Expression and Purification of INI1 Truncations

Rachel Williams

Honors Thesis

Departmental Honors Degree

Biochemistry and Molecular Biology

Ruhl Laboratory

May 6th, 2019

Abstract

INI1 (Integrase Interactor 1) is a highly-conserved, core component of the SWI/SNF remodeling complex and is characterized as a tumor suppressor. Mutations in the SMARCB1 gene/INI1 protein cause aggressive pediatric cancers called malignant rhabdoid tumors (MRT). Atypical Teratoid/Rhabdoid Tumors (AT/RT), a common form of MRT, are a pediatric cancer associated with INI1 mutations and form in the central nervous system and kidneys of infants. Although the tertiary structure of INI1 is unknown, the 385 amino acids-long protein contains an N-terminal DNA binding domain (1-186), two imperfect repeat regions (187-245 and 259-319) that could mediate protein-protein interactions, and a C-terminal coiled-coil domain (334-376). Little is understood about INI1's function, structure, regulation, or DNA binding. To determine INI1's minimal binding domain, INI1(31-125) and INI1(31-144) were expressed in *E. coli* cells. The 6X His-tagged truncations did not bind to the Ni-NTA resin, necessitating the use of binding tests. The results of this study conclude that the INI1 truncations did not bind to the Ni-NTA resin under any tested conditions, indicating that further attempts need to be made at finding the correct binding condition and ensuring that a functional 6X His-tag is present. In future studies, the purified INI1(31-125) and INI1(31-144) fragments will be used in DNA binding assays to determine if these fragments contain INI's minimum binding domain.

Table of Contents

Introduction

DNA

All the genetic information in a cell is encoded in its deoxyribonucleic acid (DNA). All known cellular life and some viruses contain DNA as their long-term information storage molecule. The human genome consists of roughly three billion base pairs that comprise 23 chromosomes [1]. DNA is a long polymer of deoxynucleotides connected by phosphodiester linkages that contains coding sequences called genes, structural segments, and regulatory sequences. The information encoded in genes instructs the cell on how to make proteins and other cellular components that are vital to life. Various proteins and enzymes interact with DNA to regulate gene expression and cellular activities. Careful regulation of gene expression is necessary to maintain a healthy, functioning cell.

Chromatin

If the DNA of one human cell was linearized and placed end-to-end, it would be approximately 6 feet (1.8 meters) in length [1]. Because of the vast amount of DNA present in each eukaryotic cell, the cell must find an efficient way to pack this information safely into a nucleus. In addition, the packing mechanism must allow DNA access to various proteins associated with gene regulation, protein expression, DNA repair, and cellular division. The cell accomplishes this feat by storing DNA in a highly-condensed structure called chromatin.

A mononucleosome is the simplest unit of chromatin and contains approximately 147 base pairs of DNA wound around a positively-charged protein octamer called a histone. Each histone complex is formed by two copies of the core histones H2A, H2B, H3, and H4, and a linker histone H1 that is bound to DNA between the nucleosomes [2]. Nucleosomes occur

roughly every 200 base pairs along the DNA strand, and the H1 linker histone is thought to facilitate the formation of nucleosome 3-dimensional structure [3].

The two main types of chromatin are euchromatin and heterochromatin, which perform different functions within the genome. Euchromatin is a less-dense, unfolded structure containing the active genes that are frequently transcribed in a genome. The nucleosomes of euchromatin are spread further apart, making the DNA more accessible for gene regulatory proteins and polymerases to bind [4]. Histones are post-translationally modified to regulate gene expression [5]. Examples of these epigenetic modifications include methylation, acetylation, and phosphorylation of the H3 core histone protein. These covalent histone modifications signal the expression or silencing of a gene. For example, acetylation, specific to euchromatin, is defined as the reversible addition of an acetyl (C_2H_3O) group to the positively-charged amino R group of lysine residues in histones [6]. The addition of an acetyl group by histone acetyl transferases (HAT) neutralizes the positive charge of the histones, reducing the binding affinity between the histones and the negatively-charged DNA. This allows the DNA to unwind from the histones and be more accessible to transcription machinery [6].

