THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

RECONSTITUTION OF THE YEAST CENTROMERIC NUCLEOSOME CORE PARTICLE

DANIEL FRIEDLAND
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Biology and Spanish
with honors in Biology

Reviewed and approved* by the following:

Song Tan
Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Stephen Schaeffer
Professor of Biology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The centromere constitutes the locus on the chromosome onto which the kinetochore assembles and microtubules attach to direct separation of sister chromatids during mitosis. Proper formation of the centromere-kinetochore complex is essential for accurate cell division as abnormalities in the cell cycle are associated with many diseases, including cancer. Thus, understanding the structures and mechanisms of the centromere-kinetochore complex may provide insight into disease pathogenesis.

Centromeric DNA, along with all nuclear DNA, is stored in a highly condensed, organized structure known as chromatin. Chromatin consists of a series of octameric protein complexes around which the DNA is wrapped. The complex of this octamer and its associated DNA is called the nucleosome core particle. Canonical nucleosome core particles are comprised of two copies each of H2A, H2B, H3, and H4, but centromeric nucleosomes contain a variant, CenH3 in place of H3. Nucleosomes serve as the substrate for several post-translational modifications, which change chromatin structure, facilitate interactions with chromatin enzymes, and regulate access to certain genes for transcription. The nucleosome serves as a fundamental model for studying gene regulation.

Unlike the centromeres of most eukaryotes, the budding yeast centromere contains only one nucleosome. This makes the yeast centromeric nucleosome an ideal model for centromere studies. In this project, a yeast centromeric nucleosome was created for potential use in binding studies with the kinetochore proteins. Additionally, parameters for further reconstitutions of this nucleosome were determined.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am extremely grateful for the guidance and support I received as an undergraduate researcher in the Tan laboratory. To Nick, thank you for serving as a mentor for me when I first started working in lab. Your commitment, patience, and knowledge provided me with a strong foundation for working independently and troubleshooting my experiments. To my fellow undergraduates, especially Turner, thank you for your support and advice over the past 2.5 years. To Kim, Jian, Ashish, Jacob, Bryan and Sanjay, thank you for answering my many questions and helping me troubleshoot. To Lauren, I truly cannot imagine what the process of my thesis lab work would have been like without your dedication, support, wit, and humor. Finally, to Song, I cannot even begin to express the gratitude I feel for the guidance you have provided me. You helped me to grow as a researcher, and you fundamentally changed the way I think about science. To all past and present members of the Tan laboratory, I wish you the best in all your future endeavors.

1. Introduction

1.1. Chromatin

Each human cell includes a copy of the genome, which contains all the genetic information necessary to encode and regulate the structures and processes of the human body. The approximately 3.2 billion base pairs (bp) of DNA in the human genome span over 2 meters in the uncondensed form, but they are stored in a nucleus with a diameter of only about 10 μm (McGinty and Tan, 2015). Rather than condensing into an unorganized bundle, eukaryotic nuclear DNA is packaged around a series of core protein complexes to form chromatin. This highly compact, organized system of genetic storage allows for the DNA template to be accessed for transcription when necessary. Chromatin structure is necessarily dynamic in order to facilitate the accessibility of certain genes according to cell needs. A variety of enzymes interact with the chromatin interface to regulate gene expression and further modulate chromatin structure via post-translational modifications of the core proteins (Luger et al, 1997). Chromatin enzymes also play a major regulatory role in DNA repair and cell cycle progression (Polo et al, 2010).

1.2. The Nucleosome Core Particle

Chromatin is composed of repeating functional units of the nucleosome core particle connected by 20-80 bp of linker DNA and stabilized by histone H1. The nucleosome core particle consists of 147 bp of DNA wrapped 1.65 times in a left-handed super-helix around a

histone octamer, which contains two copies each of histones H2A, H2B, H3, and H4 (Luger et al, 1997) (Figure 1).

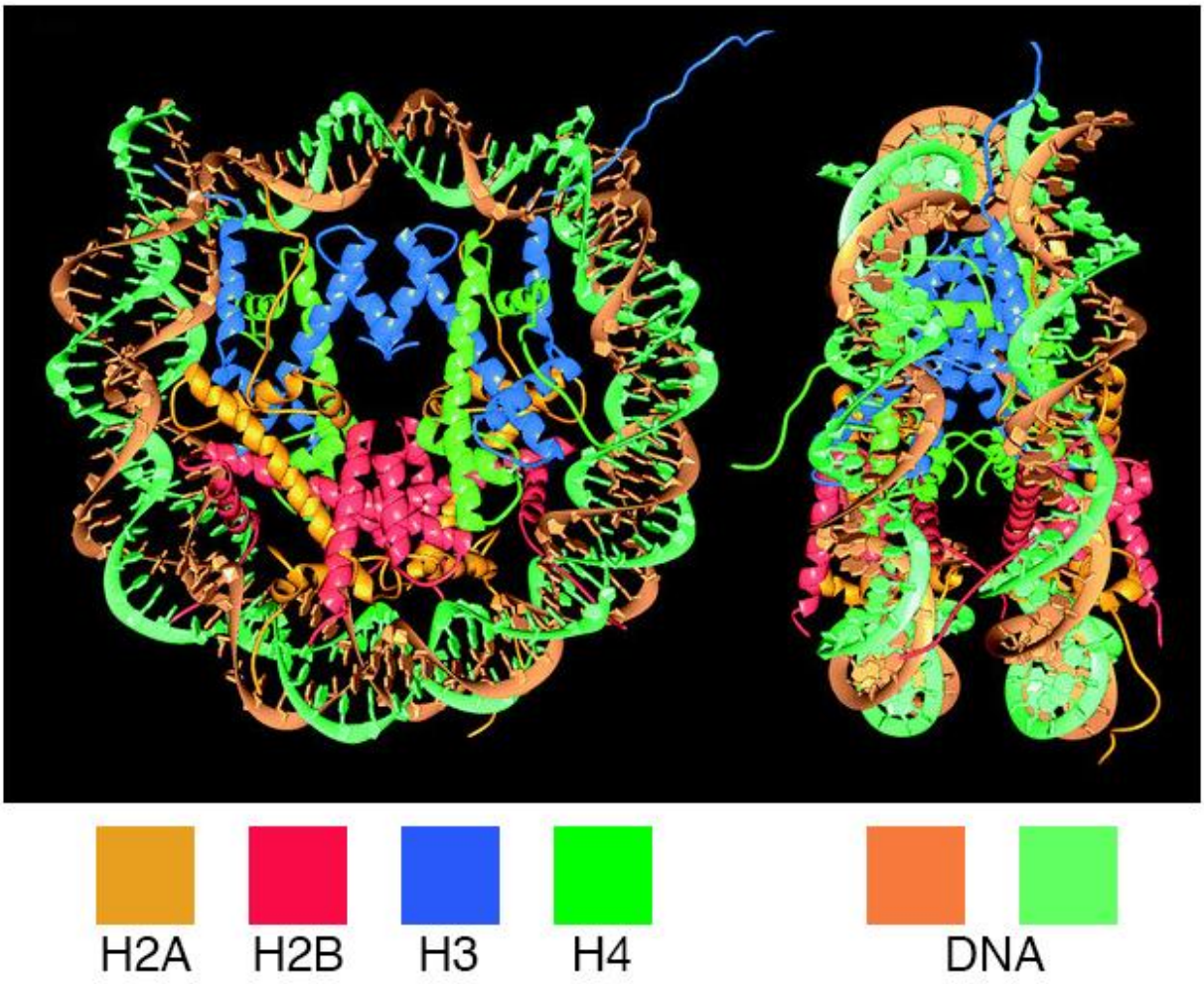

Figure 1. The Nucleosome Core Particle
Luger et al, 1997. X-ray crystal structure of the canonical nucleosome at 2.8 Å.

Two H2A-H2B heterodimers and one $(H3\text{-}H4)_2$ tetramer complex to form the histone octamer. The H3-H4 pairs interact via a four-helix bundle in the H3 histone fold domains. The tetramer binds to the H2A-H2B dimer pairs through another four-helix bundle between the H2B and H4 histone folds. The octamer is further stabilized by various α-helices in the histone folds

of the canonical histones (Luger et al, 1997). This histone fold domain is conserved among histone proteins and is marked by the helix-loop-helix-loop-helix (α1-L1-α2-L2-α3) secondary structure. Of the 147 bp of DNA wrapped around the histone octamer, 121 bp are directly bound by the histone fold domains. Each histone fold pair associates with 27-28 bp of DNA with 4 bp of DNA in between. The histone-DNA interactions occur via electrostatic interactions with the phosphodiester backbone exterior of the DNA strand (Kornberg and Lorch, 1999). Thus, canonical histone octamers generally have little DNA sequence specificity because the histone fold domains do not usually associate with the nitrogenous bases on the interior of the DNA molecule.

While they are not required for octamer core stabilization, the histone N-terminal tails play an essential role in transcription activation. Post-translational modifications, especially acetylation, have been shown to modulate chromatin structure to allow transcription machinery access to DNA (Zhang and Reinberg, 2001).

1.3. The Centromeric Nucleosome

Each chromosome contains a centromere, the locus to which the kinetochore binds and facilitates interactions with the mitotic spindle microtubules during metaphase. Proper assembly of the kinetochore on the centromere is required for the accurate segregation of sister chromatids and proper division of genetic material into the resulting daughter cells following mitosis. In contrast to the canonical nucleosome, the centromeric nucleosome contains an H3 variant known as CenH3. Prior to or during metaphase, CenH3 replaces one or both H3 histones in order to prepare for kinetochore recruitment (Dechassa et al, 2011).

CenH3 histones contain a 90-amino acid C-terminal domain that shares 60% of its sequence identity with canonical H3. While H3 is highly conserved, CenH3 shares only 45-50% sequence identity among all species (Dechassa et al, 2011). Structurally, CenH3-H4 tetramers have been shown to be more rigid than the canonical H3-H4 tetramer, though few relationships between the structure and function of the centromeric nucleosome have been elucidated (Black et al, 2004).

1.4. The Cse4 Nucleosome

Unlike the multi-nucleosome centromeres of many eukaryotes, the *Saccharomyces cerevisiae* (budding yeast) centromere contains only one nucleosome, constituting a simple system for the study of centromeric structure and mechanisms (Zhou et al, 2011). The CenH3 variant present in budding yeast centromeres is called Cse4. Three main structures have been proposed for the yeast centromeric nucleosome, all of which have been shown to be stable during part or all of the cell cycle.

In the homotypic octamer model, two of each of the canonical histones H2A, H2B, and H4 combine with two copies of Cse4 to constitute the yeast centromeric nucleosome (Figure 2). The term homotypic is employed here to differentiate this model from the heterotypic octamer model containing 2 copies each of H2A, H2B, and H4, along with one copy each of canonical H3 and Cse4. The heterotypic octamer model has been largely refuted. Centromeric DNA wraps the homotypic octamer in a left-handed super-helix (Dechassa et al, 2011). Despite the presence of conserved centromeric DNA elements, this octameric model has not been shown to bind preferentially to yeast centromeric DNA. Still, these DNA elements suggest that sequence specificity may be important for centromere function *in vivo* (Cleveland et al, 2003).

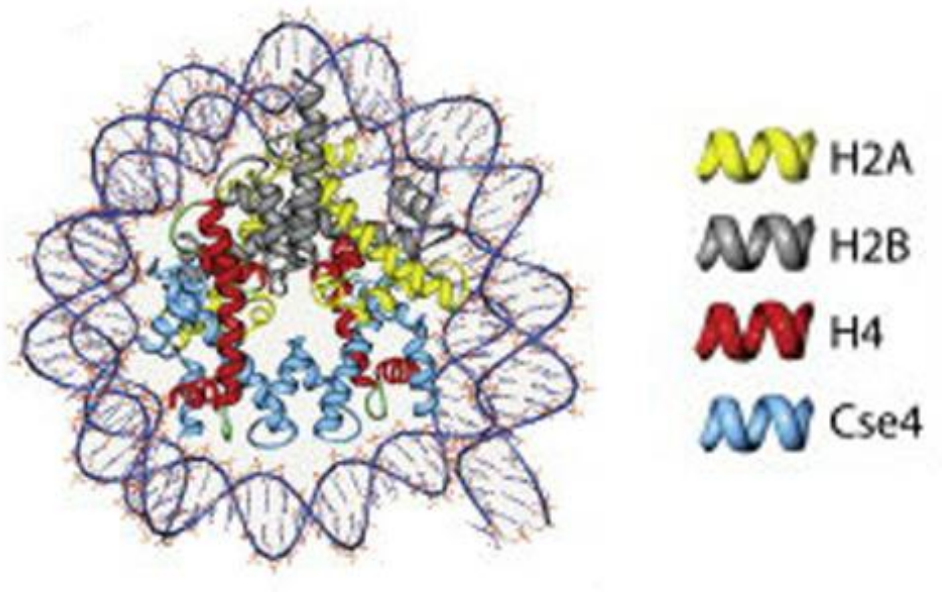

Figure 2. The Homotypic Octamer Model of the Cse4 Nucleosome
Camahort et al, 2009. Model of the Cse4 nucleosome containing two copies each of H2A, H2B, H4, and Cse4.

The hemisome model is a tetramer containing one each of the H2A, H2B, Cse4, and H4 histones (Figure 3). After previously being stabilized in *Drosophila melanogaster* via crosslinking and immunoprecipitation, the hemisome model was observed *in vivo* in budding yeast (Dalal et al, 2007). The yeast centromeric nucleosome has been shown to oscillate between the homotypic octamer and hemisome structures during different phases of mitosis. Using fluorescence correlation spectroscopy and calibrated imaging to measure the number of Cse4 histones present at the centromere throughout the cell cycle, it was shown that Cse4 abundance remains at 16, the number of chromosomes in budding yeast, during all phases but anaphase.

During anaphase, fluorescence doubled, suggesting that 32 Cse4 histones are present in the nucleus. These data support the proposition that the centromeric nucleosome exists in the hemisome form until anaphase, at which point another hemisome is recruited to form the homotypic octamer. During early anaphase, the fluorescence value required a few minutes to double, suggesting that structural modification of the nucleosome takes place largely during, rather than prior to anaphase (Manjunatha et al, 2012).

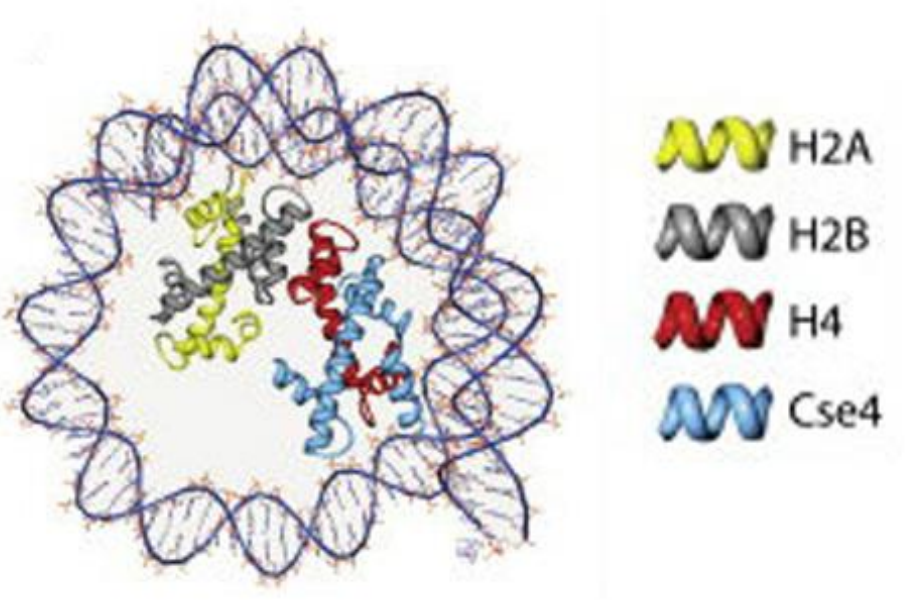

Figure 3. The Hemisome Model of the Cse4 Nucleosome
Camahort et al, 2009. Model of Cse4 nucleosome containing one copy each of H2A, H2B, H4, and Cse4.

Scm3, a protein active during the progression from G2 to M in the yeast cell cycle, has been shown to recruit Cse4 and play a role in the assembly of the Cse4 nucleosome prior to cell division (Dechassa et al, 2011). Scm3 brings together two Cse4-H4 dimers to form the (Cse4-

$H4)_2$ tetramer via a hexameric complex of two Scm3 proteins bound to the tetramer. It is thought that the Scm3 proteins deliver the tetramer to the nucleosome to combine with the pre-existing H2A-H2B dimers, although the "hexasome" model alone has also been proposed as a relevant *in vivo* structure of the centromeric nucleosome (Dechassa et al, 2014). It has been shown via chromatin immunosuppression studies that Scm3 replaces the H2A-H2B dimer *in vitro* and binds to the $(Cse4\text{-}H4)_2$ tetramer to form a stable hexasome (Mizuguchi et al, 2007) (Figure 4).

