2n8tc36m6i46scovo2e.wpengine.netdna-cdn.com/wp-content/uploads/2014/02/JGI-Bacterial-DNA-isolation-CTAB-Protocol-2012.pdf

Fenton, W. A., Kashi, Y., Furtak, K., and Horwich, A. L. Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371, 614-619 (1994).

Fenton, W. A., and Horwich, A. L. Chaperonin-mediated protein folding: fate of substrate polypeptide. Quaterly Reviews of Biophysics 36, 229-256 (2003).

Fisher, H. M., Schneider, K., Babst, M., and Hennecke, H. GroEL chaperonins are required for the formation of a functional nitrogenase in *Bradyrhizobium japonicum*. Archives of Microbiology 171, 279-289 (1999).

Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T., and Taguchi, H. A systematic survey of *in vivo* obligate chaperonin-dependent substrates. The EMBO Journal 29, 1552-1564 (2010).

Gao, Y., Thomas, J. O., Chow, R. L., Lee, G. H., and Cowan, N. J. A cytoplasmic chaperonin that catalyzes B-actin folding. Cell 69, 1043-1050 (1992).

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook, Humana Press, 571-607 (2005).

Gencic, S., Duin, E. C., and Grahame, D. A. Tight coupling of partial reactions in the acetyl-coA decarbonylase/synthase (ACDS) multienzyme complex from *Methanosarcina thermophila*. Journal of Biological Chemistry 285, 15450-15463 (2010).

Genevaux, P., Georgopoulos, C., and Kelley, W. L. The Hsp70 chaperone machines of *Escherichia coli:* a paradigm for the partition of chaperone functions. Molecular Microbiology 66, 840-857 (2007).

Georgopoulos, C. P. and Hohn, B. Identification of a host protein necessary for bacteriophage morphogenesis (the *groE* gene product). Proceedings of the National Academy of Sciences 75, 131-135 (1978).

Goldberger, R. F., Epstein, C. J., and Anfinsen, C. B. Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. Journal of Biological Chemistry 238, 628 (1963).

Golyshina, O. V., Yakimov, M. M., Luensdorf, H., Ferrer, M., Nimtz, M., Timmis, K. N., Wray, V., Tindall, B. J., and Golyshin, P. N. *Acidiplasma aeolicum* gen. nov., sp. nov., a euryarchaeon of the family *Ferroplasmaceae* isolated from a hydrothermal pool, and transfer of Ferroplasma cupricumulans to *Acidiplasma cupricumulans* comb. nov. International Journal of Systematic and Evolutionary Microbiology 59, 2815-2823 (2009).

Gong, W., Hao, B., Wei, Z., Ferguson Jr., D. J., Tallant, T., Krzycki, J. A., and Chan, M. K. Structure of the alpha2 epsilon2 Ni-dependent CO dehydrogenase component of the *Methanosarcina barkeri* acetyl-CoA decarbonylase/synthase complex. Proceedings of the National Academy of Sciences 105, 9558-9563 (2008).

Guagliardi, A., Cerchia, L., Bartolucci, S., and Rossi, M. The chaperonin from the archaeon *Sulfolobus solfataricus* promotes correct refolding and prevents thermal denaturation *in vitro*. Protein Science 3, 1436-1443 (1994).

Guglielmi, G., Mazodier, P., Thompson, C. J., and Davies, J. A survey of the heat shock response in four Streptomyces species reveals two groEL-like genes and three GroEL-like proteins in *Streptomyces albus*. Journal of Bacteriology 173, 7374-7381 (1991).

Gupta, A. J., Halder, S., Milicic, G., Hartl, F. U., and Hayer-Hartl, M. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. Journal of Molecular Biology 426, 2739-2754 (2014).

Gupta, R. S., Picketts, D. J., and Ahmad, S. A novel ubiquitous protein 'chaperonin' supports the endosymbiotic origin of mitochondrion and plant chloroplast. Biochemical and Biophysical Research Communications 163, 780-787 (1989).

Gupta, R. S. Sequence and structural homology between a mouse T-complex protein TCP-1 and the 'chaperonin' family of bacterial (GroEL, 60-65kDa heat shock antigen) and eukaryotic proteins. Biochemistry International 20, 833-841 (1990).

Gupta, R. S. Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. Molecular Microbiology 15, 1-11 (1995).