Heterochromatin is a highly-condensed form of DNA and constitutes a much smaller portion of the human genome [7]. There are two forms of heterochromatin that perform different functions. Constitutive heterochromatin contains repetitive stretches of DNA that form structural DNA components, such as telomeres and centromeres, which aid in chromosomal segregation during cell division. Constitutive heterochromatin is generally highly-methylated which renders it transcriptionally silent [7]. Facultative heterochromatin also contains genes that are silenced due to histone deacetylation or methylation. Although these genes are silenced, they influence the expression of other genes during certain periods of development and silences the genes that

are no longer needed. For example, facultative chromatin is found in inactive X chromosomes in female mammals and facilitates dose compensation to ensure that there is equal expression of X-linked genes in males and females [5].

ATP-Dependent Chromatin Remodeling

Nucleosomes are an effective way to pack and preserve DNA, but they block the binding of proteins needed for DNA repair, replication, and transcription. Chromatin remodeling is the conformational modification of chromatin to allow access for this machinery, which promotes DNA repair and gene expression. Chromatin remodeling is performed by ATP-dependent chromatin remodeling complexes that reposition nucleosomes. These complexes can remove histones and reposition the nucleosomes to expose the DNA for DNA repair, replication, and transcription machinery binding [8].

There are five identified families of chromatin remodeling complexes in eukaryotes: SWI/SNF, ISWI, INO80, SWR1, and NuRD/Mi-2/CHD. Each complex contains its own unique protein domains that are specific to the complex's function; however, each complex contains a catalytic ATPase domain that hydrolyzes ATP and uses the energy gained to perform its assigned function [9].

The SWI/SNF Chromatin Remodeling Complex

SWI/SNF (SWItch/Sucrose Non-Fermentable), first discovered in yeast, is an evolutionarily-conserved, eukaryotic multi-subunit chromatin remodeling complex that uses ATP to alter the position of nucleosomes, thus regulating transcription. The large 2MDa complex consists of at least nine protein subunits, including conserved and non-conserved proteins. The

four essential and conserved core subunits for chromatin remodeling are BRG1, INI1, BAF155, and BAF170 [10]. The other variable subunits might assist in directing specificity of the SWI/SNF complex by protein-protein interactions. The variable subunits incorporated depend on the tissue type, suggesting that these proteins have functions based on the needs of the specific cell type [10].

SWI/SNF's ATPase domain is in the BRG1 protein subunit [10]. The energy extracted from the hydrolysis of ATP allows SWI/SNF to slide the histones down the DNA strand or remove them completely. The exact mechanism of how SWI/SNF remodels chromatin is unknown due to the highly-intricate interactions between SWI/SNF, histones, DNA, and other associated proteins [11]. The BRG-1 subunit also contains a bromodomain, which is important in recognition and binding of acetylated lysine residues on histone tails [10].

The SWI/SNF complex binds to the minor groove of DNA when a specific set of DNA modifications are identified [12]. The complex possesses a large amount of DNA binding motifs because of the binding requirements of the various, non-core subunits [13]. More recent studies have suggested that SWI/SNF could also bind to DNA nonspecifically, which allows the complex to tightly bind to the nucleosome for efficient remodeling [13].

SWI/SNF still has many mysteries including its exact mechanism, its possible involvement in other cellular processes, and its associated proteins and their functions. These gaps in knowledge must be determined before researchers can begin treating SWI/SNF deficiencies. It is thought that the mammalian SWI/SNF complex acts as a tumor suppressor because various human cancer cell lines lack either the whole complex or one of its subunits [14]. In total, 20% of all human cancers contain a mutation in the SWI/SNF complex [15].

