BIOCHEMICAL CHARACTERIZATION OF THE HUMAN TELOMERE STN1-TEN1 PROTEIN COMPLEX

Esmaeel Reza Dadashzadeh
The University of Michigan
Undergraduate Thesis in Biochemistry
April 21, 2009

This thesis has been read and approved by Dr. Ming Lei. Date: 04/20/2009

Table of Contents

List of Abbreviations

aa – Amino acid(s)

BME – β-Mercaptoethanol

ds – Double stranded

DTT – Dithiothreitol

EDTA – Ethylenediaminetetraacetic acid

FPLC – Fast Protein Liquid Chromatography

IPTG – Isopropyl β-D-1-thiogalactopyranoside

LB – Luria Broth

OD – Optical Density

PMSF – Phenylmethanesulphonylfluoride

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

ss – Single stranded

Tris – trishydroxymethylaminomethane

Abstract

Telomeres are the most common solution to the end-replication problem. They are located at the ends of chromosomes and are extended by telomerase, a reverse transcriptase. Telomeres also protect chromosome ends from degradation by "capping" the single stranded 3' overhang, which would have otherwise been perceived as a double stranded DNA break and resulted in cellular senescence and apoptosis. Telomeres are not able to achieve all of its functions without its association to telomere binding proteins. The Cdc13-Stn1-Ten1 (CST) complex binds to single stranded DNA through Cdc13 and protects the chromosome. The Stn1-Ten1 complex is a negative regulator of telomerase by competing, via the N-terminal domain of Stn1, with the Est1 subunit of telomerase for binding at the same region in Cdc13. Here we report the first attempts at attaining the crystal structure of the human Stn1-Ten1 complex. After resolving solubility issues and applying extensive crystallization optimization techniques, the complex diffracted to 4.2 angstroms. Additionally, we report the attempts of solving the human Stn1 C-terminal structure (which binds to Cdc13). Both cases yielded crystals, but not a strong enough diffraction plot. We plan to continue this work by designing and testing a myriad of new constructs for both hStn1 N-terminal and C-terminal.

Chapter 1: Introduction

Part I. The End-Replication Problem and the Origins of Telomeres and Telomerase

What if one could live forever? Can such a proposition be ignored? History, religion, and fiction have often flirted with the idea of immortality, whether it is Ponce de León's quest for the Fountain of Youth, the mystery of the Holy Grail, or Achilles' infamous heel. But what was once a matter of fantasy and fancy has, in the past several decades, been moved into realm of science, theory, and practice. And all of this stems from biology's compensation for one of its own flaws.

Deoxyribonucleic acid (DNA) is the genetic, hereditary material of all living organisms. Naturally, when an organism needs to grow or replace damaged cells, an efficient means of DNA replication is necessary. The machinery for semi-conservative eukaryotic DNA replication differs from that of prokaryotes. Eukaryotic chromosomes are comprised of two antiparallel strands of DNA, and new, daughter strands of DNA can only be synthesized unidirectionally since DNA polymerases δ and ε (homologs of prokaryotic DNA polymerase III) can only extend a polynucleotide in a 5' to 3' fashion (i.e. it can only add nucleotides to a free 3'-OH group) . What's more is that DNA polymerase can only synthesize new DNA onto an existing DNA or RNA chain. Thus, the role of DNA polymerase α (DNA primase for prokaryotes) is to create an RNA primer for DNA polymerase to build onto. Working in a similar 5' to 3' fashion, DNA polymerases β and ε (homologs of prokaryotic DNA polymerase I) can eventually replace the RNA primer with DNA (1,2).

At first glance, such a model appears to work well, but what happens when the daughter strand approaches the 3' end of their respective parent strand (Fig. 1.1)? A final primer of ≈10bp will be inserted by DNA polymerase α over the 3' end of the parent, but because there will be

nothing to the 5' end of it (i.e. there is no free 3'-OH group), DNA polymerases β and ε will not be able to replace the RNA with DNA and DNA ligase will not follow with its ligation of the sugar-phosphate backbones of the primer segment with its Okazaki fragment (lagging strand). This leaves the pair of primers on the 5' ends of the daughter strands vulnerable to removal by a 5' to 3' exonuclease. Since there is no free 3'-OH group, the DNA polymerases cannot fill in the space the primer once occupied (2). Thus, we have a deletion, or a loss of DNA on our daughter strand. This biological flaw was coined in 1972 by James Watson as the "End Replication Problem" (3). Prokaryotic genera that possess plasmids and circular chromosomes will bypass this problem their DNA polymerase can simply make a full turn, returning to the 5' end of its originating primer. Thus, the context of this thesis will focus on phenomena that occur only in eukaryotic organisms and the small fraction of bacteria with linear chromosomes, such as *Streptomyces* and *Borrelia* (4).

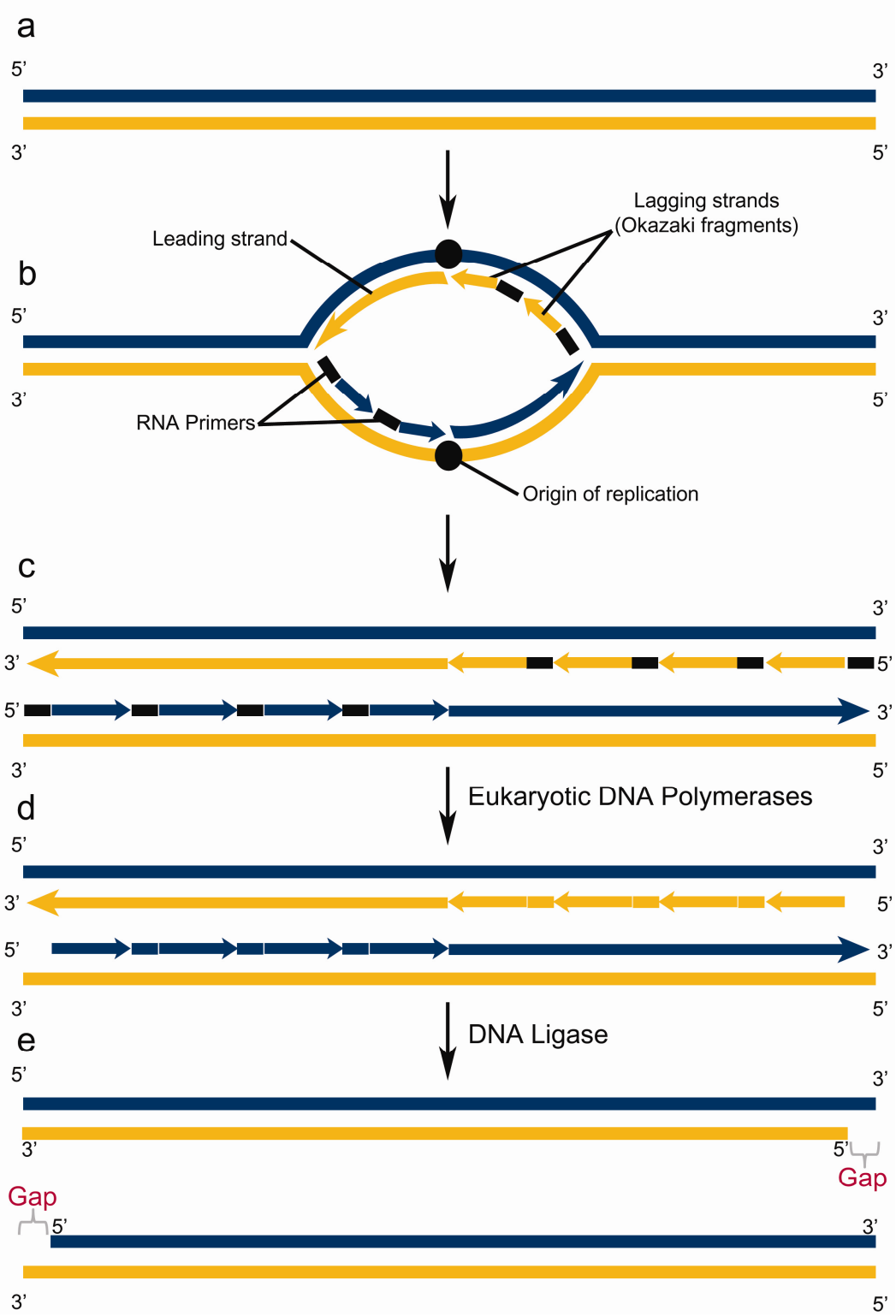

FIGURE 1.1. **The end-replication problem.** The maize and blue bars represent complimentary DNA strands. **(a)** dsDNA before replication begins. **(b)** A detailed diagram of one of the many replication forks. **(c) to (d)** The RNA fragments of the primers are removed by several eukaryotic DNA polymerases and the gaps are filled by adding the appropriate deoxyribonucleotides. **(d) to (e)** Once DNA ligase joins the Okazaki fragments, it becomes apparent that after DNA polymerase removed the final RNA primers at the 5' ends of the daughter strands, the enzyme was incapable of filling in the gaps (displayed in scarlet and grey) (2).

It is apparent that without intervention, every time a cell divides DNA would be lost. Fortunately, biology has developed a mechanism to avoid this fatal flaw. In 1978, it was discovered that the termini of the linear chromosomes of the ciliated protozoan *Tetrahymena thermophila* consist of non-coding guanine rich repeats of the hexanucleotide unit CCCCAA/GGGGTT (5). 10 years later, tandem repeats of TTAGGG were found at the 3' ends of human chromosomes (6). These guanine rich ends of eukaryotic chromosomes are collectively called the telomere region (5,6). Furthermore, the guanine rich strand is longer than its complement; therefore it protrudes as a single strand (7). The significance of this discovery was two-fold. First, the fact that chromosomes begin with this excess non-coding region signifies that it would take several generations of loss of termini DNA until a critical coding deletion is hit, depending on the starting length of the telomeric region. Second, the conserved sequence of TTAGGG hints that perhaps the telomere serves as a platform for protein binding. One would assume such to be the case; the single stranded G-rich, 3' overhang region of the telomere would otherwise be regarded as a DNA double-strand break by the cell, leading to senescence (inability for cells to replicate) or apoptosis (programmed cell death), the cell's typical responses to damaged DNA. Furthermore, the single-stranded nature of telomeres allows them to fuse with one another to form dicentric chromosomes. Such end-to-end fusions are considered erroneous by the cell, and just as the case was with a break in the DNA, senescence or apoptosis would be sure to follow (7).

The presence the non-coding telomeric region does not solve the end-replication problem—it only delays its onset. The first solution to the end-replication problem was uncovered in 1985 when the ribonucleoprotein (RNP) reverse transcriptase telomerase was discovered (8). Telomerase contains an RNA moiety which is able to bind to the single-stranded

telomeric 3' overhang and serve as a template for the extension of the telomeric repeat. Therefore, telomerase activity can be directly correlated to telomere length (9). However, telomerase maintains telomere length constant only in certain germ-line cells and totipotent embryonic stem cells (10). Pluripotent stem cells (i.e. renewal tissue such as blood, skin, and intestine) express some telomerase activity, but the telomere length of the cells of these tissue still shorten progressively over time (11). The great majority of somatic cells have minimal or no telomerase activity; the continual shortening of telomeres acts as a molecular clock where once a certain number of deletions is reached, cellular senescence or apoptosis is activated to prevent tumorigensis (12). However, there is a fine line between increasing cellular lifespan and the onset of human carcinomas. Erroneous up-regulation of telomerase in somatic cells, which occurs in 90% of human tumors, can erase this line (13).

Part II. A Mathematical Model for Cellular Senescence and Apoptosis

The original theory of cellular aging with respect to telomeres stated that there exists a direct correlation between telomere length and cellular lifespan. In other words, the longer the telomere, the more generations a cell is able to divide before it hits a critical deletion in its genome, signaling apoptosis or senescence (14). However, nearly twenty years later, the first of several telomere binding proteins were discovered, human TTAGGG binding factor 1 and 2 (TRF1 and TRF2 respectively). TRF2 "caps" telomere ends, protecting them from being recognized as breaks in the DNA (15). This new discovery suggests that perhaps it is the ability of telomeres to bind TRF2—and not the length of the telomeric repeats—that determines a cell line's lifespan.

Two mathematical models for cellular lifespan as a function of telomere loss were put forth before the discovery of TRF2 (16,17). A third model was later suggested by Natalie Arkus on the foundations of the new hypothesis that cellular apoptosis and senescence is triggered by the inability of a telomere to bind TRF2 (the binding probability of TRF2 decreases as telomeres shorten) (18). What follows is a review of the derivation of this TRF2 dependent mathematical model for cellular aging in cells lacking telomerase activity.

A. Derivation

What must be derived is an equation for the proportion of dividing, viable cells at generation *n* as a function of the increasing inability of a telomere to bind TRF2 as it shortens. The first step is to derive an equation for cellular proliferation that takes into account the end-replication problem (Fig. 1.1). Both generation 1 dsDNA products in **Figure 1.1e** have a gap, which will be defined as a deletion from this point forth. When they go on to replicate, not only

will the original deletion be passed on to the next generation, but new ones will form as a result of the end-replication problem repeating itself. Thus, we may confirm that the dynamic of telomere loss as presented in **Figure 1.2** is appropriate. Since the dynamics of a single strand of DNA is limited by its shortest telomere, be it on its 3' or 5' end, then we may define *d* as the maximum limiting deletions (i.e. if a ssDNA has two deletions on its 5' end and one deletion on its 3' end, then we may disregard the lower value, so $d = 2$).

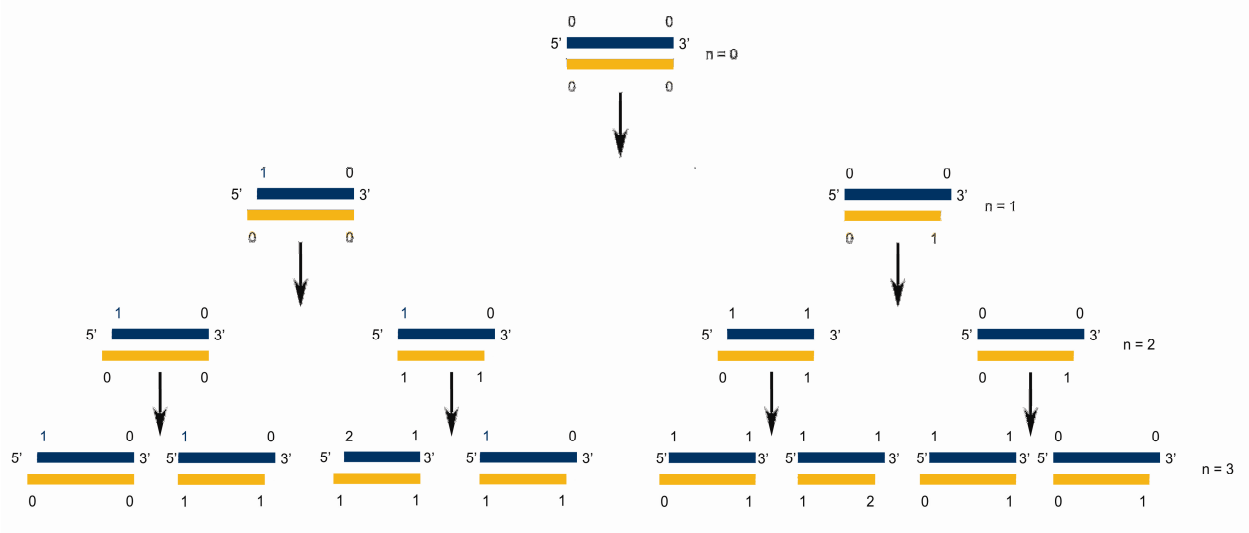

FIGURE 1.2. **The dynamics of telomere loss.** n is the generation number with the parent strand at n = 0. The number above the termini of the chromosomes represents the total number of deletions as a result of the end-replication problem.

As **Figure 1.2** indicates, the four chromosomes in generation n=2 have the following number of limiting deletions, going from left to right: (1,0), (1,1), (1,1), (0,1) where (x,y) is defined as (*d* for upper strand, *d* for lower strand). We can define the state where a chromosome has an equal number of limiting deletions on its complementary DNA strands as the even state (2*d*) and the state where one strand has one more limiting deletion than the other as the odd state (2*d*-1). In generation n=2 of **Figure 1.2**, we see that there are an equal number of chromosomes in the even state as there are in the odd state. The same holds true for generation 3 and onwards.

Furthermore, notice that the number of chromosomes in the even state at a given generation is equal to the sum of the number of chromosomes in either state from the previous generation; the same holds true for the odd state. This observation yields the following pair of equations for cells containing one chromosome:

$$\begin{aligned} \mathrm{N}_n(2d) &= \mathrm{N}_{n-1}(2d) + \mathrm{N}_{n-1}(2d-1) \\ \mathrm{N}_n(2d-1) &= \mathrm{N}_{n-1}(2d-1) + \mathrm{N}_{n-1}(2d-2). \end{aligned} \quad (1)$$

If we wish to ignore the parity of the chromosomes (i.e. distinguishing between even and odd states), the number of cells N at generation *n* with *d* limiting deletions for a cell with one chromosome is given by the sum of the two equations above:

$$\mathrm{N}_n(d) = \mathrm{N}_n(2d) + \mathrm{N}_n(2d-1). \quad (2)$$

This is the same as the sum of the number of chromosomes at generation n in the even state and the odd state. Note that since we are only dealing with cells with one chromosome for now, the terms cell and chromosome may be used interchangeably.

Equation (2) takes into account telomeric deletions as a cell proliferates, but it does not include any factors that would prevent exponential cellular proliferation. Now Michaelis-Menten rate kinetics can be used to determine a relationship for the probability that a telomere will bind TRF2 with the underlying assumption that the energy state of a bound telomere continually increases (becomes more unfavorable) as the telomere shortens. Allowing f to represent the cell's concentration of free TRF2 dimers, m to represent the concentration of telomeric, TTAGGG repeats in the cell, C to represent the concentration of TRF2-telomere complex provides

$$f + m \underset{K_r}{\overset{K_f}{\Leftrightarrows}} C \tag{3}$$

where

$$K_f = k_1 e^{-((E_b - E_m)/k_B T)}$$
$$K_r = k_2 e^{-((E_b - E_C)/k_B T)} \tag{4}$$

and K_f is the forward binding rate, K_r is the reverse binding rate, k_1 and k_2 are rate constants for the given reactions in units of per time, k_B is Boltzmann's constant, T is temperature in Kelvin, E_m is the energy state of an unbound telomere, E_C is the energy state of the TRF2-telomere complex, and E_b is the activation energy between the two states. Arkus goes on to posit that:

$$E_m = \beta, \qquad E_C = \alpha L_m, \qquad L_m = \frac{rd}{Z_0}. \tag{5}$$

The first equation above indicates that the energy state of the unbound telomere, E_m, is set to equal some constant β. The next two equations state that the energy state of the TRF2-telomere complex, E_C, is equal to an energy constant α multiplied by the percent of telomere loss, L_m, which is equal to the rate of telomere loss *r* (in units of base pairs per deletion unit) multiplied by the number of limiting deletions *d,* all together divided by the initial telomeric repeat length, Z_0 in units of base pairs. Since the experimental values for unbound and TRF2 bound telomeres have yet to be recorded, the estimated Gibbs free energy for DNA-protein binding of ΔG ≈ -$22k_BT$ will be used (ΔG = $E_C - E_m$).