Hansen, W. J., Cowan, N. J., and Welch, W. J. Prefoldin-nascent chain complexes in the folding of cytoskeletal proteins. The Journal of Cell Biology 145, 265-277 (1999).

Hawkes, R. B., Franzmann, P. D., O'hara, G., and Plumb, J. J. Ferroplasma cupricumulans sp. nov., a novel moderately thermophilic, acidophilic archaeon isolated from an industrial-scale chalcocite bioleach heap. Extremophiles 10, 525-530 (2006).

Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, R. J. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333, 330-334 (1988).

Hendrix, R. W. Purification and properties of groE, a host protein involved in bacteriophage assembly. Journal of Molecular Biology 129, 375-92(1979).

Hendrix, R. W., and Tsui, L. Role of the host in virus assembly: Cloning of the *Escherichia coli groE* gene and identification of its protein product. Proceedings of the National Academy of Sciences 75, 136-139 (1978).

Herendeen, S. L., VanBogelen, R. A., and Neidhardt, F. C. Levels of major proteins of *Escherichia coli* during growth at different temperatures. Journal of Bacteriology 139, 185-194 (1979).

Hirtreiter, A. M., Calloni, G., Forner, F., Scheibe, B., Puype, M., Vandekerckhove, J., Mann, M., Hartl, F. U., and Hayer-Hartl, M. Differential substrate specificity of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. Molecular Microbiology 74, 1152-1168 (2009).

Hoff, K. G., Silberg, J. J., and Vickery, L. E. Interaction of the iron-sulfur cluster assembly protein IscU with the Hsc66/Hsc20 molecular chaperone system of *Escherichia coli*. Proceedings of the National Academy of Sciences 97, 7790-7795 (2000).

Hoff, K. G., Ta, D. T., Tapley, T. L., Silberg, J. J., and Vickery, L. E. Hsc66 substrate specificity is directed toward a discrete region of the iron-sulfur cluster template protein IscU. Journal of Biological Chemistry 277, 27353-27359 (2002).

Hohn, T., Hohn, B., Engel, A. and Wurtz, M. Isolation and characterization of the host protein groE involved in bacteriophage Lamba assembly. Journal of Molecular Biology 129, 359-373 (1979).

Horovitz, A., Fridmann, Y., Kafri, G., and Yifrach, O. Review: allostery in chaperonins. Journal of Structural Biology 135, 104-114 (2001).

Horst, R., Bertelsen, E. B., Fiaux, J., Wider, G., Horwich, A. L., and Wuethrich, K. Direct NMR observation of a substrate protein bound to the chaperonin GroEL. Proceedings of the National Academy of Sciences 102, 12748-12753 (2005).

Horwich, A. L., Apetri A. C., and Fenton, W. A. The GroEL/GroES cis cavity as a passive anti-aggregation device. FEBS Letters 583, 2654-2662 (2009).

Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F. U. Identification of *in vivo* substrates of the chaperonin GroEL. Nature 401, 147-154 (1999).

Hutchinson, E. G., Tichelaar, W., Hofhaus, G., Weiss, H., and Leonard, K. R. Identification and electron microscopic analysis of a chaperonin oligomer from *Neurospora crassa* mitochondria. The EMBO Journal 8, 1485-1490 (1989).

Inoue, T., Takao, K, Fukuyama, Y., Yoshida, T., and Sako, Y. Over-expression of carbon monoxide dehydrogenase-I with an accessory protein co-expression: a key enzyme for carbon dioxide reduction. Bioscience, Biotechnology, and Biochemistry 78, 582-587 (2014).

Ishihama, A. T., Ikeuchi, T., and Yura, T. A novel adenosine triphosphatase isolated from RNA polymerase preparation of *Escherichia coli.* I. Copurification and separation. Journal of Biochemistry 79, 917-925 (1976).

Ishii, S., Kosaka, T., Hori, K., Hotta, Y., and Watanabe, K. Coaggregation Facilitates Interspecies Hydrogen Transfer between *Pelotomaculum thermopropionicum* and *Methanothermobacter thermautotrophicus*. Applied and Environmental Microbiology 71, 7838-7845 (2005).

Itaya, K. and Ui, M. A new micromethod for the colorimetric determination of inorganic phosphate. Clinica Chimica Acta 14, 361-366 (1966).