INI1

Integrase Interactor 1 (INI1) (also known as SNF5, BAF47, and hSNF5) is one of the four core, conserved subunits in the SWI/SNF complex. The human SMARCB1 (SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily b, member 1) gene codes for INI1 and is located on the long arm of chromosome 22 at the 22q.11.23 position. The gene contains nine exons that encode 385 amino acids and weighs approximately 47kDa [15]. INI1 is largely acidic, containing 27.5% charged residues. There are four identified isoforms in the human body but their distinct functions are unknown [15]. The isoform studied in this laboratory is Isoform B, which is 376 amino acids-long and is missing the 69^{th}-77^{th} amino acids of full-length INI1 (Isoform A). INI1 is highly-concentrated in the nucleus but is also observed in small amounts in the cytoplasm. INI1 is expressed in all tissue types with the highest prevalence in rapidly dividing tissues such as the testis, liver, lymph nodes, and brain [16].

There are four distinct, essential regions of INI1 that have been identified in all isoforms, shown in Figure 1 [17]. The potential DNA binding domain at the N-terminus spans from amino acids 1-186 and is highly basic with a pI of 9.3 [17]. The first imperfect repeat region (Rpt I) spans amino acids 186-245, while the second imperfect repeat region (Rpt II) spans amino acids 259-319. These sixty-amino-acid-long stretches are highly-charged and are thought to be involved in protein-protein interactions with c-MYC and HIV Integrase [18]. Rpt II contains a nuclear export sequence that allows INI1 to complete non-nuclear signaling[17, 19]. The C-terminal of the protein, spanning amino acids 334-376, forms a leucine-rich coiled-coil secondary structure and contains a nuclear localization sequence [18].

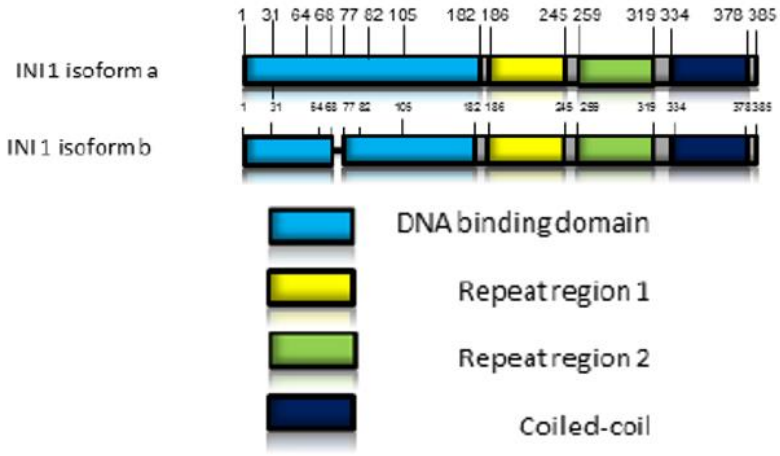


Figure 1: Domains of INI1 and their corresponding amino acid numbers [17].

The large number of aliases of INI1 indicates that INI1 is evolutionarily conserved as it has been identified in other organisms including *Drosophila* (BAF), *Saccharomyces*/yeast (SNF5), and humans (SMARCB1). Although its exact function is unknown, knockout studies of INI1 in murine models indicate that INI1 is essential for embryonic development as embryos lacking INI1 do not survive past the pre-implantation stage [20].

INI1 Mutations and Effects

Mutations of the SMARCB1 gene that cause a deletion or altered form of INI1 result in the development of rhabdoid predisposition syndrome [15]. This condition leads to increased risk of developing a highly lethal form of pediatric cancer called malignant rhabdoid tumors (MRT) [14]. MRTs are a rare but highly aggressive form of pediatric cancer that is prevalent in the kidneys, brain, and spinal cord of infants with mutations in the SMARCB1 gene. These tumors often quickly metastasize to the lymph nodes and lungs [15]. In mouse models, inactivation of the INI1 gene resulted in 100% of the test mice developing this form of cancer, strongly suggesting that functioning INI1 acts as a tumor suppressor [15]. A wide variety of mutations

have been associated with the development of MRTs including total homozygous deletions of the SMARCB1 gene, truncations of various exons, missense mutations, nonsense mutations, base-substitutions, and single-base deletions resulting in frameshifts [21, 22].