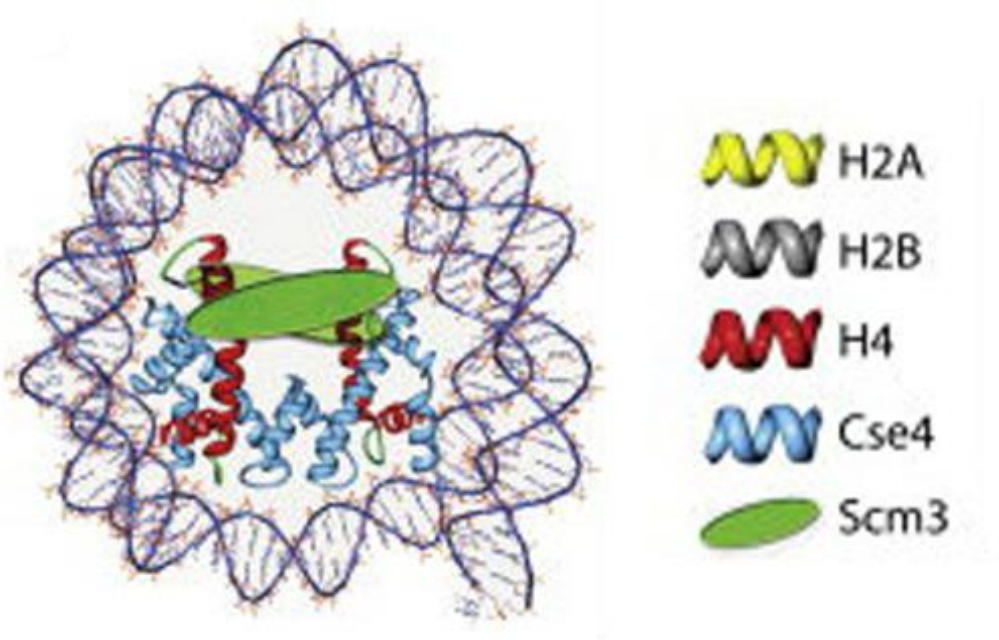

Figure 4. The Hexasome Model of the Cse4 Nucleosome
Camahort et al, 2009. Model of Cse4 nucleosome containing one copy each of H2A, H2B, H4, and Cse4.

1.5. Yeast Centromeric DNA

In most eukaryotes, including humans, the centromere is defined by epigenetic markers rather than a centromere-specific DNA sequence. While the exact mechanisms remain ambiguous, it is assumed that these epigenetic markers direct the formation of the centromeric nucleosome and facilitate proper formation of the kinetochore. In budding yeast, however, the centromere locus is defined by specific DNA sequence elements (Cleveland et al, 2003).

Budding yeast centromeres span 125 bp of DNA and contain 3 conserved DNA elements known as CDEI, CDEII, and CDEIII (Figure 5). Both CDEI and CDEIII exhibit the exact same sequence across each of the 16 yeast centromeres. The sequence of CDEII is not conserved, though it is 76-88 bp in length and exhibits about 90% A+T content across all chromosomes (Cleveland et al, 2003).

Figure 5. The Conserved DNA Elements in Budding Yeast Centromeric DNA
Cleveland et al, 2003. Approximate lengths and relative positions of CDE1, CDEII, and CDEIII.

Each of the DNA elements has been shown to associate with components of the centromeric nucleosome or the kinetochore, but only CDEI and CDEII have been shown to be essential to the proper function of the centromere. *In vivo* studies including mutations in each of the three conserved elements showed that only CDEI and CDEII mutations increased chromosomal loss (Keith and Fitzgerald-Hayes, 2000). Additionally, a CDEII insertion mutation

caused a 17-fold increase in mis-segregation of sister chromatids (Stoler et al, 1995). CDEIII, though perhaps nonessential, has been shown to initiate kinetochore assembly by binding to a tetrameric kinetochore protein complex known as CBF3 (Meraldi et al, 2006). During the process of kinetochore assembly, CDEI binds a basic helix-loop-helix zipper protein called Cbf1. Cbf1 and CBF3 associate and are thought to provide a surface to which the remaining kinetochore proteins can bind (Hemerich et al, 2000).

1.6. The pST50Tr Expression Plasmid

All cloning procedures utilized pST50Tr expression plasmids for monocistronic expression. Each pST50Tr plasmid includes a pMB1 ColE1 origin of replication, which yields ampicillin resistance. This allows for selection of cells containing the relevant expression plasmid. In addition to a promoter, transcription terminator, and translational start signal, the plasmids include affinity tags on the N-terminal and C-terminal. The Shine-Dalgarno sequence constitutes a portion of the translational start signal that facilitates ribosomal recognition and binding. The N-terminal tag contains NdeI and BamHI sites, while the C-terminal tag contains NgoMIV and BsrGI sites (Figure 6).

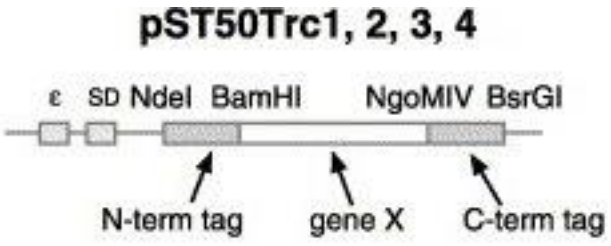

Figure 6. Modular Design of pST50Trc Vectors
Tan et al, 2005. The affinity tags are shown as well as a translational enhancer (ε) and the Shine-Dalgarno sequence (SD).

The NdeI site, marked by the CATATG sequence, includes the ATG start codon, which facilitates transcription. The BamHI site, GGATCC, follows a tobacco etch virus (TEV) NIa protease site. The TEV NIa protease prefers glycine following the site, which the GGA codon provides. The NgoMIV site, GCCGGC, encodes alanine and glycine, two small amino acids that are unlikely to affect protein function. Finally, the BsrGI site follows the TAA stop codon at the C-terminus and co-digests well with the other restriction sites. Following cloning procedures, pST50Tr plasmids were expressed in BL21(DE3)pLysS or CodonPlus(DE3) *E. coli* cells, which provide the T7 RNA polymerase under lac operator control. Protein expression can be induced via the addition of appropriate amounts of Isopropyl ß-D-1-thiogalactopyranoside (Tan et al, 2005).

1.7. Experimental Outline

The purpose of this work was to create the homotypic octameric form of the yeast centromeric nucleosome core particle. Troubleshooting of expression and purification methods has provided the Tan laboratory with parameters for future nucleosome preparations. The nucleosome core particle can be employed in future binding studies with various kinetochore proteins in an effort to further elucidate the structure and mechanism of the yeast centromere-kinetochore complex.

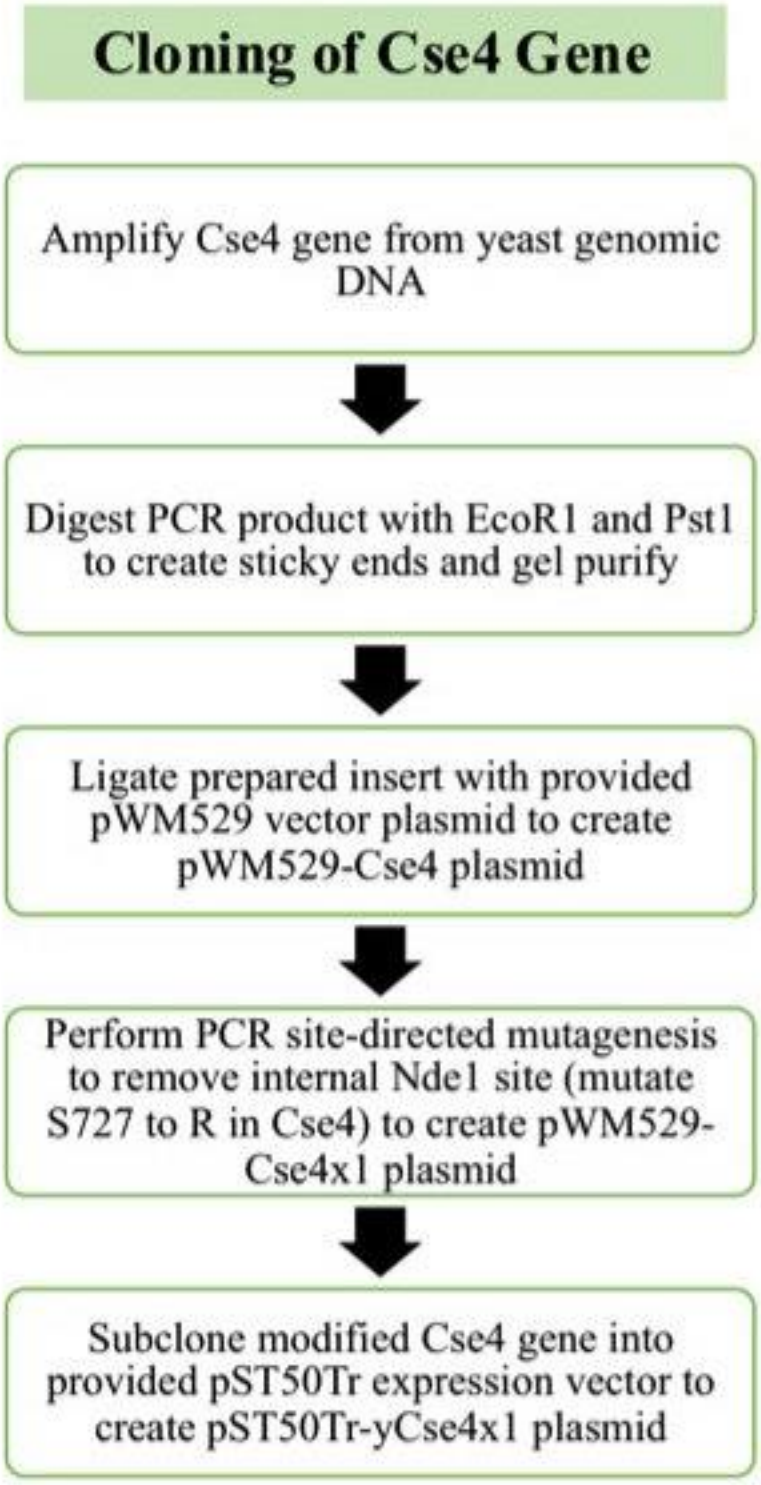

Figure 7. Schematic of Cloning Procedures
Cse4 amplification, mutagenesis, and subcloning procedures.

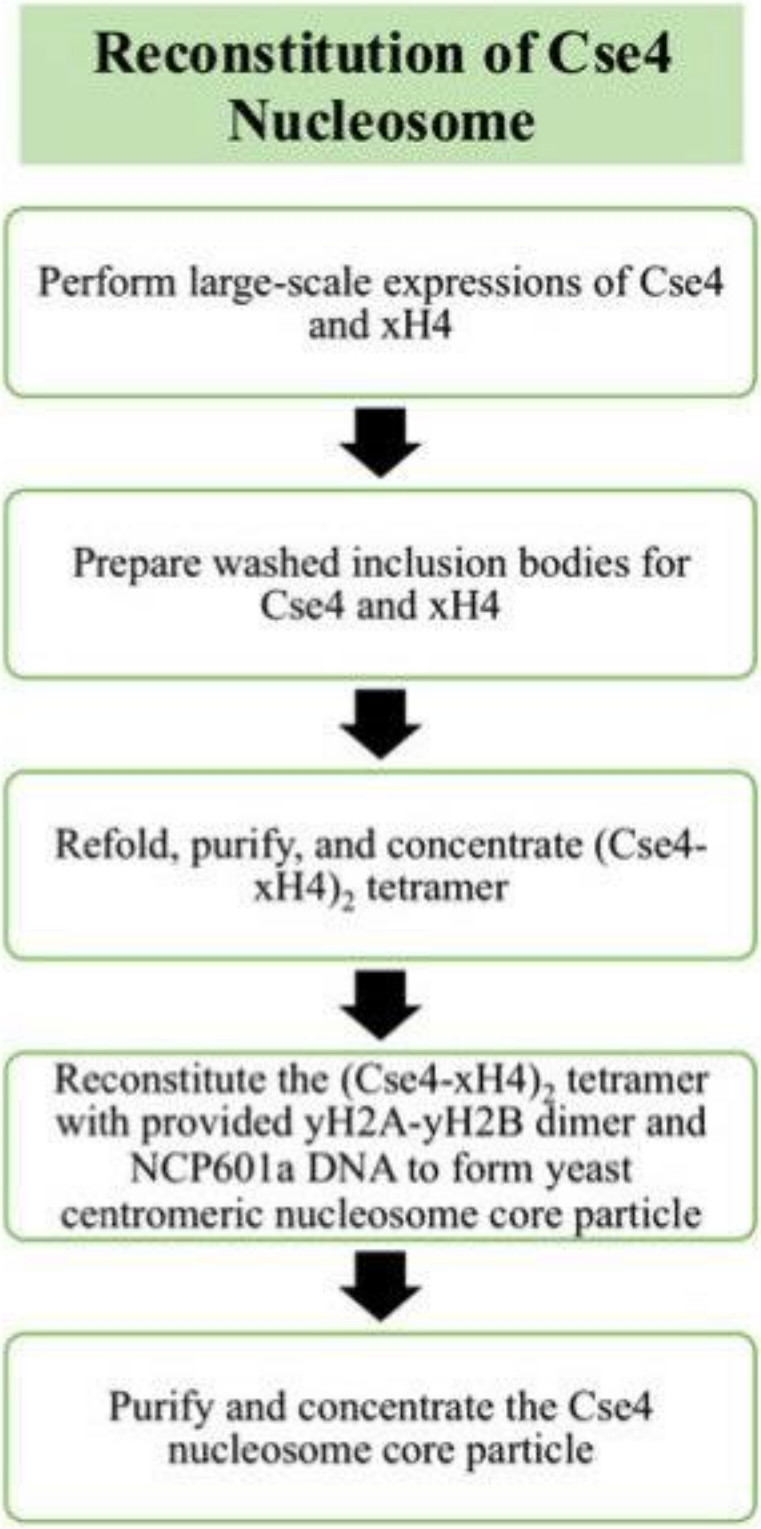

Figure 8. Schematic of Expression and Reconstitution Procedures
Protein expression, tetramer refolding, purification, and reconstitution procedures.

2. Materials and Methods

2.1. Bacteriological Methods

2.1.1. Bacteriological Strains

All cloning experiments employed the use of TG1 cells, a strain of *Escherichia coli*. BL21(DE3)pLysS cells were used for the expression of xH4, and CodonPlus(DE3) cells were used for Cse4 expression after only low levels of expression were attained with BL21(DE3)pLysS cells. Detailed information about each strain can be found in Table 1.

Table 1. *E. coli* strains for cloning and protein expression

Strain	Genotype	Extra tRNA genes
TG1	supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5 (rK - mK -) [F′ traD36 proAB lacIq ZΔM15]	None
BL21(DE3)pLysS	B F- dcm ompT hsdS(rB - mB -) gal λ(DE3) [pLysS Camr]	None
CodonPlus(DE3)	B F- ompT hsdS(rB - mB -) dcm+ Tetr gal λ endA Hte [argU ileY leuW Camr]	argU, ileY, leuW

2.1.2. Bacteriological Media

Petri dishes for all bacterial transformations contained TYE media. It contained 1.0% bacto tryptone, 0.5% yeast extract, 0.8% NaCl, and 1.5% agar. Plates used for transformation prior to plasmid preparation also included 100 µg/ml ampicillin, while those used for transformation prior to protein expression included both 100 µg/ml ampicillin and 25 µg/ml

chloramphenicol. Media was autoclaved, poured into petri dishes, and cooled at room temperature overnight. Petri dishes were labeled and stored at 4ºC prior to use.

Liquid 2xTY media was used for all bacterial transformations, plasmid preparations, and protein expressions. It contained 1.6% bacto tryptone, 1.0% yeast extract, and 0.5% NaCl. All 2xTY media was autoclaved and stored at room temperature prior to use. For plasmid preparations, 100 µg/ml ampicillin was added to the media. Both 50 µg/ml ampicillin and 25 µg/ml chloramphenicol were added to the media for protein expression.

2.2. DNA Methods

2.2.1. Agarose Gel Electrophoresis of DNA

Electrophoresis of DNA through agarose allows samples containing one or multiple DNA strands of different lengths to be spatially separated. It was used to verify correct linear amplification, ligation, transformation, and plasmid preparation during cloning experiments as well as to prepare gel-purified DNA.

Agarose gels were prepared by combining the appropriate mass of high gelling temperature (HGT) agarose with 30 ml 0.5xTBE (45 mM Tris base, 45 mM boric acid, 1 mM EDTA pH 8.0). The gel was left to sit at room temperature for 5 minutes prior to heating in a microwave for 90 seconds. 1.5 µl of 10 mg/ml ethidium bromide was mixed into the solution. The solution was cooled at room temperature until warm to the touch and poured into an agarose gel casting block. For analytical gels, a 15-well comb was installed near the top of the gel. A 10-well comb was used for preparative gels. The gel casting apparatus was left to cool for 30 minutes. Then, the comb was removed, and the gel was placed into an agarose gel

electrophoresis box. A sufficient amount of 0.5xTBE was added to the box to cover the top of the gel. Appropriate amounts of 6x gel loading buffer (GLB) (30% v/v glycerol, 0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol) were added to DNA solutions to a final concentration of 1x GLB. Samples, as well as an appropriate DNA ladder, were loaded into wells using a pipetman. The gel was run at 125 V until the bands containing bromophenol blue had reached the end of the gel. Excess 0.5xTBE was poured off of the gel, which was then viewed under UV transillumination.

2.2.2. Preparation of Insert DNA

The Cse4 gene was amplified from yeast genomic DNA via PCR amplification. The 100 µl PCR mixture contained 53.5 µl of Milli-Q® water, 20 µl of 5x Q5 reaction buffer, 10 µl of 2.5 mM dNTP, 6 µl of 100 ng/µl yeast genomic DNA template, 5 µl of 10 µM forward primer, 5 µl of 10 µM reverse primer, and 0.5 µl of µl 2 units/µl Q5 polymerase. The primers used incorporated EcoRI and PstI fragments at the ends of the amplified DNA strand to allow for sticky-ended ligation (Section 2.2.4.).