Jang, H. H., Kim, S. Y., Park, S. K., Jeon, H. S., Lee, Y. M., Jung, J. H., Lee, S. Y., Chae, H. B., Jung, Y. J., Lee, K. O., Lim, C. O., Chung, W. S., Bahk, J. D., Yun, D. J., Cho, M. J., and Lee, S. Y. Phosphorylation and concomitant structural changes in human 2-Cys peroxiredoxin isotype I differentially regulate its peroxidase and molecular chaperone functions. FEBS Letters 580, 351-355 (2006).

Jeoung, J. H., Giese, T., Gruenwald, M., and Dobbek, H. CooC1 from *Carboxydothermus hydrogenoformans* is a nickel-binding ATPAse. Biochemistry 48, 11505-11513 (2009).

Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R., and Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. Cell 159, 1042-1055 (2014).

Johnson, R. B., Fearon, K., Mason, T., and Jindal, S. Cloning and characterization of the yeast chaperonin HSP60 gene. Gene 84, 295-302 (1989).

Jones D.T., Taylor W.R., and Thornton J.M. The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences 8, 275-282 (1992).

Jungbluth SP, Lin HT, Cowen JP, Glazer BT, and Rappé S. Phylogenetic diversity of microorganisms in subseafloor crustal fluids from holes 1025C and 1026B along the Juan de Fuca flank. Frontiers in Microbiology 5,119. doi: 10.3389/fmicb.2014.00119 (2014).

Kaster, A.K., Moll, J., Parey, K., and Thauer, R. K. Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proceedings of the National Academy of Sciences 108, 2891-2986 (2011).

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10, 845-858 (2015).

Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., and Hartl, F. U. Proteome-wide Analysis of Chaperonin-Dependent Protein folding in *Escherichia coli*. Cell 122, 209-220 (2005).

Kim, P. S. and Baldwin, R. L. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annual Review of Biochemistry 51, 459-489 (1982).

Kim, S., Willison, K. R., and Horwich, A. L. Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. Trends in Biochemical Sciences 19, 543-548 (1994).

Klein, G. and Georgopoulos, C. Identification of important amino acid residues that modulate binding of *Escherichia coli* GroEL to its various cochaperones. Genetics 158, 507-517 (2001).

Klumpp, M., Baumeister, W., and Essen, L. O. Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. Cell 91, 263-270 (1997).

Klunker, D., Haas, B., Hirtreiter, A., Figueiredo, L., Naylor, D. J., Pfeifer, G., Mueller, V., Deppenmeier, U., Gottschalk, G., Hartle, F. U., and Hayer-Hartl, M. Coexistence of group I and group II chaperonins in the archaeon *Methanosarcina mazei*. The Journal of Biological Chemistry 278, 33256-33267 (2003).

Kumar, C. M. S., Khare, G., Srikanth, C. V., Tyagi, A. K., Sardesai, A. A., and Mande S. C. Facilitated oligomerization of Mycobacterial GroEL: Evidence for phosphorylation-mediated oligomerization. Journal of Bacteriology 191, 6525-6538 (2009).

Kumar, C. M. S., Mande, S. C., and Mahajan, G. Multiple chaperonins in bacteria- novel functions and non-canonical behaviors. Cell Stress and Chaperones 20, 555-574 (2015).

Kumar S., Stecher G., and Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution (2015).

Labonte, J. M., Field, E. K., Lau, M., Chivian, D., Heerden, E. V., Wommack, K. E., Kieft, T. L., Onstott, T. C., and Stepanauskas, R. Single cell genomics indicates horizontal gene transfer and viral infection in a deep subsurface Firmicutes population. Frontiers in Microbiology 6, 349. doi: 10.3389/fmicb.2015.00349 (2015).

Laksanalamai, P., Narayan, S., Luo, H., and Robb, F. T. Chaperone action of a versatile small heat shock protein from *Methanococcoides burtonii*, a cold adapted archaeon. Proteins: Structure, Function, and Bioinformatics 75, 275-282 (2008).

Laskey, R. A., Honda, B. M., Mills, A. D., and Finch, J. T. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature 275, 416-420 (1978).

Leitner, A., Joachimiak, L. A., Bracher, A., Moenkemeyer, L., Walzthoeni, T., Chen, B., Pechmann, S., Holmes, S., Cong, Y., Ma, B., Ludtke, S., Chiu, W., Hartl, F. U., Aebersold, R., and Frydman, J. The molecular architecture of the eukaryotic chaperonin TRiC/CCT. Structure 20, 814-825 (2012).