Keeping with the Michaelis-Menten rate kinetics and looking back at equation (3), the rate of formation of TRF2-telomere complex can be expressed as

$$\frac{dC}{dt} = K_f f m - K_r C \tag{6}$$

and since the concentration of free telomeric repeats and free TRF2 dimers are in a dynamic equilibrium, Arkus defined

$$M = m + C$$
$$F = f + C \tag{7}$$

where M is now the total concentration of telomeric repeats and F is the total concentration of TRF2 dimers. Rewriting equation (6) yields

$$\frac{dC}{dt} = K_f(F - C)(M - C) - K_r C \tag{8}$$

and finding the steady-state solution and solving for C provides

$$0 = K_f(F - C)(M - C) - K_r C$$
$$C = \frac{F+M+\frac{K_r}{K_f} \pm \sqrt{\left(-F-M-\frac{K_r}{K_f}\right)^2 - 4FM}}{2} \tag{9}$$

which is an expression for the total concentration of TRF2 bound telomeric repeats in the cell. Dividing Equation (9) by the total concentration of telomeric repeats, M, yields an equation for the proportion of bound telomeres:

$$\frac{C}{M} = \frac{F+M+\frac{K_r}{K_f} \pm \sqrt{\left(-F-M-\frac{K_r}{K_f}\right)^2 - 4FM}}{2M}. \tag{10}$$

Remember, M was set equal to the concentration of telomeric repeats m plus the concentration of bound TRF2-telomere complex C. However, it is important to bear in mind that in this model there is a loss of telomeres following every cell doubling (Fig. 2). Therefore, it would be more appropriate to say that $M = M_0(1\text{-}L_m)$, where M_0 is the initial total concentration of telomeric

repeats and as according to the final part of Equation (5), L_m corresponds to percent of telomere loss. Taking this new definition of *M* into account, and simplifying Equation (10) by using

$$\frac{K_r}{K_f} = K' = \frac{k_2}{k_1} e^{\frac{E_C - E_m}{k_B T}}$$

Gives us the expression for binding probability, B_d:

$$B_d = \frac{C}{M} = \frac{F+M+K' - \sqrt{\left(-F-M-K'\right)^2 - 4FM}}{2M}. \quad (11)$$

The binding probability may be considered as a fraction that decreases in value after each new generation due to telomere shortening, which then increases K_r. Recall that Equation (1) had given us a model for cell proliferation. Including the new binding probability term to the equations would convert them into a model of cell survival.

$$\begin{aligned} \mathrm{N}_n(2d) &= \mathrm{N}_{n-1}(2d)B_d + \mathrm{N}_{n-1}(2d-1)B_d \\ \mathrm{N}_n(2d-1) &= \mathrm{N}_{n-1}(2d-1)B_d + \mathrm{N}_{n-1}(2d-2)B_{d-1} \end{aligned} \quad (12)$$

As one can see, if $B_d = 1$, cell survival is 100%. Adding the above two equations together and rewriting it in a quasi-probabilistic form gives us the proportion of dividing cells at generation *n*, P_n.

$$P_n = \sum_{j=0}^{d} \frac{N'_n(j)}{2^n} \ . \quad (13)$$

The 2^n term in the denominator is just the value for the total number of cells that would have survived at generation *n* assuming that $B_d = 1$. With the key equation identified, Arkus was able to derive a polychromosomal model suitable to analyze human cells by simply applying a binomial expansion.

B. ANALYSIS

Arkus tested her model for cell viability against data from two human fibroblast cell lines—the same data used by the previous two models (16-18). A side by side comparison of the three models against the same data clearly indicates that the new model fits the best (18). What follows are three additional parameters that may be taken into account.

a. Oxidative Stress

In the presence of oxidative stress, such as organic hydroperoxide formed in lipids or superoxide anions potentially found in the mitochondria, human cells can lose up to three times as many base pairs per deletion *d (19)*. In relation to the model presented by Arkus, Equation (5) introduced the *r* value as the rate of telomere loss in units of base pairs per deletion unit and Arkus set this value at 54, the average value of base pairs lost per deletion for human cells. To account for oxidative stress in her model, Arkus at first simply altered the *r* value, and the higher it became, the sooner the cell line would go to extinction. However, next she noted that an organism is naturally exposed to more oxidative stress as it ages. This led Arkus to alter her *r* value as follows:

$$r = r_0 + \frac{O*n}{n_{max}} \qquad (13)$$

This new rate of telomere attrition *r* equals the original rate (54bp/PD) plus the attrition rate due to oxidative stress O multiplied by the fractional age of the cell n / n_{max} (18). Depending on the species and the conditions, the value of O can vary from 100 to 600 bp per deletion unit (19).

b. Over-expression of TRF2

Given the model proposed by Arkus, it may be assumed that an over-expression of TRF2 may lead to a near immortal cell line. Surprisingly, in vivo studies show that an excess of TRF2 increases the rate of telomere shortening while reducing the senescence set point (length of telomeres at senescence onset) (20). Arkus' model does not include a function for increasing the *r* value as a result of over-expression of TRF2 (18).

c. Telomerase

Arkus briefly discussed how the expression of telomerase in somatic cells, which occurs in 90% of cancer cell lines, may be accounted for by her model. In the first scenario, Arkus sets a maximum deletion *d* that when reached activates telomerase, stabilizing telomere length and maintaining the binding probability at a constant value. In the second scenario, Arkus explores the ability of telomerase to combat oxidative stress. Since telomerase can add on average between 390-420 nucleotides worth of telomeric repeats to a telomere, as long as the rate of attrition due to oxidative stress remains below 390 bp/PD, telomerase can negate the effects of such stress. In the third and final scenario, which is a modification of the first, Arkus suggests that the binding of telomerase to telomeres assures cell viability. Therefore, after a certain amount of deletions, telomerase is activated and the binding probability shoots up to 1—true immortality (18).

Part III. Cunning Cancers: Protozoan Parasites

Human malaria has reemerged as one of the world's most lethal infections, killing over one million people every year while infecting over 300 million others (21). Over 18 million individuals are debilitated by Chagas Disease in Latin America alone (22). The estimated prevalence of trypanosomiasis ranges between 50,000 and 70,000 and can be fatal if left untreated (23). Leishmaniasis is one of the most geographically widespread diseases, putting over 350 million people at risk of infection (24). Combined, hundreds of millions of people are infected with diseases caused by the parasitic protozoan *plasmodiums* and *trypanosomatids*.

On April 10th, 2009, a PubMed inquiry of the phrase "cancer telomerase" yielded 5,161 results, whereas "parasite telomerase" yielded only 15. Over the past decade research on telomeres has expanded exponentially and has led to exciting new discoveries that may potentially result in new treatments for cancer (13). However, in comparison, there is little discourse put forth by the scientific community of the potential of a telomere/telomerase approach for a new treatment for malaria, Chagas Disease, trypanosomiasis, or leishmaniasis—diseases which seldom concern Western audiences. Therefore, what follows is a literature review of the most recent findings on the telomere biology of *plasmodiums* and *trypanosomatids*.

A. *Plasmodium* Telomeres

Telomere shortening, without compensatory elongation by means of telomerase or an alternate pathway, would be a major problem for rapidly dividing cells, such as those of *Plasmodium falciparum* during the erythrocytic stage of their life history (21). In 1997, the first evidence for telomerase activity in *P. falciparum* was discovered (25). A novel characteristic of *P. falciparum* telomerase is that it is able to create new telomeres at broken chromosome ends

without the utilization of the proper telomeric template sequence. In fact, *P. falciparum* telomerase can use a variety of DNA sequences as a substrate for adding telomeric repeats. It was also found that the average length of the telomeres of *P. falciparum* remains constant during the highly proliferative erythrocytic cycle, indicating that telomerase activity must have increased in accordance to the increased rate of cellular division (25).

An appropriate question to ask would be: if an efficient inhibitor to *P. falciparum* telomerase is created, where in the *P. falciparum* life cycle would the parasite be most vulnerable? During schizogony, approximately 20,000 merozoites are released from a single hepatocyte. A single merozoite can invade an erythrocyte and undergo four to five mitotic divisions each two days before releasing 16 to 32 new merozoites into the bloodstream. This multiplying results in an average of billions of infected erythrocytes in patients with malaria (26). It is clear that there are an uncountable number of cellular divisions, and therefore, DNA replication cycles, during the life history of the parasites. If an efficient inhibitor of *P. falciparum* telomerase can be identified, then it may serve as an invaluable anti-malarial therapeutic. Currently, ddGTP and AZT-TP are the two reverse transcriptase inhibitors that are under investigation (25).

It came as a surprise to many when it was discovered that most parasite virulence factors and antigenic, cell-identity genes are located near the ends of the chromosomes (21,26). At first, one may be puzzled yet excited at this observation, after all, if the genes that make a parasite lethal and evasive to the host's immune system is located near the ends of chromosomes, then perhaps even a temporary knockout of telomerase could result in the loss of enough nucleotides on the chromosome ends to interfere with these important alleles. If these genes were located near the center of the chromosome, then it would take many more divisions without any telomere elongation to result in a phenotypic debilitation. The appropriate question to ask next is, well,

assuming that the Darwinian Theory still holds true, then what is the compensatory advantage for parasites to have their virulence and antigenic genes in subtelomeric regions? The answer is two-fold. First, via a mechanism called ectopic recombination, which occurs more frequently at chromosome ends, alleles can be shuffled around (undergo recombination with non-homologous chromosomes) and thus be mutated much more often. Mutations of antigenic genes are advantageous to the parasite in order to help it to continually evade the host's immune system (21). Second, chromosome ends are subject to the telomere position effect (TPE), which is the silencing of genes in subtelomeric regions by the extension of heterochromatin from the telomeres inward (27). Other mechanisms for gene silencing include the formation of secondary structure elements, such as t-loops, that could prevent transcription factors from arriving at its substrate via steric hindrance (Fig. 1.3). Gene silencing allows the parasite to select their preferred state while keeping the rest of their arsenal hidden from the host (21,26).

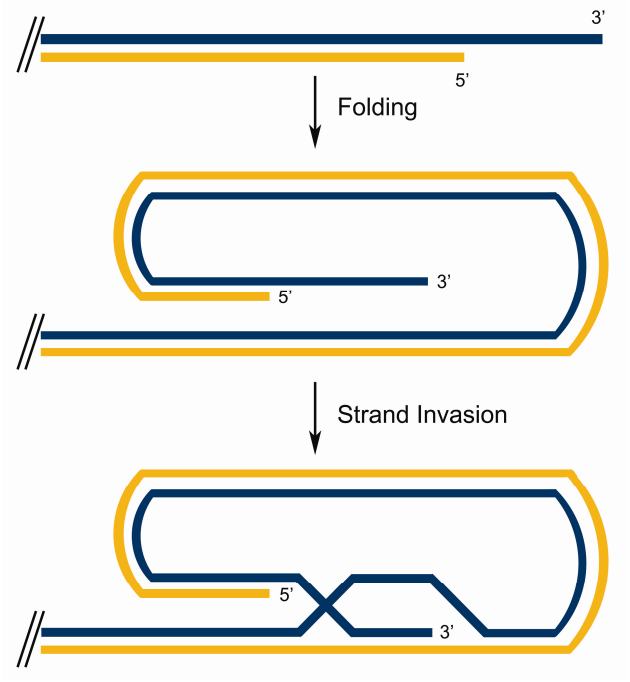

FIGURE 1.3. **Model of t-loop formation at telomeres.** T-loop structures have several functions. More commonly, they serve to protect telomeres from DNA damage and recognition as single-strand breaks by the cell. Their structure also regulates access to telomeres by proteins such as telomerase (28). For protozoan parasites, t-loops also play an important function in the silencing of antigenic and virulence genes (21).

P. falciparum does not utilize a protective glycoprotein coat. When the parasite hides within host erythrocytes, it places a number of proteins on the surface of the red blood cell. One of these proteins is the *P. falciparum* erythrocyte protein-1 (PfEMP-1), which is encoded by the *var* genes. The protein causes the erythrocyte to bind the endothelium of the blood vessel, allowing the parasite to multiply without interruption, such as potential elimination by passing through the spleen. Each *P. falciparum* parasite has up to 70 different *var* genes in its genome, and thanks to its location near the ends of chromosomes, transcriptional silencing maintains all but one of these genes in a silent state. This way, by the time the host immune recognizes one form of the protein, the parasite can switch to another and avoid the fatal immune response. This mechanism of being able to switch from one state to another at the level of transcription is formally called antigenic variation (21,23,26). *P. falciparum* serve as a model example of the benefits of linear chromosomes and the location of vital genes in subtelomeric loci.

B. *Trypanosome* Telomeres

Similarly to *Plasmodiums, Trypanosomes* have a complex life cycle which requires telomeres to be maintained at a viable length during parts of the life cycle with a high rate of cellular division and in non-proliferative yet infective stages as well. In addition, *Trypanosome* genes coding for virulence factors and variant antigens are also located near chromosome ends in subtelomeric regions (22). Upon investigating telomere length in *T. cruzi*, it was observed that at least two different strains of the parasite existed with a significant difference between their telomere lengths. Given that virulence of Chagas Disease also varies from case to case, one wonders if telomere length is somehow related to the clinical manifestations of the disease (22).

Trypanosoma brucei has a unique way of evading the host's immune system. The parasite is able to continuously switch which variant surface glycoprotein (VSG) it expresses, and only one is active at a time. There are approximately 1000 VSG genes but only 20 sites where they can be expressed. At any given time, 19 of these sites are silent. As expected, the VSG gene locus is located in subtelomeric positions (23). What is even more interesting is that during its life cycle, while *T. brucei* is in the midgut of the tsetse fly, the VSG genes are not transcribed and instead replaced by a procyclin surface glycoprotein which aids in its survival in the environment of the fly's intestine (29). The differential VSG gene expression system is activated during the parasite's final transformation into the infective metacyclic stage in the tsetse's foregut and salivary glands. Just as in the *Plasmodiums,* ectopic recombinations and transcriptional silencing are two of the underlying mechanisms in the differential VSG expression system in *T. brucei* (23). A third model has also been proposed. Unlike most other protein-coding genes, RNA polymerase I transcribes VSGs. In the nuclei of *T. brucei*, RNA polymerase I can be found in the nucleolus and in an extranucleolar structure as well where RNA transcription does not normally occur. It was observed that non-silenced VSG genes associate with this extranucleolar structure, or expression site body (ESB), while silent expression sites do not. Thus, the association of active expression sites with the ESB is another method of antigenic variation (30).

On the other hand, *Leishmania's* subtelomeric regions, unlike those of the other *trypanosomes* and *plasmodiums*, do not contain any virulence or antigenic genes; they are all found near the center of the chromosomes. Little is known as to why this is the case, and whether or not it is beneficial for the species. In order to begin deriving some answers, a study was conducted which followed the formations of telomeric clusters in *L. major* during its amastigote

(human stage) and promastigotes (insect stage) phases of its life cycle. It was observed that the clusters are more concentrated in the nucleus center in the amastigote phase. However, the functional importance of this telomeric reorganization has yet to be elucidated (24).

The discovery of a TRF2 homologue in *T. brucei*, *Trypanosoma cruzi,* and *Leishmania major* were breakthroughs in trypanosomatid telomere research. As a result, *T. brucei* is rapidly being considered as a model organism for telomere research (29). However, it is worth mentioning that in both human stem cells and human cancers, telomerase is not the only means of extending chromosomes; there are other pathways that have been collectively termed as alternative lengthening of telomeres (ALT) (Fig. 1.4) (31). Therefore, an important question to consider is if telomerase deficient *trypanosomes* infect a vertebrate host, would they lose their VSG genes rapidly due to the fact that they are located near the chromosome ends, or do they also have an alternate lengthening of telomeres pathway? Once again, the science of telomeres has drawn another fine line; this time between parasitic antigenic variance and the Achilles heel of the loci being near the end of their chromosomes. It remains to be seen if science can expose the latter and erase this line.

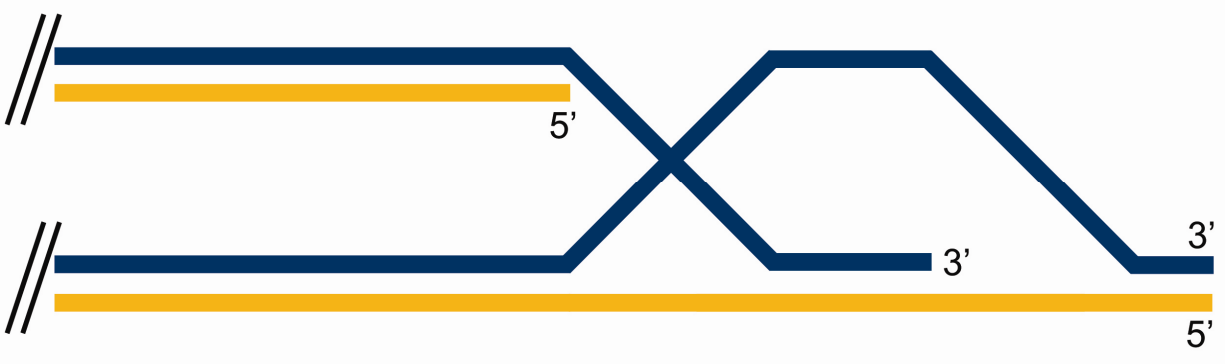

FIGURE 1.4. **Potential template for the lengthening of ALT cell telomeres via inter-telomeric recombination.** The 3’ overhang in one telomere invades the other in order to use it as a copy template to increase its telomeric DNA (32).

Part IV. Telomere Binding Proteins.

It would be incomplete to simply say that telomeres consist of tandem repeats of guanidine rich DNA which, in certain cells, is able to recruit the telomerase reverse transcriptase. In this regard **Figure 1.3** may be characterized as an oversimplification; in reality both double stranded and single stranded telomeric DNA are a Mecca for DNA binding proteins (Fig. 1.5). These proteins function to regulate the access of telomerase to the telomere, distinguish chromosome ends from double-stranded DNA breaks, and protect chromosomes from exonuclease degradation, homologous recombination and non-homologous end-joining pathways, thereby maintaining the cell's genomic integrity (33).

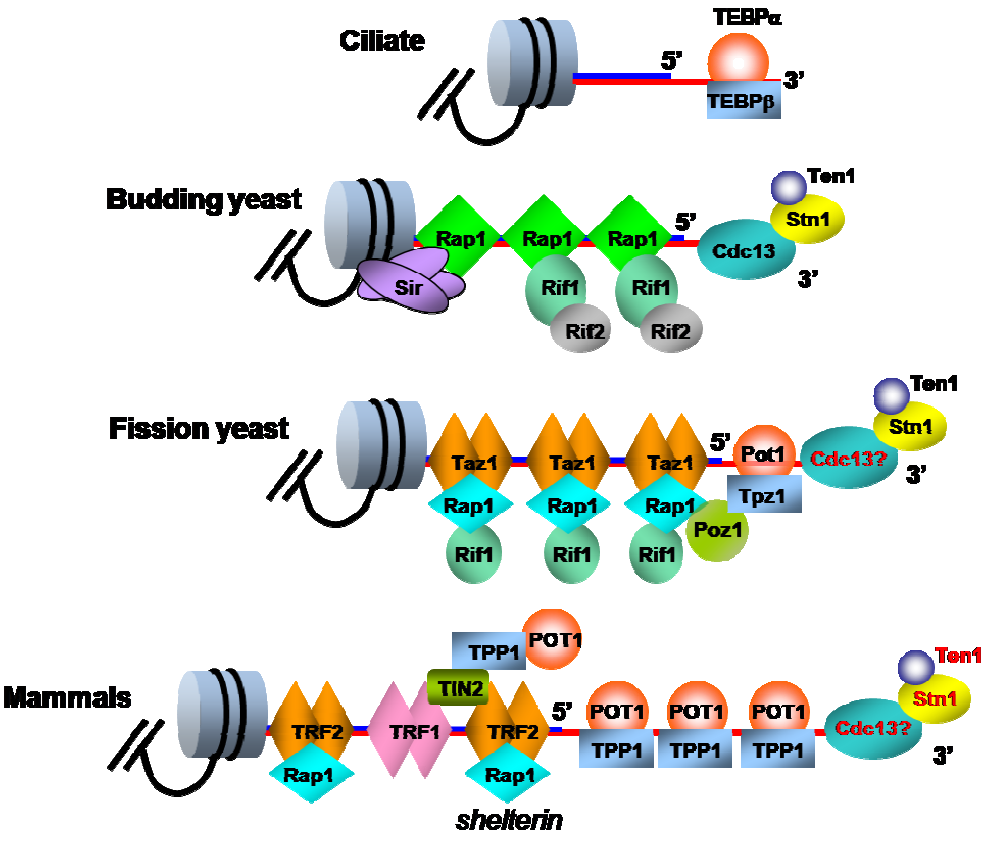

FIGURE 1.5. **Telomere ss and ds binding proteins across 4 species.** This thesis focuses in on mammalian Stn1 and Ten1.