The PCR tube containing the PCR mixture was placed into a PCR thermal cycler. The mixture was heated to 98ºC for 30 seconds to denature the DNA template. Next, the following denaturation-annealing-elongation cycle was repeated five times: 98ºC for 5 seconds, 49ºC for 30 seconds, 72ºC for 20 seconds. The previous cycle was then repeated 25 more times with an annealing temperature of 60ºC instead of 49ºC. Lastly, there was a final elongation step for 2 minutes at 72ºC.

The PCR product was run on a 1% analytical agarose gel, and the amplification of the Cse4 gene was verified. One phenol/chloroform extraction and one chloroform extraction were

performed (Section 2.2.13.), followed by ethanol precipitation (Section 2.2.14.). The pelleted DNA was resuspended in 30 µL of TE (10, 0.1) (10 mM Tris pH 8.0, 0.1 mM EDTA). The DNA solution, containing the Cse4 gene with EcoRI and PstI sites on the ends, was then digested with EcoRI and PstI to create sticky ends. The digest mixture included 3 µl of the DNA solution, 18 µl of Milli-Q® water, 3 µl of 10x NEBuffer EcoRI, 3 µl of 1 mg/ml BSA, 1 µl of 100 mM DTT, 1 µl of 20 units/µl EcoRI, and 1 µl of 20 units/µl PstI. The digest occurred at 37ºC for 2.5 hours.

2.2.3. Isolation of DNA from Agarose by Centrifugation

The digested DNA, containing the Cse4 gene with sticky ends, was run on a 1% preparative agarose gel. A filter assembly was made by poking a heated 25-gauge needle through the bottom of a 0.5 ml Eppendorf tube. A small piece of siliconized glass wool was added to cover the bottom of the tube, and the 0.5 ml Eppendorf tube was placed inside a 1.5 ml Eppendorf tube. The gel was then placed on a UV transilluminator, and the band corresponding to the desired fragment was cut out of the gel with a razor blade. The gel slice was placed into the filter assembly and centrifuged in a microcentrifuge at 7000 rpm for 5 min. The gel-purified DNA was collected in the 1.5 ml Eppendorf tube.

2.2.4. Ligation of sticky-ended DNA

The gel-purified insert DNA was ligated to a gel-purified pWM529 cloning vector plasmid prepared by Song Tan. Both the insert and vector were digested with EcoRI and PstI to create complementary sticky ends. The 10 µl ligation mixture contained 4 µl of Milli-Q® water, 1 µl of 10x T4 DNA ligase buffer, 0.5 µl of 100 mM DTT, 2.0 µl of the gel-purified vector, 1.5

µl of the gel-purified insert, and 1 µl of 40 units/µl T4 DNA ligase. The mixture was incubated at room temperature for two hours.

2.2.5. Transformation

The pWM529-Cse4 plasmid was transformed into competent TG1 cells. A 1.5 ml Eppendorf tube containing 100 µl of cells was thawed on ice. 2 µl of ligation mix were added to the cells, and the mixture was left on ice for 15 minutes. The cells were then heat shocked at 42ºC for 30 seconds and consequently returned to ice for 10 seconds. 0.5 ml of 2xTY was added to the tube, which was then placed in a shaking incubator at 37ºC for 15 minutes. Following incubation, 0.3 ml of cells was spread with a sterile inoculating loop onto a plate containing TYE and 100 µg/ml ampicillin. The plate was incubated at 37ºC for 10-18 hours.

2.2.6. PCR Screening

In order to ensure that selected colonies contained the desired plasmid, a PCR reaction was run with primers that would produce fragments of different sizes for the desired and parent (vector only) plasmids. A sterile inoculating loop was used to transfer single colonies to Eppendorf tubes containing 100 µl of Milli-Q® water. The loop was then re-streaked onto a new plate containing TYE and 100 µg/ml ampicillin. The restreaked plate was incubated at 37ºC for 10-18 hours and then stored at -8ºC. The cell suspensions contained in the Eppendorf tubes were vortexed for 5 seconds. A 133 µl stock PCR reaction mixture was prepared, which included 90.3 µl of Milli-Q® water, 14 µl of 10x Thermo Pol buffer, 14 µl of 2.5 mM dNTP, 7 µl of the

forward primer, 7 µl of the reverse primer, and 0.7 µl of 2 units/µl Pfu polymerase. 19 µl of the stock mixture were combined with 1 µl of cell suspension in a 0.5 ml PCR tube.

The PCR mixture was placed in a PCR thermocycler, and the following thermocycle was applied. The mixture was heated to 95ºC for 2 minutes, followed by 25 cycles of 95ºC for 30 seconds, 50ºC for 30 seconds, and 75ºC for 45 seconds. 6x GLB was added the PCR products, which were run on an analytical agarose gel. Two clones that exhibited bands corresponding to the desired plasmid were selected for plasmid preparation.

2.2.7. 100 ml Plasmid Preparation

A 100 ml Alkaline Lysis Plasmid Preparation procedure written by Song Tan was used to purify the desired plasmid. One colony each from the re-streak plate of the two selected clones from PCR screening was inoculated in a separate 500 ml Erlenmeyer flask containing 100 ml of 2xTY and 100 µl of 100 µg/ml ampicillin. The flasks were placed in a shaking incubator at 37ºC for 10-18 hours. Following incubation, the cells were spun at 4000 rpm for 5 minutes in separate 250 ml centrifuge bottles in a tabletop centrifuge. A pellet had formed, and the supernatant was removed. Using a 10 ml glass pipette, the cell pellet was resuspended in 5 ml of cold LYSIS buffer (50 mM glucose, 25 mM Tris-Cl pH 8.0, 10 mM EDTA Na). The resuspended cells were then poured into 50 ml polypropylene Falcon tubes. 10 ml of NaOH/SDS (0.2 M NaOH, 1% SDS) were added to the cells, and the tubes were shaken until the mixture became clear. The tubes were then placed on ice for 3 minutes. This step lysed the cells, causing them to release the desired plasmid and all other cell contents into the solution. Next, 10 ml of cold 5 M KAc / 2.5 M HAc was added to each Falcon tube to precipitate chromosomal DNA, along with some protein and RNA.

The Falcon tubes were spun at 4000 rpm for 3 minutes at room temperature in a tabletop centrifuge. The clear supernatant was carefully poured into 50 ml round-bottomed polypropylene centrifuge tubes, and 12.5 ml of isopropanol was added to precipitate the remaining nucleic acid, including the desired plasmid. The tubes were then capped, shaken, and incubated at room temperature for 5 minutes prior to being placed in a Sorvall SS-34 rotor for centrifugation. The round-bottomed tubes were spun at 13K rpm for 5 minutes at room temperature. Following centrifugation, the supernatant was poured off, and 0.5 ml of 70% ethanol was added to each pellet to remove it from the wall of the tube. Each pellet was then poured into a 1.5 ml Eppendorf tube, which was spun in a microcentrifuge for 1 minute at room temperature. The supernatant was aspirated off, and 0.15 ml of TE (10, 50) (10 mM Tris pH 8.0, 50 mM EDTA) was added to each tube. A Pipetman with a wide-ended tip was used to resuspend the pellet in the TE (10, 50). In order to degrade the remaining RNA, 1.5 µl of 10 mg/ml RNase A was added to the resuspended pellet, which was then incubated in a 37ºC water bath for 15 minutes. The tubes were vortexed periodically to break up any remaining precipitant.

Two Sephacryl S400 spun columns were prepared, one for each plasmid preparation. Siliconized glass wool was forced to the bottom of a Gilson blue pipette tip, which was placed into the top half of a 1.5 ml Eppendorf tube. This apparatus was inserted into a 5 ml polypropylene tube. The pipette tip was filled to the top with Sephacryl S400 HR resin (equilibrated in TE (10, 0.1)). The two entire columns were then centrifuged at 2000 rpm for 3 minutes at 20ºC. The liquid collected in the 5 ml polypropylene tubes was discarded.

After the RNase A digestion was complete, the digested solution underwent two phenol/chloroform extractions and one chloroform extraction. The extracted samples were pipetted onto the Sephacryl S400 HR spun columns and spun at 2000 rpm for 3 minutes at room

temperature in a tabletop centrifuge. The liquid collected in the 5 ml polypropylene tubes was transferred to 1.5 ml Eppendorf tubes.

2.2.8. UV Quantitation of DNA

The purified plasmid samples were quantitated using a ND-1000 NanoDrop Spectrophotometer. 2µl of Milli-Q® water was pipetted onto the sensor to clean it. Next, 2 µl of TE (10, 0.1) was added to the sensor and measured to provide a blank standard for measuring absorbance. The sensor was wiped clean, 2 µl of sample was added, and the absorbance was measured in the range of 220 to 350 nm. An adjusted A_{260} value was determined by subtracting the absorbance at 320 nm from the absorbance at 260 nm. The concentration was determined with Beer's Law: $A = \varepsilon cl$, where $\varepsilon = 20\ (mg/ml)^{-1}\ cm^{-1}$, $l = 1$ cm, and A = the adjusted A_{260} value. The equation was solved for "c" to find the concentration of DNA in the sample. Most purified plasmid samples were between 150 and 200 ng/µl.

2.2.9. Restriction Mapping of Purified Plasmids

Purified plasmids were digested with restriction enzymes in order to ensure that the correct insert DNA fragment had been inserted into the correct vector plasmid. One or two enzymes were selected that contained recognition sites on both the insert and vector sequences. The enzyme cutting within the insert verified that the correct insert was present in the plasmid, while the enzyme cutting within the vector ensured the presence of the correct vector. Two separate 10 µl enzyme digests were done for each plasmid sample to account for any ambiguous fragments. The digest pools contained 1 µl of the appropriate 10x NEBuffer, 1 µl of 1mg/ml

BSA, 0.5 µl of 100 mM DTT, 1 µl of the plasmid DNA sample, 0.5 µl of each enzyme (5 or 10 units/µl), and Milli-Q® water for the remaining volume (to 10 µl). Digests were run at 37ºC for one hour. Next, digested samples were run on an agarose gel to verify correct band lengths.

2.2.10. Sequencing of DNA samples

In order to further verify the presence of the desired construct, one of the purified plasmids was sequenced at the Genomics Core Facility in Chandlee Laboratory at the Pennsylvania State University. 5 µl of 0.2 µg/µl DNA sample were submitted for sequencing with 5 µl of 1 µg/µl primer. The Genomics Core Facility provided the T7, STO429, and M13 reverse primers used for this project. The individual base-pairs of the generated sequence were aligned with the theoretical sequence of the pWM529-Cse4 plasmid to verify the presence of the desired construct.

2.2.11. PCR Site-Directed Mutagenesis and DpnI digestion

The pWM529-Cse4 plasmid contained an internal NdeI recognition site, so a PCR site-directed mutagenesis was done to introduce a single bp mutation at that locus (S727 → R in Cse4 gene). Because the expression system employed in this project utilized a pST50Tr vector with NdeI and BsrGI sites at the ends, the internal NdeI site in the Cse4 gene had to be removed so the insert would not be cut during the subsequent subcloning procedures.

A 25 µl PCR mixture was prepared, which included 15.5 µl of Milli-Q® water, 5 µl of 5x Q5 reaction buffer, 2 µl of 2.5 mM dNTP, 1 µl of 10 ng/µl pWM529-Cse4, 0.5 µl of the forward primer, 0.5 µl of the reverse primer, and 0.5 µl of 2 units/µl Q5 polymerase. The primers used

were partially-overlapping in order to decrease the chance of primer dimerization and increase the probability of successful mutagenesis. The PCR reaction included the following program: 98ºC for 30 seconds, 24 cycles of 98ºC for 15 seconds, 60ºC for 15 seconds, 72ºC for 90 seconds, and a final elongation step at 72ºC for 5 minutes.

The PCR product was digested with 0.5 µl of 10 units/µl DpnI at 37ºC for 1.5 hours. DpnI digested the methylated, wild-type template DNA, leaving only the non-methylated amplified PCR product intact in solution. The digested product was then transformed into TG-1 cells (Section 2.2.5) and PCR screened (Section 2.2.6.). Selected plasmids were purified (Section 2.2.7.), restriction mapped (Section 2.2.9.), and sequenced (Section 2.2.10.) to create and verify the pWM529-Cse4x1 (S727 → R) construct.

2.2.12. Subcloning into Expression Vector

The mutagenized Cse4 gene was subcloned into a pST50Tr expression vector to facilitate subsequent protein expression steps. The insert was prepared by digesting the pWM529-Cse4x1 (S727 → R) plasmid with NdeI and BsrGI to create the sticky-ended Cse4x1 gene. The digest mixture included 16µl of Milli-Q® water, 3 µl of 10x NEBuffer 2, 3 µl of 1 mg/ml BSA, 1 µl of 100 mM DTT, 5 µl of 200 ng/µl of pWM529-Cse4x1, 1 µl of 20 units/µl NdeI, and 1 µl of 10 units/µl BsrGI. The digest was done at 37ºC for 2.5 hours. The digested product was then agarose gel purified (Section 2.2.3.). The gel purified insert was ligated to a pST50Tr expression vector with complementary NdeI and BsrGI sticky ends (prepared by Dr. Song Tan). The ligation mixture included 4 µl of Milli-Q® water, 1 µl of 10x T4 DNA ligase buffer, 0.5 µl of 100 mM DTT, 2 µl of the pST50Tr vector, 1.5 µl of the yCse4x1 insert, and 1 µl of 750 units/µl T4 DNA ligase. The ligation was run at room temperature for 2 hours. The product was then

transformed into TG-1 cells (Section 2.2.5) and PCR screened (Section 2.2.6.). Selected plasmids were purified (Section 2.2.7.), restriction mapped (Section 2.2.9.), and sequenced (Section 2.2.10.) to create and verify the pST50Tr-Cse4x1 (S727 → R) construct.

2.2.13. Phenol / Chloroform and Pure Chloroform Extractions

Phenol / Chloroform and pure chloroform extractions were used in the Preparation of Insert DNA (Section 2.2.2.) and 100 ml Plasmid Preparation (Section 2.2.7.) steps of this procedure in order to remove any protein contaminants from DNA solutions. A phenol / chloroform extraction employed a 1:1 mixture (by volume) of TE-equilibrated phenol and pure chloroform, while a pure chloroform extraction employed pure chloroform alone. The appropriate volume of the mixture was added to a sample and vortexed for 15 seconds. The mixture was then spun in a microcentrifuge at 13.3K rpm for 1 minute. The aqueous phase, which contained the extracted nucleic acids, was pipetted into a new Eppendorf tube.

2.2.14. Ethanol Precipitation of DNA

Ethanol Precipitation of DNA was used in the Preparation of Insert DNA (Section 2.2.2.) step of this procedure, as well as instances when a DNA solution needed to be concentrated or pelleted and resuspended in a new buffer. 0.1 volumes of 3 M NaAc pH 5.2 and 2.5 volumes 100% ethanol were added to the sample. The sample was vortexed for 5 seconds, left at room temperature for 10 minutes, and then spun in a microcentrifuge at 13.3K rpm for 10 minutes. The supernatant was removed, and the pellet was dried for 3 minutes. Finally, the pellet was resuspended in the desired volume of the appropriate buffer.

2.3. Protein Methods

2.3.1. SDS-PAGE Electrophoresis

2.3.1.1. SDS-PAGE Gel Preparation

5-10 pairs of long/short glass plates with spacers were loaded into a Mini-Protean II gel-pouring block (BioRad). 60 ml of 18% acrylamide separating gel mix was prepared by combining 8 ml of deionized water, 36 ml of 30%/0.5% acrylamide / bis-acrylamide, 120 µl of bromophenacyl bromide (BPB) in ethanol, 15 ml of 3 M Tris-Cl pH 8.8, 600 µl of 10% sodium dodecyl sulfate (SDS), 60 µl of tetramethylethylenediamine (TEMED), and 240 µl of 25% AMPS. A 60 ml syringe was used to inject the separating gel through the inlet port at the bottom of the gel-pouring apparatus to about 3 cm from the top of the glass plates. Water-saturated 1-butanol was pipetted onto the top of the separating gels, and the gels were left to sit until they polymerized. Residual butanol was poured off of the apparatus. Next, 20 ml of stacking gel mix was prepared by first combining 5 ml of deionized water, 10 ml of 10%/0.5% acrylamide / bis-acrylamide, and 4.8 ml of 0.5M Bis-Tris. This mixture was deaerated, and then added to 0.2 ml of 10% SDS, 15 µl of TEMED, and 80 µl of 25% AMPS. The stacking gel was added to the top of the gels. Analytical (15-well) and preparatory (10-well) combs were then inserted into the top of the gels, and the apparatus was left to polymerize. Finally, the gels were removed from the block, wrapped in wet paper towels, and stored at -8ºC prior to use.

2.3.1.2. Running an SDS-PAGE gel

A prepared SDS-PAGE gel was placed into a clamp assembly (BioRad) with the long glass plate facing the acrylic plate. The clamp assembly was then secured in the inner cooling core with the long glass plate facing outward. Spacers were used between the long glass plate and the plastic fasteners to prevent leakage. The inner cooling core was placed into the lower buffer chamber with the labeled positive and negative electrodes lined up. 1x protein gel running buffer (PGRB) (10 mM Tris base, 76 mM glycine, 0.02% SDS) was poured to the top of the upper buffer chamber (the matrix of the inner cooling core) and to the bottom of the gel in the lower buffer chamber.