Levinthal, C. Are there pathways for protein folding? Journal de Chimie Physique 65, 44-45 (1968).

Libich, D. S., Fawzi, N. L., Ying, J., and Clore, G. M. Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. Proceedings of the National Academy of Sciences 110, 11361-11366 (2013).

Lim, N. C. H., and Jackson, S. E. Mechanistic insights into the folding of knotted proteins i*n vitro* and i*n vivo*. Journal of Molecular Biology 427, 248-258 (2015).

Lopez, T., Dalton, K., and Frydman, J. The mechanism and function of group II chaperonins. Journal of Molecular Biology 427, 2919-2930 (2015).

Lopez-Garcia, P., Brouchier, C., Moreira, D., and Rodriguez-Valera, F. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. Environmental Microbiology 6, 19-34 (2004).

Lorimer, G. H. A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. The FASEB Journal 10, 5-9 (1996).

Lund, P. A. Multiple Chaperonins in bacteria- why so many? FEMS Microbiology Reviews 33, 785-800 (2009).

Luo, H. and Robb, F. T. A modulator domain controlling thermal stability in the group II chaperonins of archaea. Archives of Biochemistry and Biophysics 512, 111-118 (2011).

Macario, A. J. L., Dugan, C. B., and E. C. Macario. A *dnaK* homolog in the archaeabacterium *Methanosarcina mazei* S6. Gene 108, 133-137 (1991).

Magnabosco, C., Ryan, K., Lau, M. C., Kuloyo, O., Sherwood Lollar, B., Kieft, T. L., van Heerden, E., and Onstott, T. C. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. The ISME Journal 10, 730-741 (2016).

Marco, S., Urena, D., Carrascosa, J. L., Waldmann, T., Peters, J., Hegerl, R., Pfeifer, G., Stack-Kongehl, H., and Baumeister, W. The molecular chaperone TF55: Assessment of symmetry. FEBS Letters 341, 152-155 (1994).

Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntermann, M., Anderson, I., Mavromatis, K., Ivanova, N. N., and Kyrpides, N. C. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Research 40, D115-D122 (2012).

Martel, R., Cloney, L. P., Pelcher, L. E., and Hemmingsen, S. M. Unique composition of plastid chaperonin-60: alpha and beta polypeptide-encoding genes are highly divergent. Gene 94, 181-187 (1990).

Mazodier, P., Guglielmi, G., Davies, J., and Thompson, C. J. Characterization of the *groEL*-Like Genes in *Streptomyces albus*. Journal of Bacteriology 173, 7382-7386 (1991).

McMullin, T. W., and Hallberg, R. L. A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in *Tetrahymena thermophila*. Molecular and Cellular Biology 7, 4414-4423 (1987).

Motojima, F. and Yoshida, M. Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. The EMBO Journal 29, 4008-4019 (2010).

Nakagawa, A., Moriya, K., Arita, M., Yamamoto, Y., Kitamura, K., Ishiguro, N., Kanzaki, T., Oka, T., Makabe, K., Kuwajima, K., and Yohda, M. Dissection of the ATP-dependent conformational change cycle of a group II chaperonin. The Journal of Molecular Biology 426, 447-459 (2014).

Narberhaus, F. Negative regulation of bacterial heat shock genes. Molecular Microbiology 31, 1-8 (1999).

Nei, M. and Kumar, S. Molecular evolution and phylogenetics. Oxford University Press, New York (2000).

Neidhardt, F. C., VanBogelen, R. A., and Vaughn, V. The genetics and regulation of heat-shock proteins. Annual Review of Genetics 18, 295-329 (1984).

Okochi, M., Nomura, T., Zako, T., Arakawa, T., Iizuka, R., Ueda, H., Funatsu, T., Leroux, M., and Yohda, M. Kinetics and binding sites for interaction of the prefoldin with a group II chaperonin. The Journal of Biological Chemistry 279, 31788-31795 (2004).

Onstott, T. C., Magnasbosco, C., Aubrey, A. D., Burton, A. S., Dworkin, J. P., Elsila, J. F., Grunsfeld, S., Cao, B. H., Hein, J. E., Glavin, D. P., Kieft, T. L., Silver, B. J., Phelps, T. J., van Heerden, E., Opperman, D. J., and Bada, J. L. Does aspartic acid racemization constrain the depth limit of the subsurface biosphere? Geobiology 12, 1-9 (2014).