TRF2 has already been introduced as the "cap" of telomere ends, protecting them from being recognized as breaks in the DNA and thus preventing cellular senescence and apoptosis (15). However, TRF2 does not act alone. As indicated in **Figure 1.5**, TRF2, along with TRF1, POT1, TPP1, TIN2, and Rap1, is part of a telomeric six-protein complex called shelterin that protects human telomeres and negatively regulates telomerase activity by sequestering its telomeric DNA substrate (34). Shelterin also mediates the formation of t-loop structures (Fig. 1.3) (28).

The focus of this thesis is on the single-stranded telomere binding proteins—more specifically—the proteins Stn1 and Ten1 which along with Cdc13 forms a ternary complex (CST) essential for telomere length regulation and protection (Figs. 1.5, 1.6) (35).

A. CST Complex: Regulation of Telomere Length

In terms of telomere length, Cdc13 plays a central role in switching telomere elongation on and off (35). In a Cdc13 mutant with a missense mutation, called *cdc13-2*, the phenotype is the same as mutants with telomerase deficiency (short telomeres). This observation suggests that one of the functions of Cdc13 is to recruit telomerase to telomeres (36). Cdc13 is also able to negatively regulate telomere replication. A Cdc13 allele, called *cdc13-5,* confers a phenotype similar to mutants with up-regulated telomerase activity—extensively long 3' overhangs. Furthermore, the addition of Stn1 can suppress the *cdc13-5* mutation and halt the extension of telomeres. This observation, along with the fact that the *cdc13-2* mutation also abolished the Cdc13-Stn1 interaction, suggests that Cdc13 is able to initiate and then suspend the synthesis of telomeres (37). The mechanism by which Cdc13 coordinates both of these functions is beginning to be revealed. Cdk1-dependent phosphorylation of Cdc13 during the late S and early G2 phase

of the cell cycle increases its binding affinity with telomerase subunit protein Est1 as opposed to the Stn1-Ten1 complex, which interacts at the same domain via the C-terminal of Stn1 (the N-terminal of Stn1 interacts with Ten1, Fig 1.6). Therefore, it is a competition between Est1 and Stn1 for binding to Cdc13 that determines the fate of the telomere's length (38).

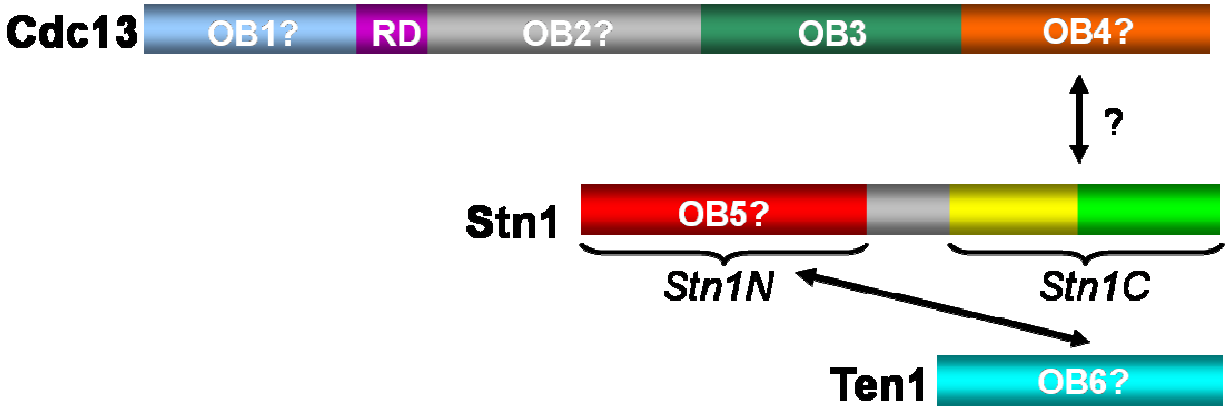

FIGURE 1.6. **Domain structure of the CST complex.**

B. CST Complex: Telomere End Protection

A second essential function of the CST complex is to protect chromosome ends from degradation (35). Mutant cells lacking functional Cdc13 are exposed to extensive resection of the 5' Cytosine rich strand (39). A similar phenotype is observed in cells with Stn1 or Ten1 mutants as well (40). In fact, cells with the lethal *cdc13*-Δ null strain can be rescued if Stn1 protein is delivered to the telomere, which hints that Cdc13's role in telomere end protection may be to deliver Stn1 to the ends of chromosomes (35). Co-overexpressing Ten1 along with Stn1 can also suppress several Cdc13 mutations, capping and protecting telomeres in a Cdc13-independent manner (41).

As their physiological role and relation to telomerase activity began to uncover, there has been an increased interest in further understanding the CST complex for the purpose of developing new therapeutic strategies for cancer treatments. However, little biochemical and structural characterizations have been made of the CST complex, and none at all for humans. In fact, when work on this thesis began, for no species has the structure of the Stn1-Ten1 complex been determined. The purpose of this thesis is to begin uncovering information on the human Stn1-Ten1 complex.

Chapter 2: Materials and Methods

Beginning with human cDNA, many steps must be taken in order to obtain crystals of the human Stn1-Ten1 complex suitable for X-ray diffraction analysis. First, we employed an *E. coli* protein expression system and purified the complex utilizing several different techniques. Then, the purified protein underwent screening for crystallization conditions, several of which were optimized by handmade crystal plates. Once all optimization parameters have been exhausted, the crystals of highest quality (large, well-defined structure) were harvested for storage and data collection at the Advanced Photon Source. X-ray diffraction analysis can potentially lead to the determination of the structure of the protein complex.

This chapter focuses on the general materials and methods utilized in our laboratory to obtain protein crystals when beginning with cDNA. The next chapter will focus specifically on the methods and results utilized for obtaining human Stn1-Ten1 crystals. The entire process may be divided into 4 parts: plasmid formation, protein growth, protein purification, and crystal growth and structural determination.

Part I: Plasmid Formation

A. POLYMERASE CHAIN REACTION (PCR)

Polymerase Chain Reaction (PCR) is used to amplify the target portion of template cDNA that codes for our protein of interest. This reaction utilizes the following reagents (the DNA polymerase enzyme is added last): 0.5 µl of AccuPrime™ *Pfx* DNA Polymerase at 2.5 U/µl (Invitrogen), 5 µl 10X AccuPrime™ *Pfx* Reaction Mix (Invitrogen), 1 µl of the forward primer (1 µM), 1 µl of the reverse primer (1 µM), 1 µl of the template (cDNA), and 41.5 µl of distilled water to arrive at a final volume of 50 µl. Each unit (U) of the AccuPrime™ *Pfx* DNA

Polymerase is defined as incorporating 10 nmol of deoxyribonucleotide into acid-insoluble material in 30 min at 74°C. The AccuPrime™ *Pfx* Reaction Mix consists of 1 mM $MgSO_4$, 0.5 mM Tris-HCl (pH 8.0), 0.5 mM KCl, 0.3 mM dNTPs, 10 µM DTT (a reducing agent that prevents DNA dimerization and unnatural intra or intermolecular disulfide bonds forming between cysteine residues in a protein), 1µM EDTA (metal ion scavenger), 0.5% glycerol, and trace amounts of stabilizers.

The PCR reaction is run according to the following programmed protocol with a constant lid temperature of 105°C throughout. The procedure begins with an initial DNA denaturation step at 95°C for 2 minutes. This is followed by 30 cycles of denaturation at 95°C for 15 seconds, annealing at 57°C for 30 seconds, and extension at 68°C for 1 minute per kb of desired PCR product. After 30 cycles, there is a final extension period at 68°C for 10 minutes. Unless noted otherwise, all primers were designed in lab and synthesized at Invitrogen with an artificial BamH I restriction site added to the 5' end of the forward primers and an artificial Xho I restriction site added to the 3' end of the reverse primers.

B. AGAROSE GEL ELECTROPHORESIS OF PCR PRODUCT

Once the PCR is complete, we test approximately 5-10% of each PCR reaction for amplification of the correct band on the basis of the DNA fragment length by examining the PCR product using agarose gel electrophoresis and a 1% agarose gel (0.5 g agarose in 50 mL 1X TBE Buffer). The samples are run at 140 V for 24 minutes and visualized using ethidium bromide in UV light.

C. PCR PURIFICATION

Once electrophoresis analysis verifies that our amplified DNA has approximately the same number of base pairs as our target sequence, we then purify the PCR products using the QIAGEN PCR Purification Kit in order to remove excess primers and other components such as nucleotides, enzymes, salts, mineral oil, and other impurities. The PCR product is loaded onto a QIAquick Spin Column which contains a silicon membrane assembly that only binds DNA in a high-salt buffer, thereby allowing the other components to be washed through. Elution with 50 µl of a low-salt buffer after a sufficient wash cycle completes the PCR purification and prepares the target DNA for the digestion of their artificially flanked restriction endonuclease sites.

D. DIGESTION OF DNA

The purified DNA fragment is then digested with the restriction endonucleases BamH I and Xho I (New England Biolabs) in order to cut the DNA at their artificially flanked restriction sites to create "sticky ends" that can bind to a vector treated with the same restriction endonucleases. 15 µl of PCR purified DNA is mixed with 5 µl of Buffer 3 (New England Biolabs), 0.5 µl Bovine Serum Albumin (New England Biolabs, inactivates contaminating nucleases and proteases in order to stabilize the restriction endonucleases), 1.5 µl of BamH I (which cuts the following sequence at the slash, 5' G/GATCC 3'), 1.5 µl of Xho I (5' C/TCGAG 3'), and 26.5 µl of distilled water to bring the total reaction volume to 50 µl. The mixture is incubated at 37°C for approximately 2-3 hours.

E. GEL EXTRACTION

DNA digestion may often result in additional fragments of DNA aside from our target sequence. Gel extraction selects for and purifies our target, digested DNA sample. All 50 µl of

the DNA digestion product is loaded onto another 1% agarose gel and run at 120 V for 20 minutes and then at 140 V for an additional 5 minutes. The sample is visualized using ethidium bromide over a short wave UV source. The band for the target DNA digestion product is excised using a razor from the agarose gel. This gel fragment is dissolved and the DNA within it is purified and recovered using a QIAGEN Gel Extraction Kit which applies the same principle as the QIAGEN PCR Purification Kit, this time washing out excess agarose and ethidium bromide amongst the other impurities. Finally, the DNA is eluted using a low-salt buffer to a final volume of 35 µl.

F. LIGATION INTO VECTOR

In order to over-express the protein, its DNA sequence must first be introduced, or transformed, into the *E. coli* protein expression system (Novagen). This transformation only occurs when the DNA is ligated into a vector, which serves as the vehicle for the transfer of foreign genetic material into the *E. coli* cells. In the course of the experiments presented in this thesis, 2 vectors were used. The first was a modified pET28b vector which codes for Sumo protein with a His_6 tag that is fused to the N-terminus of the protein we wish to express and purify (Fig. 2.1). The second was a glutathione S-transferase (GST) fusion protein expression vector, pGEX6p-1 (GE healthcare, Fig 2.2). The vectors are prepared for ligation by first digesting each with the same restriction endonucleases used to digest the ends of the DNA fragments and then purified by gel extraction.

The target DNA sequence is ligated into the desired vector using a reaction mixture consisting of 1 µl of the vector mixed with 6 µl of the insert DNA, 2 µl of T4 DNA Ligase Reaction Buffer (10x) (New England Biolabs), 0.5 µl of T4 DNA Ligase (400,000 cohesive end units/ml) (New England Biolabs), and 10.5 µl distilled water to reach a final reaction volume of

20 µl. 1X T4 DNA Ligase Reaction Buffer contains 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM Dithiothreitol and is at a pH of 7.5. For T4 DNA Ligase, a unit is defined as the amount of enzyme required to give 50% ligation of Hind III fragments of λ DNA in a total reaction volume of 20 µl in 30 minutes at 16°C in 1X T4 DNA Ligase Reaction Buffer. The mixture is incubated at 16°C for 2 hours. A negative control is also prepared by creating a mixture with an equivalent amount of water (6 µl) replacing the insert DNA.

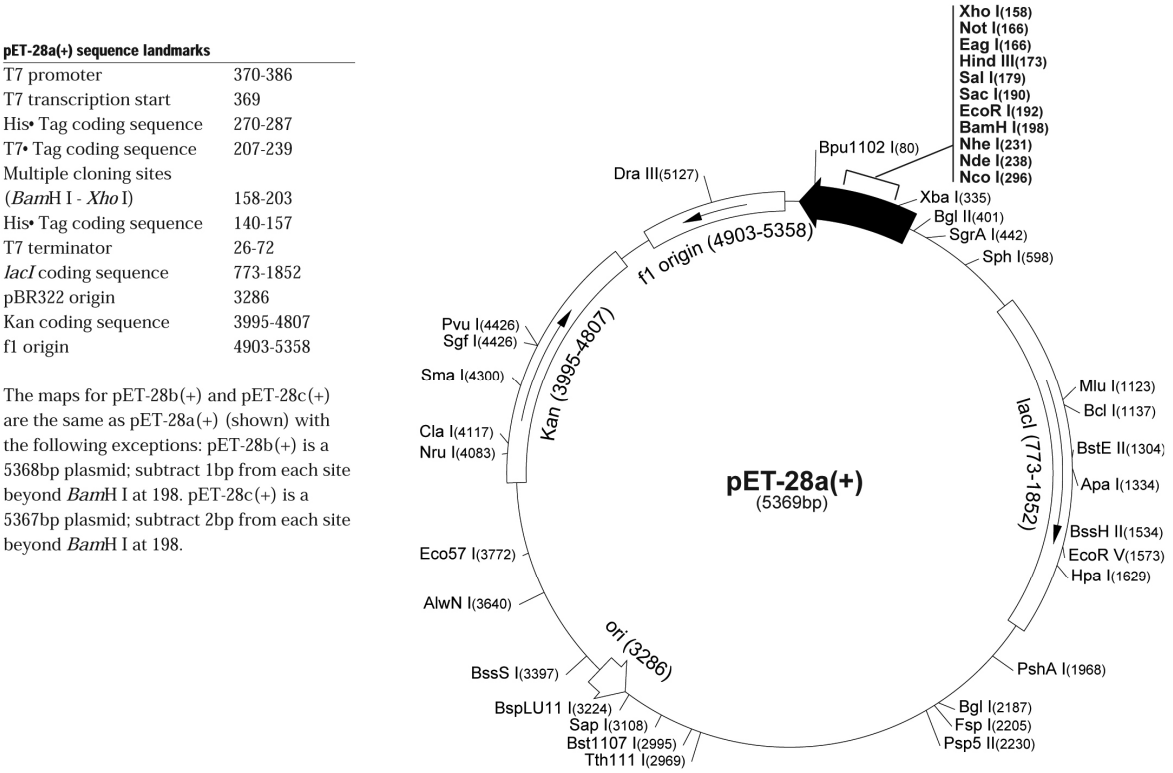

FIGURE 2.1. **pET-28a vector map.** Note the presence of a kanamycin resistance gene. The vector used in our experiments was pET-28b, which differs from pER-28a by the subtraction of 1 bp from each site beyond BamH I at bp 198 (42).

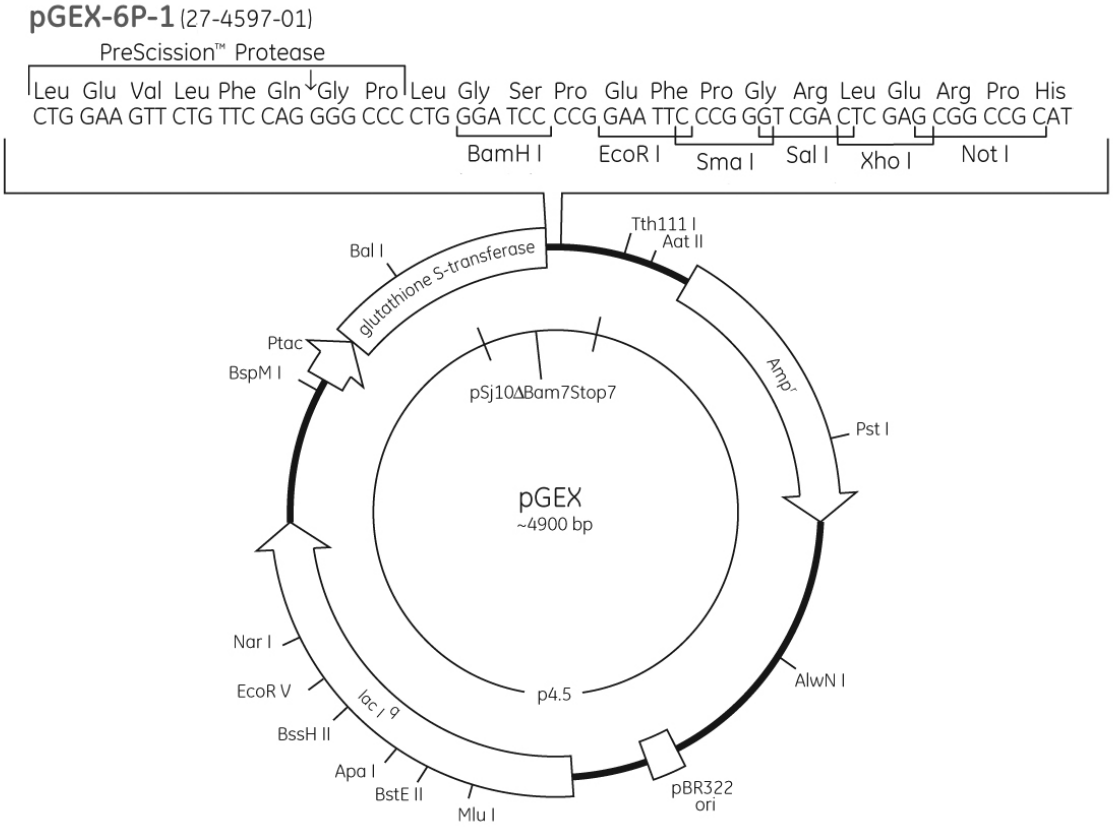

FIGURE 2.2. **pGEX-6P-1 vector map.** Note the presence of an ampicilin resistance gene (43).

G. TRANSFORMATION INTO DH5α COMPETENT CELLS

Before we can over-express the proteins, we must amplify the plasmid created from ligation and verify that the construct is correct. This is done by inserting the plasmid into DH5α competent cells, which are ideal for high efficiency transformations and increasing plasmid yield and quality. 100 µl of competent cells are added to 10 µl of the ligation product and the mixture is kept on ice for 30 minutes. Next, the mixture is heat shocked for one minute at 42°C. After the heat shock, 500 µl of LB medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 170 mM NaCl) is added to the mixture which is then shaken for 1 hour at 37°C. After shaking, the mixture is centrifuged at 3000 rpm for 1 minute. The pellet is resuspended in 200 µl of the supernatant and plated on LB plates (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) agar, 170 mM NaCl)

containing either ampicilin, kanamycin, or at times both since the pET28b plasmid confers kanamycin resistance (Fig. 2.1) and the pGEX6p-1 plasmid confers ampicilin resistance (Fig 2.2). The plates are then incubated at 37°C overnight and analyzed the next day with respect to the control. Asides from the presence of colonies, one indication that the ligation and transformation were successful would be the presence of a greater number of colonies on the ligation plates as compared to the control plates. A Mini-Prep test would provide further confirmation of correct construct formation.