A specific type of MRTs called Atypical/Teratoid Rhabdoid Tumors (AT/RT) are the result of a *de novo* germline mutation of the SMARCB1 gene. In these cases, the parents have two copies of the wildtype SMARCB1 allele, but a mutation occurs during embryogenesis due to a flaw in one of the gamete's genome [23]. These tumors are characterized as highly aggressive and typically are seen in the central nervous system of children three-years-old or younger. Due to multiple risk factors including the aggressive nature of the cancer, the young age of the patients, the harsh side-effects of chemotherapy, and the tricky placement of the tumors, there is currently no fully- effective treatment protocol for AT/RT tumors. Because of these reasons, the two-year survival rate after diagnosis is less than 20% and the average post-diagnosis lifespan is 11 months [24].

Additional studies have also linked SMARCB1/INI1 mutations to tumors such as epithelioid sarcomas, schwannomas, and synovial sarcomas [25]. Although INI1disruption has been identified as a causal factor in these cancers, the native INI1 function and the mutated INI1 mechanism are still unknown. Therefore, researchers need to determine how INI1 functions, how it is regulated, and what it affects to provide insight on the role of INI1 in cancerous cells. The eventual goal of this vast body of research is to find an effective treatment and cure for the mutation and the cancer it causes.

INI1 Binding

DNA-binding proteins function by forming noncovalent bonds with DNA to either regulate its expression or change its confirmation to make the gene targets more accessible. These noncovalent interactions include hydrogen bonding, Van Der Waals interactions, and ionic bonding, which bind the protein to the DNA. These interactions occur on the DNA-binding domain of the protein, a conserved fragment that binds to a specific sequence of modifications on the DNA strand. The protein DNA-binding domains are typically found on the exterior of the protein so it has full access to the target DNA sequence [26].

Previous research conducted in Dr. Ruhl's laboratory determined that a portion of INI1 binds to DNA, suggesting that INI1 contains a DNA-binding domain. A previous graduate student, Dustin Steele (2013-2016), started the project of determining what region of INI1 binds to DNA. To begin, he cleaved the full-length 376 amino acid-long INI1 Isoform B into two pieces. The INI1(1-186) truncation demonstrated the ability to nonspecifically bind to "naked" DNA (DNA that is not wrapped around histones) as well as mononucleosomes [17]. The INI1(187-385) truncation did not demonstrate the ability to bind to either naked DNA or mononucleosomes, narrowing the search for DNA-binding regions to the first half of the protein [17]. Further unpublished data from Steele showed that INI1(1-144) does bind to both naked DNA and mononucleosomes, while INI1(63-144) does not bind to either naked DNA or mononucleosomes [27]. To further narrow the search for the minimum DNA binding domain, plasmids of INI1(31-125) and INI1(31-144) were created. These fragments were chosen based on their presence in tumor cells of AT/RT patients and further cleavage of the previously identified binding regions.

Project Goal

The goal of this research project was to express and purify two potential DNA binding domains of INI1, INI1(31-125) and INI1(31-144), and determine if these two regions bind to DNA. If either one of these truncations is the minimum DNA-binding domain of INI1, researchers can use this information to determine the function of INI1 and understand how mutations in INI1 lead to pediatric cancer.

Methods

Protein Expression

INI1(31-125) and INI1(31-144) had previously been cloned and transformed into the pSKB3 *E. coli* vector with a Kanamycin resistance gene insert generously provided by Dr. Deng's laboratory. Two separate variants of each truncation were made: one with the 6X His-tag on the C-terminus and one with the 6X His-tag on the N-terminus. Each of the four variants was expressed with the same procedure. Cells were picked from the cryovial and placed in 100mL of autoclaved Luria Broth (1% tryptone, 0.5% yeast extract, 1% NaCl, dH_2O, and 10mg/mL kanamycin) to incubate at 30°C at 250rpm overnight. This is the overnight culture. Then, 0.3mL of Kanamycin stock solution was added to the media to kill any contaminating cell that did not have the kanamycin insert.