Pappenberger, G., Wilsher, J. A., Roe, S. M., Counsell, D. J., Willison, K. R., and Pearl, L. H. Crystal structure of the CCT gamma apical domain: implications for substrate binding to the eukaryotic cytosolic chaperonin. The Journal of Molecular Biology 318, 1367-1379 (2002).

Parsell, D. A., and Sauer, R. T. Induction of a heat shock-like response by unfolded protein in Escherichia coli: dependence on protein level not protein degradation. Genes & Development 3, 1226-1232 (1989).

Pereira, J. H., Ralston, C. Y., Douglas, N. R., Meyer, D., Knee, K. M., Goulet, D. R., King, J. A., Frydman, J., and Adams, P. D. Crystal structures of a group II chaperonin reveal the open and closed states associated with the protein folding cycle. The Journal of Biological Chemistry 285, 27958-27966 (2010).

Pereira, J. H., Ralston, C. Y., Douglas, N. R., Kumar, R., Lopez, T., McAndrew, R. P., Knee, K. M., King, J. A., Frydman, J., and Adams, P. D. Mechanism of nucleotide sensing in group II chaperonins. The EMBO Journal 31, 731-740 (2012).

Phipps, B. M., Hoffmann, A., Stetter, K. O., and Baumeister, W. A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. The EMBO Journal 10, 1711-1722 (1991).

Phipps, B. M., Typke, D., Hegerl, R., Volker, S., Hoffmann, A., Stetter, K. O., and Baumeister, W. Structure of a molecular chaperone from a thermophilic archaebacterium. Nature 361, 475- 477 (1993).

Picketts, D. J., Mayanil, C. S. K., and Gupta, R. S. Molecular cloning of a Chinese hamster mitochondrial protein related to the "chaperonin" family of bacterial and plant proteins. The Journal of Biological Chemistry 264, 12001-12008 (1989).

Popp, S. L., and Reinstein, J. Functional characteriation of the DnaK chaperone system from the archaeon *Methanothermobacter thermautotrophicus dH*. FEBS Letters 583, 573-578 (2009).

Ragsdale, S. W. Enzymology of the Wood-Ljungdahl pathway of acetogenesis. Annals of the New York Academy of Sciences 1125, 129-136 (2008).

Ramos, A. R., Grein, F., Oliveira, G. P., Venceslau, S. S., Keller, K. L., Wall, J. D., and Pereira, I. A. The FlxABCD-HdrABC proteins correspond to a novel NADH dehydrogenase/heterodisulfide reductase widespread in anaerobic bacteria involved in ethanol metabolism in *Desulfovibrio vulgaris* Hildenborough. Environmental Microbiology 17, 2288-2305 (2015).

Reissmann, S., Parnot, C., Booth, C. R., Chiu, W., and Frydman, J. Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. Nature Structural & Molecular Biology 14, 432-440 (2007).

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W. T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., Rubin, E. M., Hugenholtz, P., and Woyke, T. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499, 431-437 (2013).

Rodriguez-Quinones, F., Maguire, M., Wallington, E. J., Gould, P. S., Yerko, V., Downie, J. A., and Lund, P. A. Two of the three *groEL* homologues in *Rhizobium leguminosarum* are dispensable for normal growth. Archives of Microbiology 183, 253-265 (2005).

Rommelaere, H., Van Troys, M., Gao, Y., Melki, R., Cowan, N. J., Vandekerckhove, J., and Ampe, C. Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and seven related subunits. Proceedings of the National Academy of Sciences 90, 11975-11979 (1993).

Ruessmann, F., Stemp, M. J., Moenkemeyer, L., Etchells, S. A., Bracher, A., and Hartl, F. U. Folding of large multidomain proteins by partial encapsulation in the chaperonin TRiC/CCT. Proceedings of the National Academy of Sciences 109, 21208-21215 (2012).

Saccoccia, F., Di Micco, P., Boumis, G., Brunori, M., Koutris, I., Miele, A. E., Morea, V., Sriratana, P. Williams, D. L., Bellelli, A., and Angelucci, F. Moonlighting by different stressor: Crystal structure of the chaperone species of a 2-Cys peroxiredoxin. Structure 20, 429-439 (2012).

