Structural and Biochemical Studies on Disease Associated Proteins: Heat Shock Protein 60 and HIV Capsid

Alejandro Rodriguez

The University of Texas at El Paso

Follow this and additional works at: https://scholarworks.utep.edu/open_etd

Part of the Biochemistry Commons

Recommended Citation

https://scholarworks.utep.edu/open_etd/3539

This is brought to you for free and open access by ScholarWorks@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.
STRUCTURAL AND BIOCHEMICAL STUDIES ON DISEASE ASSOCIATED PROTEINS:

HEAT SHOCK PROTEIN 60 AND HIV CAPSID

ALEJANDRO RODRIGUEZ

Doctoral Program in Chemistry

APPROVED:

Ricardo A. Bernal, Ph.D., Chair

Zacariah L. Hildenbrand, Ph.D.

Manuel Llano, M.D., Ph.D.

Sreeprasad Sreenivasan, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School
Dedication

For my parents, to whom I owe everything I am and everything I will be.
STRUCTURAL AND BIOCHEMICAL STUDIES ON DISEASE ASSOCIATED PROTEINS:

HEAT SHOCK PROTEIN 60 AND HIV CAPSID

by

ALEJANDRO RODRIGUEZ

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry and Biochemistry

THE UNIVERSITY OF TEXAS AT EL PASO

May 2022
Abstract

Structural studies on proteins provide valuable knowledge, from information about their fundamental nature, to laying a foundation for rational drug design. Here, we study two proteins, a human mitochondrial chaperonin heat shock protein 60 (Hsp60) and the main structural HIV-1 protein capsid (CA). We focus on the structural nature of both proteins to try to understand their role in disease and elucidate the mechanism of potential drug treatments. Hsp60 is a chaperonin found in human mitochondria that is responsible for assisting client proteins reach their native, functional, state. The mechanism by which it catalyzes protein folding, however, has not been fully elucidated. Much of what is known about Hsp60 is borrowed from its bacterial homolog GroEL, but recently, data has shown that they may employ significantly different mechanisms. We investigate the mechanism by attempting to solve the structure of the chaperonin in its different conformational states using cryo-electron microscopy. Our results indicate that Hsp60 functions via a highly dynamic mechanism involving significant conformational changes. We obtained cryo-EM structures that show Hsp60 in its apo conformation forming a double ring complex, followed by an ATP bound double ring structure in complex with its co-chaperonin Hsp10, and a novel single ring ADP bound intermediate. Our data suggests that Hsp60 incorporates conformations that have not been observed in GroEL, indicating a unique and novel mechanism of action. We also use EM to study the interaction between HIV capsid protein and functionalized fullerenes, some of which have been recently shown to be potent HIV maturation inhibitors. Our studies consist of creating in-vitro assembly models to evaluate the effect of fullerenes on capsid assembly inhibition. Currently, there are no HIV treatments that target HIV capsid assembly. New targets for drug design are crucial due to the rapid emergence of drug-resistant HIV strains, resulting from a very high mutation rate in HIV.
# Table of Contents

Dedication.................................................................................................................. iii

Abstract ...................................................................................................................... v

Table of Contents ..................................................................................................... vi

List of Tables ........................................................................................................... viii

List of Figures ......................................................................................................... ix

List of Illustrations ................................................................................................. xi

HEAT SHOCK PROTEIN 60 ....................................................................................... 1

Chapter 1: Introduction - Protein Folding and Chaperones ........................................ 1

Chaperonins .............................................................................................................. 3

  Group I, Group II .................................................................................................. 4

  GroEL ...................................................................................................................... 5

Heat Shock Protein 60 (Hsp60) .............................................................................. 8

  Role of Hsp60 in Disease ...................................................................................... 9

  Project Goals and Hypothesis ............................................................................. 12

Chapter 2: Molecular Cloning .................................................................................. 14

  Background .......................................................................................................... 14

  Materials and Methods ....................................................................................... 21

  Results .................................................................................................................. 25

  Discussion ........................................................................................................... 27

Chapter 3: Protein Expression and Purification ......................................................... 29

  Background .......................................................................................................... 29

  Materials and Methods ....................................................................................... 34

  Results .................................................................................................................. 37

  Discussion ........................................................................................................... 39

Chapter 4: Functional and Structural Characterization ............................................. 41

  Materials and Methods ....................................................................................... 41

  Results .................................................................................................................. 43

  Discussion ........................................................................................................... 49
List of Tables

Table 1.1: Features of Group I and Group II Chaperonins .................................................................5
Table 1.2: Missense variations in the HSPD1 gene encoding Hsp60 [48]. ........................................10
Table 5.1: Current Antiretroviral (HIV) drugs. [76] ........................................................................59
List of Figures

Figure 1.1: Free Energy Funnel.................................................................1
Figure 1.2: (A) Group I chaperonin GroEL in complex with GroES. (B) Group II chaperonin thermosome. [5].................................................................5
Figure 1.3: (A) Domains of monomeric GroEL in the APO state. (B) Side view of tetradecameric APO GroEL. (C) Top view of tetradecameric APO GroEL. (D) Domains of monomeric GroEL in the ADP/GroES-bound state. (E) Side view of tetradecameric ADP/GroES-bound state. (F) Top view of tetradecameric ADP/GroES-bound state. [15].................................7
Figure 1.4: Position of (B) gene variations in the HSPD1 gene and (A) the HSP60 protein structure [48]...........................................................................11
Figure 2.1: Agarose gel electrophoresis of restriction enzyme digestion products. Hsp10 gene appears as a sharp band between the 0.3 kb and 0.15 kb markers. Hsp60 appears as a bright band near the 1.5 kb marker. pET-22b appears near the 5.0 kb marker and a faint smear appears near the 0.15 kb marker. .................................................................25
Figure 2.2: Transformed bacterial colonies from Hsp60 ligation product. Sufficiently spaced colonies can be observed in almost all plates but mostly on the plate with 10 µLs of cells. ..........26
Figure 3.1: Elution profile of Hsp60 in a HiTrap® Q HP column. Hsp60 eluted using a concentration of 15% Buffer B. Flow through was determined not to contain any Hsp60 and the peaks observed at a conductivity of 100 mS/cm were also determined to be contaminants by SDS-PAGE analysis.................................................................37
Figure 3.2: Elution profile of Hsp60 in a Superose® 6 Increase SEC column. Hsp60 eluted at a volume of 14.15 mL..................................................................................................................................................................................37
Figure 3.3: Elution profile of Hsp10 in a HiPrep® SP FF column. Hsp10 was the first protein to elute with a linear gradient of Buffer B. Flow through was determined not to contain any Hsp10 and the peaks observed at a conductivity of 100 mS/cm were also determined to be contaminants by SDS-PAGE analysis.................................................................38
Figure 3.4: Elution profile of Hsp10 in a Superose® 6 Increase column. Hsp10 eluted as a large single peak at 17.8 mL which was later confirmed by SDS-PAGE analysis..........................38
Figure 3.5: SDS-PAGE analysis of Hsp60 and Hsp10 after each chromatography column used. Hsp60 was observed to be >95% pure, while Hsp10 was determined to be >99% pure. ............39
Figure 4.1: MDH refolding assay. Denatured MDH shows a 90% loss of activity compared to native MDH. MDH that is refolded with Hsp60 and Hsp10 regains about 90% of the activity compared to native MDH. Lack of ATP demonstrates no Hsp60 refolding activity of MDH. ....43
Figure 4.2: Hsp60 Apo conformation 2D class-averages.................................................................44
Figure 4.3: (A) Side view and (B) top view of Hsp60 Apo conformation.................................................44
Figure 4.4: Hsp60 ATP-10 conformation 2D class-averages.................................................................45
Figure 4.5: (A) Full cryo-EM map of ATP-10 conformation. (B) Fitted coordinates with low resolution features removed.................................................................45
Figure 4.6: (A) 3.2 Å map of ATP conformation. (B) Fitted model. (C) 2D class-averages obtained from dataset .................................................................................................46
Figure 4.7: Hsp60-Hsp10 contacts in ATP bound structure.................................................................47
Figure 4.8: ATP bound nucleotide binding pocket. D51 and D86 are in close proximity to the gamma phosphate of the ATP molecule. Magnesium ion shown as a green sphere. ..................48
Figure 4.9: MDH refolding assay with D51S and D86S mutants.................................................................48
Figure 4.10: (A) ADP conformation of Hsp60 shows a novel single ring structure. (B) 2D class averaging does not show double ring complexes.

Figure 5.1: New infections in the United States by race and transmission group. [72] ..........................................

Figure 5.2: The canonical mechanism of HIV entry into host cells. [75] .........................................................

Figure 5.3: HIV genome. [76] ..................................................................................................................

Figure 5.4: (A) Illustration of the viral life cycle. (B) Domain representation of gag. (C) Structural model of gag. (D) Illustration of immature virion. (E) Illustration of fully mature HIV. [81] .................................................................

Figure 5.5: (A) Fully assembled HIV capsid cone. (B) Capsid hexamer. (C) Capsid pentamer. [81] ......................

Figure 5.6: Agarose gel electrophoresis of restriction enzyme digestion products. CA gene appears as a sharp band near the 0.76 kb marker. CA-SP1 appears as a bright band near the 0.76 kb marker. pET-22b appears near the 5.0 kb marker and a faint smear appears near the 0.15 kb marker.

Figure 5.7: Elution profile of (A) CA and (B) CA-SP1 using a protocol of tandem cation and anion exchange columns. The protein was localized in the flow through, which was collected, concentrated, and passed through a second chromatography column. The large peak seen in the elution with 100% buffer B was determined to be contaminants and none of the protein of interest. .................................................................

Figure 5.8: Elution profile of (A) CA and (B) CA-SP1 in gel filtration chromatography. A single sharp peak is observed for both samples, with CA-SP1 eluting about 0.5 mL before CA due to the slightly larger molecular weight.................................................................

Figure 5.9: SDS-PAGE analysis of purified CA after size exclusion chromatography. Sample purity is estimated to be >98%. ........................................................................................................

Figure 8.1: Protease inhibition assay. CA-SP1 is equally cut in the presence or absence of F85 fullerene when compared to the DMSO vehicle control. Bevirimat also does not show any level of protease inhibition.................................................................

Figure 8.2: CA and CA-SP1 in-vitro assembly. CA (A) and CA-SP1 (B) positive control assembly results in rod-like particles several hundred nanometers in length. CA DMSO control (C) and CA-SP1 DMSO control (D) show no change in size or shape. CA F85 (E) and CA-SP1 F85 (F) show a clear change in morphology and size..............................................................................

Figure 8.3: DLS measurement on F85 dissolved in DMSO. A mean size of around 300 nm is observed, although based on the deviation and large PDI it is a rough approximation only. .......
List of Illustrations

Illustration 1.1: Proposed Hsp60 protein folding cycle. .......................................................... 9
Illustration 2.1: Polymerase chain reaction .......................................................... 15
Illustration 2.2: Restriction enzyme digestion products. Some enzymes (e.g., HindIII, PstI) cut DNA asymmetrically resulting in single stranded overhangs, while some enzymes (EcoRV) cut DNA symmetrically resulting in blunt ends. .......................................................... 16
Illustration 2.3: DNA ligase mechanism. .......................................................... 18
Illustration 2.4: Molecular cloning schematic. Restriction enzyme digestion and DNA ligation produce a construct composed of the target gene covalently ligated to the plasmid vector of choice. .......................................................... 19
Illustration 2.5: Chemical transformation vs electroporation.................................................. 20
Illustration 3.1: IPTG induction in BL21-DE3 and pET expression system................................. 30
Illustration 3.2: Size-exclusion chromatography column separation........................................ 32
Illustration 3.3: Anion-exchange chromatography column separation........................................ 33
Chapter 1: Introduction - Protein Folding and Chaperones

The process of protein folding is a complex series of concerted steps that nascent proteins undergo to achieve their functional native state. The exact pathway that produces a native protein as it is synthesized in the ribosome has not been well established. In fact, protein folding is one of the more challenging areas of biochemistry due to the incredibly dynamic nature of proteins. The current model proposes a funnel-like pattern that illustrates a decrease in free energy as protein folding level increases, with the protein reaching an energetic minimum when it adopts its native structure [1].