H. MINI-PREP TEST FOR CORRECT CONSTRUCT FORMATION

The purpose of the DH5α transformation was to amplify the plasmid with high fidelity. The next step is to transfer the plasmid to *E. coli* BL21(DE3) (Novagen) competent cells which are specialized for growing and expressing protein. In order to do this transformation, we must first extract the plasmid from the DH5α competent cells and purify it from other components. Both of these steps can be accomplished by performing a Mini-Prep.

3 colonies from the LB plate are inoculated into 2 ml LB with 100 μM of the antibiotic present on the plate. The cells are shaken overnight at 37°C. The next day, the cells are centrifuged for 5 minutes at 4000 rpm and the supernatant is discarded. The pellet is treated with a QIAGEN Mini Prep Kit; the DH5α bacterial cultures are lysed and the lysate is applied to a QIAquick Spin Column where the plasmid adheres to the silicon membrane as the other components are washed away. Finally, the purified plasmid is eluted to a 50 μl final volume.

A small amount of the plasmid, or Mini-Prep product, can be digested with the same restriction endonucleases used as before in order to determine if the recovered plasmid is the correct one. 4 μl of the plasmid from the Mini-Prep is digested with 0.1 μl of each restriction endonuclease, 1 μl Buffer 3 (10X), 0.1 μl BSA (100X), and 4.7 μl distilled water for a final

reaction volume of 10 µl. The mixtures are incubated for 2 hours at 37°C and then examined using agarose gel electrophoresis and a 1% agarose gel. After running the samples for 20 minutes at 120 V and then for 5 minutes at 140 V, the gel can be visualized using ethidium bromide and UV light. When compared to a DNA ladder, a correct ligation would display 2 bands, one at the length of the plasmid (approximately 5300 bp for pET28b and 4900 bp for pGEX6p-1) and the other corresponding to the correct size of the target DNA insert. If all three colonies indicate correct construct formation, then the colony corresponding to the most intense band in the agarose gel is selected for transformation into BL21(DE3) competent cells.

Part II: Protein Growth and Expression

I. TRANSFORMATION INTO BL21(DE3) COMPETENT CELLS

After confirmation of correct plasmid production and isolation from the Mini-Prep, the plasmid is transformed into BL21(DE3) competent cells in order to grow and express the protein. 1 µl of plasmid from Mini-Prep is added to 25 µl of BL21(DE3) competent cells and kept on ice for 30 minutes. Next the cells are heat shocked for 1 minute at 42°C, and afterwards 600 µl of LB medium is added to the cells and the mixture is shaken for 1 hour at 37°C. After shaking, the mixture is centrifuged at 3000 rpm for 1 minute. The pellet is resuspended in 200 µl of supernatant and plated on an LB plate containing the appropriate antibiotic(s). The plate is then incubated at 37°C overnight.

J. TEST FOR PROTEIN EXPRESSION

Before continuing with mass protein growth and purification, we must be sure that the cells expressed the correct protein. Three colonies of BL21(DE3) cells are transferred from the LB plate and inoculated into 2 mL of LB medium with the appropriate antibiotic(s) (100 µM

each). The mixtures are shaken at 37°C for approximately 4 hours to allow for cell growth. Once the mixtures become cloudy, they are removed from the shaker. 1 mL of each sample is taken out and stored into a new 1.5 mL tube to serve as a non-IPTG-induced control for the protein expression test gel. To the remaining samples (approximately 1 mL left in each) protein expression is induced with 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) and 2 hours of shaking at 37°C. After shaking, 200 µl of the cells are centrifuged for 30 seconds at 8,000 rpm. 400 µl of the non-IPTG-induced cells saved earlier are also centrifuged for 30 seconds at 8,000 rpm. The supernatant is discarded and the cell pellets are resuspended in 30 µl of SDS-PAGE Loading Buffer. The mixtures are incubated for 5 minutes at 95°C and then centrifuged for another 5 minutes at 13,200 rpm. 10 µl of the supernatant is loaded and run on a 15% SDS-PAGE gel at 230 V for 42 minutes. In order to visualize the protein bands, the gel is stained with coomassie blue for 5 minutes and then de-stained with de-staining buffer (20% methanol, 10% acetic acid).

The SDS-PAGE separates proteins to distinct bands on the basis of their molecular weight (in kDa), where every 9 amino acids is approximately 1 kDa. The protein bands of the IPTG-induced samples are compared to those of the non-IPTG-induced samples. A correct and efficient expression is indicated by observing a more intense band corresponding to the molecular weight of the target protein in the IPTG-induced sample as compared to the non-IPTG-induced control.

K. MAKING GLYCEROL STOCK

If the protein expression test shows that the desired protein responds to IPTG induction, then a long-term storage glycerol stock is made with the sample that had the most protein (i.e. the lane with the most intense band). 600 µl of the non-IPTG-induced control that was set aside

earlier is added to 2 mL of LB medium with the appropriate antibiotic(s) (100 μM each). The mixtures are shaken at 37°C for 1 hour. After shaking, 1 mL of the mixture of growing cells is combined with 1 mL of 50% glycercol (25% final concentration). The stock solution is flash-frozen with liquid nitrogen and stored at -80°C to allow for future growth of cells with the correct plasmid(s) of the target protein(s).

L. GROWTH AND ISOLATION OF BL21(DE3) CELLS

From the glycerol stock, large amounts of cells can be grown in order to harvest a significant amount of the desired protein. A pipette tip's worth of glycerol stock is added to a flask containing approximately 20 mL of LB medium with the appropriate antibiotic(s) (100 μM each). The mixture is shaken at 37°C overnight. The next morning, the 20 mL of growing cells is transferred to a large flask containing 1 L of LB medium with the appropriate antibiotic(s) (100 μM each). The new 1 L mixture is shaken at 37°C and the concentration of cells is monitored by its optical density (OD) until the value reaches 0.6, at which point protein expression is induced with IPTG (0.1 mM) at room temperature (23°C) for 16 hours in the shaker. After shaking, the 1 L mixture is centrifuged for 15 minutes at 4,000 rpm. The supernatant is discarded and the cell pellet is saved at -20°C, ready for protein purification.

Part III: Protein Purification

L. LYSIS OF BL21(DE3) CELLS BY SONIFICATION

The cell pellet is resuspended in 50 mL of lysis buffer (50 mM Tris pH 8.0, 1 mM PMSF, 2.3 μM leupeptin, 1.5 μM pepstatin, 5 mM benzamidine, 10% glycerol, 400 mM NaCl, 50 mM NaH_2PO_4, 0.1 mg/ml lysozyme, 50 μl DNAse I (5 mg/ml), 500 μl $MgCl_2$ (1M)). The mixture is left on ice for 10 minutes. Next, the mixture is sonified on ice for 3, 45-second pulses with a 1

minute break between each for a total of 3 sonifications. A 3 µl sample of the sonified mixture is mixed with 5 µl of loading buffer for later analysis by SDS-PAGE (lane 1).

K. PURIFICATION OF TARGET PROTEIN BY AFFINITY COLUMN

After sonification, the mixture is centrifuged at 38,000 rpm for 1 hour. At this speed, the protein remains in the supernatant while other cellular components reside in the pellet. A 3 µl sample of this supernatant is mixed with 5 µl of loading buffer for later analysis by SDS-PAGE (lane 2). 2 ml of Ni-NTA agarose beads (QIAGEN) are prepared by transferring them into a column and allowing the beads to equilibrate by passing through them 20 ml of water and 20 ml of lysis buffer. Nickel beads are used for their ability to chelate the His_6 residues of Sumo protein. Therefore, proteins expressed from the pET28b vector with the fused His_6-tag will preferentially bind to the nickel beads while most other native bacterial proteins flow through the column.

The supernatant is mixed with the equilibrated nickel beads and tumbled vertically for 3 hours at 4°C to allow for the target protein to bind to the beads. Following this the mixture is added back to the column and the flow-through is collected. A 3 µl sample of the flow-through is mixed with 5 µl of loading buffer for later analysis by SDS-PAGE (lane 3). Next, the beads are washed with 50 ml of lysis buffer. Once again, a 3 µl sample of this first wash is mixed with 5 µl of loading buffer for later analysis by SDS-PAGE (lane 4). A second wash is conducted with a solution of 30 ml lysis buffer mixed with 30 mM imidazole, and 3 µl of the flow-through is collected and mixed with 5 µl of loading buffer for later analysis by SDS-PAGE (lane 5). Next, the protein is eluted to 1 ml fractions with a solution of lysis buffer containing 250 mM imidazole (imidazole competes with histidine for binding nickel). 1 ml fractions are collected until a Bradford assay screen indicates that protein is no longer being eluted from the column. A 3 µl sample of the second and fourth elutions are collected and each is mixed with 5 µl loading

buffer for analysis by SDS-PAGE (lanes 6 and 7, respectively). The 7 saved samples are denatured at 95°C for 5 minutes and spun down by a quick centrifugation. They are then run on an SDS-PAGE gel in order to access the quality of the purification. 6 µl of the sonification sample is loaded into lane 1, 6 µl of the supernatant is loaded into lane 2, 6 µl of the flow-through is loaded into lane 3, all 8 µl of wash 1 is loaded into lane 4, all 8 µl of wash 2 is loaded into lane 5, all 8 µl of elution 2 is loaded into lane 6, and all 8 µl of elution 4 is loaded into lane 7.

If the protein to be purified was expressed from a pGEX6p-1 plasmid, then the GST-tagged protein would have preferential binding to glutathione sepharose beads (GE Healthcare) instead of Ni-NTA agarose beads.

L. INITIAL PURIFICATION BY SIZE EXCLUSION COLUMN

In order to purify the target protein from some of the larger proteins that were also collected in the elutions, the fractions are passed through a Hiload Superdex200 column (GE Healthcare). Size exclusion columns separate proteins on the basis of their size and volume, with larger proteins eluting out first. The elution fractions are injected into an AKTA FPLC system (Amersham Bioscience) and eluted with a 150 mM NaCL, 25 mM Tris pH 8.0 buffer passing through the column. The protein is monitored as it passes through the column using UV_{280} light. The collected fractions suspected for containing the target protein is analyzed by SDS-PAGE for purity.

M. CLEAVAGE OF His_6-Sumo TAG

In the process of purifying protein, the synthetic protein tag must eventually be removed. The His_6-Sumo tag can be cleaved efficiently and specifically with ULP1 protease. First, the

concentration of protein is determined by Bradford Assay and a 50 to 1 protein to ULP1 ratio is used to achieve the cleavage.

The ULP1 protease is added to a 30 ml solution of protein (in FPLC buffer) and is incubated at 4°C overnight. The completeness and efficiency of the digestion can be confirmed by SDS-PAGE the next day. In order to separate the cleaved protein tag from the target protein, the solution can be passed through nickel beads, where the flow through should contain little Sumo protein. 5 mM DTT is added to prevent intermolecular and intramolecular disulfide bond formation (DTT cannot be added before the use of nickel beads).

In the case of GST-tagged proteins, excess protease 3C is used to cleave the fusion protein from the target protein in a similar protocol as described above.

N. FURTHER PURIFACTION BY SIZE EXCLUSION COLUMN

At this stage, most of the impurities will have a molecular weight in the range of the target protein (such as Sumo or GST). Passing the protein through a Hiload Superdex75 column (GE Healthcare) can further purify the protein from smaller impurities. The protein must be concentrated to less than 10 mL using a Millipore Amicon Concentrator with an exclusion limit of 10 kDa centrifuged at 3,800 rpm. After concentration, the solution is injected into an AKTA FPLC system and eluted with a 150 mM NaCL, 25 mM Tris pH 8.0 buffer passing through the column. The protein is monitored as it passed through the column using UV_{280} light. The collected fractions suspected for containing the target protein are analyzed by SDS-PAGE for purity.

O. PURIFICATION BY ANION EXCHANGE COLUMN

Sometimes impurities will have the same size and/or volume as the target protein, therefore debilitating size exclusion column's separation capability. When this is the case, a MonoQ column (GE Healthcare) can be used to separate proteins on the basis of a new principle, molecular charge. The protein must be concentrated to less than 10 mL using a Millipore Amicon Concentrator with an exclusion limit of 10 kDa centrifuged at 3,800 rpm. The buffer must also be changed to 50 mM NaCl, 25 mM Tris pH 8.0. After concentration and buffer switch, the solution is injected into an AKTA FPLC system and eluted through the MonoQ column. The initial buffer is a low salt, 25 mM NaCl, 25 mM Tris pH 8.0 buffer. The buffer gradually changes through a gradient to a high salt 1 M NaCl, 25 mM Tris pH.8.0 buffer in order to slowly elute the protein, that is, if it was able to bind to the column. The protein is monitored as it passed through the column using UV_{280} light. The collected fractions suspected for containing the target protein are analyzed by SDS-PAGE for purity.

Once SDS-PAGE analysis indicates that the protein is at least 95% pure, it can be concentrated to 25 μg/μl and then aliquotted into smaller tubes, 25 μl of concentrated protein in each, and flash frozen by liquid nitrogen to be stored at -80°C.

Part IV: Crystal Growth and Structural Determination

P. CRYSTALLIZATION

Crystallization conditions are initially screened by robot at the Life Science Institute at the University of Michigan using home-made crystallization kits as well as those sold by Hampton Research. This initial screen consists of 480 conditions in a sitting drop setup. The

conditions which yield crystals are then optimized by hand by altering the precipitant, buffer, and/or salt concentrations in a hanging drop setup. Once the best condition are determined, an additive screen may be prepared using 96 various additive compounds and molecules (Hampton Research). If this screen yields new conditions with better crystals, further optimization would be performed by testing varying concentrations of the appropriate additive.

Q. X-RAY DIFFRACTION

Protein crystals elected for X-ray Diffraction studies are prepared by transferring the crystal into a harvesting solution which includes a large concentration of a stable cyroprotectant and then flash freezing the crystals in liquid nitrogen. To improve the packing of the crystal, an overnight dehydration protocol may be used which consists of transferring the hanging drop to a well solution containing less water than before. X-ray Diffraction studies are performed using the X-ray beam Jeckle at the Life Sciences Institute at the University of Michigan as well as the Advanced Photon Source (Beamline 21-ID-G) at Argonne National Laboratories. The diffraction pattern is analyzed using the HKL2000 software (HKL Research, Inc.).

The dimensions of the unit cell can be determined from the diffraction spots. This is independent of the number of molecules in the unit cell. By applying Bragg's Law, the crystallographic resolution, or diffraction limit, can be determined based upon how far from the center diffraction spots can be found. Crystal symmetry, or the space group, can be determined from the systematic absences in the diffraction pattern. The ultimate goal of X-ray crystallography analysis is to create a 3D electron density map specific to our protein so that, given its amino acid sequence, the residues can be "threaded" through the map. However, diffraction data from the native crystal can only calculate the "magnitude" of the electron density map, but not where it is in the unit cell (the 3D aspect). This is commonly known as the phase

problem. If the native crystal diffracts to a good resolution, such as less than 3 angstroms, then by applying multiple isomorphous replacement (for instance, growing our protein in selenium-methionine) the phase problem can be solved because the resulting isomorphous crystal produces a slightly different diffraction pattern. From this, a 3D electron density map can be obtained and the structure of the protein can be revealed (44).

Chapter 3: Determining the Structure of the Human Stn1-Ten1 Complex

Previously, Ke Wan of Dr. Ming Lei's laboratory attempted to purify full-length human Stn1 protein (368 amino acids). However, the protein was not soluble and therefore purification was not possible. By examining the secondary structure prediction of hStn1, and taking into account that hStn1 homologs in other species bind to Ten1 (122 amino acids) via their N-terminus, Ke Wan decided to create a new hStn1 construct consisting of the amino acids 2-185 (Fig. 3.1). However, this construct expressed even worse solubility during the purification steps.

Our first goal was to solve this solubility problem for hStn1 and successfully purify the hStn1-hTen1 complex. We suggested three plans for tackling this problem:

Plan A: Develop a co-purification protocol for the hStn1 and hTen1 proteins with a His_6-Sumo tag fused to each.

Plan B: Develop a co-expression test for a GST tagged hStn1 and a His_6-Sumo tagged hTen1.

Plan C: Develop a co-expression test for a His_6-Sumo tagged hStn1 and a GST tagged hTen1.

In order to assess all three plans, the DNA sequences for hStn1 and hTen1 were inserted into both pET28b and pGEX6p-1 vectors. The previous hStn1 construct of amino acids 2-185 was trimmed by 5 amino acids for the following experiments. Therefore, for the rest of this chapter, hStn1 will refer to human Stn1 amino acids 2-180 (the N-terminal region). The methods used to obtain the following results are the same as those presented in Chapter 2. Changes to the protocol will be noted below.

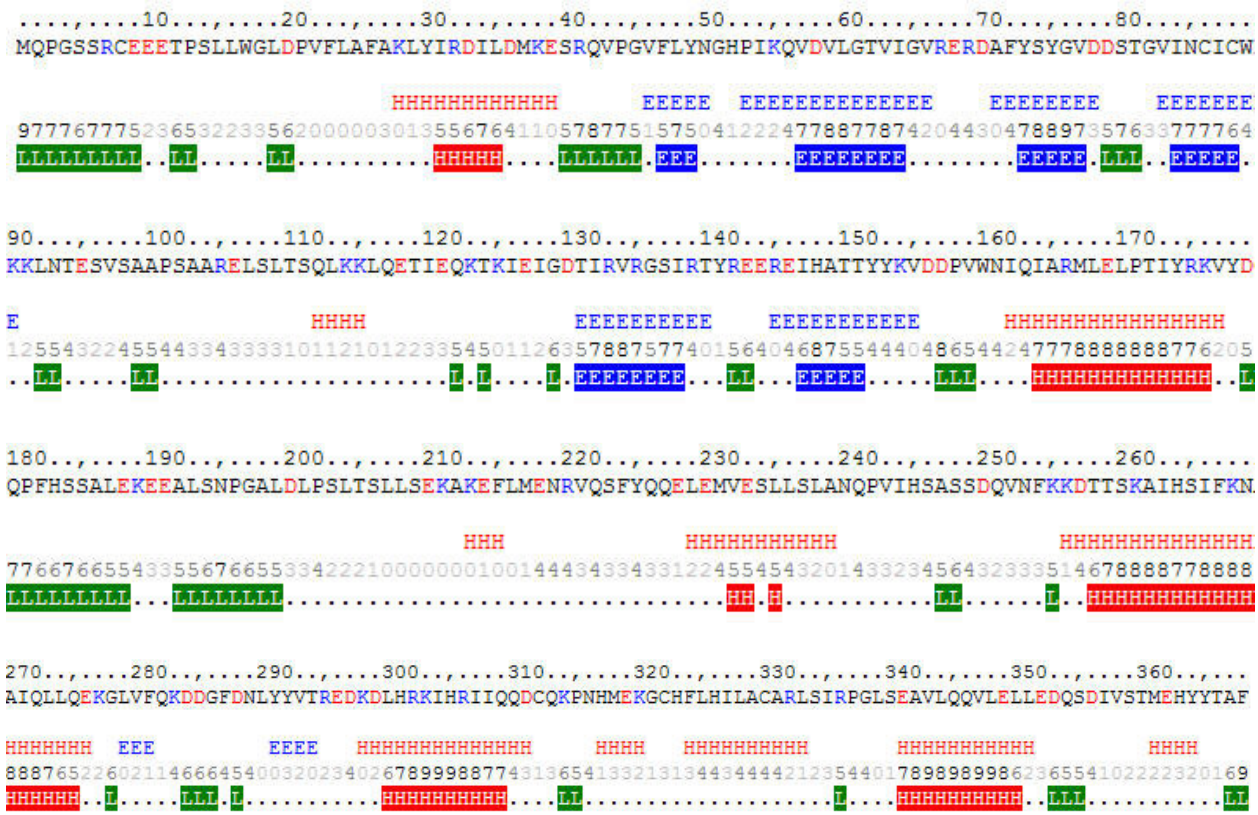

FIGURE 3.1. **Secondary structure prediction for human Stn1.** "H" represents a helix, "E" represents a beta strand, and "L" represents a lack of either secondary structures. The number between the letter designations represents the confidence level, with 9 being the highest. Basic, positively charged amino acids are in blue and acidic, negatively charged amino acids are in red.