Samples were mixed in a 1:1 ratio with protein gel loading buffer (PGLB) (125 mM Bis-Tris pH 6.8, 20% glycerol (v/v), 4% SDS (v/v), 15% 2-mercaptoethanol (v/v), and 0.04% bromophenol blue (w/v)). Samples were boiled for two minutes and then loaded into the wells of the gel using a Hamilton syringe. The gel was run at 10 W until the bromophenol blue dye ran off the bottom.

The gel was removed from the clamp assembly and freed from the glass plates with a razor blade. It was then placed into a gel box, covered in FIX solution (45% ethanol, 9% glacial acetic acid), and left to sit on a rocker for 5 minutes. The FIX solution was then poured off, and the gel was covered in STAIN solution (45% ethanol, 9% glacial acetic acid, 0.5% Coomassie Blue R (w/v)) and left to sit for 5 minutes. After the STAIN solution was poured off and the gel was rinsed with deionized water, the gel was covered in DESTAIN solution (7% ethanol, 5% glacial acetic acid). A folded paper towel was placed in the box with the gel, which was left to

soak in DESTAIN in a 60ºC rotating water bath until blue bands were visible on the gel. The gel was then soaked in deionized water until it was dried in cellophane for at least 24 hours.

2.3.2. 100 ml Small-Scale Protein Expression

The T7 expression system was used to express all proteins in the project (Studier, 1990). The pST50Tr-Cse4x1 (S727 → R) expression vector was transformed into CodonPlus(DE3) cells (Section 2.2.5.). Both a 37ºC and a 28ºC expression were done. Two 500 ml Erlenmeyer flasks containing 100 ml of 2xTY, 100 µl of 50 µg/ml ampicillin, and 100 µl of 25 µg/ml chloramphenicol were inoculated with five colonies each from the transformation plate. Both flasks were placed in a shaking incubator at 37ºC. The OD_{600} was checked periodically (Section 2.3.9.) until one flask reached 0.05-0.15. That flask was then moved to a 28ºC shaking incubator. The OD_{600} was monitored until each flask reached 0.5-0.9, at which point 100 µl of 0.2M isopropyl ß-D-1-thiogalactopyranoside (IPTG) was added to induce expression. 250 µl of uninduced sample was collected, centrifuged at 13.3K rpm for 1 minute, and resuspended in 50 µl of PGLB. 125 µl samples from each flask were taken, spun down, and resuspended in PGLB every hour for four hours.

Three hours after induction, 50 ml of the culture was poured into a 50 ml Falcon tube and centrifuged at 4000 rpm for 10 minutes at room temperature in a tabletop centrifuge. The supernatant was poured off, and the pellet was resuspended in 10 ml of T100 (20 mM Tris-Cl pH 8.0, 100 mM NaCl, 0.5 mM EDTA Na, 1 mM benzamidine, 10 mM 2-mercaptoethanol). The solution was stored at -20ºC. An SDS-PAGE gel was run with lanes containing the uninduced sample, samples at each 1-hour time point, and molecular weight markers. Expression was determined to be strongest at 37ºC.

2.3.3. Solubility Test for Overexpressed Protein

The 10 ml of frozen resuspended cells were thawed in lukewarm warm. The solution was then poured into a 50 ml glass beaker and sonicated twice for 10 seconds at 40% power, 50% cycle. The beaker was stored on ice between sonication steps. After sonication, 25 µl of the solution was combined with 25 µl of PGLB for SDS-PAGE. 0.5 ml of the sample was placed in an Eppendorf tube and spun at 13.3K rpm in a microcentrifuge for 5 minutes at room temperature. The supernatant was transferred into a new Eppendorf tube, while the pellet was resuspended in 0.5 ml of T100. Samples for SDS-PAGE were prepared for the pellet and supernatant samples as shown above. An SDS-PAGE gel was run with lanes containing the whole cell extract, pellet, and supernatant from each of the 28ºC and 37ºC expressions. The bands were inspected to determine the solubility of the protein. The mutagenized Cse4 protein was largely insoluble at 37ºC.

2.3.4. Large-Scale Protein Expression

As in the Small-Scale Protein Expression (Section 2.3.2.), the pST50Tr-Cse4x1 (S727 → R) expression vector was transformed into CodonPlus(DE3) cells (Section 2.2.5.). For the expression of xH4, a pET3a-xH4 expression plasmid prepared by Karolin Luger was transformed into BL21(DE3)pLysS cells (Secion 2.2.5.). Two 500 ml Erlenmeyer flasks containing 100 ml of 2xTY, 100 µl of 50 µg/ml ampicillin, and 100 µl of 25 µg/ml chloramphenicol were inoculated with five colonies each from the transformation plate. The flasks were placed in a 21ºC shaking incubator overnight. The OD_{600} was checked periodically through the next day until one pre-culture flask reached between 0.1 and 1. In order to prepare

for large-scale expression, 12 1 L flasks were prepared by adding 500 ml of 2xTY, 0.5 ml of 50 mg/ml ampicillin, and 0.5 ml of 25 mg/ml chloramphenicol, and 3 ml of the pre-culture to constitute a 6 L protein expression.

The 1 L flasks were transferred to a shaking incubator at 37ºC, and the OD_{600} of some of the flasks was monitored until it reached 0.5-0.9. Expression was then induced by adding 0.5 ml of 0.2M IPTG to each flask. A 250 µl sample of preinduced culture was collected, spun down, and resuspended in 50 µl of PGLB for the SDS-PAGE gel. After 3 hours, 125 µl samples from four of the flasks were collected, spun down, and resuspended in 50 µl of PGLB, as well. An SDS-PAGE gel was run with lanes containing an induced sample and one sample from each of four flasks 3 hours after induction of expression.

3 hours after expression induction, cells from the flasks were poured into 500 ml centrifuge bottles and spun at 7000 rpm for 5 minutes at room temperature in a SLA3000 rotor in an RC5C Plus floor centrifuge. The supernatant was poured off, and cells were resuspended in 150 ml of T100 and poured into a 250 ml Nalgene polypropylene bottle. The cells were then flash-frozen with liquid nitrogen and stored at -20ºC.

2.3.5. Preparation of Washed Inclusion Bodies

Inclusion bodies were prepared for xH4 and Cse4. The frozen resuspended cells were thawed in a 30ºC water bath. The cells were then divided between four beakers and sonicated four times at 70% power, 50% cycle for 10 seconds. The beakers were stored on ice between sonication steps. One 10 µl sample of sonicated sample was collected and combined with 10 µl of PGLB. After sonication, the cells were transferred to 50 ml polypropylene tubes and centrifuged at 18K rpm for 20 minutes at room temperature in an SLA3000 rotor in an RC5C

Plus floor centrifuge. The supernatant was poured into a flask, and a 10 µl sample was collected and combined with 10 µl of PGLB. The pellet was resuspended in a total of 180 ml of TRITON buffer (20 mM Tris-Cl pH 8.0, 0.5 mM EDTA Na_2, 100 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM benzamidine, 1% TRITON X-100). Another 10 µl sample was collected and combined with 10 µl of PGLB. The tubes were centrifuged again at 18K rpm for 10 minutes at room temperature in the same rotor and centrifuge as above. The supernatant was transferred to a flask, and a 10 µl sample of the supernatant was collected and combined with 10 µl of PGLB. The TRITON extraction was repeated three more times, and samples were collected from the supernatant after each centrifugation step. Finally, the pellet was resuspended in 30 ml of WASH buffer (20 mM Tris-Cl pH 8.0, 0.5 mM EDTA Na_2, 100 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM benzamidine) in a single 50 ml polypropylene tube. A 10 µl sample of the resuspended WASH pellet was collected and combined with 10 µl of PGLB. Tubes were spun in the above rotor and centrifuged at 18K rpm for 10 minutes at room temperature. The supernatant was again transferred to a flask, and a 10 µl sample was collected and combined with 10 µl of PGLB. The washed inclusion bodies were stored at -20ºC. An SDS-PAGE gel was run with lanes containing the crude whole cell extract, crude extract supernatant, whole cell extract in TRITON, whole cell extract in WASH, and cell extract supernatant for each of the four TRITON washes. The majority of the desired protein was present in the whole cell extracts and not the supernatants.

2.3.6. Determination of Cse4:xH4 Ratio for Tetramer Refolding

Each inclusion body was resuspended in 30 ml of WASH buffer. 200 µl of each resuspension was added to a 1.5 ml Eppendorf tube. A Pipetman was used to resuspend any chunks in the resuspension. The following dilutions in WASH buffer were prepared for each

resuspended pellet: 1:500, 2:500, 4:500, 6:500, 8:500, 10:500. 10 µl of each dilution was combined with 10 µl of PGLB. An SDS-PAGE gel was then run to compare the relative intensities of bands for the Cse4 and xH4 dilutions and to determine the ratio of Cse4:xH4 to use in refolding.

2.3.7. Small- and Large-Scale Refolding and HPLC Purification of Tetramer

Small-scale refolding and purification experiments were done to optimize conditions for the large-scale version. To prepare for dialysis, 1 ml of each resuspended WASH pellet was transferred to an Eppendorf tube. The tubes were centrifuged at 13.3K rpm for 3 minutes at room temperature in a microcentrifuge. The supernatant was removed, and 30 µl of dimethyl sulfoxide (DMSO) was added to each pellet. After the pellets were left to sit for 5 minutes, they were resuspended in 1 ml each of TG0 (20 mM Tris-Cl pH 8.0, 10 mM DTT, 7 M guanidine HCl). The resuspended pellets were then centrifuged at 13.3K rpm for 3 minutes at room temperature with the solubilized proteins contained in the supernatant. Based upon the determined Cse4:xH4 ratio, the appropriate volumes of each solubilized histone and 2 volumes of TG0 were added to 7 kDa dialysis tubing. The titration was dialyzed overnight against 1L of H20-urea (10 mM HEPES pH 7.0, 20 mM NaCl, 10 mM 2-mercaptoethanol, 2M urea). The buffer was changed in the morning and left to dialyze for at least 3 more hours.

Following dialysis, 100 µl from the small-scale titration was pipetted from the dialysis tubing into one Eppendorf tube with the remainder being transferred into separate Eppendorf tubes. A 10 µl sample collected from the separate Eppendorf tube was combined with 10 µl of PGLB to constitute the “whole” sample. The Eppendorf tube containing the 100 µl sample and the Eppendorf tubes containing the remaining dialyzed sample were centrifuged at 13.3K rpm for

5 minutes at 10ºC in a microcentrifuge. The supernatant was removed from the tube with the 100 µl sample, and the pellet was resuspended in 100 µl of the dialysis buffer. A 10 µl sample was collected from this resuspension and combined with 10 µl of PGLB to constitute the "pellet" sample. A 10 µl sample was collected from the supernatant of one of the other Eppendorf tubes and combined with 10 µl of PGLB to constitute the "input" sample. The samples collected for SDS-PAGE and the Eppendorf tubes containing the spun titration samples were stored at 4ºC.

High Performance Liquid Chromatography (HPLC) with a Source S1 cation-exchange column was done to refold and purify the $(Cse4\text{-}xH4)_2$ tetramer. The column was equilibrated with H20-urea, and protein was eluted off of the column using a stepwise gradient of increasing NaCl concentration. H20-urea and H1500-urea (10 mM HEPES pH 7.0, 1500 mM NaCl, 10 mM 2-mercaptoethanol, 2M urea) were used to create the gradient. The flow rate over the column was 4.0 ml/min, and 4 ml fractions were collected. Following the salt gradient, the column was washed with 0.2 M NaOH and 20% acetic acid to elute any residual protein. Throughout the HPLC run, the UV 280 nm absorbance was recorded on a chromatogram to facilitate determination of fractions containing protein. 10 µl of each protein-containing fraction was combined with 10 µl of PGLB. The "whole," "pellet," "input," and fraction samples were run on an SDS-PAGE gel to determine which fractions contained bands corresponding to both Cse4 and xH4 with few extraneous bands corresponding to impurities.

After conditions were optimized, a large-scale refolding and purification was done with the remainder of the inclusion body material. The above procedure was followed with the exception of proportionally higher amounts of DMSO and TG0 being used. For the large-scale purification, the appropriate fractions were pooled, pipetted into 7 kDa dialysis tubing, and dialyzed against H20 (10 mM HEPES pH 7.0, 20 mM NaCl, 10 mM 2-mercaptoethanol)

overnight to remove the urea. The dialysis buffer was changed once the next morning and left to dialyze for at least 3 hours.

2.3.8. Concentration of Purified Proteins

The purified $(Cse4\text{-}xH4)_2$ tetramer was concentrated to 11.20 mg/ml to prepare for reconstitution of the Cse4 nucleosome. Following dialysis, the protein was transferred from the dialysis tubing to a 50 ml Falcon Tube. The concentration of the dialyzed protein was determined via UV-Visible Spectrophotometry (Section 2.3.9.). A 30,000 MWCO Amicon concentration device was used to concentrate the tetramer. In order to prime the concentrator, 20 ml of Milli-Q® water was loaded and spun at 3500 rpm for 5 minutes at room temperature in a tabletop centrifuge. The filtrate was discarded, and 20 ml of the dialysis buffer was added to the top of the concentrator. It was spun again at 3500 rpm for 5 minutes at room temperature in a tabletop centrifuge. The filtrate was discarded.

First, a small-scale concentration was done in order to ensure high yield. 5 ml of dialyzed protein was loaded, and the concentrator was centrifuged at 3500 rpm at room temperature until 0.5 ml remained above the filter. The remaining 0.5 ml of protein was quantitated, and the yield was calculated. 60 µl of concentrated protein was saved for dynamic light scattering (DLS), which determined the protein's aggregation state. DLS measures a solution's polydispersity, which refers to the size heterogeneity of the particles in solution. A Viscotek 802 DLS instrument was used, which injected beams of light into the solution that scattered upon contact with particles. The intensity of that light scattering was measured using OMNISIZE software, and a polydispersity value was determined. Because the tetramer was theoretically the only species in solution, a low polydispersity was expected.

Barring a high polydispersity or low recovery, a large-scale concentration was done to concentrate the remainder of the dialyzed protein. The sample was loaded sequentially into the concentrator, which was centrifuged at 3500 rpm at 10ºC until the desired concentration was reached. The final concentration was verified and recorded, and the concentrated protein was transferred from the concentrator to an Eppendorf tube. To store the concentrated protein, glycerol was added to the solution to 20%. The contents were thoroughly mixed, and the tube was flash-frozen in liquid nitrogen and stored at -80ºC.

2.3.9. UV Quantitation of Protein and Nucleosomes

In order to determine the OD_{600} or A_{280} of a protein solution or A_{260} of a nucleosome solution, a Cary 50 Bio UV-Visible Spectrophotometer was used. For OD_{600} determinations in protein expressions, the spectrophotometer was blanked against 2xTY. For A_{280} determinations during protein concentrations and A_{260} determinations during nucleosome concentrations, it was blanked against the appropriate dialysis buffer. The absorbance was measured in the range of 220 to 350 nm. An adjusted A_{280} or A_{260} value was determined by subtracting the absorbance at 320 nm from the absorbance at 280 nm or 260 nm. The concentration was determined with Beer's Law: $A = \varepsilon cl$, where ε varied depending on the protein, $l = 1$ cm, and A = the adjusted A_{280} or A_{260} value. The equation was solved for "c" to find the concentration of protein in the sample.

2.4. Reconstitution of the Yeast Centromeric Nucleosome

2.4.1. Nucleosome Reconstitution and Gradient Dialysis

Two 2 mg + 2mg (~2 mg protein + ~2 mg DNA) nucleosome reconstitutions were performed. The first was a stoichiometric 2.60 dimer:tetramer ratio with a stoichiometric 1.56 DNA:octamer ratio. The second was a 2.80 dimer:tetramer ratio with a stoichiometric 1.56 DNA:octamer ratio. For each reconstitution, appropriate amounts of $(Cse4:xH4)_2$ tetramer, histone dimer (provided by Michael Doyle) and NCP601a DNA were combined in a 5 ml Eppendorf tube. The contents were mixed, and the tube was placed on ice for 1 hour. Then, the contents were loaded into 7 kD dialysis tubing.

A gradient dialysis was set up using a peristaltic pump and the following two buffers: RB-low (10 mM Tris-Cl pH 7.5, 1 mM EDTA Na_2, 1 mM DTT, 0.25 M KCl) and RB-high (10 mM Tris-Cl pH 7.5, 1 mM EDTA Na_2, 1 mM DTT, 2 M KCl). 2 L of RB-low buffer was added to 2 L plastic beaker, and 500 mL of RB-high was added to a 1 L glass beaker. The dialysis tubing containing the nucleosome components was placed into the RB-high beaker. A line connecting the RB-low beaker to the RB-high beaker was used to pump RB-low buffer into the RB-high beaker. Another line connecting the RB-high beaker to waste transferred contents of the RB-high beaker to waste. The dialysis was left to run overnight.