Figure 1.1: Free Energy Funnel
A protein’s native state represents its functional, catalytically active form, and failure to adopt its native structure has almost certainly negative implications. These can range from the more “benign,” such as a mere loss of function of a single protein, to the more detrimental ones such as protein aggregation. The latter often manifest as large insoluble fibrils that have been linked to neurodegenerative diseases such as Alzheimer’s and Parkinson’s disease [2]. Most proteins can reach their native state on their own with thermodynamics being the main driving force behind their assembly. However, a fraction of proteins, some of which are indispensable for proper cellular function and whose absence leads to disease, cannot fold on their own. These proteins require the assistance of specialized protein complexes known as chaperonins, a subset of proteins belonging to the broader family of proteins known as chaperones, to reach their native state. Molecular chaperones are protein complexes that aid in the proper folding of nascent peptides, or in the re-folding of misfolded proteins. It is estimated that about 20%-30% of mammalian proteins interact with a molecular chaperone during their lifetime, highlighting the importance of chaperones in proteostasis. Although they are grouped under the same “class” of proteins, the mechanism of action observed in different chaperones is very distinct. In fact, chaperones can be broken down into further subclasses with some carrying out de novo folding of proteins, others mainly refolding denatured proteins, and some not folding client proteins, but instead binding to them to prevent irreversible aggregation [3]. The drastically different chaperones are often called “heat shock proteins” (HSPs) as a larger group due to the shared characteristic of being upregulated during cellular stress and are differentiated by their molecular weight (e.g. Hsp27, Hsp60, Hsp70) [4].
**CHAPERONINS**

Chaperonins form macromolecular protein complexes that aid in the proper folding of their client proteins. Their substrates can be either newly synthesized proteins or preexisting proteins that have been denatured in the cell due to environmental conditions and stressors. Nascent proteins that require the assistance of a chaperonin to reach their native state would not be functional if a chaperonin failed to assist during its folding pathway. Instead, these proteins remain trapped in a conformation that provides a local energy minimum rather than a global energy minimum. Partially folded pathways can be deleterious in two ways. First, without reaching their native state most proteins are not catalytically active or functional. This loss of function can affect cellular processes which can compromise the overall state of the cell. Second, proteins that are only partially folded are more likely to exist with exposed hydrophobic regions that would normally be buried in their native conformation. These exposed regions can begin to interact with other misfolded proteins increasing the probability of forming irreversible protein aggregates that can be harmful to the cell. Equally, proteins that have become misfolded through denaturation due to some form of cellular stress can form irreversible aggregates posing a similar risk to cellular stability [5]. Examples of such aggregates are protein fibrils seen in the cells of patients suffering from a neurodegenerative disease such as Alzheimer’s [2]. In cells, chaperonin-mediated protein folding helps reduce the number of misfolded proteins by catalyzing the folding of client proteins in a secluded and more controlled environment. For these reasons, chaperonins are thought to be incredibly important factors in cellular homeostasis. In fact, it has been demonstrated that chaperonin absence or loss-of-function results in cell death in mice, yeast, and bacterial models [6-8]. It has also been shown that chaperonins are not only necessary during stress conditions, but also for normal growth in “optimal” conditions. GroEL
knockout strains are unable to grow, even at temperatures below heat-shock [8]. Chaperonins carry out protein folding cycles in an ATP-dependent manner. ATP binding, hydrolysis, and ADP release, in turn, prompt conformational changes that drive the catalytic folding cycles of chaperonins [8-13].

**Group I, Group II**

Chaperonins have largely been categorized into two distinct groups, group I and group II, based on several key differences. Group I chaperonins, perhaps most notably GroEL from *E. coli* with most of them being found in bacteria, assemble into homotetradecameric complexes composed of two rings with a hollow center that are stacked against each other [14-16]. (Figure 1.2-A) The inter-ring subunits sit across each other in a staggered fashion, where there is not a direct overlap between the subunits of the two rings [14, 17]. Group I chaperonins also require the assistance of a co-chaperonin, GroES in the case of GroEL. The co-chaperonin acts as a “lid” that fully sequesters the substrate inside of the GroEL cavity. Upon ATP hydrolysis and folding of the client protein, the co-chaperonin lid falls off allowing the substrate to exit the cavity and another misfolded substrate to enter [17-19].

Group II chaperonins are found in eukaryotes and archaea, such as the eukaryotic TriC chaperonin. These can form homo or hetero oligomers that range anywhere from seven to nine subunits per ring. The full complexes, therefore, range from tetradecameric to octadecameric. Unlike their group I counterparts, group II chaperonins have subunits that sit directly over one another in-register. (Figure 1.2-B) Interestingly, chaperonins of this group do not require the presence of a co-chaperonin as they have a built-in lid that closes and opens the main cavity via significant conformational changes seen during the protein folding cycle [5, 20-24]
Table 1.1: Features of Group I and Group II Chaperonins

<table>
<thead>
<tr>
<th>Feature</th>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td>Bacteria, <em>Homo sapiens</em></td>
<td>Archaea and eukaryotes</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>Cytoplasmic and endosymbiotic organelles</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td><strong>Substrate</strong></td>
<td>Promiscuous</td>
<td>Substrate-specific</td>
</tr>
<tr>
<td><strong>Subunits per ring</strong></td>
<td>7</td>
<td>7–9</td>
</tr>
<tr>
<td><strong>Oligomeric organization</strong></td>
<td>Homo-oligomeric</td>
<td>Hetero-oligomeric</td>
</tr>
<tr>
<td><strong>Co-chaperonin</strong></td>
<td>Required</td>
<td>Not Required</td>
</tr>
<tr>
<td><strong>Inter-ring Interactions</strong></td>
<td>Out of register (1:2)</td>
<td>In-register (1:1)</td>
</tr>
<tr>
<td><strong>Ring separation</strong></td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*Items in bold demonstrate similarities between 4-EL chaperonin and Group I and II chaperonins.

**GroEL**

The most widely studied chaperonin is the GroEL chaperonin, along with its co-chaperonin GroES, from *E. coli*. This is due to its high yield when obtained from bacterial culture as well as its remarkable stability even after chromatography purification. Recombinant GroEL and GroES also retain catalytic activity *in-vitro* [14, 19].
GroEL forms a tetradecameric complex composed of two symmetrical heptameric rings that sit against each other in a back-to-back fashion, forming a large double-barrel complex with a hollow cavity in the center of each ring. The architecture of the monomeric subunits can be separated into three domains: apical, intermediate, and equatorial. (Figure 1.3-A) The apical domain contains the ring opening where substrates are thought to interact via hydrophobic interactions between hydrophobic amino acids of GroEL and the exposed hydrophobic patches of misfolded proteins. Upon binding, the client proteins enter the inner cavity of GroEL. The equatorial domain is responsible for nucleotide binding and intra-ring communication. The intermediate domain mostly undergoes conformational changes in response to nucleotide binding and dissociation, driving most of the activity inside of the barrel and the overall conformational changes observed in structural studies [12, 14, 15, 19, 25]. GroES acts like a cap for the otherwise exposed center of the GroEL barrel and encapsulates the substrate inside of the cavity to provide an isolated, energetically favorable environment that is more conducive to proper protein folding than the cytosol [16].

The two rings on GroEL do not function simultaneously but instead alternate in a negative-cooperativity type of mechanism [12]. The actual catalysis is driven by the binding and subsequent hydrolysis of ATP, followed by the release of ADP from the nucleotide-binding pocket. The ring that is actively carrying out a folding reaction is known as the cis ring, while the inactive ring is called the trans ring. For protein folding to occur, ATP must bind to each subunit of the cis ring, while the nucleotide-binding pockets of the trans ring are occupied by ADP. It is important to note that every subunit of GroEL gets occupied by ATP or ADP, meaning the cis ring has seven molecules of ATP while the trans ring has seven molecules of ADP. Upon substrate binding in the apical domain and entry into the central cavity, the barrel is closed by
GroES and ATP is hydrolyzed, providing the energy necessary to catalyze the folding reaction [26]. Subsequently, after ATP is hydrolyzed in the cis ring, GroES dissociates from GroEL, and the substrate is released. Here, there is a concomitant exchange of ADP for ATP in the trans ring which now becomes the active cis ring and another round of refolding can occur there [18]. This mechanism has been colloquially described as a “two-stroke” mechanism and is a salient characteristic of GroEL-mediated protein folding [19].
The GroEL/ES chaperonin system and its mechanism has become the stereotype to describe other chaperonins. In fact, GroEL/ES system is the face of the Group I chaperonins, all of which share major characteristics. Group II chaperonins have a similar mechanism overall but display some key differences such as number of subunits, inter-ring connectivity, and perhaps most significantly, the absence of a co-chaperonin [20, 21, 24, 27-29]. Recently, however, new studies have elucidated the mechanisms of novel bacteriophage-encoded chaperonins that share Group I and Group II chaperonin characteristics as well as a new distinct ring-separation feature not observed in the two groups [23, 24, 30].

**Heat Shock Protein 60 (Hsp60)**

In humans, heat shock protein 60 (Hsp60) is a chaperonin responsible for folding mitochondrial proteins along with its co-chaperonin Hsp10 [31, 32]. The HSPD1 gene that encodes for Hsp60 has been localized to chromosome two in humans. The HSPE1 gene that encodes for the co-chaperonin Hsp10 is localized in a head-to-head position with respect to HSPD1 with the DNA region separating the two genes acting as a bi-directional promoter that, under normal conditions, shows twice the transcriptional activity of Hsp60 than that of Hsp10. Heat shock conditions show to upregulate the transcription in both directions by a factor of about 12, highlighting the importance of the Hsp60/Hsp 10 chaperonin system for cellular stability [33]. Hsp60 contains a mitochondrial localization signal that targets it for mitochondrial import via the TOM/TIM complex [22, 34]. In the mitochondria, Hsp60 forms complexes to support protein folding, however, the nature of these complexes has not been fully elucidated. It has been reported that the APO form of the protein can alternate between single heptameric rings and tetradecameric double rings (like GroEL) [30, 35-38]. Upon ATP and Hsp10 binding, it has been shown that Hsp60 adopts an “American football” conformation as it can bind two molecules of
Hsp10 simultaneously, one on each ring opening; these complexes have not been seen in GroEL [9, 13, 37, 39-42]. Furthermore, the ADP conformation has been shown to separate into single heptameric rings of Hsp60 with Hsp10 bound [42]. Interestingly, cryo-EM structures have also shown ADP-bound Hsp60 to form American football complexes, similar to the ATP conformation [43]. Our proposed protein folding cycle (Illustration 1.1) involves a highly dynamic set of intermediates that include most of the previously reported conformations. At the moment, however, further high-resolution APO, ATP and ADP structures are required to fully support one oligomeric state hypothesis over the other.

Illustration 1.1: Proposed Hsp60 protein folding cycle.

Role of Hsp60 in Disease

Chaperonins are ubiquitously expressed across all types of organisms, from bacteriophages to humans [6, 28, 33, 44-47]. It has been demonstrated that they are indispensable for cellular viability as their absence results in cell death. It has been found that mtHsp60/Hsp10 can replace GroEL/GroES in bacteria even though it is believed they are mechanistically distinct, showing the universality of their function [39, 48]. Similar loss-of-function experiments have been done in yeast where it was found that truncating the c-terminus of Hsp60 resulted in a complete lack of growth. In this work it was also determined that there is a defined boundary
between functionality of Hsp60 and cell viability in that partially functional Hsp60 can support cellular growth, although to a lesser extent, while non-functional Hsp60 is lethal to the cells [7]. Drosophila survival has also been shown to be dependent on Hsp60 availability [47]. Additionally, in-vivo studies have demonstrated that Hsp60 is essential for the survival of mouse embryos, and that the deletion of Hsp60 in adult mouse cardiomyocytes leads to heart failure and significantly perturbs mitochondrial protein homeostasis and mitochondrial function [6, 49].

Table 1.2: Missense variations in the HSPD1 gene encoding Hsp60 [48].

<table>
<thead>
<tr>
<th>Variation</th>
<th>Disease association</th>
<th>Growth in genetic complementation assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.As29Gly</td>
<td>MitCHAP60</td>
<td>Slow, temperature-sensitive</td>
</tr>
<tr>
<td>p.Leu46Phe</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>p.Val98Ile</td>
<td>SPG13</td>
<td>No growth</td>
</tr>
<tr>
<td>p.Arg142Lys</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>p.Lys156Arg</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>p.Asn184Ser</td>
<td></td>
<td>Unaffected</td>
</tr>
<tr>
<td>p.Asn265Ser</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>p.Leu291Val</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>p.Arg370His</td>
<td></td>
<td>Not tested</td>
</tr>
<tr>
<td>p.Asp379Gly</td>
<td></td>
<td>Unaffected</td>
</tr>
<tr>
<td>p.Gln461Glu</td>
<td>SPG13</td>
<td>Impaired</td>
</tr>
<tr>
<td>p.Gly559Asp</td>
<td></td>
<td>Unaffected</td>
</tr>
<tr>
<td>p.Gly563Ala</td>
<td></td>
<td>Unaffected</td>
</tr>
</tbody>
</table>
Several genetic mutations in the mtHsp60 encoding HSPD1 gene have been associated with human diseases. (Table 1.2) (Figure 1.4) Two of these diseases include SPG13 and MitCHAP-60, both neurodegenerative diseases with different symptoms and modes of inheritance [48, 50, 51]. The first disease-causing mutation associated with mtHsp60, V98I, was linked to SPG13, an autosomal dominant form of hereditary spastic paraplegia (HSP). HSP symptoms tend to appear as leg stiffness (spasticity) and weakness, although symptoms and symptom progression vary depending on the onset of the disease. Degeneration of spinal cord nerves is also a characteristic feature of HSPs. It is hypothesized that the genetic dominance of
SPG13 is due to the deleterious effect mutant mtHsp60 has on wild-type mtHsp60; that is, V98I mtHsp60 compromises the integrity and function of wild-type mtHsp60 protein complexes. It was found that Gro-EL knockout bacteria remained viable if expression of wild-type mtHsp60 was induced, however, mutant mtHsp60 (V98I) was unable to support bacterial cell growth, highlighting the effect of the V98I mutation on mtHsp60 function [52-55]. MitCHAP-60 is an autosomal-recessive disease, unlike SPG13, best described as a hypomyelinating leukodystrophy caused by the D29G missense mutation [56]. Although MitCHAP-60 is fundamentally similar to SPG, major differences characterize the former: MitCHAP-60 is an early onset disease that is better described as a lethal hypomyelinating neurodegenerative disease. Considering “pure” SPGs are characterized solely by lower extremity spasticity and “complicated” SPGs are accompanied by other, more severe symptoms, MitCHAP-60 can be viewed as a complicated SPG, while SPG13 is best described as a pure SPG [53, 57]. Additionally, it was found that the D29G mtHsp60 is also unable to support the growth of GroEL knockout bacteria, similar to V98I mtHsp60, although it was not established if the mtHsp60 loss of function was due to a mechanistically similar cause in the two variants [56].

**PROJECT GOALS AND HYPOTHESIS**

Mutations in the *HSPD1* gene, which encodes for Hsp60 in humans, have been linked to neuropathies such as hereditary spastic paraplegia type 13 (SPG13) and mitochondrial Hsp60 chaperonopathy (MitCHAP-60) neither of which has a cure or treatment. Since no definitive high-resolution structure exists for Hsp60 it is difficult to understand the direct link between the point mutations and the disease phenotype. Therefore, high-resolution structures will help understand the mutations at an atomic level and rational drug design can be implemented. This
will allow us to have a more focused approach to target a specific problem on the protein rather than using random drug screens to find candidates.