Part I: Plasmid Formation

A. POLYMERASE CHAIN REACTION

The DNA template for hTen1 was provided by a collaborating lab group. cDNA for hStn1 was purchased from OriGene (Ref. ID: NM_024928). The complete nucleotide sequence is presented in **Figure 3.2**. Due to the presence of two distinct BamH I cutting sites within the sequence coding for amino acids 2-180, a forward primer had to be designed flanked with a new restriction endonucleases recognition site, in this case Bgl II (5' A/GATCT 3'). We were able to maintain the Xho I site for the reverse primer, and both BamH I and Xho I were used for the forward and reverse primers of hTen1, respectively.

PCR was used to amplify hStn1 and hTen1 separately using the cDNA templates and primers mentioned above. The reaction products were run on an agarose gel (not pictured). When compared to the DNA ladder, the hStn1 lane resided slightly over 500 bp (hStn1 amino acids 2-180 contains 537 bp) and the hTen1 lane resided at around 350 bp (hTen1 has 363 bp), indicating that the PCR was successful. The remaining PCR product not used for the gel electrophoresis analysis was purified to remove excess nucleotides, primers, and template DNA.

FIGURE 3.2. **Annotated hStn1 nucleotide sequence.** The green box highlights the start codon. The blue boxes highlight the BamH I recognition sites. The yellow box highlights amino acid 180. The red box highlights the opal stop codon after amino acid 368. The orange arrow represents the forward primer which anneals to the complimentary strands and carries a Bgl II recognition site. The purple arrow represents the reverse primer which anneals to the sense strand above and carries an Xho I recognition site and an ochre stop codon.

B. DNA DIGESTION

The purified PCR products were then digested for 2 hours at 37°C. hTen1 was digested with the restriction endonucleases BamH I and Xho I while hStn1 was digested with Bgl II and Xho I. The digested DNA was allowed to separate using agarose gel electrophoresis. The DNA bands corresponding to the correct size of the hStn1 and hTen1 fragments were visualized with a short wave UV light and excised for purification via gel extraction.

C. LIGATION INTO VECTORS AND TRANSFORMATION INTO DH5α COMPETENT CELLS

hStn1 and hTen1 were each ligated into both pET28b and pGEX6p-1 vectors and then transformed into DH5α cells. A negative control was also prepared for each ligation (4 in total). The 2 ligations of each protein into the pET28b vector were spread over kanamycin LB plates whereas the 2 ligations of each protein into the pGEX6p-1 were spread over ampicilin LB plates. The next day, we observed that each of the 4 plates had more colonies than 4 controls.

3 colonies from each of the 4 plates (12 in total) were selected for inoculation into LB media. After centrifugation, Mini-Prep was performed to purify the plasmids from the rest of the bacterial components. A small portion of the plasmids were then digested with the appropriate restriction endonucleases and run on an agarose gel electrophoresis to check for the correct DNA construct. For hTen1, all 6 colonies inoculated yielded the correct band at around 350 bp, therefore the sample corresponding to the most intense band was chosen for each vector for transformation into BL21(DE3) *E.Coli* cells (not pictured). For hStn1, only several of the inoculated colonies yielded the correct DNA construct. Lane 3 was chosen as the representative hStn1 pGeX6p-1 plasmid and lane 5 was selected as the hStn1 pET28b plasmid for transformation into BL21(DE3) cells (Fig 3.3).

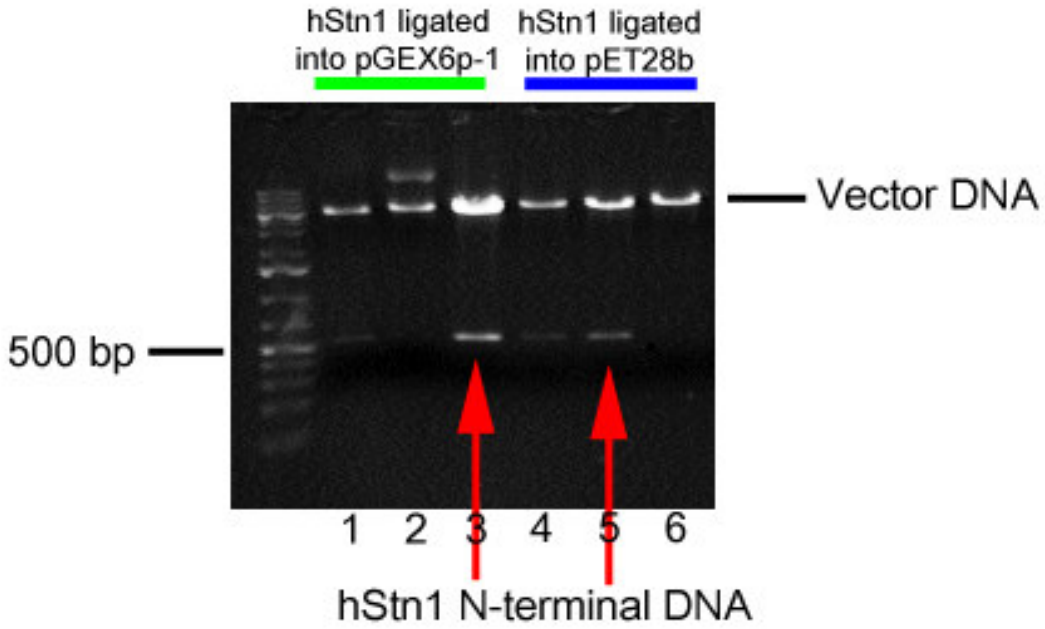

FIGURE 3.3. **Results of hStn1 (aa 2-180) Mini-Prep digestion.** For the pGEX6p-1 insertion, lane 3 is the most intense, and it corresponds to a DNA fragment length of slightly over 500 bp (actual size is 537 bp). Lane 5, which corresponds to the same protein just ligated into a PET28b vector, was selected for the same criteria.

Up to this point, the following 4 plasmids have been purified and verified for correct construct: hStn1 in pGEX6p-1, hStn1 in pET28b, hTen1 in pGEX6p-1, and hTen1 in pET28b. We assessed **Plan A** first.

Plan A: Develop a co-purification protocol for the hStn1 and hTen1 proteins with a His_6-Sumo tag fused to each.

The hStn1 and hTen1 pET28b plasmids were separately transformed into BL21(DE3) competent cells which were plated on kanamycin LB plates at 37°C overnight. A protein expression test verified that each protein was expressed in the BL21(DE3) cells in response to IPTG induction, therefore a glycercol stock was prepared for each. 3 L of hTen1 and 2 L of hStn1 were grown and the cell pellets were all mixed together as the first step of the co-purification protocol. A 1 L excess of hTen1 was used hypothesizing that it would aid in the

previously noted poor solubility of hStn1. However, the SDS-PAGE analysis of the purification displayed that hStn1 still suffered from poor solubility in solution despite an additional 1 hour protein binding time pre-ultracentrifugation (Fig 3.4). **Plan A** was abandoned.

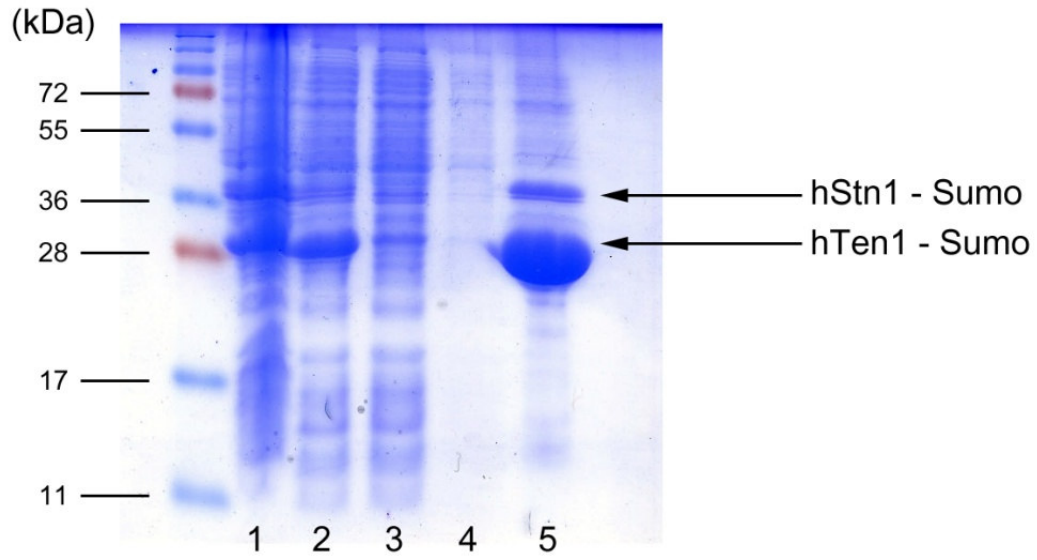

FIGURE 3.4. **Results of His_6-Sumo tagged hStn1-hTen1 co-purification.** Lane 1 is whole cell after sonification. Lane 2 is the supernatant protein mixture post-ultracentrifugation. Lane 3 is the flow-through of the nickel beads. Lane 4 is the flow-through of the first wash. Lane 5 is a sample of the beads. No elutions were collected. The molecular weight of hStn1 (aa 2-180) is 20 kDa and with Sumo fused we add another 17 kDa for a total of 37 kDa, which corresponds well to its band in lane 5. The molecular weight of hTen1 is 13 kDa, and with Sumo the fused protein weighs a total of 30 kDa, also corresponding well to its band. However, there are two clear signs that this purification is far from ideal. First of all, lane 5 indicates that the ratio of hTen1 to hStn1 is worse than the 3 L to 2 L ratio we began with. Furthermore, it is apparent in lane 2 that that not much of the hStn1 protein was soluble, as its band loses intensity from before ultracentrifugation to after (compare the 37 kDa band in lane 1 with lane 2).

Plan B: Develop a co-expression test for a GST tagged hStn1 and a His_6-Sumo tagged hTen1.

The following protocol was used to co-express both GST-tagged hStn1 and His_6-tagged hTen1 in the same *E. coli* cells. In a single tube with 60 µl BL21(DE3) competent cells, 2 µl of hStn1/pGEX6p-1 Mini-Prep product and 2 µl of hTen1/pER28b Mini-Prep product were added and the mixture was kept on ice for 30 minutes. Then the cells were heat shocked for 1 minute at 42°C, after which 600 µl of LB medium was added to the cells and the mixture was shaken for 1

hour at 37°C. After shaking, the samples were centrifuged at 3000 rpm for 1 minute. The pellet was resuspended in 200 µl of supernatant and plated on LB plates containing both ampicilin and kanamycin. Theoretically, only cells with both plasmids would be able to grow on the plate. The plates were then incubated at 37°C overnight, and the next day 3 colonies were selected to test for protein expression. The SDS-PAGE gel displayed a band at 30 kDa corresponding to hTen1 with Sumo (13 kDa + 17 kDa, respectively) but in both the control and IPTG-induced lanes, no band appeared at 48 kDa, which would have corresponded to hStn1 (aa 2-180) with GST (20 kDa + 28 kDa, respectively). This co-expression test was not successful and therefore **Plan B** was abandoned. The rest of this chapter will focus on the results from **Plan C**.

Plan C: Develop a co-expression test for a His_6-Sumo tagged hStn1 and a GST tagged hTen1.

Part II: Protein Growth and Expression

D. CO-TRANSFORMATION INTO BL21(DE3) COMPETENT CELLS

The Mini-Prep product plasmid DNAs for hStn1/pET28b and hTen1/PGEX6p1 were co-transformed into the same BL21(DE3) cells. After shaking, the cell pellets were resuspended in 200 µl supernatant and were plated onto LB plates with both ampicilin and kanamycin. The next day, a protein expression test was performed on 4 colonies acquired from the LB plate to determine if the cells could co-express both proteins. SDS-PAGE was used to visualize the protein expression test (Fig. 3.5). All 4 lanes indicated that upon IPTG induction, both proteins were expressed, even though their bands appear very close on the gel since hStn1 – Sumo is about 37 kDa (20 kDA + 17 kDa, respectively) and hTen1 – GST is about 41 kDa (13 kDa + 28 kDa, respectively). Glycerol stock was created with the cells from the colony represented by lane 2. 4 L of the co-transformed BL21(DE3) were grown until the O.D. value reached 0.6 and

induced with 100 µM IPTG at room temperature overnight. The solution was centrifuged and the cell pellets were saved at -20°C.

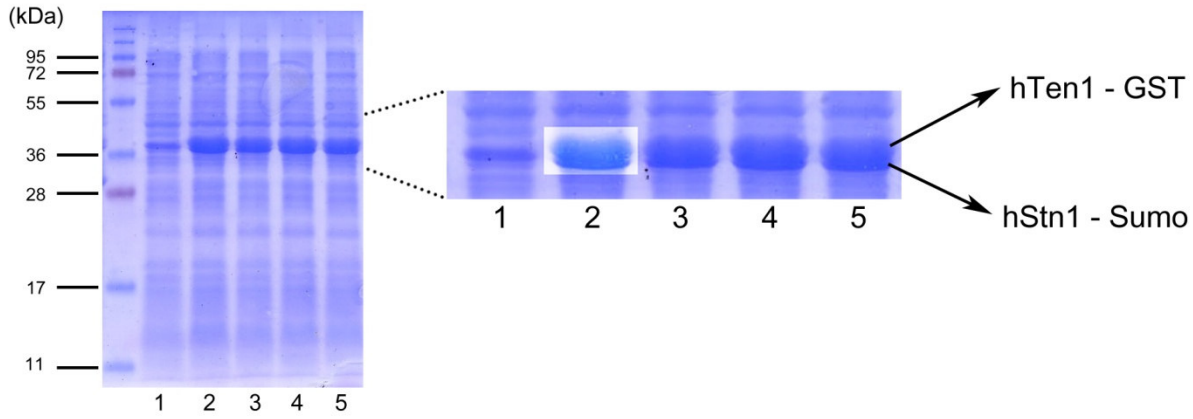

FIGURE 3.5. **Results of hStn1-Sumo/hTen1-GST co-expression test.** Lane 1 is the control and lanes 2-5 represent 4 different colonies from the LB plate. Unfortunately, since the molecular weights of the two fusion proteins are so similar (37 kDa and 41 kDa, respectively) it is difficult to see if both bands exists. In lane 2 of the enlarged portion of the gel, the contrast was altered in order to see both bands. Compared to the control lane, the IPTG-induced cells only showed an increase of protein expression in the area of interest. Glycercol stock was created from the cells represented by lane 5.

Part III: Protein Purification

E. PURIFICATION OF hStn1-hTen1 COMPLEX

The cell pellets were lysed and after ultracentrifugation, the supernatant was incubated with 7 ml nickel beads for 4 hours by vertical tumbling in 4°C. Next, the flow-through was collected and the beads were washed twice before being eluted with imidazole (Fig. 3.6). 16 1ml-elutions were collected and concentrated to 10 ml with a total yield of 240 mg, a good portion of which can be considered as junk protein. The protein was run through an FPLC Hiload Superdex200 size exclusion column for an initial round of purification (Fig. 3.7). The single non-aggregate peak was analyzed by SDS-PAGE (Fig. 3.8, lanes 2-13).

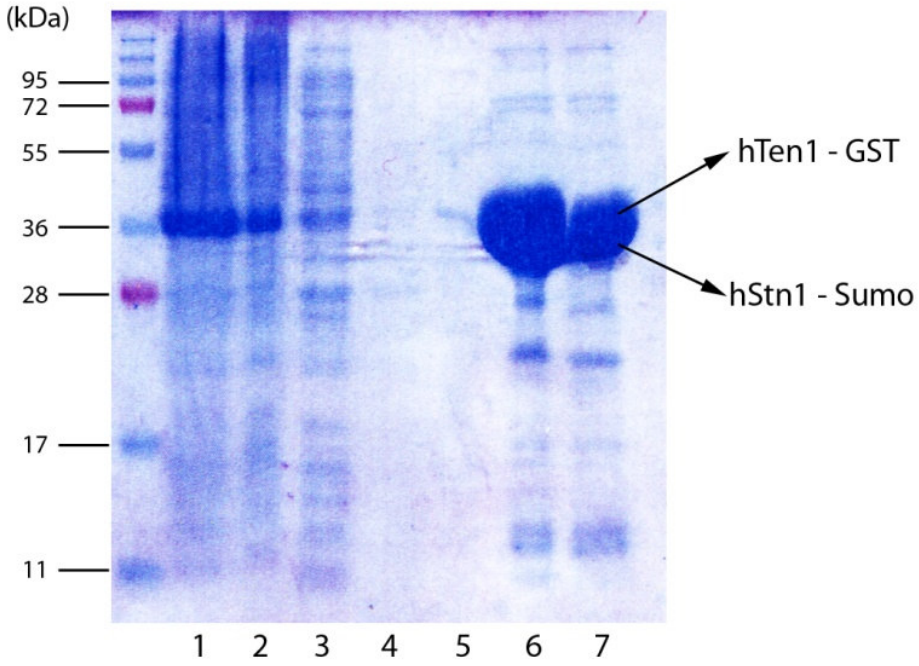

FIGURE 3.6. **Results of hStn1-Sumo/hTen1-GST purification.** Lane 1 is whole cell after sonification. Lane 2 is the supernatant protein mixture after ultracentrifugation. Lane 3 is the flow-through of the nickel beads. Lane 4 is the flow-through of the first wash. Lane 5 is the flow-through of the second wash. Lane 6 represents the second elution. Lane 7 represents the fourth elution. Once again, the sizes of the proteins are too similar to be distinguishable, which makes it difficult to assess lane 2 and protein solubility. Nevertheless, the protein expression in the elution fractions is substantial and further treatment with ULP1 would easily reveal which proteins were purified here.

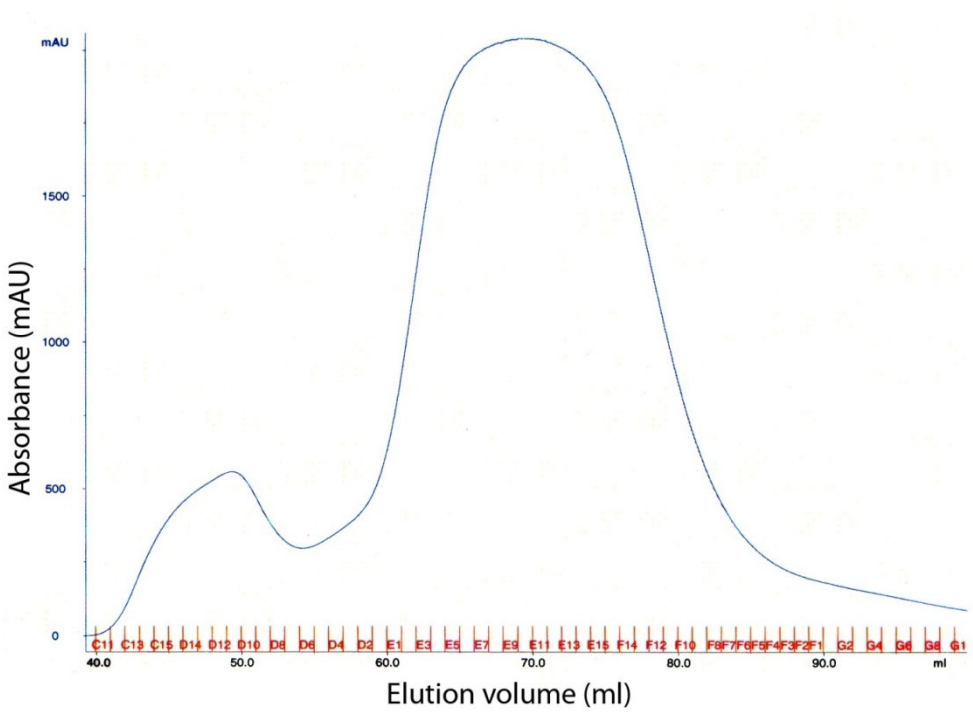

FIGURE 3.7. **Profile of initial hStn1-Sumo/hTen1-GST purification through HiLoad Superdex200 size exclusion column.** The peak from 55 to 90 ml was analyzed by SDS-PAGE (Fig. 3.8).