2.4.2. Source Q HPLC Purification of the Nucleosome Core Particle

Following dialysis, the sample was transferred from dialysis tubing into Eppendorf tubes. The tubes were centrifuged at 13.3K rpm for 3 minutes at 10ºC in a microcentrifuge. 4 µl of the supernatant were combined with 10 µl of 10% sucrose and 6 µl of Milli-Q® water to constitute

the input sample. HPLC with a Source Q10 anion-exchange column was done to purify the reconstituted nucleosome core particle. The column was washed with two column volumes of water prior to the start of the salt gradient. The nucleosome core particle was eluted from the column using a KCl gradient. The gradient was created with TEK250 (10 mM Tris-Cl pH 7.5, 0.5 mM EDTA Na_2, 250 mM KCl) and TEK 1000 (10 mM Tris-Cl pH 7.5, 0.5 mM EDTA Na_2, 1000 mM KCl) buffers.

Based upon the generated chromatogram, fractions containing protein were run on a pre-run 10% native gel in 0.2x TBE (18 mM Tris base, 18 mM boric acid, 0.4 mM EDTA pH 8.0) at 5W for 1 hour in order to visualize the NCP601a DNA wrapped around the octamer core. Samples were prepared by combining 10 µl from the HPLC fraction with 10 µl of 10% sucrose. After the run, the native gel was stained with ethidium bromide. The fractions corresponding to lanes containing DNA were pooled and added to 7 kD dialysis tubing. The pool was dialyzed against 500 ml of NCP storage buffer (10 mM Potassium Cacodylate pH 6.5, 0.1 mM EDTA Na_2) over night. The buffer was changed once the next day and allowed to dialyze for at least 2 hours.

2.4.3. Concentration of the Reconstituted, Purified Nucleosome Core Particle

The concentration of protein in the dialyzed solution was determined via UV-Visible Spectrophotometry (Section 2.3.9.). 2 µl sample was added to 998 µl of 0.1 M NaOH to create a 1:500 dilution. The pool was transferred from dialysis tubing into a 15 mL Falcon tube and added sequentially into a primed 10,000 MW Amicon concentrator. In order to prime the concentrator, 5 ml of Milli-Q® water was loaded and spun at 3500 rpm for 5 minutes at room temperature in a tabletop centrifuge. The filtrate was discarded, and 5 ml of the used dialysis

buffer (NCP storage buffer) was added to the top of the concentrator. It was spun again at 3500 rpm for 5 minutes at room temperature in a tabletop centrifuge. The filtrate was discarded.

The concentrator containing the purified nucleosome was centrifuged at 3500 rpm at 10ºC until the desired concentration of about 10 mg/ml was reached. The concentration was verified using UV-Visible Spectrophotometry. The concentrated sample was pipetted into an Eppendorf tube and stored at -8ºC.

3. Results and Discussion

3.1. Cloning of Cse4

3.1.1. Amplification of Cse4 Gene from Yeast Genomic DNA

The Cse4 gene was amplified from yeast genomic DNA using STO4643 and STO4642 primers (see Appendix). Following agarose gel purification, the amplified product was digested with 20 units/µl EcoRI and 20 units/µl PstI in order to create sticky ends for subsequent ligation into the pWM529 cloning vector. The digested product was analyzed on an agarose gel, and the bright band in Figure 9 corresponds to the expected 709 base pair (bp) product.

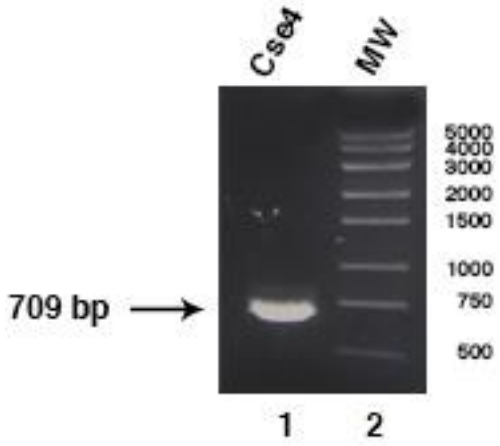

Figure 9. Digested Cse4 Gene
Cse4 with sticky ends for ligation.

3.1.2. Subcloning and PCR Site-Directed Mutagenesis of Cse4

The amplified and digested Cse4 gene was ligated with a pWM529 cloning vector via sticky-ended ligation. The sequence of the purified pWM529-Cse4 construct was verified. The ligation of Cse4 into a cloning vector allowed for the gene to be modified further prior to being subcloned into an expression vector.

The serine at position 727 in the Cse4 gene was changed to arginine to remove the internal NdeI site. Because the expression system employed in this procedure included digestion with NdeI, the NdeI site within the Cse4 gene had to modified so it would not be cut during this digestion. This modification was made using PCR site-directed mutagenesis with STO4644 and STO4645 primers (see Appendix). The sequence of the purified pWM529-Cse4x1 construct contained the desired mutation. The x1 notation is used here to differentiate the mutagenized Cse4 gene from the wild-type Cse4 gene.

3.1.3. Subcloning of Cse4 gene into Expression Vector

The pWM529-Cse4x1 construct was digested with NdeI and BsrGI to create sticky ends complementary to the sticky ends in the pST50Tr expression vector and to isolate the Cse4x1 gene from the cloning vector. The digested product was run on an agarose gel. The 692 bp fragment corresponds to the Cse4 gene lacking the internal NdeI site (Figure 10), while the larger 2024 bp band represents the pWM529 cloning vector.

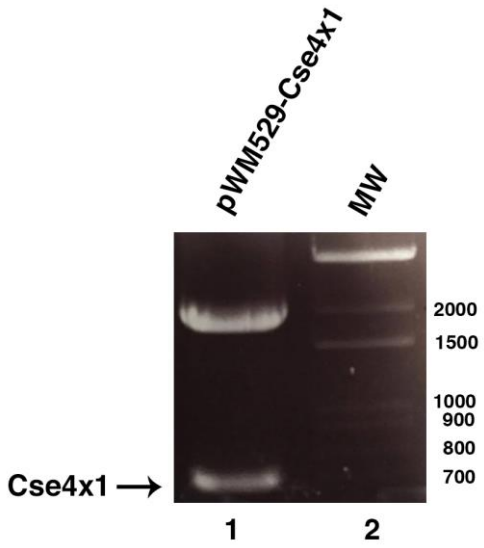

Figure 10. Cse4 Lacking Internal NdeI Site
Mutagenized Cse4 gene with sticky ends.

The Cse4x1 fragment was gel purified and ligated with a pST50Tr expression vector via sticky-ended ligation. The sequence of the purified pST50Tr-Cse4x1 plasmid was verified. This expression vector was used for subsequent expression of the Cse4 protein. While the Cse4 protein expressed and utilized in this procedure exhibits one amino acid difference from the wildtype Cse4 protein, its structure and function are negligibly affected.

3.2. Expression of Cse4 and xH4

3.2.1. Small-Scale Expression of Cse4

Cse4 was not expressed in BL21(DE3)pLysS cells at either 28ºC and 37ºC (Figure 11).

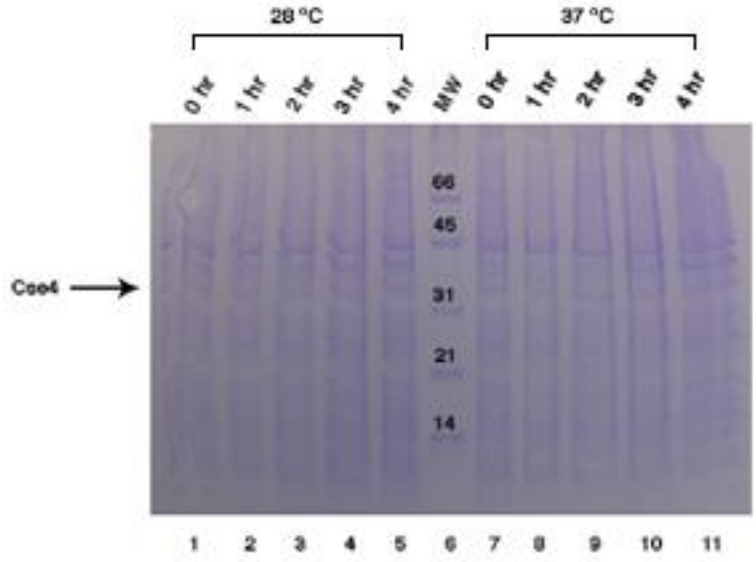

Figure 11. Small-Scale Expression of Cse4 in BL21(DE3)pLysS Cells
Cse4 runs on SDS-PAGE at about 34 kD. No Cse4 expression at either 28ºC or 37ºC.

In most cases, BL21(DE3)pLysS cells are ideal for protein expression because this cell line is low in protease activity, includes a λ prophage that encodes the T7 RNA polymerase, and contains a plasmid that expresses a T7 lysozyme that prevents leaky expression. However, use of BL21(DE3)pLysS cells produced only low levels of Cse4 expression. A potential issue may have been codon usage bias, the phenomenon in which part of a subcloned gene contains an abundance of codons for which the host organism does not have a sufficient pool of corresponding tRNAs. CodonPlus(DE3) cells contain the argU gene that encodes the tRNA for the AGA arginine codon, the ileY gene that encodes the tRNA for the AUA isoleucine codon,

and the leuW gene that encodes the tRNA for the CUA leucine codon. These codons are all present in the Cse4 sequence, but BL21(DE3)pLysS cells lack these extra tRNA genes. Thus, CodonPlus(DE3) cells were employed in an attempt to circumvent the codon usage bias problem because these cells contains extra tRNA genes. Cse4 expression levels with CodonPlus(DE3) cells were markedly higher (Figure 12), so it is likely that codon usage bias was the main reason for low levels of expression in BL21(DE3)pLysS cells.

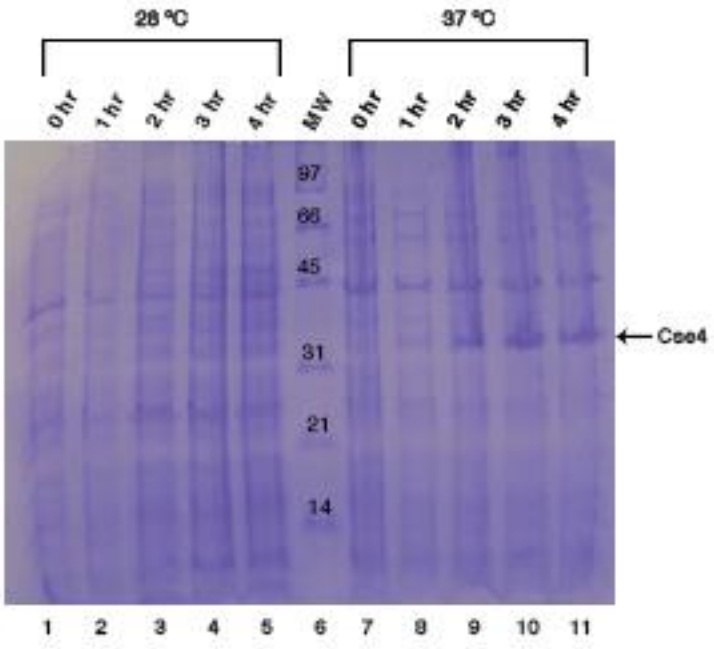

Figure 12. Small-Scale Expression of Cse4 in CodonPlus(DE3) Cells
High-level expression of Cse4 at 37ºC.

Cse4 was expressed at a high level in CodonPlus(DE3) cells at 37ºC but not at 28ºC. Also, high levels of Cse4 expression were observed 3 hours following induction, so these parameters were used for the large-scale expression of Cse4. In order to determine whether the protein was expressed in the soluble or insoluble form, samples of the whole protein solution

were run on an SDS-PAGE gel alongside pelleted and supernatant samples following expression (Figure 13).

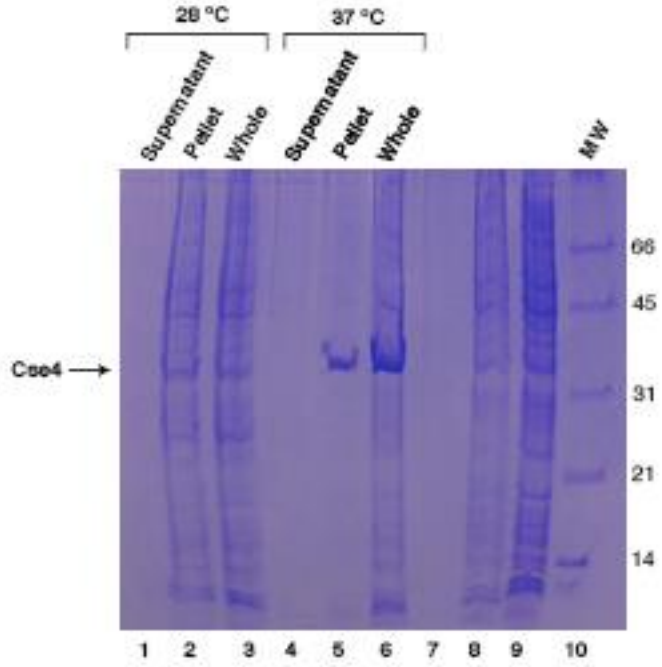

Figure 13. Solubility Test of Cse4
Cse4 is contained in the pellet sample.

Because the 37ºC supernatant lane contained no band and a thick band was observed in the 37ºC pellet lane, Cse4 was determined to be insoluble. The solubility of a protein determines the ways in which it can be purified. Insoluble histones can be readily prepared and purified from inclusion bodies, which is what was done in this procedure.

3.2.2. Large-Scale Expressions of Cse4 and xH4

Large-scale expressions of both Cse4 and xH4 were performed at 37ºC and harvested 3 hours after induction. xH4 refers to the *Xenopus* H4 histone, but it is identical to yH4, the yeast H4 histone. The notation "xH4" will be used to describe the H4 histone included in this study. Cse4 was expressed in CodonPlus(DE3) cells, while xH4 was expressed in BL21(DE3)pLysS cells. Parameters for xH4 expression were previously determined by other members of the laboratory. Levels of expression for both Cse4 and xH4 matched the level of expression in the small-scale Cse4 expression shown in Figure 12.

3.3. Purification of Cse4 and xH4

3.3.1. Inclusion Body Preparation of Cse4 and xH4

Inclusion bodies of Cse4 and xH4 were prepared using sonication, four TRITON buffer washes, and one WASH buffer wash. Inclusion bodies result from aggregation of proteins following high-level expression in cells. In addition to the over-expressed protein, these aggregates contain additional endogenous cellular components that contaminate the protein of interest. TRITON is a detergent that serves to remove these contaminants from the inclusion body. Centrifugation allows the insoluble protein to refold, while the contaminants remain in the supernatant. This TRITON wash was done four times in an effort to remove the vast majority of cellular contaminants. The final step includes resuspension of the inclusion body in WASH buffer, which serves to remove excess TRITON X-100 from the sample.

For the Cse4 inclusion body preparation, Cse4 was present in each of the whole cell extracts. About 10% of the Cse4 remained in the crude extract and TRITON supernatants, but

the majority of Cse4 was retained in the inclusion body (Figure 14). Fairly equal amounts of Cse4 were lost in each of the supernatants.

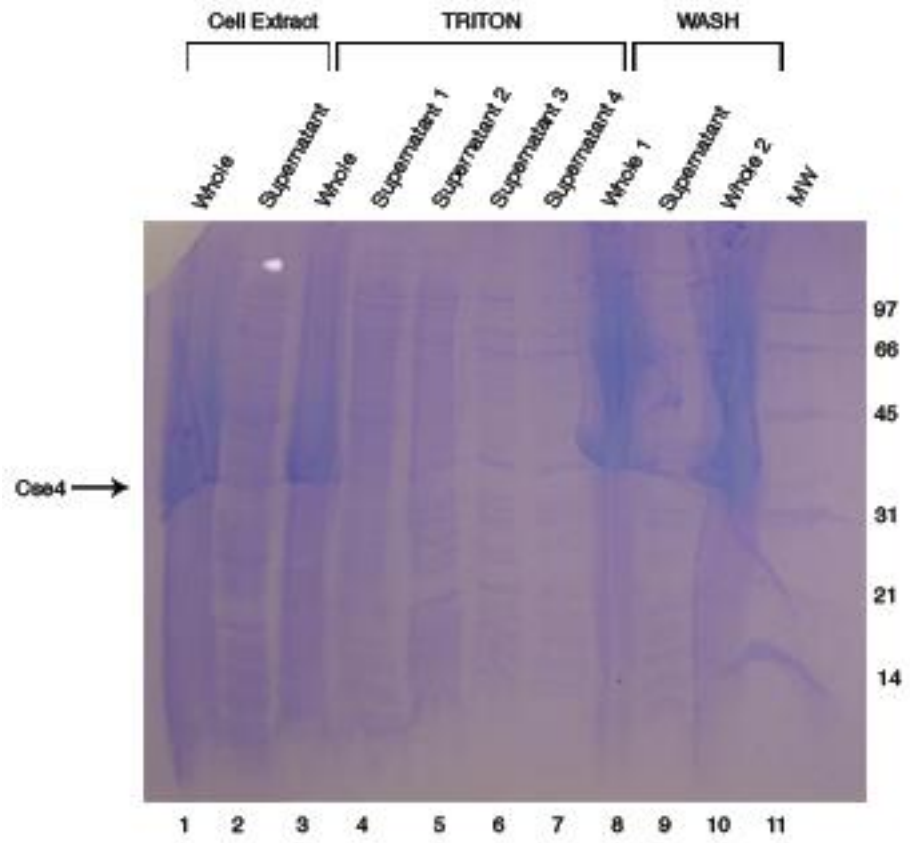

Figure 14. Inclusion Body Preparation of Cse4
Most Cse4 retained in the whole samples.

Like Cse4, the majority of xH4 remained in the inclusion body. Only about 5% of the protein was removed in the crude extract and TRITON supernatants. However, the majority of xH4 lost was retained in the crude extract supernatant, while almost none was retained in the TRITON supernatants (Figure 15).