Our first goal is to generate clones of HSPD1 and HSPE1 in bacterial expression vectors. Second, we will obtain pure recombinant Hsp60 and Hsp10 from bacterial cultures purified using a tandem liquid chromatography approach. We will then characterize our proteins functionally and structurally using cryo-electron microscopy (cryo-EM) with the goal of obtaining high-resolution structures that will be fitted with atomic coordinates.

Our working hypothesis is that high resolution cryo-EM structures of Hsp60 will evidentiate a folding mechanism distinct to that of GroEL composed of a double ring APO conformation, an American football ATP-bound conformation, and an Hsp10-bound single ring Hsp60 in the ADP-bound state.
Chapter 2: Molecular Cloning

BACKGROUND

Molecular cloning is the process of copying a target gene of interest and inserting it into a vector of choice. This construct is then introduced into the desired organism for the purposes of protein expression [58]. Once expressed, the proteins can be studied with regard to the phenotype that they produce in the live organism, the effect the expression of the protein has on the regulation of other genes, or outside of the organism, in-vitro, by lysing the organism and extracting the target protein. The process of molecular cloning almost universally begins with polymerase chain reaction (PCR). Here, a double strand of DNA is amplified by several orders of magnitude to obtain a highly concentrated and potentially modified gene. PCR works by repeated rounds of DNA denaturation using high temperatures, primer annealing by lowering the temperature, and DNA extension done by a DNA polymerase enzyme [59]. These three steps are usually cycled 25-35 times to obtain a high yield of amplified DNA. (Illustration 2.1) Key aspects of this reaction include the use of a thermostable DNA polymerase that can withstand the DNA denaturation stage, the use of predesigned DNA primers that allow the initiation of the synthesis of a sister DNA strand, and the availability of free single nucleotides that serve as building blocks for the synthesized DNA strands. PCR is also used to introduce mutations to a target piece of DNA in a process called site-directed mutagenesis. Here, the DNA primers used in the PCR reaction are synthesized containing the desired nucleotide variations and incorporated into the synthesized DNA due to the fact that the DNA primers themselves become incorporated to the daughter DNA strands.
Once the specific gene of interest has been amplified, restriction enzymes are used to cut the ends of the newly synthesized DNA to create single stranded DNA overhangs that will allow for the insertion of the processed DNA into a vector, typically a plasmid, that has been treated with the same restriction enzymes [60]. Restriction enzymes, or restriction endonucleases, are enzymes that catalyze the hydrolysis of double stranded DNA. These enzymes recognize a specific sequence, typically a palindrome 5-10 base pairs in length, which allows them to bind to the DNA double strand. Once bound, the enzyme induces a double stranded break of the DNA backbone, resulting in a complete break of the DNA into two individual pieces. An important feature of restriction endonucleases is that most of them induce an asymmetric break in the DNA strand resulting in single base pair overhangs, commonly referred to as “sticky ends,” although some of them induce symmetric DNA breaks resulting in “blunt ends.” (Illustration 2.2)
Illustration 2.2: Restriction enzyme digestion products. Some enzymes (e.g., HindIII, PstI) cut DNA asymmetrically resulting in single stranded overhangs, while some enzymes (EcoRV) cut DNA symmetrically resulting in blunt ends.

In molecular cloning, the gene of interest is amplified using primers that will incorporate the desired restriction sites on the 5’ and 3’ ends of the gene. The restriction sites are important for several reasons, and the selection of them should be carefully planned and mapped. First, the restriction sites should also be available on the destination vector that will be used to generate the final clone. In the vast majority of cases the vector is in the form of a DNA plasmid. These plasmids will the gene of interest and are usually inserted into bacteria for propagation and storage. The transformed bacteria will replicate the plasmid and allow for high yield extraction of it when extracted and purified using silica columns. The highly concentrated and purified plasmid can then be inserted into the organism that will express the gene of interest [58]. A large variety of plasmids are now commercially available and contain various features that should be
considered when choosing one. Almost universal features of plasmids include a multiple cloning site region that contains several restriction sites to choose from when inserting a synthesized gene of interest, an antibiotic resistance gene that is used as a selection marker, and an origin of replication that is used by bacteria to propagate the plasmid [61]. Plasmids vary in the number and type of tags that they contain. These tags are segments of DNA that will be incorporated into the transcribed gene of interest and subsequently the translated protein that usually serve as localization signals to shuttle the protein of interest to a desired area in the cell, or as affinity tags that allow for the purification of the protein of interest using affinity chromatography. Examples of tags include the pelB leader sequence which directs the expressed protein to the periplasm in *E. coli* [62], and the 6xHis tag which allows for the selective purification of the tagged protein using a nickel containing column in affinity chromatography [63]. When inserting the gene of interest into a plasmid vector, the multiple cloning site region should be examined to determine which restriction sites will be used. Usually, two different sites are chosen to ensure that the gene of interest is inserted unidirectionally, with the 5’ end of the gene being next to the promoter region found in the plasmid and the 3’ end of the gene being next to the terminator region. The amplified gene and the plasmid vector are both digested with the desired restriction enzymes to ensure that the resulting overhangs are complimentary to each other. If blunt end digestions are used, directionality and overall insertion efficiency is drastically reduced.

Once both the amplified gene and the plasmid vector have been digested, they are mixed in certain ratios and linked covalently using a ligase enzyme. DNA ligases catalyze the covalent joining of two strands of DNA via the formation of a new phosphodiester bond that links the backbones of the DNA strands [64]. (Illustration 2.3) DNA ligases require ATP to create the phosphodiester linkage.
Illustration 2.3: DNA ligase mechanism.

Molecular cloning results in a construct that is composed of a plasmid vector and a gene of interest that have been covalently linked together. (Illustration 2.4) The gene insertion is directional to ensure that the transcription of the gene is done in the 5’ to 3’ direction. The construct can then be inserted into high plasmid copy bacteria for plasmid storage and propagation or expression systems that are suitable for the target protein.
Plasmids containing a gene of interest are inserted into an organism for propagation, usually bacteria, or for protein expression. With many proteins bacterial expression is the system of choice due to the low cost, high yield, and ease of use of bacterial cultures. To introduce a plasmid into bacteria, however, it is necessary to prepare cells to be able to uptake exogenous DNA, or to purchase such bacteria. Bacteria that have been prepared in this way are referred to as “competent cells.” Two types of competent cells exist, chemically competent cells and electrocompetent cells. Chemically competent cells are treated with divalent cations to neutralize the negative lipopolysaccharide (LPS) layer found in most gram-negative bacteria. This allows the DNA to interact more closely with the outer layer of bacteria and mitigates the otherwise repulsive force between the negatively charged DNA and the negative LPS layer [65]. Electrocompetent cells do not require special preparation but do need to be put in a salt-free
environment. Chemical transformation utilizes heat shock to permeabilize the cell membrane. This allows the DNA to flow into the cell where it is permanently trapped once the cell repairs its membrane. Electroporation employs applying a current to the cells to disrupt the cell membrane and create pores. DNA can flow freely into the cell and similarly to chemical transformation gets trapped once the cell repairs its membrane. (Illustration 2.5) Cells that successfully uptake the plasmid DNA become resistant to the antibiotic that the antibiotic resistance gene targets. This allows for the simple screening of transformants by growing the product of the transformation reaction in a growth medium that contains the desired antibiotic. Bacterial colonies that grow on this medium can be grown in liquid culture and stored for future use.

Illustration 2.5: Chemical transformation vs electroporation.
**MATERIALS AND METHODS**

cDNA for the genes *HSPD1* and *HSPE1* which encode for human Hsp60 and Hsp10, respectively, were amplified using PCR (Q5® DNA Polymerase, NEBiolabs) under the recommended thermocycling conditions. Q5® 2X Master Mix, which contains the Q5® enzyme as well as 10 µM of free dNTPs, was used for all PCR reactions. 0.5 µM of forward and reverse primers obtained from IDT™ along with 10 pmol of Hsp60 cDNA were also included in the PCR reaction. Thermocycling was set up with a 98°C denaturation for 10 seconds, followed by annealing at 67°C for 25 seconds, and an extension at 72°C for 40 seconds for a total of 35 cycles. Restriction sites for NdeI and XhoI were added to the 5’ and 3’ ends, respectively, of the gene for cloning into a pET-22b vector. The PCR product was run on a 0.8% agarose gel containing SybrSafe® DNA dye at a 1X concentration at 120 V for 30 minutes and visualized using blue light with a blue light filter. The PCR products were gel extracted and purified using a PureLink™ gel extraction kit. Briefly, the resulting DNA band from the agarose gel was excised using a steel blade and weighted. The gel slice was dissolved using Gel Solubilization Buffer at a 5:1 ratio (v/w) of buffer to agarose gel. The gel slice was incubated at 50°C for 10 minutes. One gel slice volume of pure isopropanol was added to the dissolved gel slice. The dissolved gel slice was added to a silica column and centrifuged at 12,000xg for 1 minute. 600 µLs of Wash Buffer were used to wash the bound DNA and centrifuged at 12,000xg for 1 minute. The wash buffer was discarded, and the column was spun down again at 18,000xg for 2 minutes to completely remove trace amounts of wash buffer. DNA was eluted using Elution Buffer that was preheated to 65°C. The column was incubated with the elution buffer for 2 minutes and centrifuged at 12,000xg for 2 minutes. The eluted DNA was collected, and the DNA concentration was measured using a spectrophotometer by measuring absorption at 260 nm. The DNA was stored
in -20°C long term. The gene for Hsp10 was amplified, extracted, and purified using the same procedure except that the extension time in the PCR reaction was reduced to 12 seconds, and the agarose percentage used for electrophoresis was increased to 2%. Two micrograms of each gene as well as two micrograms of pET-22b (EMD Millipore®) were digested for 4 hours at 37°C with NdeI and XhoI (NEBiolabs®). The digestion reaction was composed of 1 µL of each restriction enzyme, 2 µg of DNA, water to a total volume of 50 µLs, and Cutsmart® buffer at a 1X concentration. The digestion products were run on a 1% agarose gel at 120 V for 20 minutes, excised with a steel blade, and gel extracted using the previously described protocol. Ligation reactions were done using T4 DNA Ligase® from NEBiolabs®. Each reaction was done at a total volume of 20 µLs containing plasmid DNA, gene DNA, 1 µL of ligase enzyme, ligase buffer at a concentration of 1X containing ATP, and water. Each gene was ligated with 100 ng of digested plasmid at a molar ratio of 3:1 overnight at 16°C. The ligation reaction was terminated at 65°C for 15 minutes. One tube of One Shot™ Chemically Competent cells was thawed on ice for 10 minutes. 2 µLs of each ligation reaction was added to its respective tube of competent cells and mixed by light flicking. The cells were allowed to incubate with the ligation product for 30 minutes on ice. The cells were then heat shocked in a water bath preset to exactly 42°C for exactly 30 seconds. The cells were returned to the ice for 5 minutes. 250 µLs of room temperature S.O.C. media were added to each tube of cells and incubated at 37°C for 1 hour while shaking horizontally at 250 rpm. After incubation, the transformed cells were plated on LB agar plates containing 200 mg/mL ampicillin. Four plates were prepared for each transformation reaction with 10 µLs, 25 µLs, 50 µLs, and 100 µLs of the transformed cells and spread with a sterile spreader. Different volumes were used to ensure evenly spread colonies on one of the plates. The plates were inverted and incubated overnight at 37°C. Colony growth was observed
on all plates and 3 colonies were selected from each transformation reaction based on size and morphology to inoculate liquid cultures. 50 mL cultures were made for each colony selected containing 200 mg/mL ampicillin in a baffled 250 mL flask with a cap. The cultures were grown overnight at 37°C with shaking at 220 rpm. 25 mL of each culture was transferred to a sterile 50 mL conical tube and centrifuged in a tabletop centrifuge at 5250xg for 15 minutes at 4°C. Plasmid DNA was extracted and purified from the cells using a PureLink™ HiPure Plasmid Midiprep kit. The harvested cells were resuspended in 4 mL of Resuspension Buffer containing RNAse A. The cells were then lysed using 4 mL of Lysis Buffer and inverting 10 times. After a 5-minute incubation at room temperature, 4 mL of Precipitation Buffer were added, the samples were inverted 10 times, and incubated on ice for 10 minutes. The cell lysate was centrifuged at 18,000xg for 15 minutes at 4°C. During this time, a midi purification column was equilibrated with 5 mL of Equilibration Buffer allowed to pass with gravity. The supernatant was added to the midi column and allowed to pass through the column by gravity. 10 mL of wash buffer were added to the column and allowed to pass by gravity two separate times. The DNA was then eluted using 5 mL of Elution Buffer that had been preheated to 65°C. The eluted DNA was collected in a clean tube and precipitated by adding 3.5 mL of pure isopropanol. The precipitated DNA was centrifuged at 18,000xg for 30 minutes at 4°C. The supernatant was decanted, and the pellet was washed with 5 mL of 70% ethanol. The washed DNA was again centrifuged at 18,000xg for 15 minutes at 4°C. The ethanol was decanted, and the resulting pellet was allowed to air dry overnight. The DNA pellets were resuspended in 100 µLs of TE buffer and stored at -20°C. The plasmids were sequenced to confirm the presence of our target gene of interest in each plasmid. Since three plasmids were purified for Hsp60 and three for Hsp10, the plasmid with the highest concentration and correct sequence was used for the rest of the experiments. One tube of
BL21-DE3 competent cells was used for each plasmid. Frozen tubes of competent cells were thawed on ice for 10 minutes. 50 ng of plasmid DNA were added to their respective tubes and incubated on ice for 30 minutes. Cells were heat shocked for 10 seconds in a water bath set to 42°C and immediately transferred back to an ice bucket for 5 minutes. 950 μLs of S.O.C. media were added to each transformation reaction and incubated at 37°C with shaking horizontally at 220 rpm for one hour. Four plates were prepared for each transformation reaction with 10 μLs, 20 μLs, 40 μLs, and 80 μLs of the transformed cells and spread with a sterile spreader. Different volumes were used to ensure evenly spread colonies on one of the plates. The plates were inverted and incubated overnight at 37°C. Colony growth was observed on all plates and 5 colonies from each transformation reaction were selected based on size and morphology to inoculate liquid cultures for protein expression check. 25 mL cultures were prepared with 2XYT containing 200 mg/mL ampicillin and were incubated at 37°C with shaking at 220 rpm until an OD\textsubscript{600} of 0.6 was reached in each culture. At this point, IPTG was added to a final concentration of 1 mM and allowed to continue incubating for 2 more hours. After the incubation, 1 mL samples of each culture were collected and centrifuged at 18,000xg for 1 minute at room temperature. The resulting cell pellets were resuspended in 100 μLs of deionized (DI) water and 2 μLs of resuspended cells were added to 10 μLs of 4X Lammeli Buffer. 10% SDS-PAGE gels were run at 200 V for 45 minutes and stained with Coomassie blue staining buffer. The colonies with the highest expression level of Hsp60 and Hsp10, respectively, were reinoculated in 2XYT, grown to an OD\textsubscript{600} of 0.4, mixed 1:1 with 50% glycerol, aliquoted, and frozen at -80°C for long-term storage.
RESULTS