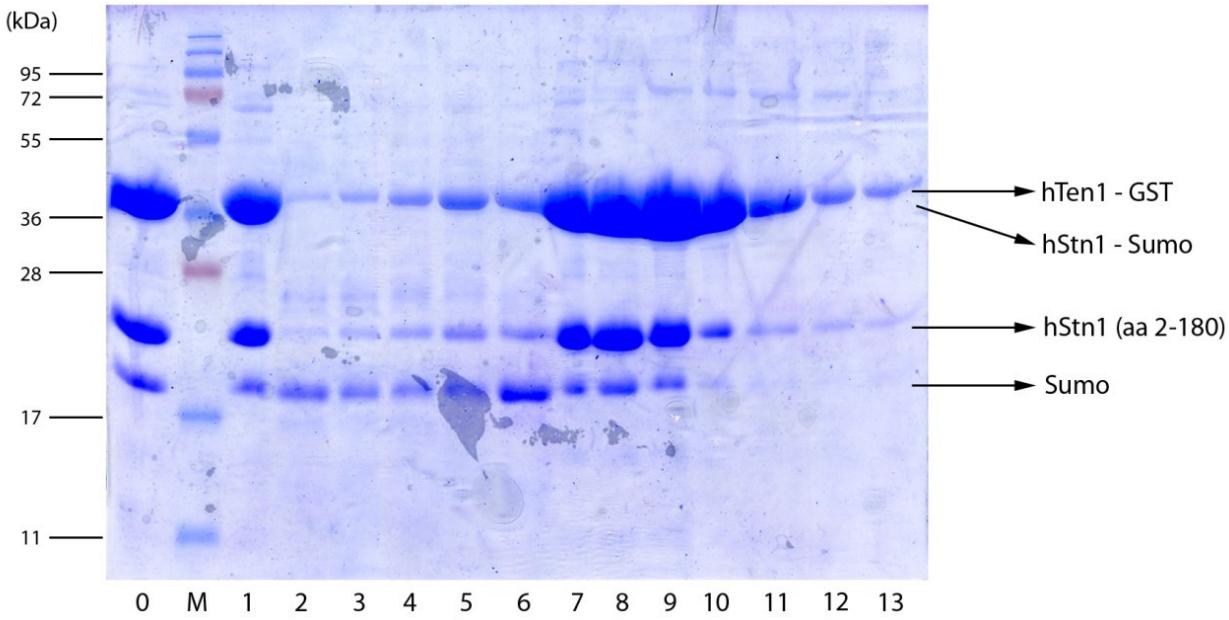

FIGURE 3.8. **Result of hStn1-Sumo/hTen1-GST HiLoad Superdex200 column.** Lane 0 is a sample of the 16 collected elutions mixed together. Lane 1 is a sample of the elutions after being concentrated to 10 ml. Lanes 2-13 analyze the peak from 55 to 90 ml in the HiLoad Superdex200 column (Fig. 3.7). Note that the gel, including lanes 0 and 1, was run 12 hours after the elutions were collected (this may account for the partial Sumo digestion in lanes 0 and 1). Sumo protein is characterized as appearing slightly over the 17 kDa ladder mark.

As indicated by the gel (Fig. 3.8), some of the hStn1 His_6-Sumo tag had already been cleaved. Complete removal of the His_6-Sumo tag was achieved by combining the collected fractions (32 mL) from the Hiload Superdex200 column with 5 mM DTT, 6 ml glutathione sepharose beads, and excess ULP1 protease. The mixture was tumbled horizontally for 6 hours at 4°C.

Theoretically speaking, the beads should bind the GST protein, which is fused to hTen1. As indicated by the single, non-aggregate peak for the Hiload Superdex200 size exclusion column, we were able to confirm that hTen1 formed a complex with hStn1 (the single peak had a maximum absorbance at approximately 70.5 ml, which corresponds to a protein complex molecular weight of 79.6 kDa according to the column's calibration plot; the combined molecular weight of hStn1–Sumo and hTen1–GST is approximately 78 kDa). Therefore the beads bind to GST, which is fused to hTen1, which binds hStn1, which is fused to Sumo protein. The ULP1 protease can efficiently and specifically cleave the Sumo protein. After the 5 hour binding and digesting time, the mixture was poured into a column and the flow-through (which should only be the cleaved SUMO protein) was collected (Fig 3.9, lane 3). A 50 ml wash was conducted with FPLC buffer (Fig 3.9, lane 2) and a sample of the beads was collected for SDS-PAGE analysis (Fig 3.9, lane 1). The beads were resuspended to 35 ml in FPLC buffer and an excess amount of protease 3C was added to remove the GST-tag. The digestion was allowed to occur overnight in 4°C by horizontal tumbling. The next day, the flow through was collected (Fig 3.9, lane 5) and a sample of the beads were collected for SDS-PAGE analysis (Fig 3.9, lane 4).

Finally, the hStn1-hTen1 complex in the collected flow-through was further purified by passage through an FPLC Hiload Superdex75 size exclusion column (Fig. 3.10) and the fractions

were characterized by SDS-PAGE (Fig. 3.11). Having achieved greater than 95 percent purity, the peak fractions were concentrated to a final concentration of 25 mg/ml. The protein complex was then aliquotted into 25 µl fractions and flash-frozen with liquid nitrogen for storage in -80°C.

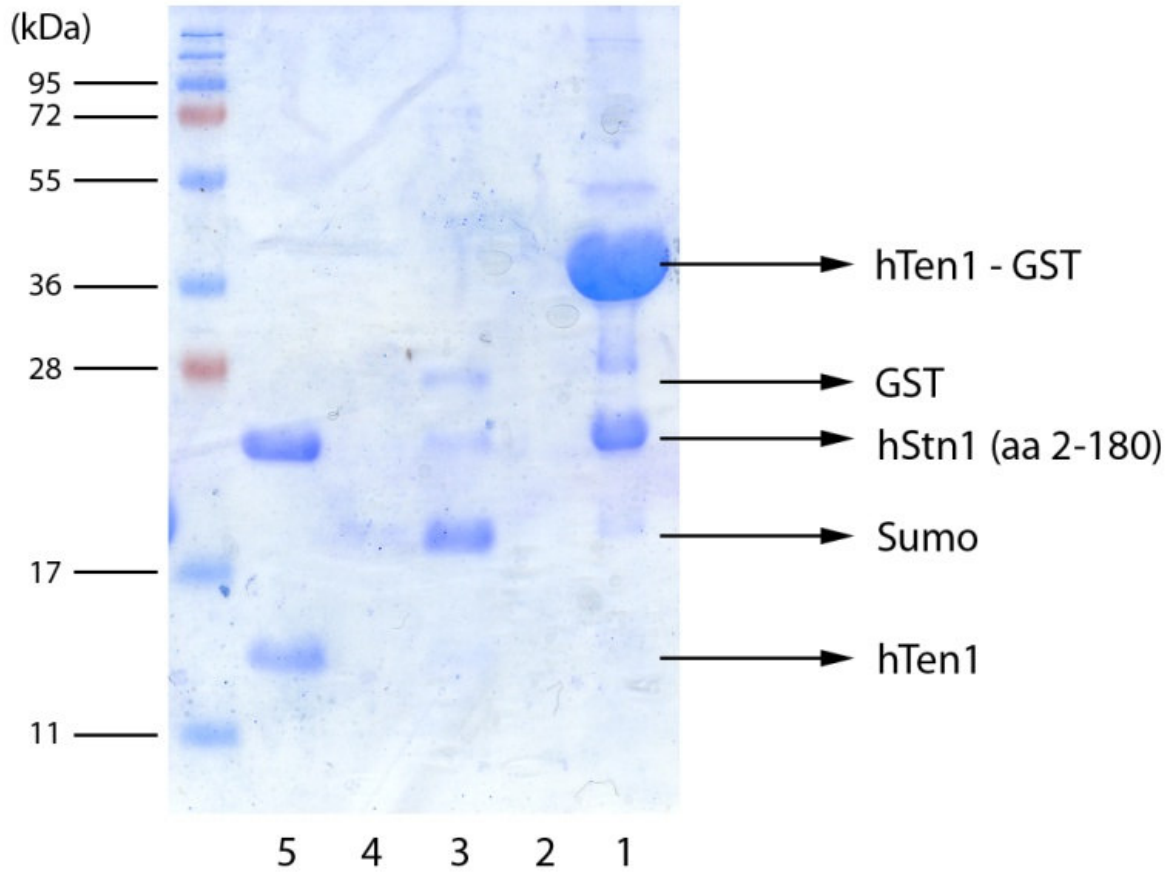

FIGURE 3.9. **Analysis of hStn1-Sumo/hTen1-GST ULP1 and 3C protease digestion.** Lane 3 is the flow-through of the glutathione sepharose beads column after ULP1 digestion. Lane 2 is the wash the followed after flow-through collection. Lane 1 is a sample of the glutathione sepharose beads after the wash. Lane 5 is the flow-through collected after overnight protease 3C digestion. Lane 4 is a sample of the same beads after a wash (this lane should contain GST protein alone, the lack of a GST protein band here at around 28 kDa indicates that perhaps not enough beads were collected for the SDS-PAGE analysis).

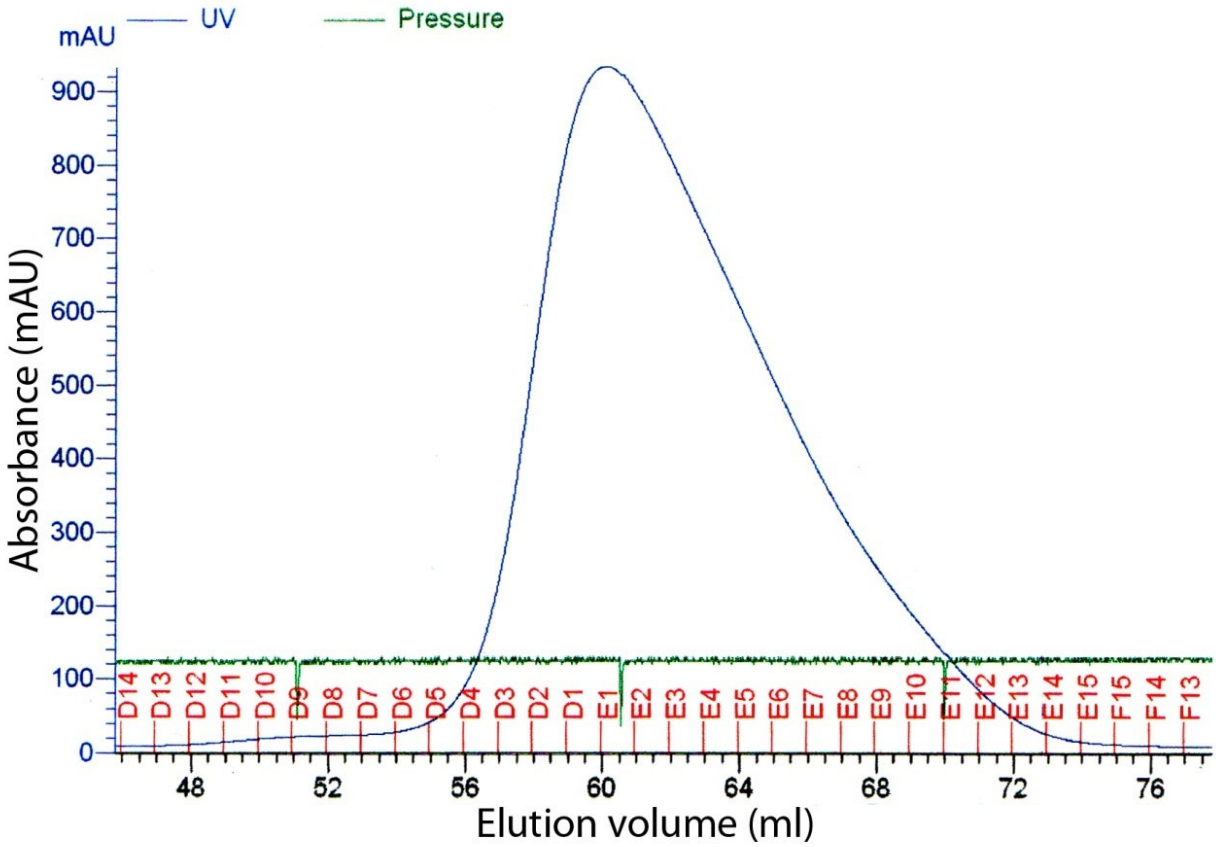

FIGURE 3.10. **Profile of secondary hStn1-hTen1 purification through Hiload Superdex75 size exclusion column.** The lone peak from 56 to 72 ml was analyzed by SDS-PAGE (Fig. 3.11).

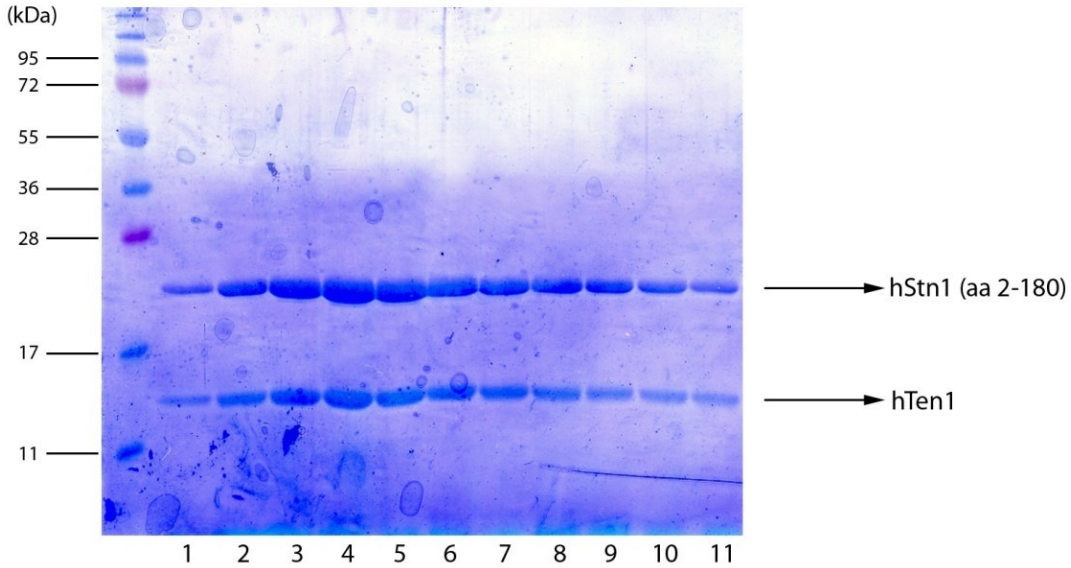

FIGURE 3.11. **Result of hStn1-hTen1 HiLoad Superdex75 column.** Lanes 10 and 11, corresponding to elution volumes 68 and 70 ml respectively, had trace amounts of Sumo contamination. These fractions were not collected.

Part IV: Crystal Growth and Structural Determination

F. CRYSTALLIZATION OF hStn1- hTen1

Using the purified aliquots of hStn1-hTen1 purified complex, 480 crystallization conditions were screened by sitting drop in 5, 96-well trays by robot at the University of Michigan Life Sciences Institute. After re-examining the trays over a month, 2 conditions yielded protein crystals. The first condition was 15% ethanol (precipitant), 0.1 M citrate pH 5.5 (buffer), and 0.2 M $LiSO_4$ (salt) (Fig. 3.12) and the second condition was 20% PEG 200 (precipitant), 0.05 M HEPES pH 7.0 (buffer), 0.2 M KCl (salt), and 0.025 M $MgSO_4$ (salt) (Fig. 3.13).

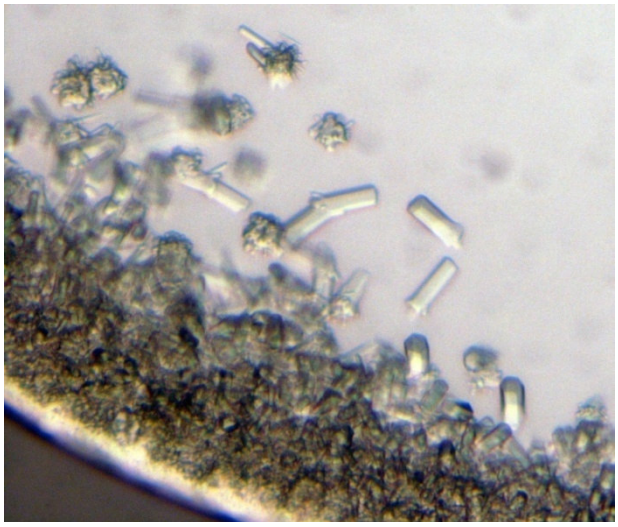

FIGURE 3.12. **Representative crystals of hStn1-hTen1 complex protein.** Crystals were grown in a 15% ethanol precipitant well solution and were magnified from a 96-well tray utilizing a sitting drop for crystallization.

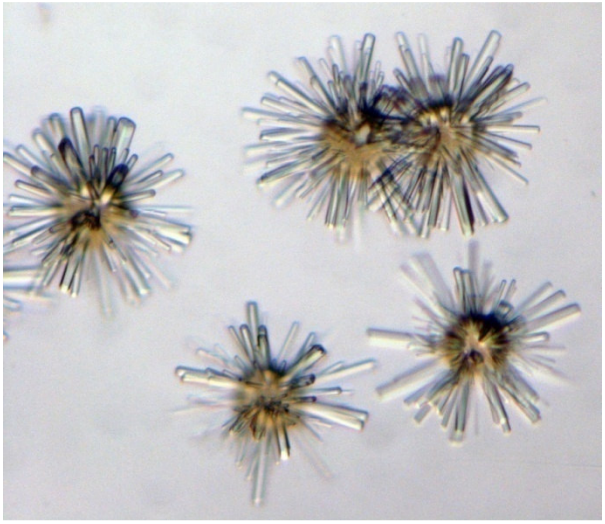

FIGURE 3.13. **Representative crystals of hStn1-hTen1 complex protein.** Crystals were grown in a 20% PEG 200 precipitant well solution and were magnified from a 96-well tray utilizing a sitting drop for crystallization.

Further crystallization trays were created by hand in order to optimize both crystallization conditions with the aim of growing larger crystals with a more defined shape. Over 10, 24-well crystal trays were prepared by altering parameters such as salt concentration, precipitant concentration, pH, temperature, and protein concentration. Midway through the process, Dr.

Yuting Yang collected X-ray diffraction data of the crystals at the University of Michigan Life Sciences Institute. The PEG 200 crystals yielded no spots and the ethanol crystals only yielded a few, therefore we focused on further optimizing only the latter condition. Finally, an optimal crystallization condition was found which consisted of a well solution of 10% ethanol, 0.1 M citrate pH 5.5, and 0.4 M Li_2SO_4 (Fig. 3.14).