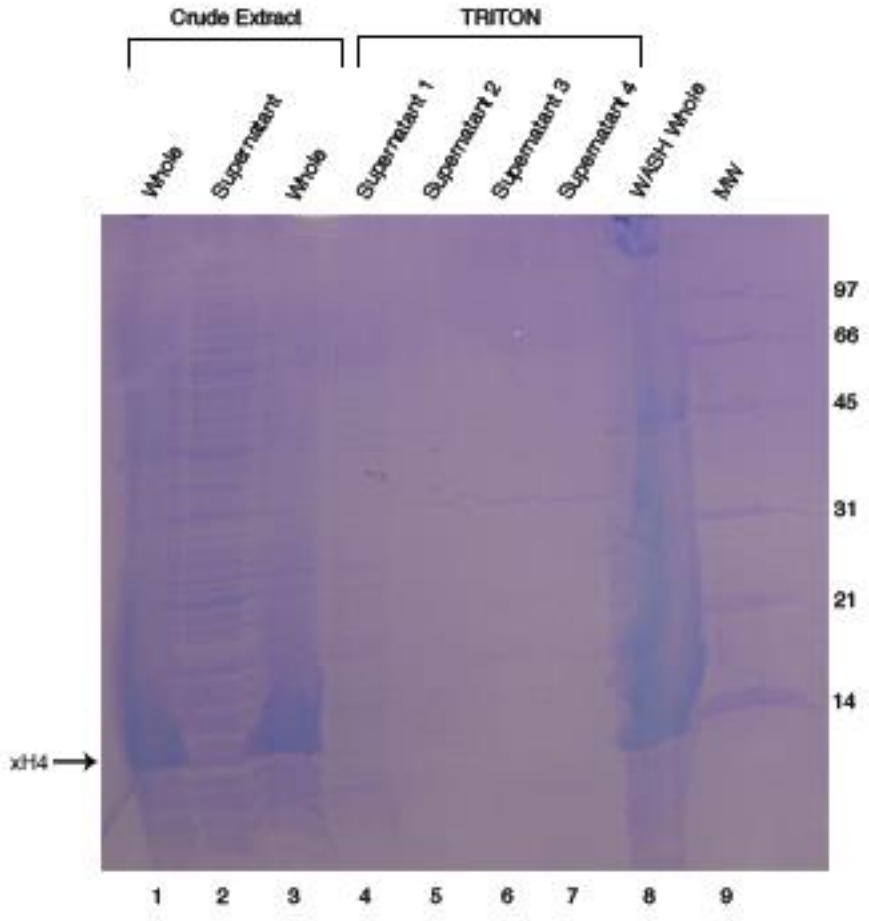

Figure 15. Inclusion Body Preparation of xH4
Most xH4 retained in the whole samples.

3.4. Refolding, Purification, and Concentration of the Tetramer

3.4.1. Determination of Cse4:xH4 Ratio from Small-Scale Dilutions

The ratio of Cse4:xH4 for refolding and purification was partially determined by running a series of small-scale dilutions in WASH buffer on an SDS-PAGE gel. The intensity of the Cse4 and xH4 bands were different, so many HPLC runs were done using a combination of Cse4:xH4 ratios, including 0.2:1, 0.3:1, 0.4:1, 0.6:1, 0.8:1, and 1:1 (Figure 16). The 0.2:1 Cse4:xH4 ratio yielded the most protein eluting during the salt gradient, so this ratio was chosen to determine the relative amounts of each resuspended, solubilized inclusion body to be used in refolding and purification of the tetramer.

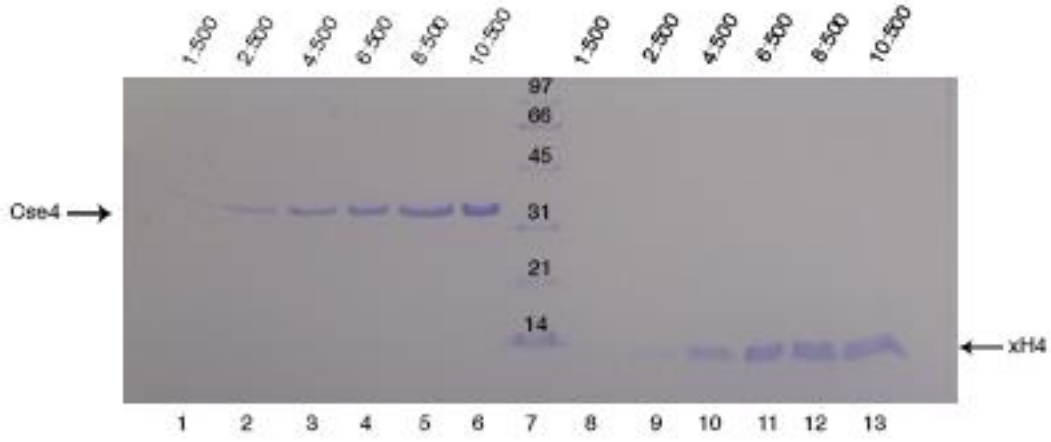

Figure 16. Small-Scale Dilutions of Cse4 and xH4
Samples of Cse4 and xH4 inclusion bodies in various dilutions with WASH buffer.

The band intensities in Figure 16 were not expected based upon the relative expression levels and inclusion body preparations for Cse4 and xH4. Cse4 and xH4 seemed to express at similar levels, yet slightly more Cse4 was lost than xH4 during the inclusion body preparation.

Thus, it was expected that the bands corresponding to xH4 would be slightly thicker than those corresponding to Cse4. However, determinations on the basis of band intensity are certainly not exact. In making the small-scale dilutions of each inclusion body, the inclusion bodies were resuspended in WASH buffer, but the xH4 inclusion body settled on the bottom of the tube faster than did the Cse4 inclusion body. While samples for dilutions were pipetted within a few seconds of resuspension, it is possible that the Cse4 dilutions were more concentrated because the Cse4 inclusion body remained resuspended for longer.

3.4.2. Source S1 HPLC with H100 at pH 7.5

Samples of each inclusion body were solubilized in preparation for dialysis. Inclusion bodies were resuspended in 30 ml of WASH buffer, and 1 mL of each resuspension was collected, centrifuged, and aspirated, leaving small pellets of each inclusion body. 30 µl of DMSO and 1 ml of TG0 were added to pellet, which was then centrifuged, leaving the majority of each histone in the supernatant of the solution. TG0 contains guanidine-HCl, a strong base that facilitates protein denaturation. As the inclusion body protein associations were broken by guanidine-HCl, the protein aggregation was reduced, and the histone was solubilized. Based upon the 0.2:1 Cse4:xH4 ratio determined by several HPLC runs, 250 µl of solubilized Cse4 was combined with 525 µl of solubilized xH4 and 1550 µl of TG0 in 7 kDa dialysis tubing. The solution was dialyzed in H100 buffer, which contained 10 mM HEPES pH 7.5, 100 mM NaCl, and 10 mM 2-mercaptoethanol. High Performance Liquid Chromatography (HPLC) was done with a Source S1 cation-exchange column in an attempt to refold and purify the tetramer. H100 and H1500 buffers were used at a pH of 7.5 for the low and high-salt HPLC buffers, respectively.

In cation-exchange chromatography, the sample protein initially binds to an anionic column. Once all sample has been loaded onto the column, a salt gradient passes over the column. The cation component of the salt in solution displaces the sample protein as it flows over the column. The concentration of salt required to displace each protein varies, which allows separation of a sample protein from most protein contaminants. Fractions are collected during the salt gradient, which separate the proteins eluting from the column at different salt concentrations.

Elution of the protein is measured throughout the HPLC run via UV-Visible spectrophotometry and displayed on a chromatogram (Figure 17). The blue line on each chromatogram refers to the UV_{280} absorbance of proteins. Blue peaks signify protein elution. The fractions containing high protein peaks were collected and run on an SDS-PAGE gel in order to determine the identity of the eluted protein in each fraction. Here, fractions 1A1, 1A4, 1B4, 1B5, and 1B6 were selected (Figure 18).

Fractions 1A1 and 1A4 represent the protein collected while the input sample was being applied to the column. The vast majority of Cse4 and xH4 eluted in these fractions, suggesting that the tetramer did not bind to the column and was not properly purified.

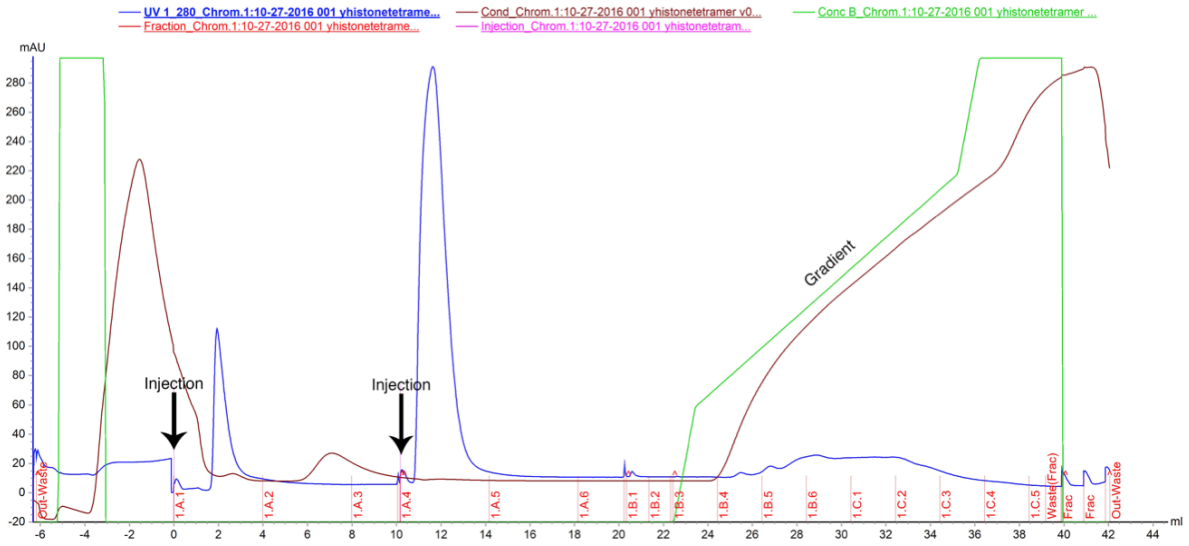

Figure 17. Chromatogram of Source S1 HPLC with H100 at pH 7.5
Most protein eluted immediately following injection prior to salt gradient.

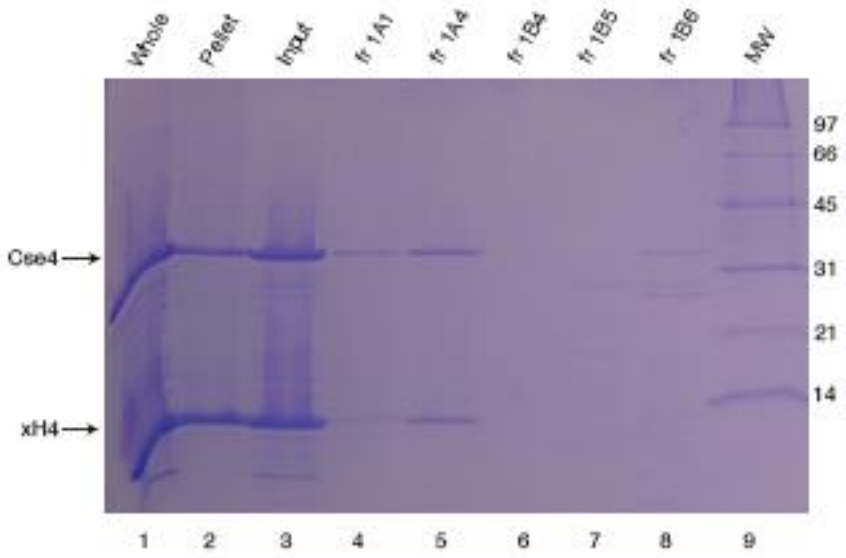

Figure 18. Tetramer Elution for H100 at pH 7.5
SDS-PAGE gel of HPLC fractions containing protein peaks.

3.4.3. Source S1 HPLC with H20 at pH 7.5

In order to facilitate initial binding of the tetramer complex, the concentration of the low salt buffer was decreased to account for the possibility that the tetramer elutes at a concentration lower than 100 mM NaCl. Thus, the low salt and dialysis buffer was changed to H20 (10 mM HEPES pH 7.5, 20 mM NaCl, 10 mM 2-mercaptoethanol), and HPLC was repeated as in Section 3.4.2. Again, no protein eluted during the salt gradient, and most of the protein did not bind to the column.

3.4.4. Source S1 HPLC with H20 at pH 7.0

The pH of the dialysis buffer ad HPLC buffers was manipulated in an attempt to promote tetramer binding to the column. Each protein and protein complex has an isoelectronic point (pI), which is the pH at which the protein exhibits a net charge of zero. When the pH of the solution containing the protein is above the pI, the protein will exhibit a negative charge. The protein will be positively charged when the pH of the solution is below the protein's pI. The pI of a complex as large as a tetramer gives little useful information for specifying HPLC parameters because the complex contains several potential points of contact with the column. Each subunit of the complex may exhibit a slightly different charge and react differently to changes in pH. Still, decreasing the pH of the dialysis buffer and HPLC buffers would make more portions of the tetramer positively charged, which would facilitate column binding. Thus, the HEPES buffer was reduced to pH 7.0. H20 (10 mM HEPES pH 7.0, 20 mM NaCl, 10 mM 2-mercaptoethanol) and H1500 (10 mM HEPES pH 7.0, 1500 mM NaCl, 10 mM 2-mercaptoethanol) were used. HPLC was repeated as in Section 3.4.2. The chromatogram is shown in Figure 19.

Only small peaks were observed in fractions 1A1 and 1A4, suggesting that most of the protein did bind to the column. Additionally, some protein eluted during the salt gradient, marked by the green and brown lines. However, the largest peak occurred during the column wash with 0.2 M NaOH and 20% acetic acid. Several fractions were run on an SDS-PAGE gel (Figure 20).

Fractions 1B4-1C3 contained protein that eluted during the gradient. Based upon the thickness of the bands in those lanes in comparison to the bands in the input lane, only approximately 30% of the tetramer eluted during the gradient. The remaining tetramer eluted during the column wash, which indicated that the tetramer was binding too tightly to the column to elute even at 1500 mM NaCl. This refolding and purification procedure was repeated, but these results were unable to be replicated. All repetitions exhibited no tetramer eluting during the salt gradient.

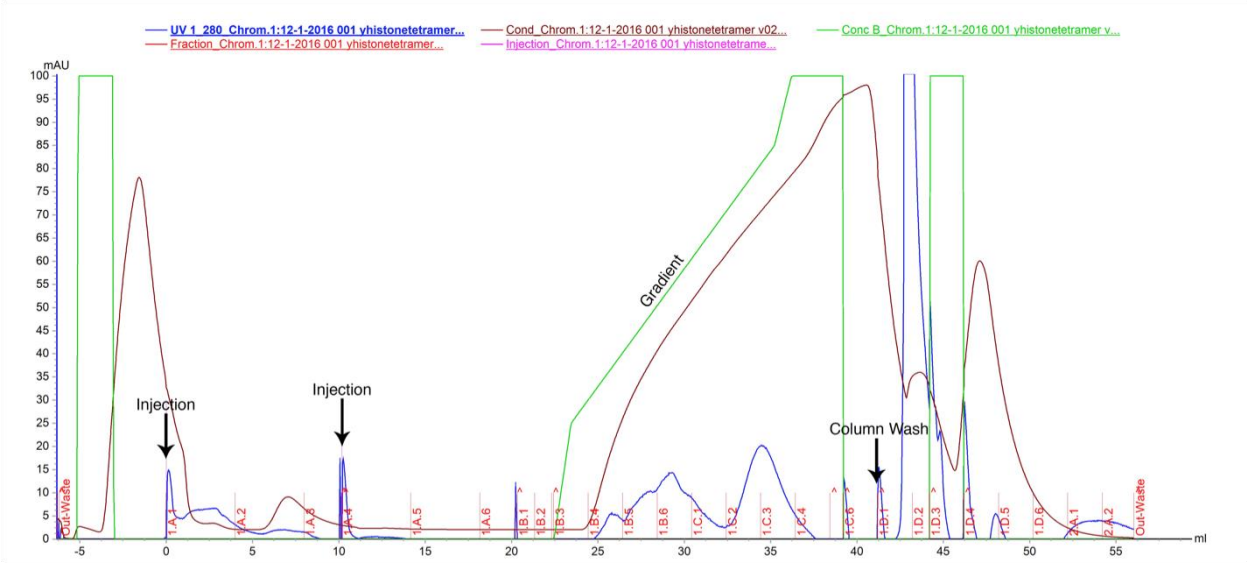

Figure 19. Chromatogram of Source S1 HPLC with H20 at pH 7.0
Most protein eluted during the column wash (1D1-1D3) with acetic acid and sodium hydroxide.