PCR amplification of the HSPD1 and HSPE1 genes resulted in bright bands at around 1.5 kb and 300 kb, respectively. After restriction enzyme digestion, the size of the two genes remained unchanged, appearing at around 1.5 kb and 300 kb for Hsp60 and Hsp10, respectively. The digested pET-22b plasmid appeared to be around 5 kb in length after digestion, and a faint band near 150 bp appeared. (Figure 2.1)

![Agarose gel electrophoresis of restriction enzyme digestion products. Hsp10 gene appears as a sharp band between the 0.3 kb and 0.15 kb markers. Hsp60 appears as a bright band near the 1.5 kb marker. pET-22b appears near the 5.0 kb marker and a faint smear appears near the 0.15 kb marker.](image)

After transforming competent cells with the ligation product of Hsp60, plates with four different volumes of cells were prepared to ensure that at least one plate contained colonies that did not overlap or touch. The plate with 10 µLs of transformed cells showed spaced colonies of a good size that were suitable for inoculation. The cells were plated on plates containing 200
mg/mL therefore all bacterial growth observed must have been resistant to the antibiotic used.

(Figure 2.2) The plates that resulted from the transformation of Hsp10 were virtually indistinguishable from the plates of Hsp60 shown.

Figure 2.2: Transformed bacterial colonies from Hsp60 ligation product. Sufficiently spaced colonies can be observed in almost all plates but mostly on the plate with 10 µLs of cells.
DISCUSSION

The amplification of our target genes produced products that were very near their expected size based on their theoretical length. The gene that encodes for Hsp60 is 1647 bp in length and the observed band was closest to the 1.5 kb maker. Because the DNA fragments run in a logarithmic manner, and no other by product was seen in the amplification, it is safe to say that the resulting product is most likely our gene of interest. Likewise, Hsp10 was amplified and resulted in a band that is very close to the expected size of 309 bp. After digestion with NdeI and XhoI, it is almost impossible to see the fragment of DNA that is removed from the ends of the genes because it is only about 3-4 bp in length. Because of this, Figure 2.1 shows bands that are practically identical to the bands produced from the PCR reaction, but this is expected. However, the DNA band that is seen for the cut pET-22b plasmid gives us more information because it appeared very close to the theoretical size of 5.4 kb. Uncut plasmids are circular and supercoiled, and they usually do not run according to size because of this. Once cut, plasmids do run true to size because they have been linearized. Furthermore, the faint smear that appears on the bottom of the gel appears to be the cut fragment of DNA that resides between the NdeI and XhoI restriction sites on the multiple cloning site region of pET-22b. This is evidence that the restriction enzyme reaction most likely was successful for all three samples. Furthermore, we did not observe fragmentation of the genes or the plasmid, something that would be indicative of a problem like non-specific digestions such as the ones seen as a product of star activity. Although ligation reactions do not produce quantitative or qualitative results, it is the success of the transformation reaction that we can use as evidence of a successful ligation. Had our ligation reactions been unsuccessful, the DNA that is inserted into the bacterial cells during transformation would be linear. Linear DNA is degraded inside of bacteria if it is taken up at all.
If degraded, the cell would not be resistant to the antibiotic used as the antibiotic resistance gene would have been lost along with the plasmid. Therefore, the presence of colonies not only evidentiate a successful transformation, but it also indirectly tells us that the ligation reaction was most likely successful as well.
Chapter 3: Protein Expression and Purification

BACKGROUND

Recombinant protein is expressed using a system of expression vectors and expression organisms that have key features to allow for the controlled induction of protein synthesis. Bacterial expression systems rely heavily on the use of operons to regulate the expression of a target protein. The most commonly used bacterial expression lines utilize the T7 promoter system to tightly regulate protein synthesis. BL21-DE3 is a genetically modified strain of *E. coli* that is perhaps the most widely used T7 bacterial expression line. This strain of *E. coli* contains the DE3 λ prophage in its genome, with the prophage containing the gene for T7 RNA polymerase (RNAP). Importantly, all genes carried in this prophage are downstream and under the control of the lacUV5 promoter, which is itself regulated by the lac operon. The lacUV5 promoter is a more potent variant of the native lacZ promoter found in *E. coli*, therefore better suited for high-yield protein expression [66]. Expression systems that contain the gene for RNAP also require the use of a suitable expression vector. Commercially available pET vectors are now the industry standard for these purposes, as they incorporate a T7 promoter upstream of their cloning site, as well as a downstream T7 terminator. Here, expression is controlled due to the specificity of the T7 promoter for RNAP, the former not being recognized by any of the endogenous RNA polymerases in *E. coli*. Therefore, transcription of the target gene is indirectly regulated through the regulation of the expression of RNAP, and because RNAP expression is controlled via the lac operon, we can trace the expression of our protein of interest back to the regulation that the lac operon provides. In the lacUV5 promoter, which can be recognized by native *E. coli* polymerases, transcription of the downstream RNAP gene is inhibited by the repressor protein encoded by the lacI gene, where the repressor protein binds tightly to the
operator region of the operon. Removal of the repressor protein is achieved by the addition of lactose or more commonly a non-hydrolysable analog in isopropyl β-d-1-thiogalactopyranoside (IPTG). IPTG binding induces a conformational change in the repressor protein that allows it to release the operator region, allowing the unencumbered transcription of the genes downstream of the lacUV5 promoter. This initiates the production of RNAP which can later recognize the T7 promoter upstream of the target gene of interest in the plasmid vector and synthesize the mRNA that will be translated into our target protein. An extra operator sequence is located downstream of the T7 promoter in most pET vectors, serving the same regulatory purpose that the operator has in the lacUV5 promoter, but the addition of IPTG to the cell culture allows for the removal of repressor proteins bound to operator sites downstream from, both, the lacUV5 promoter in the bacterial genome, and in the T7 promoter in our plasmid vector. (Illustration 3.1) The presence of this extra operator site is an example of one of the methods employed to prevent “leaky” or premature expression of our target protein of interest in bacterial expression systems [67]. This type of expression is generally unwanted and can compromise the health of the cell culture if the recombinant protein being expressed is toxic to the host cell.

Illustration 3.1: IPTG induction in BL21-DE3 and pET expression system.
Recombinant proteins obtained from non-native expression systems are almost always extracted from the host cells and studied in-vitro. During protein extraction, a medley of proteins is extracted along with the protein of interest, most, if not all, of which are considered contaminants in the study of the target protein. Because of this, it is necessary to purify the target protein from the crude extract obtained following cell lysis. Liquid chromatography is by far the most widely used method of protein purification due to its ease of use, efficiency, and ability to provide purity levels of ~99% [68]. Chromatography systems involve a series of pumps and detectors that allow the accurate detection of proteins in real time as well as other conditions such as conductivity, pH, pressure, etc. The separation, however, occurs in columns that can be purchased pre-packed, or packed manually with resins that have been functionalized to have various properties. The chemical nature of the resin employed will dictate what type of separation one is to expect. Common types of separation include size-exclusion chromatography (SEC), and ion-exchange chromatography (IEC).

SEC involves the separation of proteins based on their molecular weight and to some extent their shape. The gel-like resin used in SEC is composed of spherical beads, commonly agarose based, of different sizes. The particles of different sizes create small pores that act as a molecular sieve, where the smaller proteins are trapped in the pore channels more frequently, and the larger proteins traveling more freely. This results in size-based separation where large proteins are eluted through the column faster than smaller proteins. (Illustration 3.2) Because the separation relies almost entirely on the pores created in the resin, the bead size of the material used is of high importance and responsible for the separation capabilities of the column [69].
Illustration 3.2: Size-exclusion chromatography column separation.

IEC is a chromatography technique that separates proteins based on their charge. Here, the resin consists of a material, usually agarose beads, which have been functionalized on their surface to contain a permanent charge. Two types of IEC exist, anion-exchange and cation-exchange. In anion-exchange chromatography, negatively charged proteins bind to the resin of the column, while positively charged proteins do not interact, flowing through the column, effectively separating them from their negatively charged counterparts. (Illustration 3.3) Because of this, the resin in anion-exchange columns is positively charged, with groups such as quaternary ammoniums covering the surface of the beads. Cation-exchange columns are the opposite, binding positively charged proteins by having strong negative charges such as methyl sulfonates covalently added to the surface of their resin [70]. Elution of bound proteins in IEC is more complex than SEC, involving various steps and being heavily dependent on the conditions of the buffers used. Typically, protein samples are loaded using a buffer of low ionic strength (0-50 mM NaCl) to allow for the binding of proteins to the column. The column is washed with the
same buffer to remove unbound or weakly bound proteins before the elution of the bound proteins. Next, NaCl is introduced via second buffer in a linear or a step gradient. The increasing ionic strength results in the elution of ionically bound proteins on the resin. This occurs as the ions in the buffer compete for the charge on the column resin, effectively releasing the bound proteins. Since proteins display different levels of charge, the proteins that are less negative will elute first, requiring less salt for their elution. Because of this, it is common to introduce salt over a broad volume to better separate eluting proteins based on smaller charge differences and avoid co-eluting proteins of differing charge. Salt is typically added until 1-2 M concentrations are reached to ensure the complete elution of bound proteins. The column is usually re-equilibrated with low salt buffer for storage.

Illustration 3.3: Anion-exchange chromatography column separation.

Chromatography purification is visualized by tracking the elution of proteins using UV absorption at 280 nm. This information is displayed graphically in the form of a chromatogram that can be analyzed further. Chromatograms also display information like sample conductivity, indicative of salt content, pH, and system information such as column pressure, and temperature.
**MATERIALS AND METHODS**

Hsp60 cultures were grown from a 1 mL frozen glycerol stock that was thawed and added directly into 1 L of 2XYT media containing 200 mg/mL ampicillin. The culture was incubated at 37°C with shaking at 220 rpm in a baffled 2.5 L flask. Cells were grown until an OD$_{600}$ of 0.6-0.8 was reached, at which point the culture was induced with the addition of IPTG to a final concentration of 0.5 mM. Expression was allowed to occur for an additional 3 hours at 37°C. Cells were then harvested by centrifugation at 6000xg for 30 minutes at 4°C. The spent media was discarded, and the resulting pellets were resuspended in lysis buffer containing 25 mM Tris pH 7.5, 50 mM NaCl, and 50 mM EDTA. Lysis buffer was added to a ratio of 5 mL of lysis buffer per gram of cell pellet. DTT was added to a final concentration of 1 mM to the resuspended cells. Lysozyme purchased from Sigma-Aldrich® was added to a concentration of 0.25 mg/mL of resuspended cells. Cells were incubated with lysozyme at 4°C in a rotating shaker for 1 hour. The cells were then frozen in a -20°C freezer and thawed at room temperature a total of 3 cycles. MgCl$_2$ was added to the lysed cells to a final concentration of 60 mM along with DNase also purchased from Sigma-Aldrich® to a final concentration of 0.1 mg/mL. The cell lysate was incubated with DNase for 1 hour at 4°C in a rotating shaker. Streptomycin sulfate was added to the lysate to a final concentration of 3% w/v. The cell lysate was again incubated for 1 hour at 4°C in a rotating shaker. Cell lysates were clarified by centrifugation at 18,000xg for 30 minutes at 4°C. The clarified lysate was treated with a saturated ammonium sulfate solution to a final concentration of 30% v/v. The protein was allowed to precipitate for 30 minutes at 4°C in a rotating shaker. Ammonium sulfate precipitates were then centrifuged at 12,000xg for 30 minutes at 4°C to collect the precipitated protein. The supernatant was transferred to a clean tube and additional ammonium sulfate was added to bring the total
concentration to 50%. The protein was allowed to precipitate for 30 minutes at 4°C in a rotating shaker. The supernatant was discarded, and protein pellet was resuspended in 5 mL of Buffer A (25 mM Tris, pH 8.0) per gram of protein pellet. The resuspended protein (crude extract) was then flash frozen with liquid nitrogen and stored for future purification.

Hsp10 cultures were grown and treated the same with some minor differences. AEBSF was added to the resuspended cells in lysis buffer to a final concentration of 1 mM to prevent protein degradation that was observed. Also, the ammonium sulfate treatment of Hsp10 began with an initial treatment of 40% ammonium sulfate, followed by an increase to 60% ammonium sulfate. Finally, the protein pellets were resuspended in Buffer C (25 mM Bis-Tris, pH 6.5, 50 mM NaCl), aliquoted, and frozen with liquid nitrogen.