Sahlan, M., Zako, T., Tai, P. T., Ohtaki, A., Noguchi, K., Maeda, M., Miyatake, H., Dohmae, N., and Yohda, M. Thermodynamic characterization of the interaction between prefoldin and Group II chaperonin. The Journal of Molecular Biology 399, 628-636 (2010).

Saitou, N. and Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406-425 (1987).

Sant'Anna, F. H., Lebedinsky, A. V., Sokolova, T. G., and Gonzalez, J. M. Analysis of three genomes within the thermophilic bacterials species *Caldanaerobacter subterraneus* with a focus on carbon monoxide dehydrogenase evolution and hydrolase diversity. BMC Genomics 16, 757 (2015).

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spuge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Research 29, 2994-3005 (2001).

Schneider T. D., Stephens R. M. Sequence Logos: A New Way to Display Consensus Sequences. Nucleic Acids Research 18, 6097-6100 (1990).

Sela, M., White, F. H., and Anfinsen, C. B. Reductive Cleavage of Disulfide Bridges in Ribonuclease. Science 125, 691 (1957).

Shimamura, T., Koike-Takeshita, A., Yokoyama, K., Masui, R., Murai, N., Yoshida, M., Taguchi, H., and Iwata, S. Crystal structure of the native chaperonin complex from *Thermus thermophilus* revealed unexpected asymmetry at the cis-cavity. Structure 12, 1471-1480 (2004).

Shomura, Y., Yoshida, T., Iizuka, R., Maruyama, T., Yohda, M., and Miki, K. Crystal structure of the Group II chaperonin from *Thermococcus* strain KS-1: steric hindrance by the substituted amino acid, and inter-subunit rearrangement between two crystal forms. The Journal of Molecular Biology 335, 1265-1278 (2004).

Shtilerman, M., Lorimer, G. H., and Englander, S. W. Chaperonin function: folding by forced unfolding. Science 284, 822-825 (1999).

Siegart, R., Leroux, M. R., Scheufler, C., Hartl, F. U., and Moarefi, I. Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. Cell 103, 621-632 (2000).

Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. Nature Reviews Neuroscience 4, 46-90 (2003).

Spang, A., Poehlein, A., Offre, P., Zumbraegel, S., Haider, S., Rychlik, N., Nowka, B., Schmeisser, C., Lebedeva, E. V., Rattei, T., Boehm, C., Schmid, M., Galushko, A., Hatzenpichler, R., Weinmaier, T., Daniel, R., Schleper, C., Spieck, E., Streit, W., and Wagner, M. The genome of the ammonia-oxidizing *Candidatus* Nitrosphaera gargensis: insights into metabolic versatility and environmental adaptations. Environmental Microbiology 14, 3122-3145 (2012).

Sparrer, H., Rutkat, K., and Buchner, J. Catalysis of protein folding by symmetric chaperone complexes. Proceedings of the National Academy of Sciences 94, 1096-1100 (1997).

Stan, G., Brooks, B. R., Lorimer, G. H., and Thirumalai, D. Identifying natural substrates for chaperonins using a sequence-based approach. Protein Science 14, 193-201 (2005).

Stan, G., Brooks, B. R., Lorimer, G. H., and Thirumalai, D. Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. Proceedings of the National Academy of Sciences 103, 4433-4438 (2006).

Sternberg, N. Properties of a mutant of *Escherichia coli* defective in bacteriophage λ head formation (*groE*). Journal of Molecular Biology 76, 1-23 (1973).

Sternlicht, H., Farr, G. W., Sternlicht, M. L., Driscoll, J. K., Willison, K., and Yaffe, M. B. The t-complex polypeptide 1 complex is a chaperonin for tubulin and actin *in vivo*. Proceedings of the National Academy of Sciences 90, 9422-9426 (1993).

Studier, F. W. Protein production by auto-induction in high density shaking cultures. Protein Expression and Purification 41, 207-234 (2005).

Svetlichny, V. A., Sokolova, T. G., Gerhardt, M., Ringpfeil, M., Kostrikina, N. A., and Zavarzin, G. A.. *Carboxdyothermus hydrogenoformans* gen. nov. sp., a CO-utilizing thermophilic anaerobic bacterium from hydrothermal environments of Kunashir Island. Systematic and Applied Microbiology 14, 254-260 (1991).