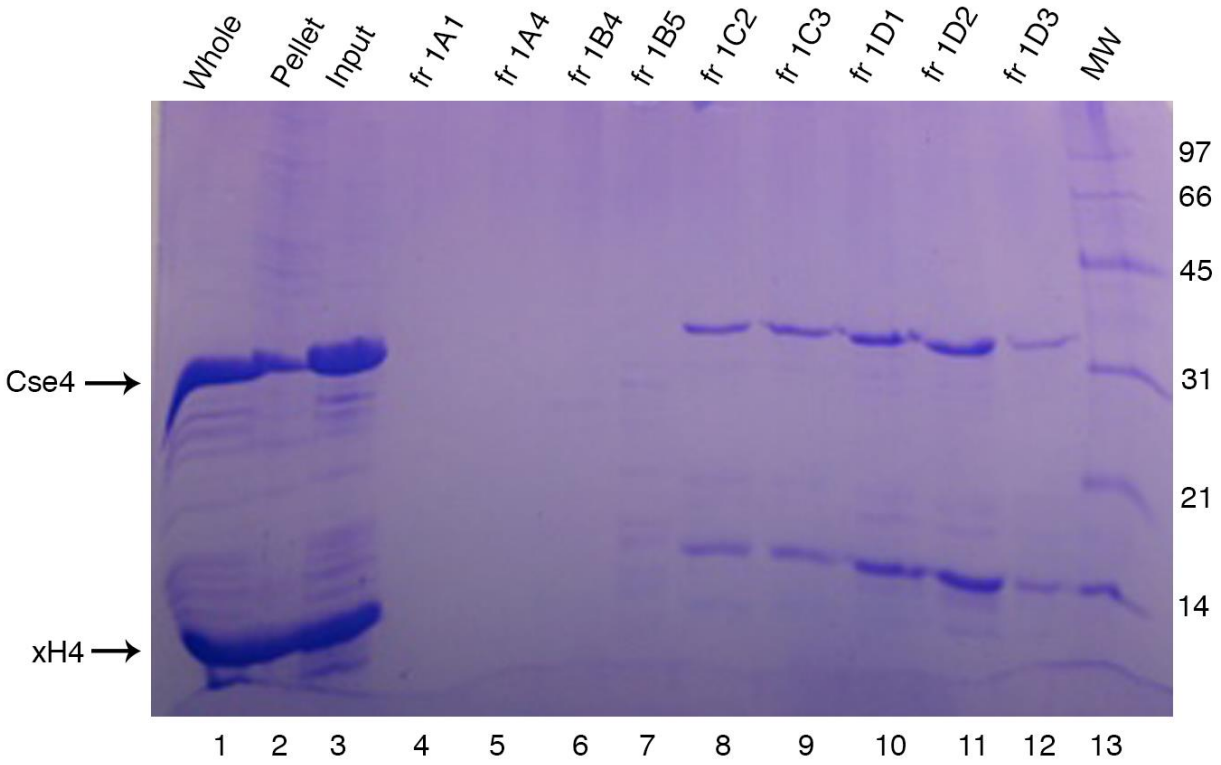

Figure 20. Tetramer Elution for H20 at pH 7.0
SDS-PAGE gel of HPLC fractions containing protein peaks.

3.4.5. Source S1 HPLC with H20 at pH 7.0 and 2M Urea

In addition to the decrease in NaCl concentration and the reduction of the HEPES pH, 2M urea was added to all buffers to loosen the binding of tetramer to the column. Urea at concentrations of 1-2 M has been shown to act as a mild denaturant and facilitate elution (Shukla et al, 2007). By incorporating 2 M urea into the dialysis and HPLC buffers, the number and strength of column contacts was likely reduced. H20 (10 mM HEPES pH 7.0, 20 mM NaCl, 10 mM 2-mercaptoethanol, 2 M urea) was used for the dialysis and low-salt buffer, and H1500 (10 mM HEPES pH 7.0, 1500 mM NaCl, 10 mM 2-mercaptoethanol, 2 M urea) was used for the high-salt buffer. HPLC was repeated as in Section 3.4.2., and approximately 70% of the input protein eluted during the salt gradient (Figures 21 and 22).

With the above parameters set, large-scale refolding and HPLC was done using H20 (10 mM HEPES pH 7.0, 20 mM NaCl, 10 mM 2-mercaptoethanol, 2 M urea) as the low-salt and dialysis buffers and H1500 (10 mM HEPES pH 7.0, 1500 mM NaCl, 10 mM 2-mercaptoethanol, 2 M urea) as the high-salt buffer. Fractions eluting during the gradient were collected and added to 7 kDa dialysis tubing and dialyzed overnight in H20 without added urea. This dialysis was done in order to remove excess urea from the tetramer and restore its native state.

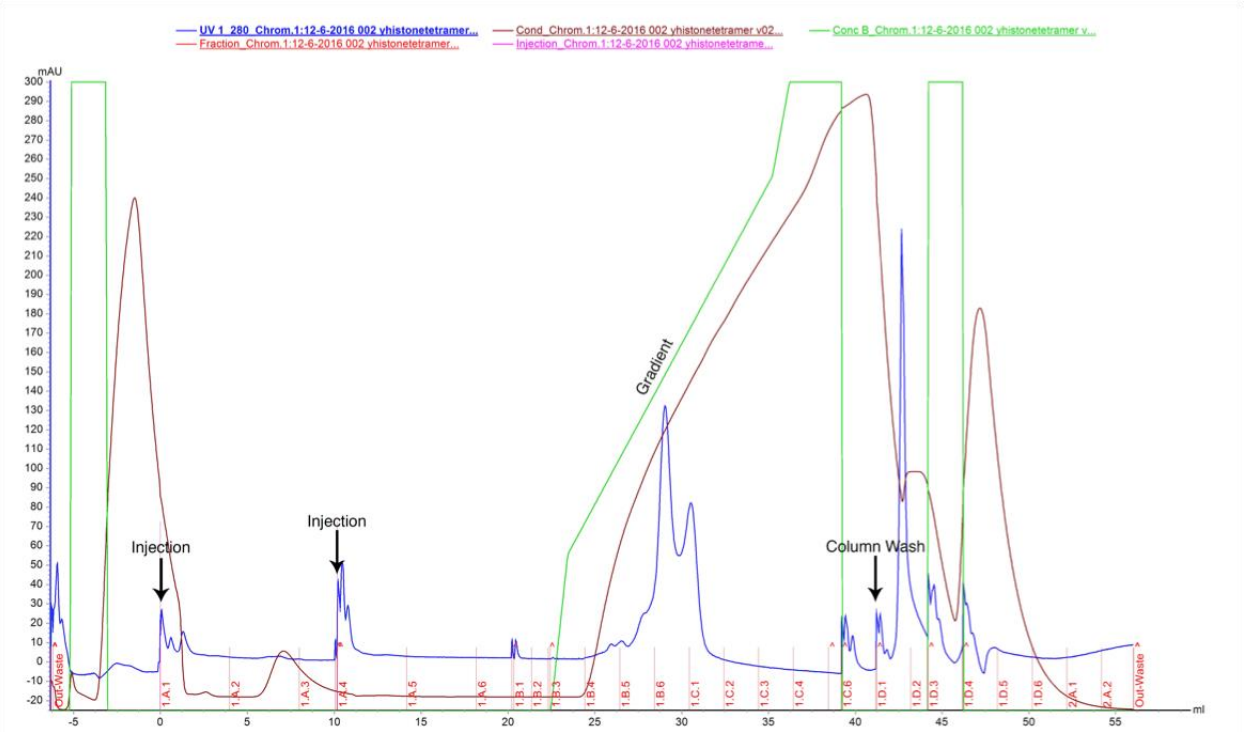

Figure 21. Chromatogram of Source S1 HPLC with H20 at pH 7.0 with 2 M urea
Protein eluted during both the salt gradient (1B5-1C1) and the column wash (1D1-1D2).

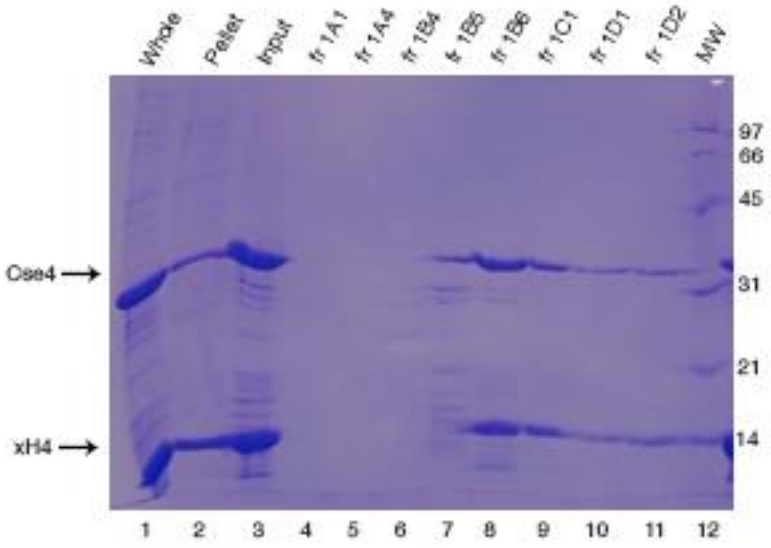

Figure 22. Tetramer Elution for H20 at pH 7.0 with 2 M urea
SDS-PAGE gel of HPLC fractions containing protein peaks.

3.4.6. Concentration of the Tetramer

After the urea was dialyzed out of the purified tetramer, the sample was concentrated to 11.20 mg/ml. UV-Visible spectrophotometry was used to quantitate the protein following concentration. The adjusted A_{280} value and Beer's law were used to determine the concentration of tetramer. The concentration was multiplied by the volume of the concentrated solution to determine the amount of tetramer present. A total of 14.2 mg of purified and concentrated tetramer was collected.

3.4.7. Dynamic Light Scattering of the Concentrated Tetramer

Dynamic light scattering (DLS) determines the distribution of differently sized particles in solution by injecting beams of light. When the light hits a particle, it scatters and undergoes constructive or destructive interference with surrounding particles. The intensity of this scattering changes as it interacts with other particles over time. This change in intensity is measured, which gives a polydispersity value and other information about the particle size distribution within the solution. Polydispersity refers to the particle size heterogeneity in a solution. DLS data comes in two primary forms, an intensity distribution and a mass distribution. The intensity distribution shows the molecular weight of each particle based on its scattering intensity, while the mass distribution converts the scattering intensity data to masses with the assumption that all particles are perfectly spherical (Malvern, 2011). DLS was performed on the concentrated tetramer to determine the aggregation state of the sample. Figure 23 and Table 2 show the intensity distribution and its associated data.

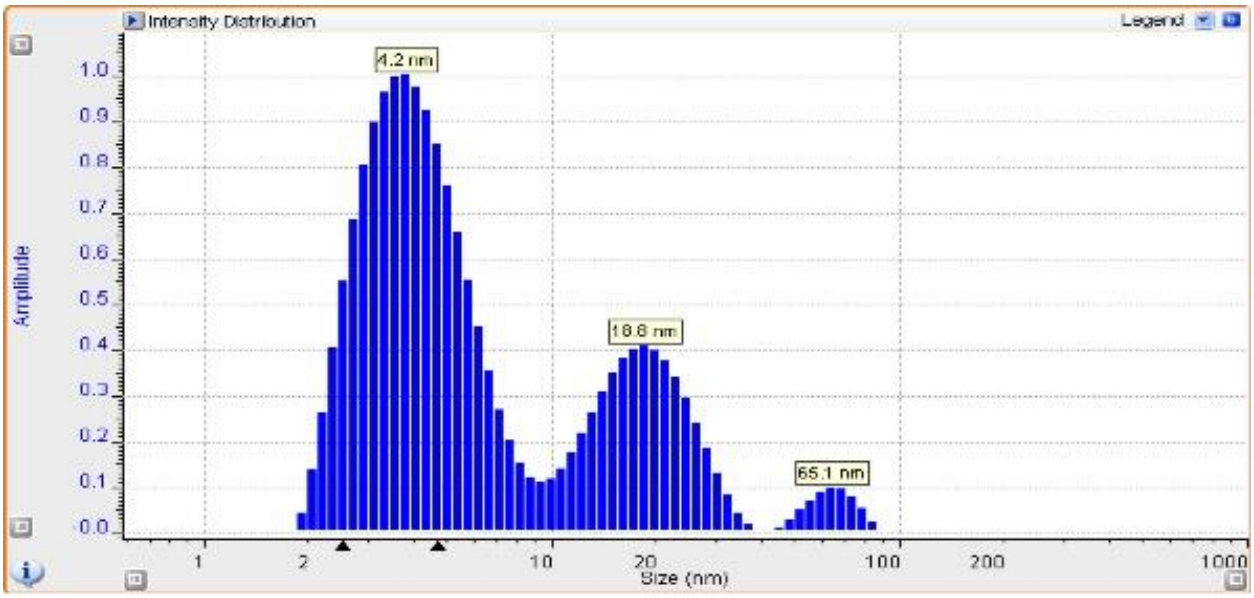

Figure 23. DLS Intensity Distribution
Polydispersity of the tetramer solution based upon intensity of light scattering.

Table 2. DLS Intensity Distribution Data

Peak	% Area	Rh (nm)	Position	Polydispersity (%)	MW (kD)
1	63.8	4.15	3.8	34.1	95.64
2	24.1	18.76	18.62	30.5	3374.4
3	2.6	65.11	64.57	14.2	6.38E+04
4	9.5	2530.19	2511.89	12.3	3.63E+08

According to the DLS intensity distribution, there were several species of different sizes within the tetramer solution. Peak 1, which likely corresponds to the tetramer (MW 75.9 kD), has a polydispersity of 34.1%. The monodisperse tetramer should have a polydispersity of 20% or less, which accounts for DLS measurement errors. This high polydispersity suggests that the

solution contained tetramer in addition to other particles. Peaks 2-4 have much larger associated molecular weights, which shows that protein aggregates were likely present in the solution. The mass distribution displays similar data that further support this analysis (Figure 24 and Table 3).

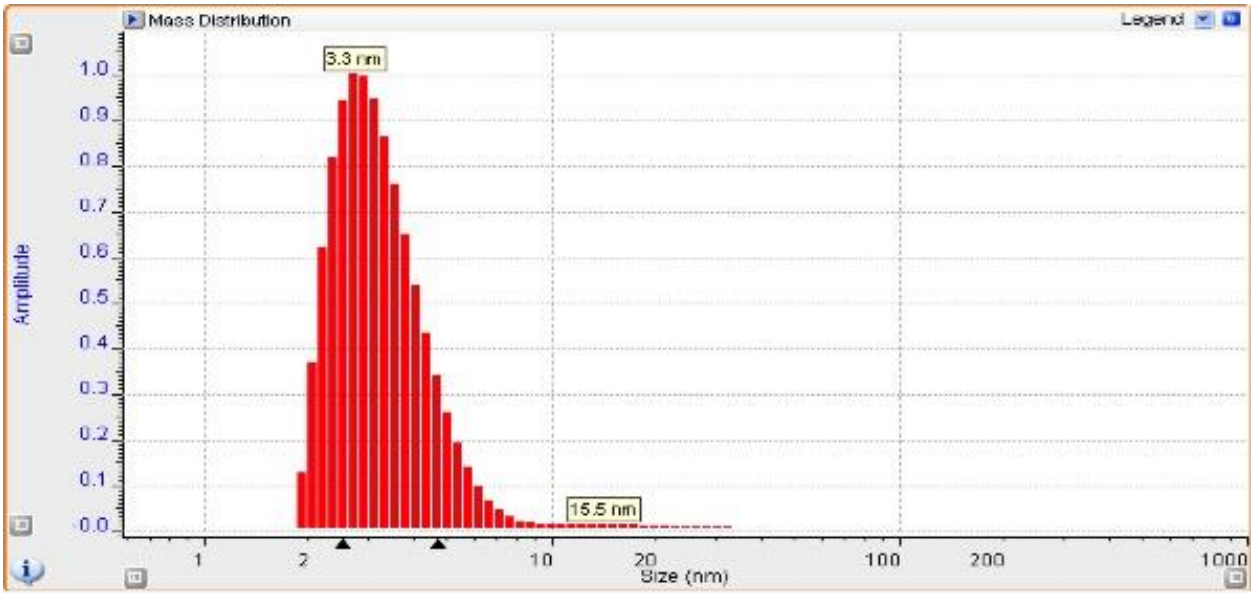

Figure 24. DLS Mass Distribution
Polydispersity of the tetramer solution based upon converted mass calculations.

Table 3. DLS Mass Distribution Data

Peak	% Area	Rh (nm)	Position	Polydispersity (%)	MW (kD)
1	99.1%	3.29	2.69	31.5	55.03
2	0.9%	15.55	14.13	29.2	2164.09

Based upon the DLS mass distribution, fewer aggregates were present in the tetramer solution than shown in the intensity distribution. Still, a high polydispersity of 31.5% was found for peak 1, which corresponds to the tetramer. Monomers of Cse4 and xH4 and/or other small impurities are likely present, as well. In order to reduce aggregation, the salt concentration of the dialysis buffer could be increased. This would break up some of the aggregation and yield a greater proportion of active tetramer.