Hsp60 was purified by dissolving 1 mL of crude extract in 9 mL of Buffer A and loaded onto a pre-equilibrated 5 mL HiTrap™ Q HP column. The column was washed with 4 column volumes of Buffer A before eluting. The elution of Hsp60 was done using a step gradient of 5% Buffer B (25 mM Tris, pH 7.5, 1 M NaCl) for four column volumes, followed by an increase to 15% Buffer B for six column volumes. Buffer B concentration was increased to 100% Buffer B for two column volumes to wash the column and subsequently equilibrated with two column volumes of Buffer A. Fractions for the elution at 15% Buffer B were collected, pooled, and concentrated to <500 µls using an Amicon® Ultra centrifugal filter unit with a molecular weight cut off of 10 kDa. The concentrated protein was aliquoted and frozen. An aliquot of purified Hsp60 was thawed and MgCl₂ and ATP were added to a concentration of 10 mM and 4 mM, respectively. This was incubated at 30°C for one hour and loaded onto a 5 mL HiTrap® Desalting column equilibrated with Buffer E (25 mM Tris, pH 7.5, 1 M NaCl, 5 mM EDTA). The buffer exchanged protein was incubated at room temperature in this buffer for at least one
hour before concentrating to <500 µLs and loading onto a Superose® 6 Increase gel filtration column. Fractions were collected for the run after the 10 mL mark. Fractions containing Hsp60 were pooled and concentrated to a protein concentration of >10 mg/mL. Protein concentrations were determined using a Pierce™ BCA Assay. SDS-PAGE analysis was used to determine the point of elution of our protein of interest throughout the purification procedure. All SDS-PAGE gels for Hsp60 were made to a final acrylamide concentration of 10% and were run at 200 V for 45 minutes.

Hsp10 was purified by dissolving 1 mL of crude extract in 9 mL of Buffer C and loaded onto a pre-equilibrated 20 mL HiPrep™ SP FF column. The column was washed with 4 column volumes of Buffer C before eluting. The elution of Hsp10 was done using a linear gradient of increasing Buffer D (25 mM Bis-Tris, pH 6.5, 1 M NaCl). Buffer D concentration was increased until Hsp10 fully eluted, then increased to 100% for two column volumes to wash the column and subsequently equilibrated with two column volumes of Buffer C. Fractions for the elution were collected, pooled, and concentrated to <500 µLs using an Amicon® Ultra centrifugal filter unit with a molecular weight cut off of 10 kDa. The concentrated protein was loaded onto a Superose® 6 Increase gel filtration column. Fractions were collected for the run after the 10 mL mark. Fractions containing Hsp10 were pooled and concentrated to a protein concentration of >10 mg/mL. Protein concentrations were determined using a Pierce™ BCA Assay. SDS-PAGE analysis was used to determine the point of elution of our protein of interest throughout the purification procedure. All SDS-PAGE gels for Hsp10 were made to a final acrylamide concentration of 15% and were run at 200 V for 45 minutes.
RESULTS

During IEC using a HiTrap® Q HP column, Hsp60 was observed to elute at a concentration of 15% Buffer B. (Figure 3.1) Another peak was observed at a 5% Buffer B concentration, but upon SDS-PAGE analysis it was determined that it was a very small fraction of Hsp60 containing a lot of contaminants and therefore was excluded from collection.

![Figure 3.1: Elution profile of Hsp60 in a HiTrap® Q HP column. Hsp60 eluted using a concentration of 15% Buffer B. Flow through was determined not to contain any Hsp60 and the peaks observed at a conductivity of 100 mS/cm were also determined to be contaminants by SDS-PAGE analysis.](image)

Hsp60 was loaded onto a Superose® 6 Increase SEC column and was seen to elute at 14 mL as a sharp single peak. (Figure 3.2) More peaks were observed at 16 mL and 19 mL, but these were determined to be contaminants when analyzed using SDS-PAGE.

![Figure 3.2: Elution profile of Hsp60 in a Superose® 6 Increase SEC column. Hsp60 eluted at a volume of 14.15 mL.](image)
Hsp10 was first purified using IEC in a HiPrep® SP FF column. The first peak observed during the elution was determined to be Hsp10 after SDS-PAGE analysis. (Figure 3.3) The linear gradient was stopped, and the concentration of Buffer D increased to 100% to wash the column. All other peaks observed, including flow through was found to be contaminants upon SDS-PAGE analysis.

[Image: Figure 3.3: Elution profile of Hsp10 in a HiPrep® SP FF column. Hsp10 was the first protein to elute with a linear gradient of Buffer B. Flow through was determined not to contain any Hsp10 and the peaks observed at a conductivity of 100 mS/cm were also determined to be contaminants by SDS-PAGE analysis.]

Hsp10 containing fractions collected from IEC were pooled, concentrated, and loaded onto a Superose® 6 Increase column where it was observed to elute as a sharp, single peak at around 17.8 mL. (Figure 3.4) SDS-PAGE analysis confirmed that this peak contained our protein and the other smaller peaks observed downstream were only contaminants.

[Image: Figure 3.4: Elution profile of Hsp10 in a Superose® 6 Increase column. Hsp10 eluted as a large single peak at 17.8 mL which was later confirmed by SDS-PAGE analysis.]
SDS-PAGE analysis of protein samples after each step in the purification scheme was done to assess the purity level after each column. (Figure 3.5) Hsp60 runs as a 60 kDa protein, between the 50 kDa and 75 kDa markers, while Hsp10 runs as an 11 kDa protein, between the 10 kDa and 15 kDa markers on a 12.5% SDS-PAGE gel. The relative purity of the samples was approximated by comparing the intensity of the band of interest against total visible protein in the gel. Hsp60 was estimated to be >95% pure, and Hsp10 was observed to be >99% pure. While Hsp10 was determined to be ~99% after the first round of chromatography, SEC was used mainly as a way to buffer exchange the sample, as well as an analytical tool to approximate complex size, therefore not shown on SDS-PAGE.

![Figure 3.5: SDS-PAGE analysis of Hsp60 and Hsp10 after each chromatography column used. Hsp60 was observed to be >95% pure, while Hsp10 was determined to be >99% pure.](image)

**DISCUSSION**

After purification, the yield of Hsp60 was calculated to be about 5 mg per liter of culture, and around 8 mg per liter of culture for Hsp10. The final purity of both samples was >95%
which is suitable for cryo-EM studies. Crystallographic studies may have required a higher purity, especially for Hsp60, at which point a third chromatography column was introduced, but we moved forward with the obtained results shown. We concluded, based on the elution volume, that we in fact purified Hsp60 in its complex form and not its monomeric form. The assembly of the complex was done after IEC with the addition of ATP and magnesium, and skipping this step resulted in a shift in the elution volume of Hsp60 indicative of a smaller species. The purity of Hsp10 was very high after IEC and the use of SEC, although not imperative, was useful in buffer exchanging the sample into the final buffer used in cryo-EM while getting an estimate of the size of the protein. We concluded here that we were obtaining Hsp10 in a complex form as well, more specifically a heptamer, which we think is able to self-assemble with no need for special conditions given that we obtained a homogenous sample of complexed Hsp10 with no action required. Purified samples were used directly as obtained after purification since no affinity tags were used in the purification, so there was no need for a tag cleavage step. The stability of the purified protein, although not quantitatively measured, is thought to be high as purified protein was shown to retain activity even days after purification. This suggests that our purification protocols were not intrusive or damaging to our proteins. Finally, from the chromatograms we observed that both SEC elutions show sharp single peaks that are indicative of a homogenous sample. This is especially important for structural studies such as cryo-EM as protein heterogeneity is one of the most difficult obstacles to overcome. This allowed us to proceed with structural studies without the need for further optimization of buffer conditions.
Chapter 4: Functional and Structural Characterization

Materials and Methods

Malate dehydrogenase (MDH) was used as a substrate to assay the refolding activity of purified recombinant Hsp60 and Hsp10. MDH Assay Kit (Sigma®) MDH enzyme was denatured with HCl to a final concentration of 5 mM at room temperature for 10 minutes. The HCl was neutralized by diluting the denatured MDH 50-fold in assay buffer (50 mM Hepes, pH 7.5, 50 mM NaCl, 10 mM MgCl₂). Hsp60 was added to the denatured MDH to a 20:1 molar ratio and Hsp10 was added to the reaction mixture to a 2:1 molar ratio to Hsp60. ATP was added to a final concentration of 5 mM. The reaction was incubated at 37°C for 1 hour to allow for refolding of MDH. 10 µL of the refolding reaction were mixed with 2 µL of substrate, 10 µL of developer, and 28 µL of MDH assay buffer, all components of the MDH Assay Kit, in a 96 well plate and the absorbance at 450 nm was read in a plate reader set to 37°C for one and half hours. The same procedure was used to evaluate the activity of the D51S and D98S mutants.

DLS measurements were obtained using a Malvern® Zetasizer set to measure a scattering angle of 173°. All samples were measured at room temperature. The complexes were assembled by mixing Hsp60 and Hsp10 to a final molar ratio of 1:1. Final Hsp60 concentration was 1 mg/mL. For the ATP and ADP conformations, the respective nucleotides were added to a final concentration of 5 mM. Measurements were run in triplicates.

Cryo-EM grids were prepared by mixing Hsp60 and Hsp10 to a final molar ratio of 1:1. Final Hsp60 concentration was 1 mg/mL. For the ATP and ADP conformations, the respective nucleotides were added to a final concentration of 5 mM right before freezing the grid as to minimize incubation time. Quantifoil® 300 mesh R 1.2/1.3 copper grids were glow discharged using a plasma discharger for 1 minute at 25 mA. 3 µL of the protein mixture were added to the
carbon side of the grid and blotted for 4 seconds with Whatman® filter paper. The grids were prepared in high humidity and plunged into a pool of liquid nitrogen cooled liquid ethane using a manual plunge freezer. Grids were stored in liquid nitrogen until visualized.

Cryo-EM data was collected on a Titan Krios TEM (Thermo Fisher®) equipped with a K3® direct electron detector at a nominal magnification of 60,000x in super-resolution mode. Movies were recorded in frames to a total dose of 40 e⁻/Å².

ATP, ATP-10, and Apo movies were aligned using motion correction in Cryosparc. ADP movies were aligned with MotionCor2. CTF estimation, particle picking, and ab-initio model building were done in Cryosparc for all datasets. 3D refinement and CTF refinement were done in Relion. Global refinements for ATP and ATP-10 datasets were done in Phenix. Model building and refinement were done with Coot. Cryo-EM electron density maps and refined atomic models were visualized in UCSF Chimera.

Negative stain-EM grids were prepared using the same ratio of Hsp60 and Hsp10 as above but using a final Hsp60 concentration of 0.05 mg/mL. 400 mesh copper grids on a continuous carbon support (EMS®) were glow discharged as described above. 3 µL of sample were added to the carbon side of the grids, incubated for 30 seconds at ambient temperature, and blotted using filter paper before adding 3 µL of 2% uranyl acetate two separate times. Grids were stored in storage boxes overnight to allow for complete drying before visualization.

Negative stain grids were imaged using a Jeol® 3200FS TEM using an accelerating voltage of 300 kV at a nominal magnification of 100,000x. Images were recorded on a Gatan® UltraScan 4000 CCD using 1 second exposures.
RESULTS

MDH refolding was used to assess the *in-vitro* activity of purified Hsp60 and Hsp10. (Figure 4.1) We observe a loss of MDH activity of about 90% in the denatured MDH sample when compared to the native MDH sample. However, incubation with Hsp60 and Hsp10 shows a recovery of MDH activity of about 90%, similar to the activity level seen in the native MDH sample. Denatured MDH regains no activity when incubated with Hsp60 and Hsp10 but no ATP. Hsp60, Hsp10 and ATP with no MDH show no background level of activity in this assay.

![Graph showing MDH refolding assay](image)

Figure 4.1: MDH refolding assay. Denatured MDH shows a 90% loss of activity compared to native MDH. MDH that is refolded with Hsp60 and Hsp10 regains about 90% of the activity compared to native MDH. Lack of ATP demonstrates no Hsp60 refolding activity of MDH.
Cryo-EM reconstructions of Hsp60 in its various conformations show significant structural changes. Hsp60 in its Apo conformation (4.8 Å) adopts a double-ring structure with an open apical domain. (Figure 4.3) The resolution obtained does not allow for fitting of atomic coordinates in this model, but the overall dimensions of the complex are shown. Class averages obtained from the dataset show very clear and distinct top views, tilts, and side views of the complex. (Figure 4.2)

Figure 4.2: Hsp60 Apo conformation 2D class-averages.

Figure 4.3: (A) Side view and (B) top view of Hsp60 Apo conformation.
A reconstruction of ATP bound intermediate without Hsp10 (ATP-10) was obtained with a local resolution of 3.2 Å along the equatorial domain. Atomic coordinates were only fitted along this area due to the low resolution of the apical domain. (Figure 4.5) 2D class-averages of ATP-10 dataset show a blurred apical domain, indicating disorder in that region. (Figure 4.4) Interestingly, high-resolution features are seen up until the flexible hinge region of the intermediate domain.

Figure 4.4: Hsp60 ATP-10 conformation 2D class-averages.

Figure 4.5: (A) Full cryo-EM map of ATP-10 conformation. (B) Fitted coordinates with low resolution features removed.
A 3.2 Å reconstruction of the ATP conformation of Hsp60 in complex with Hsp60 was obtained and atomic coordinates were fitted to this model. (Figure 4.6) The structure obtained

Figure 4.6: (A) 3.2 Å map of ATP conformation. (B) Fitted model. (C) 2D class-averages obtained from dataset.
shows Hsp60 in complex with Hsp10 forming an “American football” conformation. The atomic model suggests that Hsp60-Hsp10 interactions are mainly hydrophobic in nature with several leucines, isoleucines, and valines making up most of the contacts. (Figure 4.7) The inter-ring contacts appear to be composed of a hydrogen bond occurring between S464 of one ring and S464 on the opposite ring, as well as a L465 interacting with the beta and gamma carbons of E468 of the opposing ring. However, the structure obtained was missing density along the center of the rings, an intrinsically disordered methionine-glycine repeat tail that is thought to form the “floor” of the cavity in the barrel, and where more inter-ring interactions may be occurring. The resolution of our structure was high enough to obtain densities for the ATP molecule bound in the nucleotide binding pocket. Upon inspection, we can fit the gamma-phosphate of the ATP molecule and localize it to be in close proximity to D52 and D87. (Figure 4.8)
Figure 4.8: ATP bound nucleotide binding pocket. D51 and D86 are in close proximity to the gamma phosphate of the ATP molecule. Magnesium ion shown as a green sphere.