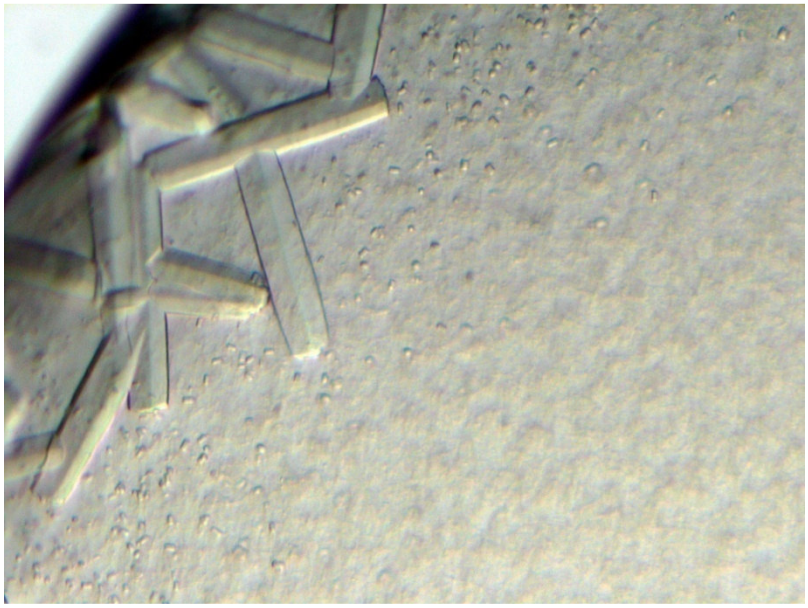

FIGURE 3.14. **Optimized crystals of hStn1-hTen1 complex protein.** Crystals were grown in a 10% ethanol precipitant well solution and were magnified from a 24-well tray utilizing a hanging drop for crystallization.

With the optimal condition at hand, a 96-additive screen was conducted in order to elucidate new conditions that may affect crystal size and internal protein packing within the unit cell. Additives were added to the hanging drop in a 1 to 1 to 0.1 ratio of well solution to protein to additive. The following additives yielded comparable crystals: NaCl (salt), glycine (amino acid), taurine (linker), guanidine HCl (chaotrope), KCl (salt), calcium (divalent cation), and strontium (divalent cation) (Fig. 3.15). However, the additive 0.1 M spermine tetra-HCl yielded the largest crystals with the most well defined shape and best response to a color filter (Fig. 3.16).

4 to 5 crystals from the conditions listed above were harvested for X-ray diffraction, along with native, pre-additive crystals, some of which underwent an overnight dehydration to improve crystal packing. While harvesting these crystals, several optimistic observations were made that differed from the previous X-ray diffraction collection. The crystals were heavier, had a harder surface, and were more brittle since the use of ethylene glycol, glycercol, and PEG 200 resulted in cracking. The cyroprotectant sodium formate did not harm the crystals; therefore the crystals were harvested by transferring them to a solution containing 5% ethanol, 0.1 M citrate pH 5.5, 5% PEG 200, and 5 M sodium formate and flash freezing the crystals in liquid nitrogen for data collection and storage.

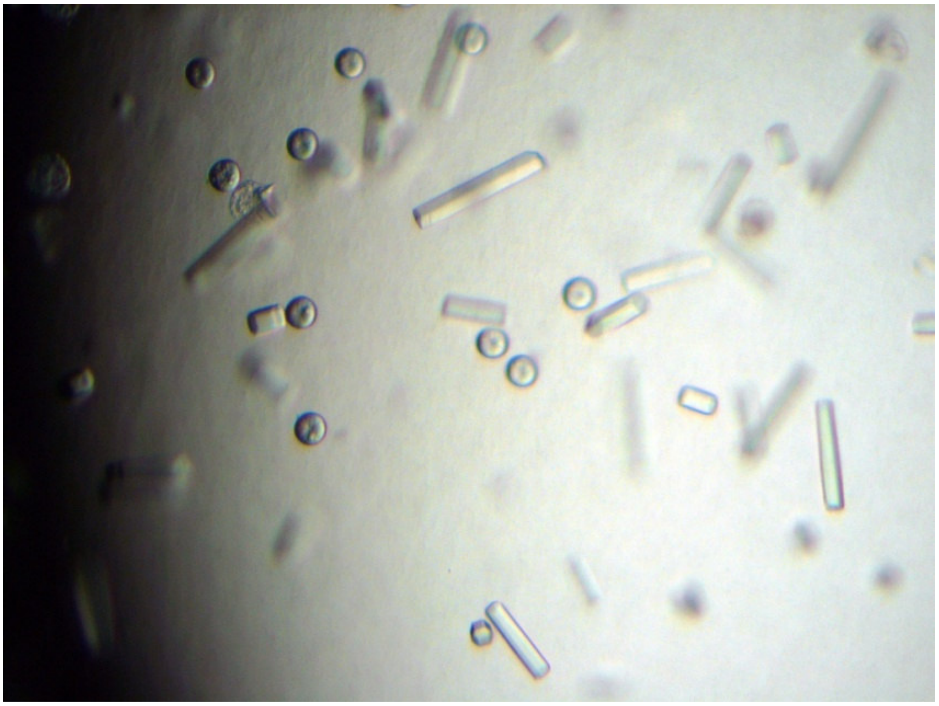

FIGURE 3.15. **Crystals of hStn1-hTen1 complex protein in strontium additive.** Crystals were grown in a 10% ethanol precipitant well solution and were magnified from a 24-well tray utilizing a hanging drop for crystallization.

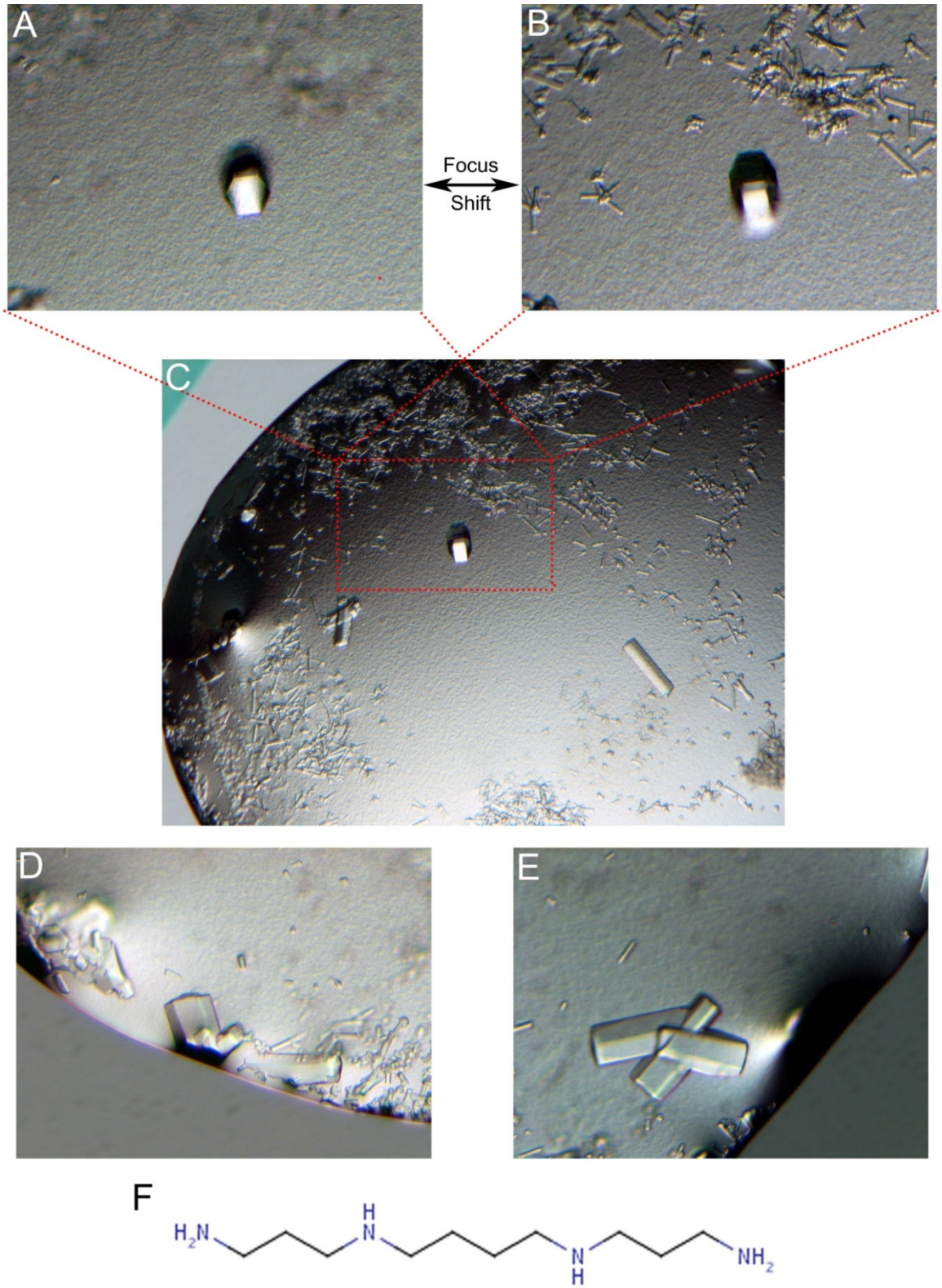

FIGURE 3.16. **Optimized crystals of hStn1-hTen1 complex protein in spermine tetra-HCl additive.** Crystals were grown in a 10% ethanol precipitant well solution and were magnified from a 24-well tray utilizing a hanging drop for crystallization. Image A and B are magnified from the dotted box in image C and only differ by a change in focus. Image D and E were found in other areas of the same drop. Image F shows molecular structure of spermine.

G. X-RAY DIFFRACTION

X-ray diffraction data was collected at the Advanced Photon Source (Beamline 21 ID-G) and was processed by Dr. Yuting Yang using the program HKL2000 (Fig. 3.17) (45). Every additive conditions diffracted to a resolution and pattern similar to that of the native condition (Fig. 3.17a), except for the spermine additive, which improved diffraction from 5.5 Å to 4.2 Å (Fig. 3.17b). From the spermine diffraction pattern, we were able to determine that the crystals belong to space group P6 (hexagonal) with unit cell parameters refined to a = b = 203.07 Å and c = 58.32 Å. Unfortunately, we are not able to solve for the structure of hStn1- hTen1 at 4.2 Å. Alternate constructs of hStn1 must be tested in order to improve what appears to be poor internal crystal packing.

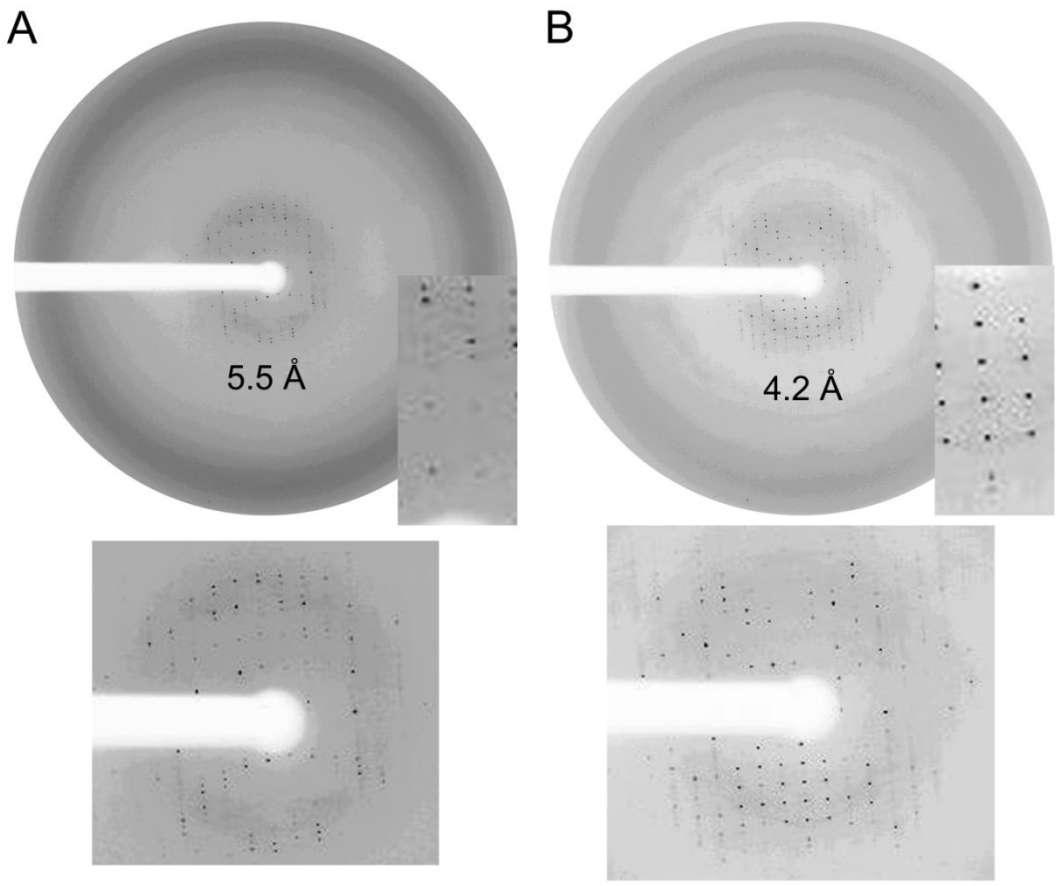

FIGURE 3.17. **Diffraction improvement of the hStn1-hTen1 complex crystals.** Image A is the diffraction pattern of native crystals. Image B shows an over 1 angstrom improvement in the diffraction of the crystals in spermine tetra-HCl additive. Enlarged images show the quality of the diffractions.

Chapter 4: Alternative hStn1 N-terminal Constructs

In order to improve the internal crystal structure of the hStn1-hTen1 complex for X-ray diffraction data collection, we attempted to purify several new constructs of the complex. By referring back to the hStn1 secondary structure prediction and sequence alignment, two new hStn1 constructs were elected, hStn1 amino acids 12-180 and hStn1 amino acids 25-180 (Figs. 3.1, 4.1).

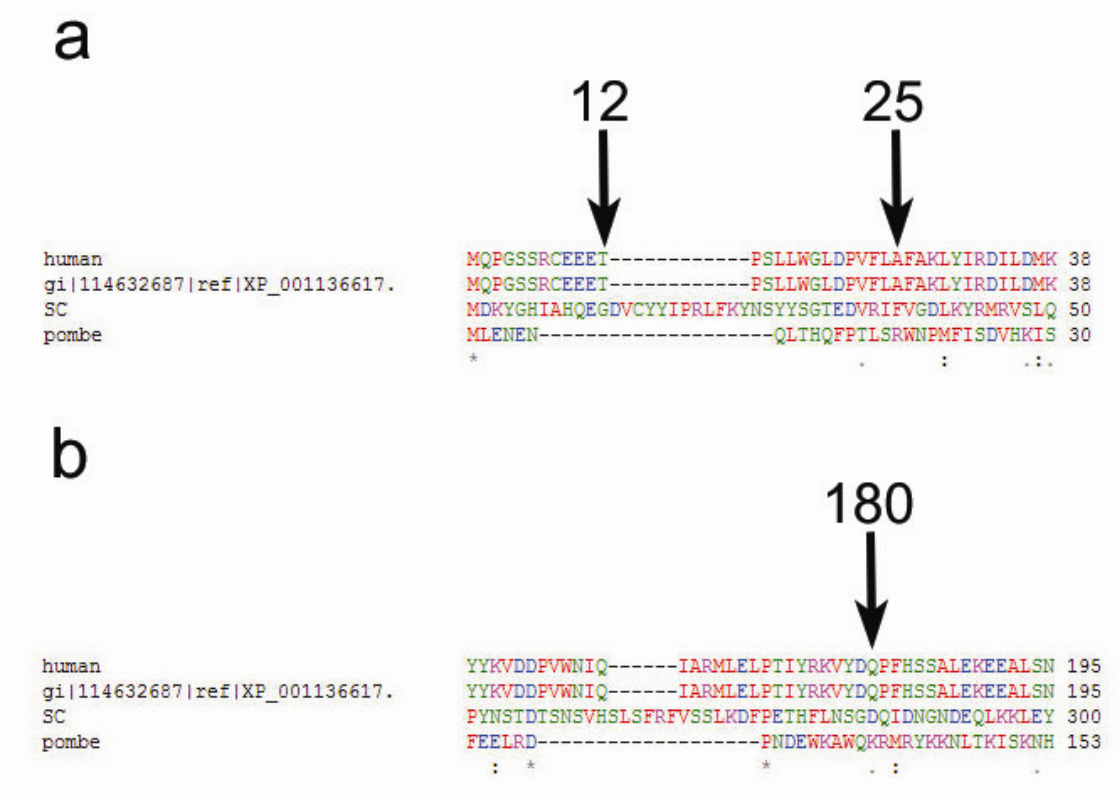

FIGURE 4.1. **Stn1 regions of interest ClustalW sequence alignment.** Human, *Pan* (chimpanzee), budding yeast, and fission yeast Stn1 sequences were aligned, respectively. The color code of the amino acids is as follows: red represents small hydrophobic residues, blue represents acidic residues, magenta represents basic residues, and green represents hydroxyls and amines. **(a)** N-terminal of Stn1, with the two new starting positions for the hStn1 constructs pointed out. **(b)** The 180th amino acid was still used as the final residue for the new constructs.

Using a similar co-expression protocol presented in the previous chapter, we were able to create plasmids for hStn1-hTen1 complexes with both new constructs. We began with the purification of hStn1 (aa 12-180)–hTen1 complex. Once again, the procedure was analogous to

before, except for the passage through a MonoQ column before the final purification via the Hiload Superdex75 column (Fig. 4.2). Throughout the process, there were degradation bands at approximately 11 kDa on the SDS-PAGE gels that we were not able to separate from the rest of the complex. The protein was saved for later usage in biochemical assays. Since the purity of the sample was less than 95%, we did not continue with its crystallization. Due to the fact that it is an even shorter construct, it was presumed that the hStn1 (aa 25-180)-hTen1 complex would face similar problems, therefore it was not purified.

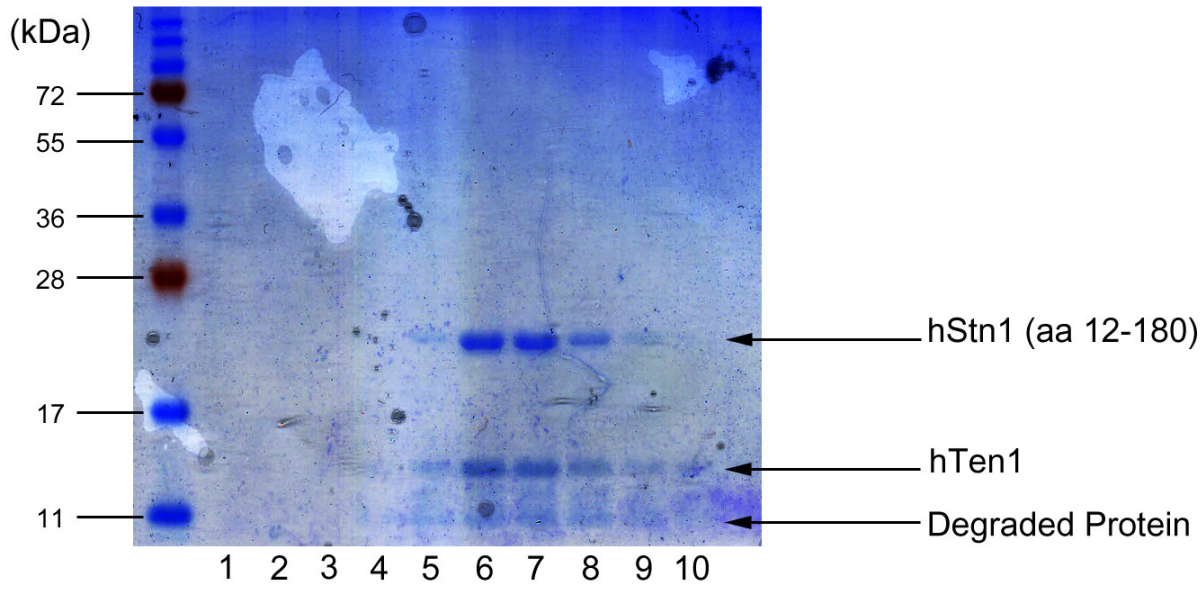

FIGURE 4.2. **Result of hStn1-hTen1 HiLoad Superdex75 column.** Despite the fact that the degraded protein lies under the hTen1 band, one cannot conclude on this lone observation that it is the degradation product of hTen1.