Svetlichnyi, W., Dobbek, H., Meyer-Klaucke, W., Meins, T., Thiele, B., Roemer, P., Huber, R., Meyer, O. A functional Ni-Ni-[4Fe-4S] cluster in the monomeric acetyl-CoA synthase from *Carboxydothermus hydrogenoformans*. Proceedings of the National Academy of Sciences 101, 446-451 (2004).

Svetlichnyi, V., Peschel, C., Acker, G., and Meyer, O. Two membrane-associated NiFeS-carbon monoxide dehydrogenases from the anaerobic carbon-monoxide-utilizaing

eubacterium *Carboxydothermus hydrogenoformans*. Journal of Bacteriology 183, 5134-5144 (2001).

Techtmann, S. M. 2009. A genomic and biochemical charactertization of carbon monoxide utilizing thermophilic bacteria. Dissertation, University of Maryland Baltimore.

Techtmann S. M., and F. T. Robb. Archaeal-like chaperonins in bacteria. Proceedings of the National Academy of Sciences 107, 20269-20274 (2010).

Techtmann, S. M., Lebedinsky, A. V., Colman, A. S., Sokolova, T. G., Woyke, T., Goodwin, L., and Robb, F. T. Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenase. Fronties in Microbiology 3, doi: 10.3389/fmicb.2012.00132 (2012).

Teixeira, F., Castro, H., Cruz, T., Tse, E., Koldewey, P., Southworth, D. R., Tomas, A. M., Jakob, U. Mitochondrial peroxiredoxin functions as a crucial chaperone reservoir in *Leishmani infantum*. Proceedings of the National Academy of Sciences E616-E624, (2015).

Tilly, K., Murialdo, H., and Georgopoulos, C. Identification of a second *Escherichia coli* groE gene whose product is necessary for bacteriophage morphogenesis. Proceedings of the National Academy of Sciences 78, 1629-1633 (1981).

Tomlinson, G. A., Jahnke, L. L., and Hochstein, L. I. *Halobacterium denitrificans* sp. nov., an extremely halophilic denitrifying bacterium. International Journal of Systematic Bacteriology 36, 66-70 (1986).

Trent, J. D., Nimmesgern, E., Wall, J. S., Ulrich Hartl, F., and Horwich, A. L. A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. Nature 354, 490-493 (1991).

Tsytlonok, M. and Itzhaki, L. S. The how's and why's of protein folding intermediates. Archives of Biochemistry and Biophysics 531, 14-23 (2013).

Vainberg, I. E., Lewis, S. A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H. L., Cowan, N. J. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. Cell 93, 863-873 (1998).

Venetianer, P., and Straub, F. B. The enzymatic reactivation of reduced ribonuclease. Biochimica et Biophysica Acta 67, 166 (1963).

Waldinger, D., Subramanian, A. R., and Cleve, H. The polymorphic human chaperonin protein HuCha60 is a mitochondrial protein sensitive to heat shock and cell transformation. European Journal of Cell Biology 50, 435-441 (1989).

Wang, Y., Zhang, W. Y., Zhang, S., Li, J., Li, Z. F., Tan, A. G., Zhang, T. T., Wu, Z. H., Liu, H., and Li, Y. Z. Mechanisms involved in the functional divergence of duplicated GroEL chaperonins in *Myxococus xanthus* DK1622. PLOS Genetics 9, e1003306 (2013).

Warnecke, T. Loss of the DnaK-DnaJ-GrpE chaperone system among the Aquificales. Molecular Biology and Evolution 29, 3485-3495 (2012).

Wasmund, K., Schreiber, L., Lloyd, K. G., Petersen, D. G., Schramm, A., Stepanauskas, R., Jorgensen, B. B., and Adrian, L. Genome sequencing of a single cell of the widely distributed marine subsurface Dehalococcoidia phylum Chloroflexi. The ISME Journal 8, 383-397 (2014).

Wasserfallen, W., Noelling, J., Pfister, P., Reeve, J., and Conway de Macario, E. Phylogenetic analysis of 18 thermophilic *Methanobacterium* isolates supports the proposals to create a new genus, *Methanothermobacter* gen. nov., and to reclassify several isolates in three species, *Methanothermobacter thermautotrophicus* comb. nov., *Methanothermobacter wolfeii* comb. nov., and *Methanothermobacter marburgensis* sp. nov. International Journal of Systematic and Evolutionary Microbiology 50, 43-53 (2000).