The effect of the D51 and D86 residues on ATP hydrolysis was assessed by mutating the sites to serines. An MDH refolding assay was done with these mutants, and it was observed that D86S shows a lower activity level than even the denatured MDH negative control. (Figure 4.9) D51S shows a higher activity level compared to the denatured MDH control. A t-test showed that the difference between the two mutants compared to the denatured MDH control is significant in both cases.

Figure 4.9: MDH refolding assay with D51S and D86S mutants.
A 5.1 Å reconstruction of Hsp60 in the ADP bound conformation was obtained. This structure reveals a single ring conformation where the two rings that make the preceding ATP conformation are no longer present. The limited resolution did not allow for fitting of atomic coordinates; however, the overall conformation can be observed in the cryo-EM model, and no double ring complexes were observed in the class averaging of over 80,000 particles. (Figure 4.10).

![Figure 4.10](image.png)

Figure 4.10: (A) ADP conformation of Hsp60 shows a novel single ring structure. (B) 2D class averaging does not show double ring complexes.

**Discussion**

Refolding of acid denatured MDH *in-vitro* demonstrates that catalytically active Hsp60 and Hsp10 can be obtained through our expression and purification methods. We observed a significant increase in MDH activity when denatured MDH was incubated with Hsp60 and Hsp10 for one hour, which, when compared to the MDH activity of the denatured MDH sample, shows that the nearly complete loss of activity turned into an almost complete gain of function as compared to the native MDH sample. A control reaction without ATP was included and showed no increase in activity, with a signal almost identical to the denatured MDH sample. This, along with the last control with no MDH, suggest that the increase in MDH activity in our
The experimental sample is exclusively a result of the incubation of denatured MDH with our chaperonin. Knowing a purified protein is catalytically active enables structural studies as structures obtained will be of a functional protein and not of an inactive artifact.

The Apo conformation was not of a high enough resolution to fit atomic coordinates, but valuable information was still gathered. Our model shows a clear double ring complex that has an open apical domain. This can indicate that the apical domain of Hsp60 needs to be in an open conformation presumably to accommodate for substrate binding and entering the cavity. Hsp60 in the presence of ATP shows that nucleotide binding precedes Hsp10 binding as we were able to fit atomic coordinates for the ATP molecule in our structure. Interestingly, a disordered apical domain is observed in our model. This can perhaps be occurring to allow some flexibility in the region of Hsp60 that comes in direct contact with Hsp10. A rigid apical domain would not be suitable for the conformational changes required to allow Hsp10 to bind. It should be noted that the loss of resolution, and therefore the increase in flexibility, occurs in a loop found in the intermediate domain of Hsp60. This region in homologs such as GroEL is thought to be a flexible hinge that facilitates the required conformational changes to happen. This, along with the fact that the resolution along the equatorial domain is very high, indicate that the intermediate and apical domains are the flexible regions of the protein, at least in the Apo and ATP conformations.

The high-resolution ATP football complex highlights many of the features of the Hsp60 structure. Here, we see that the interaction between Hsp60’s hydrophobic apical domain and Hsp10 is in fact hydrophobic in nature itself. It can be observed that a series of leucines, isoleucines and valines drive the interaction between Hsp10 and the apical domain of Hsp60. It is noteworthy that very strong interactions do not exist along these contacts, possibly because
Hsp10 release is just as important as its binding. A very stable and strong interaction between Hsp60 and Hsp10 would likely trap the folded substrate inside the cavity indefinitely, or long enough to make the catalytic cycle inefficient in nature. The inter-ring contacts seen in our structure are few, consisting only of a single hydrogen bond per neighboring subunits and perhaps a weak hydrophobic interaction between L465 and the beta and gamma carbons of E468. However, as previously stated, there are parts of Hsp60 that our cryo-EM data did not result in electron density. This region is the methionine-glycine rich tail that extends from the equatorial domain into the folding cavity. There could be further inter-ring interactions occurring in that area that we were not able to see due to this information being absent in our dataset or lost during processing.

One of the most important features of this chaperonin is the nucleotide binding pocket, and more specifically, the mechanism of ATP hydrolysis. Our structure is the first to date to use real ATP to obtain this conformation. Previously, work done on this protein relied on non-hydrolysable analogs of ATP, but this made it impossible to map the actual position of the gamma phosphate of ATP that is involved in the hydrolysis of ATP into ADP. We localized the gamma phosphate to be a few angstroms away from two aspartates that may be responsible for the nucleophilic attack on the gamma phosphate to begin the hydrolysis reaction. To determine if one of these residues is directly involved in this mechanism, we mutated both sites separately into serines. This was done to preserve the polar nature of the amino acid residue, while substituting the potential nucleophile with an extremely weak one in the form of an OH group. Here, we see that the D86S mutant, which is 2 Å closer to the gamma phosphate than the D51S mutant, shows an even lower activity than the negative MDH control itself. We attributed this to the idea that this mutant may perhaps be able to bind substrate and trap it inside of the cavity but
is unable to progress through the natural mechanism. If true, this would lower the effective concentration of free MDH in solution, thereby lowering the activity of MDH below the negative control. The D51S mutant also showed significantly impaired refolding ability, however, the data suggests that the difference between it and the denatured MDH control is in fact statistically significant. A possible explanation for this can be that the mutation does in fact disrupt the binding affinity of ATP to the nucleotide binding pocket but does not abolish the ability to hydrolyze it. Because of this, we believe D86 is directly involved in the hydrolysis of ATP, likely acting as a nucleophile that attacks the phosphorus atom of the gamma phosphate to initiate hydrolysis.

Finally, a novel single ring intermediate was observed in the ADP conformation. More recently, other single ring intermediates have been observed in some chaperonins. The benefit of ring separation may be an expansion of the cavity in the chaperonin, potentially allowing for larger substrates to be folded. Higher resolution models are necessary to fit atomic coordinates to this intermediate to fully understand the conformational changes that occur between the transition from the ATP football complex to the ADP single ring.

**Future Work**

Further improvement of the Apo and ADP structures is necessary to obtain an atomic model of these key intermediates. Data collection is currently underway using preparation methods that are slightly different than the ones used for the work presented here. The ATP-10 model also needs to be confirmed through further data collection. This is to confirm that the lack of resolution along the apical domain is in fact a result of intrinsic flexibility along the intermediate domain of Hsp60 when ATP binds, or if poor sample quality during this data collection resulted in the loss of resolution observed. Finally, a double aspartate residue may help
determine whether the remaining levels of activity seen in the D51S mutant is due to this residue not being involved in ATP hydrolysis but rather having a kinetic effect on the rate of ATP binding to the nucleotide binding pocket.
Although great efforts have produced significant advances in the treatment of human immunodeficiency virus (HIV) it remains among one of the largest incurable infections across the world. The prevalence of the virus in terms of number of infections has largely decreased in developed countries but is still ravaging through lower income areas across the world. It is estimated that in 2019 about 38 million people were living with the infection across the world with 1.7 million new infections and 690,000 deaths worldwide [71]. Since the beginning of the epidemic around 50 years ago, more than 75 million infections have been reported, and about 32.7 million deaths. Of the current 38 million infected people, only about 25.4 million were undergoing some type of antiretroviral treatment [71]. That translates to about 35% of the infected population, or 13 million people being left without treatment. More than half of the active infections (20.7 million) are found in eastern and southern Africa where the socioeconomic climate is most unstable [71]. According to the CDC data, there are around 1.2 million active cases in the United States today [72]. The numbers and rate of infection currently affects southern states more than any other region, and although the rates across the nation have decreased significantly from the start of the epidemic, they have remained relatively steady for almost a decade. Infections in the country also affect minority groups disproportionately. (Figure 5.1)
HIV is a retrovirus that mainly targets activated CD4 T lymphocytes of the immune system via interactions with CD4 or with the CCR5 or CXCR4 chemokine coreceptors. Other cells displaying CD4 or chemokine coreceptors can also be infected by HIV [73]. HIV infections start with the Env complex on the surface of HIV interacting with the CD4 receptor on the surface of the host cell [74]. Env is comprised of three gp41-gp120 heterodimers, with gp120 interacting via a surface loop with CD4. Coreceptor binding follows and a fusion peptide on gp41 is inserted into the host membrane where a fusion pore forms and the viral membrane fuses with the host membrane to complete the infection [75]. (Figure 5.2)
Upon successful fusion with the host cell, the virus begins the process of preparing its genome for integration into the host genome in order to propagate indefinitely. The genome of HIV consists of three main structural polyproteins as well as an assortment of regulatory proteins. The gag reading frame contains the matrix, capsid, and nucleocapsid proteins, pol contains the protease, reverse transcriptase, RNase H, and integrase proteins, and env contains the gp120 and gp41 surface proteins [76]. (Figure 5.3)

First, the RNA genome of HIV is reverse transcribed into DNA by the viral enzyme "reverse transcriptase." Because the genome of the virus is made of RNA and neither the viral reverse transcriptase enzyme nor the host RNA polymerase that synthesizes the genome of newly formed viruses have proofreading activity, mutations arise very rapidly, and this is exponentiated
by a very high viral turnover rate. This results in a larger than usual amount of genetic variation in HIV; this in turn allows the viral progeny to become resistant to many of the anti-HIV drugs used in patients [77]. This enzyme is encoded in the Pol region of the HIV genome and is incorporated into the virion via a fusion of Gag-Pol that occurs due to a frameshift in the p6 region of Gag seen in 5% of Gag translations [77]. This creates a DNA version of the RNA HIV genome that can then be integrated into the host genome to be subsequently transcribed and translated to form more copies of the virus inside of the host [78].

The integration of viral DNA into the host genome is mainly catalyzed by the HIV enzyme “integrase”. Upon reverse transcription of the RNA genome, the DNA travels to the nucleus where it is covalently attached to a chromosome of the host’s genome. It is not well understood if the virus has preferred integration sites, but studies show that very often, viral genomes integrate near active transcription units [78]. If true this would increase the transcriptional activity of the integrated provirus, increasing the propagation of the mature virus onto other cells.

The HIV provirus is responsible for propagating itself by mass producing HIV virions and releasing them so that they can carry out a similar life cycle in other cells. Here, the provirus is transcribed and translated using host enzymes and ribosomes to produce many copies of all of the proteins encoded in the HIV genome. Once sufficient copies of all proteins have been produced, assembly of new HIV particles can begin (Figure 6.4 A). This process is mediated mostly by the structural proteins in the gag gene. Gag is expressed as a polyprotein which is linked to the inner plasma membrane of the host cell via a myristoyl group that anchors it upon the interaction of matrix protein and phosphatidyl inositol bisphosphate [79] (Figure 5.4 B & C). High local concentrations of membrane-bound gag bud out of the cell to form an immature
virion bound by the host plasma membrane (Figure 5.4 D). The resulting virion is non infective and requires further processing to achieve its final form [80]. The immature virion undergoes “maturation” in which HIV protease cleaves the gag polyprotein into individual proteins that assemble into their final forms [81] (Figure 5.4 E). The mature virus is now infectious and able to enter other cells and complete more rounds of propagation.

Figure 5.4: (A) Illustration of the viral life cycle. (B) Domain representation of gag. (C) Structural model of gag. (D) Illustration of immature virion. (E) Illustration of fully mature HIV. [81]
CAPSID ASSEMBLY

The replication of HIV viruses is a complex and concerted series of steps that must occur with a high level of precision to ensure that the resulting virions are able to further infect other cells and propagate. In fact, every available antiviral drug targets a single enzyme in the viral life cycle. (Table 5.1)

Table 5.1: Current Antiretroviral (HIV) drugs. [76]

<table>
<thead>
<tr>
<th>Nucleoside/nucleotide reverse transcriptase inhibitors (NRTI)</th>
<th>Non-nucleoside reverse transcriptase inhibitors (NNRTI)</th>
<th>Protease inhibitors (PI or PI)</th>
<th>Integrate inhibitors (INI)</th>
<th>Entry inhibitors</th>
</tr>
</thead>
</table>

*Year of approval in parentheses.
**Not approved in Germany.

Interestingly, a widely studied mechanism that has not produced viable drug targets to date is capsid assembly. Upon cleavage of capsid from matrix and SP1, its flanking partners, HIV capsid monomers assemble into a helical lattice of hexagonal hexamers [82]. Twelve pentagonal pentamers are incorporated in the assembled capsid and confer curvature at both extremes to close the cone-like polymer [83]. (Figure 5.5) The fully assembled capsid encases the genome of the virus along with the bound nucleocapsid protein. It has been shown that altering the morphology or stability of the assembled capsid is enough to abolish viral infectivity [83, 84].
Recently, functionalized fullerene compounds were shown to inhibit HIV replication by 99%, and pull-down assays showed that the fullerene compounds were interacting with capsid. Previously, it was thought that fullerene inhibition of HIV was a result of C$_{60}$ binding to the catalytic pocket of the HIV protease and inhibiting its function, but the same study disproved that idea [85]. Because capsid is strictly a structural protein with no known moonlighting functions, the fullerene compounds may be directly altering the assembly of the HIV capsid cones, resulting in a non-infective virus, although there is no evidence for this as of now.