Chapter 5: hStn1 C-terminal

While working with the hStn1-hTen1 complex, we attempted to solve for the structure of the hStn1 C-terminal domain as well. By referring back to the hStn1 secondary structure prediction and sequence alignment, we decided to first purify the construct hStn1 amino acids 200-368 (Figs. 3.1, 5.1). All constructs were ligated into a pET-28b vector.

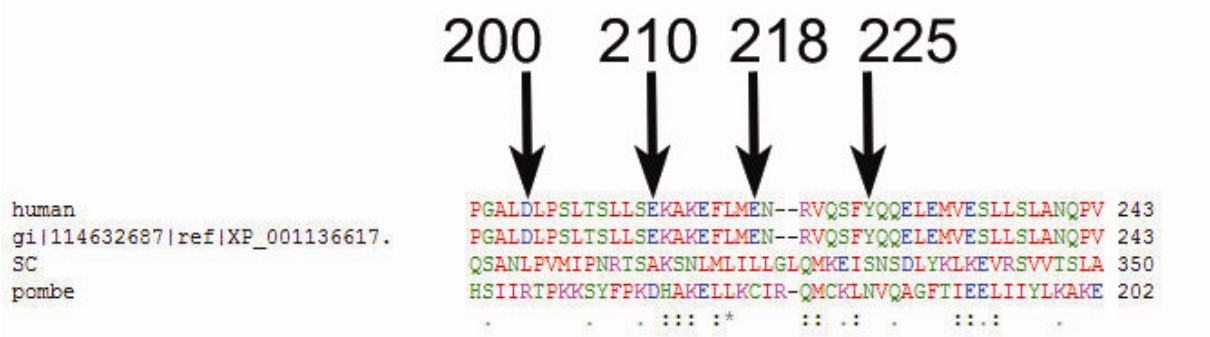

FIGURE 5.1. **Stn1 region of interest ClustalW sequence alignment.** Human, *Pan* (chimpanzee), budding yeast, and fission yeast Stn1 sequences were aligned, respectively. The color code of the amino acids is as follows: red represents small hydrophobic residues, blue represents acidic residues, magenta represents basic residues, and green represents hydroxyls and amines. The starting N-terminal residues for the 4 constructs purified are pointed out.

Using a similar protocol as before, we were able to purify the protein and obtain crystals, albeit small and poorly defined ones that were not able to diffract. Ke Wan attempted to purify a new construct, hStn1 (aa 225-368), however the protein was poorly soluble. Presuming that a 25 amino acid jump to this new construct was too large, we decided to try 2 other constructs, hStn1 (aa 210-368) and hStn1 (aa 218-368). Unfortunately, by analyzing the SDS-PAGE gels of the purification of both constructs, it was apparent that their solubility was poor and thus obtaining crystals that diffract well would be even more unlikely compared to before (Fig. 5.2). Since the original construct, hStn1 (aa 200-368), by comparison displayed better solubility, hStn1 (aa 210-368) and hStn1 (aa 218-368) were not purified further.

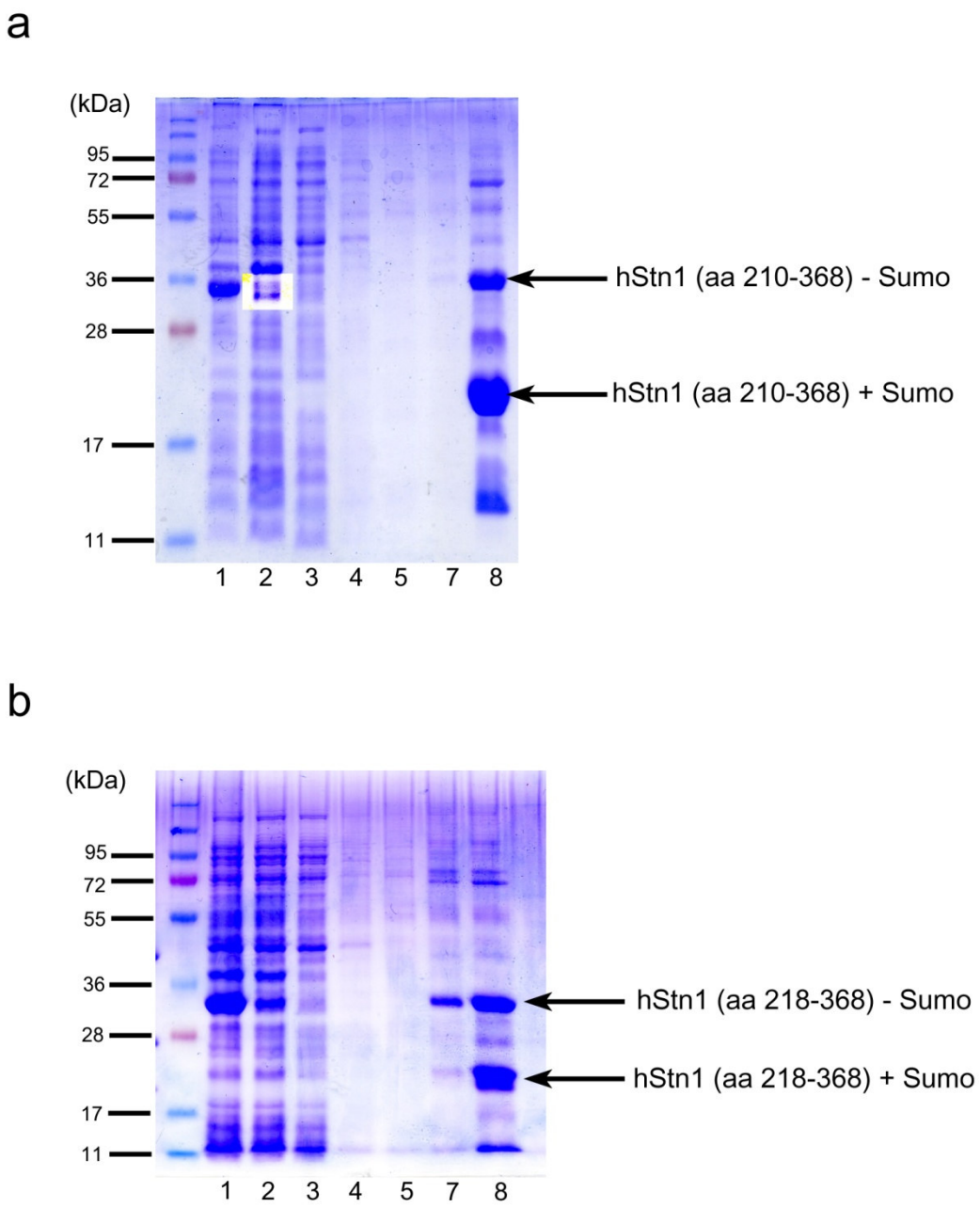

FIGURE 2.2. **Purification of hStn1 (aa 210-368) and hStn1 (aa 218-368).** Both purifications were abandoned due to their poor solubility, which was determined by analyzing lane 2, the supernatant after ultracentrifugation. Since the protein band from lane 1 (lysed cells) corresponding to the target protein was not as intense in lane 2, this indicates that some of the protein was not soluble within the supernatant. Also, elution lanes 8 in both gels display a large amount of premature Sumo digestion, indicating another instability within the protein. In **(a)**, the region in lane 2 where one expects an intense band as in lane 1 has its contrast adjusted for improved visualization.

Chapter 6: Discussion

While we were able to solve the solubility problem of hStn1 N-terminal and grow crystals that diffracted to 4.2 angstroms, further attempts must be made to improve this resolution so that the structural determination of the hStn1-hTen1 can be made. We faced a similar problem with hStn1 C-terminal. Since the optimization parameters of their crystal growth have been exhausted, more constructs must be created and purified. It is often reported that a change by only a few amino acids results in drastic improvement in diffraction. A limited proteolysis was conducted for both hStn1 N-terminal and hStn1 C-terminal but no digestions took place. I suggest that since we were able to diffract the hStn1-hTen1 complex to less than 5 angstroms, very minor changes should be made to the hStn1 N-terminal construct, on the scale of only 2-5 amino acids. Also, we should alter the construct of the C-terminus of hStn1 N-terminal at amino acid 180 since we have yet to do so. Likewise, for hStn1 C-terminal, constructs should be designed similar to that of hStn1 (aa 200-368), which at the least was able to yield crystals.

During the course of the work of this thesis, the first structures for the Stn1-Ten1 complex were solved and analyzed for *S. cerevisae* (budding yeast) and *S. pombe* (fission yeast) by Jia Sun, a member Dr. Ming Lei's laboratory (Fig 6.1) (46). Early analysis indicates that the complexes are structurally conserved, hinting that perhaps they are functionally conserved as well. While we continue to work on solving the structure for the human Stn1-Ten1 complex, one thing is certain; analysis of the yeast structures will answer many questions with regards to the CST complex, but even more questions will appear. Only if we continue to expand our knowledge on telomere binding proteins and the CST complex can we then begin to erase the line. The fine line that separates laboratory results from the treatment of human diseases.

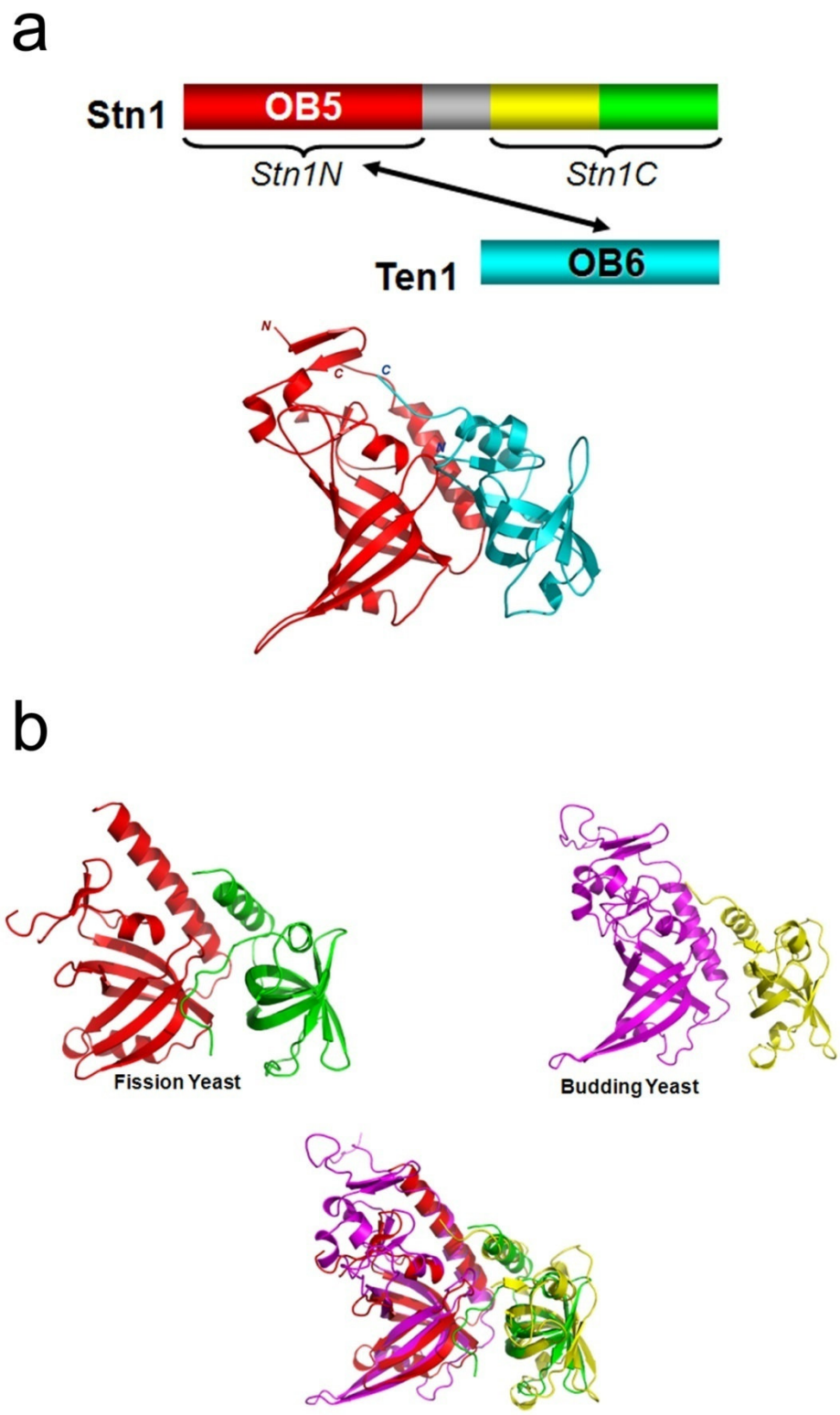

FIGURE 6.1. **Crystal structures of fission yeast and budding yeast Stn1-Ten1 complex. (a)** Domain structure of the Stn1-Ten1 complex along with its crystal structure for budding yeast. The colors of the domains correspond to the color of the protein structure. **(b)** Superimposition of the Stn1-Ten1 complex for budding yeast and fission yeast (46).

Acknowledgements

First and foremost, I would like to thank everyone in Dr. Lei's laboratory for their unremitting support and help. I want to thank Dr. Lei for giving me this amazing opportunity to conduct research in his laboratory and for letting me stay for awhile longer so that I may finish what I have already started. Under his guidance, I not only matured as a scientist, but as a student and as a person, and for that I am forever grateful. I want to thank Ke Wan, with whom I worked very close with over the past year. Ke took away so much of his time to teach me anything and everything I needed to know about the lab, from ligations to transformations and even some Chinese phrases. I want Dr. Yuting Yang for sharing her expertise on X-ray crystallography with me and for her patience as she taught me how to harvest crystals. I want to thank Dr. Yong Chen for guiding me through FPLC usage and trouble shooting and Dr. Wei Deng always going out of her way to give me a helping hand. I want to thank Jia Sun, Dr. Zhixiong Zeng, and Feng Wang for always stopping whatever they were in the midst of doing to answer any of my lingering questions and for their advice and support during times of frustration. I would also like to thank my fellow undergraduates in the lab. I want to thank Austin Reese for all of his help, especially for his guidance on how to use the ClustalW sequence alignment. I want to also thank Sean Tamir for timely comic relief. I will miss everyone very much.

I would like to thank my friends and family. I want to thank my parents and my sister for always praying that my crystals grow bigger. I want to thank my fellow biochemists—Jesse, Rob, Roshan, and Jason—with whom I have had many great memories as we fought through our biochemistry concentration. I also would like to thank Marzieh for all of her emotional support; she was always the first person I would call to cheer me up after a bad result. Finally, I want to thank Omeid Seirafi-pour. His constant motivation drove me to finish my thesis strong.

References

1. Bell, S. P., and Dutta, A. (2002) *Annu. Rev. Biochem.* **71**, 333-374
2. Hubscher, U., Maga, G., and Spadari, S. (2002) *Annu. Rev. Biochem.* **71**, 133-163
3. Watson, J. D. (1972) *Nature New Biology* **239**, 197-201
4. Casjens, S. (1998) *Annu. Rev. Genet.* **32**, 339-377
5. Blackburn, E. H., and Gall, J. G. (1978) *Journal of Molecular Biology* **120**, 33-53
6. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J. R. (1988) *Proc Natl Acad Sci U S A.* **85**, 6622-6626
7. Cervantes, R. B., and Lundblad, V. (2002) *Current Opinion in Cell Biology* **14**, 351-356
8. Greider, C. W., and Blackburn, E. H. (1985) *Cell* **43**, 405-413
9. Harrington, L. A., and Greider, C. W. (1991) *Nature* **353**, 451-454
10. Wright, W., Piatyszek, M., Rainey, W., Byrd, W., and Shay, J. (1996) *Developmental Genetics* **18**, 173-179
11. Forsyth, N., Wright, W., and JW, S. (2002) *Differentiation* **69**, 188-197
12. Cong, Y.-S., Wright, W. E., and Shay, J. W. (2002) *Microbiol. Mol. Biol. Rev.* **66**, 407-425
13. Cech, T. R. (2004) *Cell* **116**, 273-279
14. Olovnikov, A. M. (1973) *Journal of Theoretical Biology* **41**, 181-190
15. Chong, L., Steensel, B. v., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P., and Lange, T. d. (1995) *Science* **270**, 1663-1667
16. Arino, O., Kimmel, M., and Webb, G. F. (1995) *Journal of Theoretical Biology* **177**, 45-57
17. Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W., and Harley, C. B. (1992) *Journal of Molecular Biology* **225**, 951-960
18. Arkus, N. (2005) *Journal of Theoretical Biology* **235**, 13-32
19. von Zglinicki, T. (2002) *Trends in Biochemical Sciences* **27**, 339-344
20. Karlseder, J., Smogorzewska, A., and de Lange, T. (2002) *Science* **295**, 2446-2449
21. Scherf, A., Figueiredo, L. M., and Freitas-Junior, L. H. (2001) *Current Opinion in Microbiology* **4**, 409-414
22. Freitas-Junior, L., Porto, R., Pirrit, L., Schenkman, S., and Scherf, A. (1999) *Nucl. Acids Res.* **27**, 2451-2456
23. Barry, J. D., Ginger, M. L., Burton, P., and McCulloch, R. (2003) *International Journal for Parasitology* **33**, 29-45
24. Dossin, F. d. M., Dufour, A., Dusch, E., Siqueira-Neto, J. L., Moraes, C. B., Yang, G. S., Cano, M. I., Genovesio, A., and Freitas-Junior, L. H. (2008) *PLoS ONE* **3**, e2313
25. Bottius, E., Bakhsis, N., and Scherf, A. (1998) *Mol. Cell. Biol.* **18**, 919-925
26. Figueiredo, L., and Scherf, A. (2005) *Chromosome Research* **13**, 517-524
27. Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., and Lanzer, M. (1998) *The EMBO Journal* **17**, 5418-5426
28. de Lange, T. (2004) *Nat Rev Mol Cell Biol* **5**, 323-329
29. Li, B., Espinal, A., and Cross, G. A. M. (2005) *Mol. Cell. Biol.* **25**, 5011-5021
30. Navarro, M., and Gull, K. (2001) *Nature* **414**, 759-763
31. Bryan, T. M., and Reddel, R. R. (1997) *European Journal of Cancer* **33**, 767-773
32. Henson, J. D., Neumann, A. A., Yeager, T. R., and Reddel, R. R. (2002) *Oncogene* **21**, 598-610

33. Blackburn, E. H. (2001) *Cell* **106**, 661-673
34. de Lange, T. (2005) *Genes & Development* **19**, 2100-2110
35. Pennock, E., Buckley, K., and Lundblad, V. (2001) *Cell* **104**, 387-396
36. Nugent, C., Hughes, T., Lue, N., and Lundblad, V. (1997) *Science* **274**, 249-252
37. Chandra, A., Hughes, T. R., Nugent, C. I., and Lundblad, V. (2001) *Genes & Development* **15**, 404-414
38. Li, S., Makovets, S., Matsuguchi, T., Blethrow, J. D., Shokat, K. M., and Blackburn, E. H. (2009) *Cell* **136**, 50-61
39. Wellinger, R. J., Ethier, K., Labrecque, P., and Zakian, V. A. (1996) *Cell* **85**, 423-433
40. Grandin, N., Damon, C., and Charbonneau, M. (2001) *The EMBO Journal* **20**, 1173-1183
41. Petreaca, R. C., Chiu, H.-C., Eckelhoefer, H. A., Chuang, C., Xu, L., and Nugent, C. I. (2006) *Nat Cell Biol* **8**, 748-755
42. Figure from Novagen Catalog # 69865-3.
43. Figure from GE Healthcare Catalog # 27-4597-01.
44. Serdyuk, I. N., Zaccai, N. R., and Zaccai, J. (2007) *Methods in Molecular Biophysics*, Cambridge University Press, New York
45. Otwinowski, Z., Minor, W., and Charles W. Carter, Jr. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. in *Methods in Enzymology*, Academic Press. pp 307-326
46. Jia Sun, unpublished data.