3.5. Purification and Concentration of the Reconstituted Nucleosome Core Particle

3.5.1. Source Q10 HPLC

Reconstitutions were performed with assistance from Bryan Tornabene. The concentrated tetramer was used in two nucleosome reconstitutions, along with provided yH2A-yH2B dimer and NCP601a DNA. The 601 DNA sequence contains 147 bp of DNA that exhibit high affinity for histone octamers. It also includes a biotin group on the 5' end, which is useful in nucleosome binding assays. For the first reconstitution, 1.12 mg of tetramer, 1.07 mg of dimer, and 2.24 mg of DNA were combined to constitute a 2.60 stoichiometric dimer:tetramer ratio reconstitution. The second reconstitution utilized a 2.80 dimer:tetramer ratio and included 1.12 mg of tetramer, 1.16 mg of dimer, and 2.24 mg of DNA. These components were pooled, stored on ice for 1 hour, and loaded into 7 kDa dialysis tubing. The reconstitutions underwent a gradient dialysis with RB-low and RB-high buffers. The reconstituted nucleosome core particle was purified via Source Q10 anion-exchange HPLC. In anion-exchange chromatography, the column is positively charged and binds negatively charged proteins and complexes. A salt gradient is used to elute the desired protein from the column as the salt's anions displace the bound protein. The

chromatogram for the 2.60 dimer:tetramer reconstitution is shown in Figure 25. The chromatogram for the 2.80 dimer:tetramer reconstitution is not shown because it is almost identical to that shown in Figure 25.

Fractions 1A5-1B6, which contain DNA peaks peaks, from both reconstitution experiments were run on a 10% native gel to determine which fractions contained the band corresponding to the NCP601a DNA around the octameric protein complex (Figure 26). The gels for both reconstitutions were essentially identical. Only fractions 1A5, 1A6, 1B1, and 1B2 contained the nucleosome bound to NCP601a DNA. Extraneous bands likely correspond to DNA that has not been incorporated into the nucleosome core particle. Thus, each of the four fractions likely contains the yeast centromeric nucleosome core particle along with extraneous DNA strands. Fractions 1A5-1B2 were pooled, loaded into 7 kDa dialysis tubing, and dialyzed into NCP storage buffer.

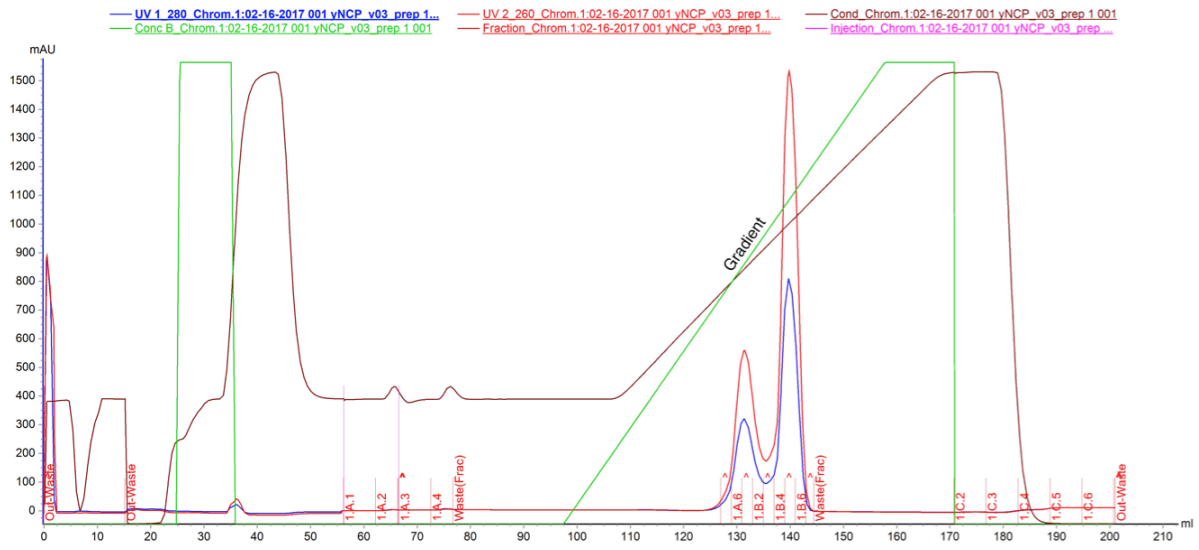

Figure 25. Chromatogram of Source Q10 HPLC
The nucleosome eluted during the salt gradient.

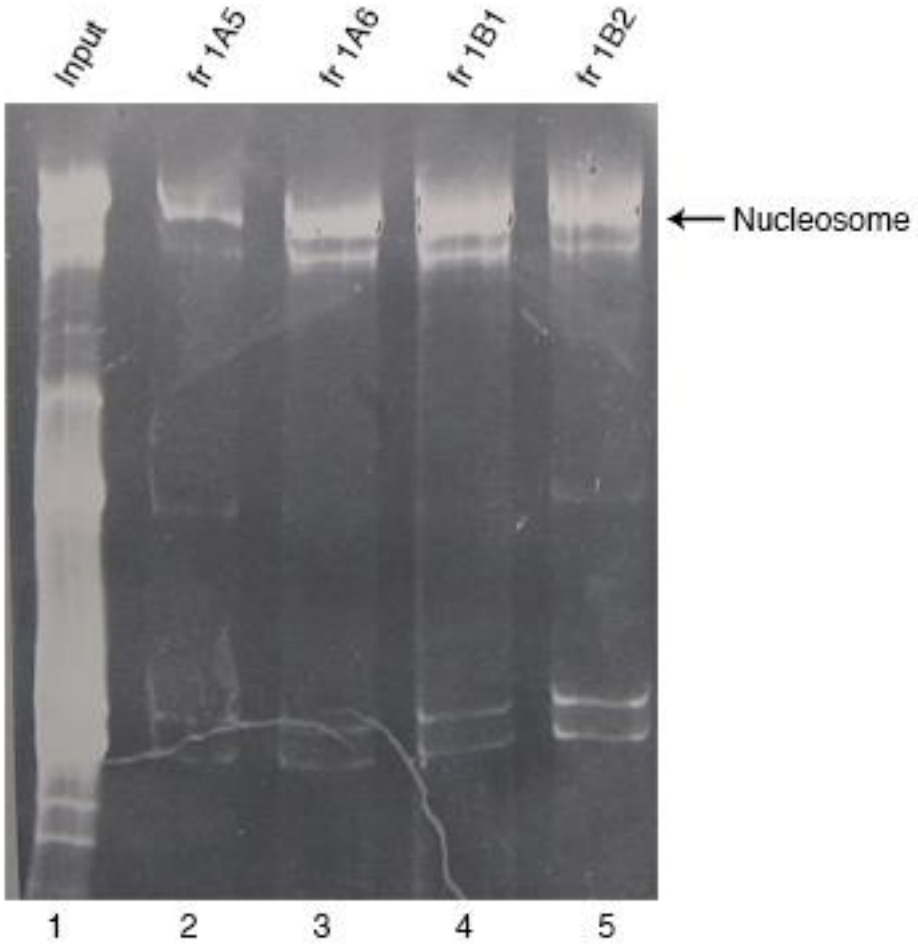

Figure 26. HPLC Fractions Containing the Nucleosome Core Particle
Fractions 1A5-1B2 contained the yeast centromeric nucleosome.

3.5.2. Concentration of the Reconstituted NCP

The purified nucleosome core particle with a 2.60 dimer:tetramer ratio was concentrated to 8.47 mg/ml. A total of 0.5 mg of the nucleosome core particle was collected, which constitutes a yield of only 11.2%. The purified nucleosome core particle with a 2.80 dimer:tetramer ratio was concentrated to 9.13 mg/ml. Again, a total of 0.5 mg of the nucleosome core particle was collected, constituting a yield of only 11.1%.

The yields of the nucleosome core particles were extremely low in comparison to the 40-70% yields for most nucleosome preparations in the Tan laboratory. It is likely that much of the NCP601a DNA did not wrap around the free histone octamers in solution and eluted after the nucleosome during Source Q10 HPLC. Much of the DNA contained in the fractions under the second DNA peak in the chromatogram (Figure 25) was unbound to the octamer and was not pooled following HPLC. This decreased the concentration of the DNA measured by UV-Visible spectroscopy following concentration of the nucleosome core particle. Based upon the DLS results of the tetramer, some aggregated tetramer was present in the solution from which the ~1 mg of tetramer was acquired. Thus, it is likely that the aggregated tetramer failed to reconstitute with dimer or DNA. In order to determine the polydispersity of the nucleosome core particle solutions, DLS could have been done. Unfortunately, due to low yields, a sufficient amount of the sample for DLS testing was lacking.

4. Conclusions

4.1. Reconstitution of the Yeast Centromeric Nucleosome Core Particle

The homotypic octameric form of the yeast centromeric nucleosome core particle was successfully reconstituted. It contains two copies of H2A, H2B, Cse4, and H4 around which NCP601 DNA is wrapped.

4.2. Further Troubleshooting

Although this study has set many parameters for the reconstitution of the yeast centromeric nucleosome core particle, further improvements are necessary to improve yield. Prior to concentrating the tetramer, dialyzing into a high salt buffer may reduce the aggregation suggested by the DLS data. Additionally, merely skipping the dialysis step prior to tetramer concentration may prevent aggregation. Following Source S1 HPLC, the purified tetramer was dialyzed in order to remove urea, a mild denaturant. However, if the urea remained in solution, the tetramer would be less likely to aggregate, though the tetramer may not be retained in its native state. The urea could potentially be dialyzed out of solution during the nucleosome reconstitution. Because the urea may slightly denature the dimer complex and prevent formation of the octamer, concentrations of urea lower than 2 M could be used in an attempt to prevent tetramer aggregation while still retaining the possibility of nucleosome reconstitution. Lastly, refolding of the $(Cse4\text{-}xH4)_2$ tetramer has been shown to be inefficient with yields of less than 0.05% (Camahort et al, 2009). Perhaps purifying the Cse4 and xH4 histones individually prior to nucleosome reconstitution would circumvent the potential issue of tetramer instability.

4.3. Applications

The prepared nucleosomes can be employed in binding studies, such as High-throughput Interactions by Fluorescence Intensity (HI-FI) assays. HI-FI allows quantification of binding strength between two proteins or complexes. Fluorescent labels could be added to different loci in the prepared nucleosomes, and dequenching or fluorescence resonance energy transfer (FRET) assays could be used to determine where and how strongly the nucleosome interacts with various kinetochore proteins.

While the prepared nucleosomes can be useful for elucidating the structures and mechanisms of the yeast centromere-kinetochore complex, they do not take into account the relevance of yeast centromeric DNA. Instead of using NCP601a DNA, yeast centromeric DNA can be used in further reconstitutions so that nucleosomes more closely resemble those observed *in vivo*. Furthermore, interactions between the conserved domain elements and both the histones and the kinetochore proteins can be studied.

Appendix - Oligonucleotide Primer Sequences

STO 4642 – Reverse primer for amplification of Cse4 gene from yeast genomic DNA

CGGGCTGCAGTGTACATTAAATAAACTGTCCCCTGATTCTT

STO 4643 – Forward primer for amplification of Cse4 gene from yeast genomic DNA

CGGAATTCCATATGTCAAGTAAACAACAATGGGTTA

STO 4644 – Forward primer for site-directed mutagenesis of Cse4 gene

GCGAACTCATTCTTATGCCTTAGACAGATATGTTA

STO 4645 – Reverse primer for site-directed mutagenesis of Cse4 gene

CTAAGGCATAAGAATGAGTTCGCACTGGT

REFERENCES

Black, Ben E, Daniel R Foltz, Srinvas Chakravarthy, and Karolin Luger. "Structural determinants for generating centromeric chromatin." *Nature* 430.6999 (2004): 578-582.

Camahort, Raymond, Manjunatha Shivaraju, Mark Mattingly, Bing Li, Shima Nakanishi, Dongxiao Zhu, Ali Shilatifard, Jerry L. Workman, and Jennifer L. Gerton. "Cse4 Is Part of an Octameric Nucleosome in Budding Yeast." *Molecular Cell* 35.6 (2009): 794-805.

Cleveland, Don W, Yinghui Mao, and Kevin F Sullivan. "Centromeres and Kinetochores: From Epigenetics to Mitotic Checkpoint Signaling." *Cell* 112.4 (2003): 407-421.

Dalal, Yamini, Hongda Wang, Stuart Lindsay, and Steven Henikoff. "Tetrameric Structure of Centromeric Nucleosomes in Interphase *Drosophila* Cells." *PLOS Biol* 5.8 (2007): e218.

Dechassa, Mekonnen Lemma, Katharina Wyns, Ming Li, Michael A. Hall, Michelle D. Wang, and Karolin Luger. "Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes." *Nature Communications* 2.313 (2011).

Dechassa, Mekonnen Lemma, Katharina Wyns, and Karolin Luger. "Scm3 deposits a (Cse4-H4)2 tetramer onto DNA through a Cse4-H4 dimer intermediate." *Nucleic Acids Research* 42.9 (2014): 5532-5542.

Hemmerich, Peter, Tanja Stoyan, Gerhard Wieland, Marianne Koch, Johannes Lechner, and Stephan Diekmann. "Interaction of yeast kinetochore proteins with centromere-protein/transcription factor Cbf1." *PNAS* 97.23 (2000): 12583-12588.

Keith, Kevin C., and Molly Fitzgerald-Hayes. "CSE4 Genetically Interacts With the *Saccharomyces cerevisiae* Centromere DNA Elements CDEI and CDEII but Not CDEIII: Implications for

the Path of the Centromere DNA Around a Cse4p Variant Nucleosome." *Genetics* 156.3 (2000): 973-981.

Kornberg, Roger D., and Yahli Lorch. "Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome." *Cell* 98.3 (1999): 285-294.

Luger, Karolin, Armin W. Mäder, Robin K. Richmond, David F. Sargent, and Timothy J. Richmond. "Crystal structure of the nucleosome core particle at 2.8 Å resolution." *Nature* 389 (1997): 251-260.

Malvern Instruments. "Dynamic Light Scattering. Common Terms Defined." 2011: 1-6.

Manjunatha, Shivaraju, Jay R. Unruh, Brian D. Slaughter, Mark Mattingly, Judith Berman, and Jennifer L. Gerton. "Cell cycle coupled structural oscillation of centromeric nucleosomes in yeast." *Cell* 150.2 (2012): 304-316.

McGinty, Robert K., and Song Tan. "Nucleosome Structure and Function." *Chemical Reviews* 115 (2015): 2255-2273.

Meraldi, Patrick, Andrew D McAinsh, Esther Rheinbay, and Peter K Sorger. "Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins." *Genome Biology* 7.R23 (2006).

Mizuguchi, Gaku, Hua Xiao, Jan Wisniewski, M. Mitchell Smith, and Carl Wu. "Nonhistone Scm3 and Histones CenH3-H4 Assemble the Core of Centromere-Specific Nucleosomes." *Cell* 129.6 (2007): 1153-1164.

Polo, Sophie E, Abderrahmane Kaidi, Linda Baskcomb, Yaron Galanty, and Stephen P Jackson. "Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodeling factor CHD4." *The EMBO Journal* 29.18 (2010): 3130-3139.

Shukla, Abhinav A., Priyanka Gupta, and Xuejun Han. "Protein aggregation kinetics during Protein A chromatography: Case study for an Fc fusion protein." *Journal of Chromatography A* 1171.1-2 (2007): 22-28.

Stoler, Sam, Kevin C. Keith, Kathie E. Curnick, and Molly Fitzgerald-Hayes. "A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis." *Genes & Development* 9 (1995): 573-586.

Studier, FW. "Use of T7 RNA polymerase to direct expression of cloned genes." *Methods Enzymol.* 185 (1990): 60-89.

Tan, Song, Ronald C. Kern, and William Selleck. "The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*." *Protein Expression and Purification* 40.2 (2005): 385-395.

Zhang, Yi, and Danny Reinberg. "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails." *Genes & Development* 31.4 (2001): 2343-2360.

Zhou, Zheng, Hanqiao Feng, Bing-Rui Zhou, Rodolfo Ghirlando, and Kaifeng Hu. "Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3." *Nature* 472.7342 (2011): 234-237.

Academic Vita of Daniel Friedland

Email: dbf5060@psu.edu

EDUCATION:

The Pennsylvania State University, 8/2013 – 5/2017
Schreyer Honors College
Bachelor of Science, Biology (Neuroscience option)
Bachelor of Arts, Spanish
Minor, Violin Performance
Minor, Philosophy

IES Abroad Madrid Language & Area Studies 9/2015 – 12/2015

ACADEMIC HONORS: Dean's List (all semesters), Schreyer Academic Excellence Scholarship, Penn State Provost Award Scholarship

RESEARCH EXPERIENCE:

Song Tan Laboratory, Center for Eukaryotic Gene Regulation, University Park, PA 1/2015 - Present
Undergraduate Researcher

- Extended knowledge of biochemistry, molecular biology, and epigenetics
- Completed thesis: Reconstitution of the Yeast Centromeric Nucleosome

CLINICAL EXPERIENCE:

WellSpan York Hospital, York, PA 7/2016 - 8/2016
Pre-medical Summer Internship

- Worked alongside Family Medicine residents and attending physicians
- Interpreted for Spanish-speaking patients
- Completed project on depression awareness in both English and Spanish

WellSpan York Hospital, York, PA 6/2015 - 8/2015
Volunteer – Patient Aide

- Worked for over 100 hours with nurses and nurse aides to deliver optimal patient care
- Transported patients to and from appointments
- Assisted patients in entering and exiting cars
- Stocked IV fluid bags, blankets, gloves, sheets, gowns
- Interacted with patients before and after surgery
- Delivered ice chips and fluids to patients
- Cleaned medical equipment
- Answered call lights

WellSpan Surgery & Rehabilitation Hospital, York, PA 12/2013 - 3/2016
WellSpan York Hospital, York, PA
Shadowing

- Observed Daniel F. Motter, D.O. (Hospitalist)
- Gained insight into treatment plans, common medications, physician-patient interactions, and the roles of various medical personnel in the hospital setting
- Observed Robert Sterling, Jr., M.D. (Neurologist)