**PROJECT GOALS AND HYPOTHESIS**

HIV continues to be one of the most widespread epidemics we face across the world, especially in the less developed countries. Due to the high mutation rate of the virus, drug resistant strains are emerging at an alarmingly high frequency. To date, there are no drugs that utilize capsid assembly as a potential drug target. Elucidating the interaction between capsid and fullerene compounds that strongly inhibit viral replication can highlight a potential target for subsequent drug development.
Our first goal is to generate clones of HIV capsid (CA) and HIV capsid with the adjacent spacer peptide SP1 (CA-SP1) in bacterial expression vectors. Second, we will obtain pure recombinant CA and CA-SP1 from bacterial cultures purified using a tandem liquid chromatography approach. We will then assay the protease processing of CA-SP1 \textit{in-vitro} in the presence of functionalized fullerene compounds. Finally, we will study the effect of fullerene compounds on CA and CA-SP1 assembly \textit{in-vitro} using negative stain electron microscopy (ns-EM).

We hypothesize that functionalized fullerenes will not inhibit proteolytic processing of CA-SP1 but rather inhibit the proper assembly of capsid \textit{in-vitro} via hydrophobic interactions between the fullerene carbon cage and the hydrophobic c-terminal domain of CA/CA-SP1.
Chapter 6: Molecular Cloning

Materials and Methods

cDNA for the genes CA and CA-SP1 was amplified using PCR (Q5® DNA Polymerase, NEBiolabs) under the recommended thermocycling conditions. Q5® 2X Master Mix, which contains the Q5® enzyme as well as 10 µM of free dNTPs, was used for all PCR reactions. 0.5 µM of forward and reverse primers obtained from IDT™ along with 10 pmol of CA cDNA were also included in the PCR reaction. Thermocycling was set up with a 98°C denaturation for 10 seconds, followed by annealing at 69°C for 25 seconds, and an extension at 72°C for 16 seconds for a total of 35 cycles. Restriction sites for Ndel and XhoI were added to the 5’ and 3’ ends, respectively, of the gene for cloning into a pET-22b vector. The PCR product was run on a 0.8% agarose gel containing SybrSafe® DNA dye at a 1X concentration at 120 V for 30 minutes and visualized using blue light with a blue light filter. The PCR products were gel extracted and purified using a PureLink™ gel extraction kit. Briefly, the resulting DNA band from the agarose gel was excised using a steel blade and weighted. The gel slice was dissolved using Gel Solubilization Buffer at a 5:1 ratio (v/w) of buffer to agarose gel. The gel slice was incubated at 50°C for 10 minutes. One gel slice volume of pure isopropanol was added to the dissolved gel slice. The dissolved gel slice was added to a silica column and centrifuged at 12,000xg for 1 minute. 600 µLs of Wash Buffer were used to wash the bound DNA and centrifuged at 12,000xg for 1 minute. The wash buffer was discarded, and the column was spun down again at 18,000xg for 2 minutes to completely remove trace amounts of wash buffer. DNA was eluted using Elution Buffer that was preheated to 65°C. The column was incubated with the elution buffer for 2 minutes and centrifuged at 12,000xg for 2 minutes. The eluted DNA was collected, and the DNA concentration was measured using a spectrophotometer by measuring absorption at 260 nm. The
DNA was stored in -20°C long term. The gene for CA-SP1 was amplified, extracted, and purified using the same procedure. Two micrograms of each gene as well as two micrograms of pET-22b (EMD Millipore®) were digested for 4 hours at 37°C with NdeI and XhoI (NEBiolabs®). The digestion reaction was composed of 1 µL of each restriction enzyme, 2 µg of DNA, water to a total volume of 50 µLs, and Cutsmart® buffer at a 1X concentration. The digestion products were run on a 1% agarose gel at 120 V for 20 minutes, excised with a steel blade, and gel extracted using the previously described protocol. Ligation reactions were done using T4 DNA Ligase® from NEBiolabs®. Each reaction was done at a total volume of 20 µLs containing plasmid DNA, gene DNA, 1 µL of ligase enzyme, ligase buffer at a concentration of 1X containing ATP, and water. Each gene was ligated with 100 ng of digested plasmid at a molar ratio of 3:1 overnight at 16°C. The ligation reaction was terminated at 65°C for 15 minutes. One tube of One Shot™ Chemically Competent cells was thawed on ice for 10 minutes. 2 µLs of each ligation reaction was added to its respective tube of competent cells and mixed by light flicking. The cells were allowed to incubate with the ligation product for 30 minutes on ice. The cells were then heat shocked in a water bath preset to exactly 42°C for exactly 30 seconds. The cells were returned to the ice for 5 minutes. 250 µLs of room temperature S.O.C. media were added to each tube of cells and incubated at 37°C for 1 hour while shaking horizontally at 250 rpm. After incubation, the transformed cells were plated on LB agar plates containing 200 mg/mL ampicillin. Four plates were prepared for each transformation reaction with 10 µLs, 25 µLs, 50 µLs, and 100 µLs of the transformed cells and spread with a sterile spreader. Different volumes were used to ensure evenly spread colonies on one of the plates. The plates were inverted and incubated overnight at 37°C. Colony growth was observed on all plates and 3 colonies were selected from each transformation reaction based on size and morphology to
inoculate liquid cultures. 50 mL cultures were made for each colony selected containing 200 mg/mL ampicillin in a baffled 250 mL flask with a cap. The cultures were grown overnight at 37°C with shaking at 220 rpm. 25 mL of each culture was transferred to a sterile 50 mL conical tube and centrifuged in a tabletop centrifuge at 5250xg for 15 minutes at 4°C. Plasmid DNA was extracted and purified from the cells using a PureLink™ HiPure Plasmid Midiprep kit. The harvested cells were resuspended in 4 mL of Resuspension Buffer containing RNAse A. The cells were then lysed using 4 mL of Lysis Buffer and inverting 10 times. After a 5-minute incubation at room temperature, 4 mL of Precipitation Buffer were added, the samples were inverted 10 times, and incubated on ice for 10 minutes. The cell lysate was centrifuged at 18,000xg for 15 minutes at 4°C. During this time, a midi purification column was equilibrated with 5 mL of Equilibration Buffer allowed to pass with gravity. The supernatant was added to the midi column and allowed to pass through the column by gravity. 10 mL of wash buffer were added to the column and allowed to pass by gravity two separate times. The DNA was then eluted using 5 mL of Elution Buffer that had been preheated to 65°C. The eluted DNA was collected in a clean tube and precipitated by adding 3.5 mL of pure isopropanol. The precipitated DNA was centrifuged at 18,000xg for 30 minutes at 4°C. The supernatant was decanted, and the pellet was washed with 5 mL of 70% ethanol. The washed DNA was again centrifuged at 18,000xg for 15 minutes at 4°C. The ethanol was decanted, and the resulting pellet was allowed to air dry overnight. The DNA pellets were resuspended in 100 µLs of TE buffer and stored at -20°C. The plasmids were sequenced to confirm the presence of our target gene of interest in each plasmid. Since three plasmids were purified for CA and three for CA-SP1, the plasmid with the highest concentration and correct sequence was used for the rest of the experiments. One tube of BL21-DE3 competent cells was used for each plasmid. Frozen tubes of competent cells were
thawed on ice for 10 minutes. 50 ng of plasmid DNA were added to their respective tubes and incubated on ice for 30 minutes. Cells were heat shocked for 10 seconds in a water bath set to 42°C and immediately transferred back to an ice bucket for 5 minutes. 950 μLs of S.O.C. media were added to each transformation reaction and incubated at 37°C with shaking horizontally at 220 rpm for one hour. Four plates were prepared for each transformation reaction with 10 μLs, 25 μLs, 50 μLs, and 100 μLs of the transformed cells and spread with a sterile spreader. Different volumes were used to ensure evenly spread colonies on one of the plates. The plates were inverted and incubated overnight at 37°C. Colony growth was observed on all plates and 5 colonies from each transformation reaction were selected based on size and morphology to inoculate liquid cultures for protein expression check. 25 mL cultures were prepared with 2XYT containing 200 mg/mL ampicillin and were incubated at 37°C with shaking at 220 rpm until an OD₆₀₀ of 0.6 was reached in each culture. At this point, IPTG was added to a final concentration of 1 mM and allowed to continue incubating for 2 more hours. After the incubation, 1 mL samples of each culture were collected and centrifuged at 18,000xg for 1 minute at room temperature. The resulting cell pellets were resuspended in 100 μLs of deionized (DI) water and 2 μLs of resuspended cells were added to 10 μLs of 4X Lammeli Buffer. 10% SDS-PAGE gels were run at 200 V for 45 minutes and stained with Coomassie blue staining buffer. The colonies with the highest expression level of CA and CA-SP1, respectively, were reinoculated in 2XYT, grown to an OD₆₀₀ of 0.4, mixed 1:1 with 50% glycerol, aliquoted, and frozen at -80°C for long-term storage.
RESULTS

PCR amplification of the CA and CA-SP1 genes resulted in bright bands at around 700 bp. After restriction enzyme digestion, the size of the two genes remained unchanged, appearing again at around 700 kb for both CA and CA-SP1. The digested pET-22b plasmid appeared to be around 5 kb in length after digestion, and a faint band near 150 bp appeared. (Figure 6.1)

Figure 6.1: Agarose gel electrophoresis of restriction enzyme digestion products. CA gene appears as a sharp band near the 0.76 kb marker. CA-SP1 appears as a bright band near the 0.76 kb marker. pET-22b appears near the 5.0 kb marker and a faint smear appears near the 0.15 kb marker.

After transforming competent cells with the ligation product of CA, plates with four different volumes of cells were prepared to ensure that at least one plate contained colonies that did not overlap or touch. The plate with 10 µLs of transformed cells showed spaced colonies of a good size that were suitable for inoculation. The cells were plated on plates containing 200 mg/mL therefore all bacterial growth observed must have been resistant to the antibiotic used. The plates that resulted from the transformation of CA-SP1 were virtually indistinguishable from the CA plates.
**DISCUSSION**

The amplification of our target genes produced products that were very near their expected size based on their theoretical length. The gene that encodes for CA is 693 bp in length and the observed band was closest to the 0.76 kb maker. Because the DNA fragments run in a logarithmic manner, and no other by product was seen in the amplification, it is safe to say that the resulting product is most likely our gene of interest. Likewise, CA-SP1 was amplified and resulted in a band that is very close to the expected size of 745 bp. After digestion with NdeI and XhoI, it is almost impossible to see the fragment of DNA that is removed from the ends of the genes because it is only about 3-4 bp in length. Because of this, Figure 7.1 shows bands that are practically identical to the bands produced from the PCR reaction, but this is expected. However, the DNA band that is seen for the cut pET-22b plasmid gives us more information because it appeared very close to the theoretical size of 5.4 kb. Uncut plasmids are circular and supercoiled, and they usually do not run according to size because of this. Once cut, plasmids do run true to size because they have been linearized. Furthermore, the faint smear that appears on the bottom of the gel appears to be the cut fragment of DNA that resides between the NdeI and XhoI restriction sites on the multiple cloning site region of pET-22b. This is evidence that the restriction enzyme reaction most likely was successful for all three samples. Furthermore, we did not observe fragmentation of the genes or the plasmid, something that would be indicative of a problem like non-specific digestions such as the ones seen as a product of star activity. Although ligation reactions do not produce quantitative or qualitative results, it is the success of the transformation reaction that we can use as evidence of a successful ligation. Had our ligation reactions been unsuccessful, the DNA that is inserted into the bacterial cells during transformation would be linear. Linear DNA is degraded inside of bacteria if it is taken up at all.
If degraded, the cell would not be resistant to the antibiotic used as the antibiotic resistance gene would have been lost along with the plasmid. Therefore, the presence of colonies not only evidentiate a successful transformation, but it also indirectly tells us that the ligation reaction was most likely successful as well.
Chapter 7: Protein Expression and Purification

**MATERIALS AND METHODS**

The protocols used to grow cultures, extract protein, and purify recombinant protein via chromatography were the same for CA and CA-SP1 unless noted otherwise. A 1 mL frozen glycerol stock was thawed and added directly into 1 L of 2XYT media containing 200 mg/mL ampicillin. The culture was incubated at 37°C with shaking at 220 rpm in a baffled 2.5 L flask. Cells were grown until an OD$_{600}$ of 0.6-0.8 was reached at which point the culture was induced with the addition of IPTG to a final concentration of 0.5 mM. Expression was allowed to occur for an additional 3 hours at 37°C. Cells were then harvested by centrifugation at 6000xg for 30 minutes at 4°C. The spent media was discarded, and the resulting pellets were resuspended in lysis buffer containing 25 mM Tris pH 7.5, 50 mM NaCl, and 50 mM EDTA. Lysis buffer was added to a ratio of 5 mL of lysis buffer per gram of cell pellet. DTT was added to a final concentration of 1 mM to the resuspended cells. Lysozyme purchased from Sigma-Aldrich® was added to a concentration of 0.25 mg/mL of resuspended cells. Cells were incubated with lysozyme at 4°C in a rotating shaker for 1 hour. The cells were then frozen in a -20°C freezer and thawed at room temperature a total of 3 cycles. MgCl$_2$ was added to the lysed cells to a final concentration of 60 mM along with DNase also purchased from Sigma-Aldrich® to a final concentration of 0.1 mg/mL. The cell lysate was incubated with DNase for 1 hour at 4°C in a rotating shaker. Streptomycin sulfate was added to the lysate to a final concentration of 3% w/v. The cell lysate was again incubated for 1 hour at 4°C in a rotating shaker. Cell lysates were clarified by centrifugation at 18,000xg for 30 minutes at 4°C. Clarified lysates were treated differently for CA and CA-SP1. The CA lysate was treated with a saturated ammonium sulfate solution to a final concentration of 20% v/v. The protein was allowed to precipitate for 30
minutes at 4°C in a rotating shaker. The clarified lysate for CA-SP1 was treated similarly except that this one was treated with ammonium sulfate only to a concentration of 18% v/v. Ammonium sulfate precipitates were then centrifuged at 12,000xg for 30 minutes at 4°C to collect the precipitated protein. The supernatants were discarded, and protein pellets were resuspended in 5 mL of Buffer A (25 mM Tris, pH 8.0) per gram of protein pellet. The resuspended protein (crude extract) was then flash frozen with liquid nitrogen and stored for future purification.