Weaver, J. and Rye, H. S. The C-terminal tails of the bacterial chaperonin GroEL stimulate protein folding by directly altering the conformation of a substrate protein. The Journal of Biological Chemistry 289, 23219-23232 (2014).

Williams, W. T., Codoner, F. M., Toft, C., and Fares, M. A. Two chaperonin systems in bacterial genomes with distinct ecological roles. Trends in Genetics 26, 47-51 (2010).

Woese, C. R. and Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proceedings of the National Academy of Sciences 74, 5088-5090 (1977).

Woese, C. R., Kandler, O., and Wheelis, M. L. Towards a natural system of oragnisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences 87, 4576-4579 (1990).

Wright, P. E. and Dyson, J. H. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of Molecular Biology 293, 321-331 (1999).

Wu, M. et al. Life in Hot Carbon Monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. PLoS Genetics 1, e65. DOI:10.1371/journal.pgen.0010065 (2005).

Wyganowski, K. T., Kaltenbach, M., and Tokuriki, N. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. Journal of Molecular Biology 425, 3403-3414 (2013).

Xu, Z., Horwich, A. L., and Sigler, P. B. The crystal structure of the asymmetric GroEL-ES-(ADP)7 chaperonin complex. Nature 388, 741-750 (1997).

Yamamoto, Y. Y., Abe, Y., Moriya, K., Arita, M., Noguchi, K., Ishii, N., Sekiguchi, H., Sasaki, Y. C., and Yohda, M. Inter-ring communication is dispensable in the reaction cycle of group II chaperonins. Journal of Molecular Biology 426, 2667-2678 (2014).

Yang, D., Ye, X., and Lorimer, G. H. Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. Proceedings of the National Academy of Sciences E4298-E4305 (2013).

Yifrach, O., and Horovitz, A. Nested Cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. Biochemistry 34, 5303-5308 (1995).

Yoshida, T., Kawaguchi, R., Taguchi, H., Yoshida, M., Yasunaga, T., Wakabayshi, T., Yohda, M., and Maruyama, T. Archaeal group II chaperonin mediates protein folding in the cis-cavity without a detachable GroES-like co-chaperonin. Journal of Molecular Biology 315, 73-85 (2002).

Zeikus, J. G., and Wolfe, R. S. *Methanobacterium thermoautotrophicus* sp. n., an anaerobic, autotrophic, extreme thermophile. Journal of Bacteriology 109, 707-713 (1972).

Zhang, J., Baker, M. L., Schroeder, G. F., Douglas, N. R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C. J., Levitt, M., Ludtke, S. J., Frydman, J., and Chiu, W. Mechanism of folding chamber closure in a group II chaperonin. Nature 463, 379-384 (2010).

Zhang, K., Wang, L., Liu, Y., Chan, K. Y., Pang, X., Schulten, K., Dong, Z., and Sun, F. Flexible interwoven termini determines the thermal stability of thermosomes. Protein Cell 4, 432-444 (2013).

Zhalaina, K. V., Dias, R., Leonard, M. T., Dorr de Quadros, P., Camargo, F. A. O., Drew, J. C., Farmerie, W. G., Daroub, S. H., and Triplett, E. W. Genome sequence of *Candidatus* Nitrosphaera evergladensis from group I.1b enriched from Everglades soil reveals novel genomic features of the ammonia-oxidizing Archaea. PLoS ONE 9, e101648 (2014).

Zhilina, T. N., and Svetlichnaya, T. P. Ultrafine structure of *Methanohalobium evestigatum*, an extremely halophilic methanogenic bacterium. Microbiology 58, 248-253 (1989).

Zhuang, W. Q., Yi, S., Bill, M., Brisson, V. L., Feng, X., Men, Y., Conrad, M. E., Tang, Y. J., and Alvarez-Cohen, L. Incomplete Wood-Ljungdahl pathway facilitates one-carbon metabolism in organohalide-respiring *Dehalococcoides mccartyi*. Proceedings of the National Academy of Sciences 111, 6419-6424 (2014).

Zweig, M., and Cummings, D. J. Cleavage of head and tail protein during bacteriophage T5 assembly: Selective host involvement in the cleavage of a tail protein. Journal of Molecular Biology. 80, 505-518 (1973).