CA and CA-SP1 were purified using the same protocol unless noted otherwise. 1 mL of crude extract was dissolved in 9 mL of buffer A and loaded onto a pre-equilibrated set of 2 ion exchange columns stacked directly on top of each other. The first column was a 5 mL HiTrap™ SP HP and the second column a 5 mL HiTrap™ Q HP. The flow through was collected and concentrated to <500 μls using an Amicon® Ultra centrifugal filter unit with a molecular weight cut off of 10 kDa. The concentrated protein was then loaded onto a Superdex® 75 gel filtration column. The single peak observed was pooled and concentrated as described above, to a protein concentration of >20 mg/mL. Protein concentrations were determined using a Pierce™ BCA Assay. SDS-PAGE analysis was used to determine the point of elution of our protein of interest throughout the purification procedure. All SDS-PAGE gels were made to a final acrylamide concentration of 10% and were run at 200 V for 45 minutes.
RESULTS

Neither CA nor CA-SP1 bound to the ion exchange columns at pH 8.0, therefore the flow through was collected while contaminants remained bound to the column. (Figure 7.1)

Figure 7.1: Elution profile of (A) CA and (B) CA-SP1 using a protocol of tandem cation and anion exchange columns. The protein was localized in the flow through, which was collected, concentrated, and passed through a second chromatography column. The large peak seen in the elution with 100% buffer B was determined to be contaminants and none of the protein of interest.

Pooled flow through was loaded onto a gel filtration column and a single large, sharp peak appears at 17.6 mL in the elution for CA and 16.6 mL for CA-SP1. It can be observed that CA elutes about 1 mL after CA-SP1. (Figure 7.2)
Figure 7.2: Elution profile of (A) CA and (B) CA-SP1 in gel filtration chromatography. A single sharp peak is observed for both samples, with CA-SP1 eluting about 0.5 mL before CA due to the slightly larger molecular weight.

Fractions from all chromatography runs were analyzed further using SDS-PAGE. 10% acrylamide gels were used to visualize protein for both CA and CA-SP1. The SDS-PAGE analysis of CA after size exclusion chromatography shows the fractions corresponding to the single observed peak to be a highly pure sample of CA at the correct molecular weight. (Figure 7.3) Fractions for CA-SP1 (not shown here in SDS-PAGE) were nearly identical in purity, but slightly larger in size due to the additional weight the SP1 peptide adds to this construct. In both cases, protein purity was estimated to be >98% using only SDS-PAGE analysis.
DISCUSSION

The protocols for protein extraction and purification were very similar for CA and CA-SP1. It is not surprising that this is the case since the 1 kDa SP1 peptide that differentiates the two is known to be an inert spacer that does not affect CA stability or behavior. Our cell cultures incorporated an ammonium sulfate pre-purification step that allowed us to obtain a very high purity using only 2 columns without using an affinity tag for purification. At the pH used for tandem cation-anion exchange chromatography we determined using SDS-PAGE that both CA and CA-SP1 did not bind to either column. Therefore, we used these columns to bind contaminants and collect the flow through were we determined our protein eluted. Contaminants were observed in the single peak obtained during this step in the purification which was the elution at 100% buffer B, or 1 M NaCl. Since this gave us good results, we did not attempt to change the buffer conditions to force our protein of interest to bind to either ion exchange chromatography column. Size exclusion chromatography showed a very large single and sharp
peak, indicative of a large amount of a single protein species. Here, the elution volume for CA and CA-SP1 varied by 0.5 mL which was expected due to the larger molecular weight of CA-SP1. Our SDS-PAGE analysis showed a protein purity >98% by Coomassie blue staining. It also shows that the peak observed in the chromatogram does in fact correspond to our protein of interest. Because our next experiments involved electron microscopy, we determined that the purity obtained with this simple protocol was sufficient for the subsequent experiments.
Chapter 8: *In-vitro* Capsid Processing and Assembly Inhibition

**Materials and Methods**

*In-vitro* protease inhibition using fullerene compound F85 was assessed by conducting a protease cleavage reaction on purified CA-SP1 in the presence of fullerene. Bevirimat, a well-studied protease inhibitor that only works with assembled HIV-Gag was used as a negative control. 20 μL reactions were set up in buffer containing 25 mM Tris pH 7.5, 150 mM NaCl. CA-SP1 was used at a final concentration of 20 μM and HIV protease purchased from Sigma-Aldrich® was used at a final concentration of 4 μM. Fullerene-containing reactions had CA-SP1 pre-incubated with fullerene for 15 minutes prior to the addition of the protease. HIV protease was mixed with CA-SP1 as well as fullerene-treated CA-SP1 and incubated for 30 minutes at 37°C. Several molar ratios of fullerene and bevirimat to CA-SP1 were assessed. After incubation, 2 μg of each sample were loaded on a 15% SDS-PAGE gel and run for 1 hour at 200 V. All gels were stained using Coomassie brilliant blue.

To assess fullerene capsid assembly inhibition *in-vitro*, purified CA and CA-SP1 were diluted to a final concentration of 1 mg/mL in assembly buffer containing 25 mM Tris, pH 8.0, and 2 M NaCl at 37°C for 2 hours. A set of assembly reactions were also prepared with an equimolar ratio of fullerene to CA and CA-SP1 along with vehicle controls containing DMSO. After incubation, 3 μLs of each sample were taken and placed on the carbon side of a glow-discharged lacey carbon copper EM grid. The samples were allowed to adsorb for 1 minute and excess sample was blotted with filter paper. The samples were stained by adding 3 μLs of 2% uranyl acetate two separate times, blotting with filter paper in between. The stained grids were allowed to air dry and imaged in a Jeol-3200 FS TEM. Images were acquired at a nominal magnification of 40,000x with a 1 second exposure in a Gatan® CCD camera.
DLS measurements were obtained using a Malvern® Zetasizer using a fixed angle of 173° at ambient temperature. Fullerene F85 was analyzed in pure DMSO at a concentration of 100 µM.

RESULTS

SDS-PAGE analysis of the protease inhibition assay shows that all samples of CA-SP1 were proteolytically cut in the presence of HIV protease. The DMSO vehicle control shows nearly identical levels of cleavage as the samples containing F85 or Bevirimat. (Figure 8.1)

![Protease inhibition assay](image.png)

**Figure 8.1:** Protease inhibition assay. CA-SP1 is equally cut in the presence or absence of F85 fullerene when compared to the DMSO vehicle control. Bevirimat also does not show any level of protease inhibition.

CA and CA-SP1 *in-vitro* assembly produced long rod-like assemblies such as the ones that have been described previously [82]. The DMSO vehicle controls did not result in any significant morphological changes when compared to the positive control assemblies. However, obvious differences were observed in the CA and CA-SP1 assemblies that contained F85. Here, the morphology seems to turn the long filament-like assemblies into fragments of themselves.
The occurrence of assemblies overall in the F85 samples is also much lower than the positive control or the DMSO control (not depicted in this result).

Figure 8.2: CA and CA-SP1 in-vitro assembly. CA (A) and CA-SP1 (B) positive control assembly results in rod-like particles several hundred nanometers in length. CA DMSO control (C) and CA-SP1 DMSO control (D) show no change in size or shape. CA F85 (E) and CA-SP1 F85 (F) show a clear change in morphology and size.
DLS measurements on pure F85 show an unusually large size of approximately 300 nm. This measurement is further complicated by the high PDI value of the sample, indicating a lack of homogeneity and possible aggregation. (Figure 8.3)

![DLS measurement graph](image)

Figure 8.3: DLS measurement on F85 dissolved in DMSO. A mean size of around 300 nm is observed, although based on the deviation and large PDI it is a rough approximation only.

**DISCUSSION**

Our *in-vitro* protease inhibition assay using F85 corroborates the previously published data that HIV inhibition by F85 works in a protease-independent manner [85]. Bevirimat was used as a control in the form of an established protease inhibitor that is known to be inactive when used with CA-SP1 as its substrate. Since comparable protease activity was observed in the F85 containing samples as the bevirimat and DMSO vehicle samples we concluded that F85 has no protease inhibition activity that is specific to the processing of CA-SP1 into CA and SP1. We
did not investigate, however, if, like bevirimat, F85 only shows protease inhibition using full length Gag polyprotein assembled into immature virus-like particles (VLPs).

Using pure recombinant CA and CA-SP1 we were able to establish an assay to determine whether F85 is instead inhibiting HIV progression \textit{in-vivo} through the inhibition of proper capsid assembly. Using high protein and salt concentrations we were able to observe capsid assembly \textit{in-vitro} using both CA and CA-SP1 constructs. The observed assemblies undoubtedly vary from real viral capsids, but they share key structural features that make them a viable method of studying capsid assembly. More specifically, the helical lattice that is formed in our \textit{in-vitro} assay is the same as the fullerene-like lattice that is observed in actual capsids. In other words, the hexamer-of-hexamer assembly is the same even if the dimensions of the assembly vary. This key difference between capsid assemblies \textit{in-vivo} \textit{vs} \textit{in-vitro} can be attributed to the lack of pentameric complexes \textit{in-vitro}. When immature virions are processed for maturation, conical capsid cores are formed when 12 capsid pentamers are incorporated on the edges of the cone. These induce a curvature that allows the capsid core to close, and \textit{in-vitro}, we cannot control for the integration of these pentamers into our assemblies and therefore we obtain an abnormally long spiral capsid lattice that most likely stops growing when the concentration of free CA in solution drops. Knowing that aside from the lack of the 12 pentameric subunits, our assemblies are structurally representative of real capsid cores, we preformed the assembly in a 1:1 molar ratio of CA or CA-SP1 to F85. We observed that the DMSO vehicle had no impact on the morphology of the capsid assemblies, while a significant change in size and morphology can be seen in the assemblies formed in the presence of F85. The overall amount of assembly, albeit morphologically different, was also significantly reduced in the F85 containing samples when compared to the positive control and the DMSO control. No significant difference was observed.
when comparing CA assemblies to CA-SP1 assemblies, indicating that the SP1 peptide may not necessarily play a role in the interaction between CA and F85, at least in the assembly inhibition capacity. Something we observed during the experiment was the high degree of fullerene precipitation and sedimentation. This can become problematic as fullerene precipitation may effectively decrease the bioavailability of fullerene in solution. This may contribute to the fact that a small degree of morphologically compromised assemblies was still observed.

To have a better idea of the behavior of F85 in solution we analyzed a 100 μM sample in pure DMSO, a solvent that is presumably much better suited for the compound than the aqueous environment used in our assembly assay. Using DLS we observed that practically all of the fullerene exists as a type of aggregate in the 300 nm size range. The PDI for this sample was also very high, this being indicative of a highly polydisperse and likely aggregated species. This may further support the claim that a low concentration of bioavailable F85 is in solution at any given time, so the molar ratios are not 1:1 but rather something that is much lower in fullerene than what is calculated. This may allow for partial capsid assembly, especially if a fullerene to protein ratio that is even close to 1:1 or higher is necessary to abolish all capsid assembly.

**Future Work**

Having observed a morphological difference in the capsid assemblies containing F85 we would like to explore other types of fullerene compounds to corroborate this hypothesis. Furthermore, the hydrophobic nature of fullerene compounds makes them difficult to use in the environment that proteins are used in. Therefore, a similar fullerene compound with a few functional groups that aid with solubility may be an important next step in this work. As shown previously, we believe that the actual bioavailable concentration of fullerene in solution is much lower than what is calculated because a significant amount of the compound is forming large
aggregates that are not amenable to our capsid assembly inhibition assays. Furthermore, evaluating these compounds with other constructs derived from the Gag polyprotein may be necessary to determine if capsid assembly inhibition can only occur in the more complex assemblies seen in immature virion particles. Other proteins to consider are, for example, nucleocapsid and matrix. Both of these structural proteins play very different roles in the HIV life cycle and interaction of fullerene with these proteins may also be occurring.
References


42. Wang, J., et al., *MitCHAP-60 and Hereditary Spastic Paraplegia SPG-13 Arise from an Inactive hsp60 Chaperonin that Fails to Fold the ATP Synthase β-Subunit*. Scientific Reports, 2019. 9(1).


Vita

Alejandro Rodriguez was born and raised in El Paso, Texas. He attended Hanks High School and continued his education at the University of Texas at El Paso, where he graduated *suma cum laude* and received a Bachelor of Science degree in Cellular and Molecular Biochemistry in May of 2016. As an undergraduate student he joined the laboratory of Dr. Ricardo Bernal where he got his first experience in a research lab. He became a graduate student of Dr. Bernal where he used cryo-electron microscopy to study protein structure and function.

During his time as a graduate student, Alejandro enjoyed teaching undergraduate level labs in different fields, from structural biochemistry to organic chemistry. He presented his work in several conferences and won awards for some of his presentations. He considered attending research conferences one of the best parts of being a graduate student. During this time, he authored several publications, from structural biochemistry reviews to papers in inorganic chemistry through collaborations with other labs. He enjoyed using electron microscopy to learn about fields other than his own.

Alejandro is currently living in El Paso as he gets ready to graduate. He lives with his wife and their many dogs. He is very close to his family and spends as much time as possible with them. In his free time, Alejandro likes to watch practically all sports, but his favorites are baseball, soccer, and Formula 1. He feels incredibly lucky to have had the chance to pursue a degree in his institution under the mentorship of Dr. Bernal and in the company of all his fellow lab members. He hopes to be able to continue conducting impactful research in the future.

Contact email: alex.rod93.ar@gmail.com