Ultraviolet-visible (UV) Spectroscopy was used to read the optical density (OD) at λ=600 nm, which quantifies the concentration of cells in the medium by measuring the amount of light that is deflected by the cells. A higher OD number signifies a higher concentration of cells. One mL of overnight culture was added to 400mL of fresh Luria Broth and the OD was read. If the sample was below an OD of 0.06, another milliliter of overnight culture was added until the desired OD was reached. An OD of 0.06 suggests that the *E. coli* were about to enter the log phase of growth. The flasks were incubated at 37°C and 250rpm until the OD reached 0.6. An OD of 0.6 was chosen to ensure that the *E. coli* were in the logarithmic phase of growth as this phase is ideal for protein expression. At this point, the flasks were removed from the shaker and allowed to cool to room temperature. Once cool, the flasks were transferred to a refrigerated shaker at 18°C and 250rpm. The INI1 truncation expression was induced by adding 400µL of

isopropyl-β-D-1-thiogalactopyranoside (IPTG) to each flask. The flasks were left in the refrigerated shaker overnight to allow protein synthesis.

The flasks were removed from the refrigerated shaker 18-22 hours after induction. The cells were spun down into a pellet using a GSA rotor made for the DuPont Sorvall RC-5B Superspeed Centrifuge at 4,000rpm for 30 minutes. The supernatant was discarded, and the pellets were stored at -80°C.

Protein Purification

All protein truncations were purified in a similar manner. Thawed pellets were suspended in 20mL of Buffer A (20mM Tris, 500mM NaCl, 10% Glycerol, 20mM Imidazole, pH 7.8), vortexed until dissolved in buffer, then placed on ice. To rid the sample of some possible contaminants, lysozyme and DNase were added. Ten milligrams of dehydrated lysozyme were dissolved in 1mL of dH_2O and left on ice for 40 minutes before adding to each pellet. Each pellet received 5µL of Invitrogen Amplification Grade DNase I before sonication. Each pellet sample was poured into a 50mL glass beaker and kept on ice. The samples were sonicated in 10-second cycles for a total time of two minutes. The combination of salt in Buffer A and the bombardment of sound waves produced by the sonicator burst the cell membranes, releasing all cellular contents, including the protein of interest. The samples were centrifuged in a SA-600 rotor made for the DuPont Sorvall RC-5B at 12,000rpm for 30 minutes at 4°C. The supernatant containing the cellular contents was collected in 50mL falcon tubes and stored at -80°C until further purification. A 50µL sample of each supernatant was collected in a 1mL microfuge tube, labeled as “cell extract,” and placed in the -80°C freezer.

The 6X poly-histidine tail (6X His-tag) on either terminal was added because of its high affinity and specificity to Ni-NTA agarose resin for affinity column chromatography. The Ni-NTA resin was pre-equilibrated with Buffer A. After equilibration, 200µL of Ni-NTA resin was transferred to each falcon tube of supernatant (about 25mL) using a p1000 pipetman with roughly ½ cm of the tip cut off to not break the resin molecules. The supernatant and resin were rocked for one hour at 4°C, then poured over a 25mL gravity flow column to collect the resin and allow the liquid to flow through. If the 6X His-tag is not hidden within the protein, it should bind to the resin and stay on the column while proteins without the 6X His-tag or a similar highly-positive region flow through into the collection tube. The column was washed with 100mL of 4°C Buffer A to wash off any unbound protein that was caught in the resin. A portion of the flow-through was pipetted back into the binding falcon tube to wash off any remaining resin and was poured back on the column. All flow through and wash samples were collected and stored in -80°C. Ten milliliters of Buffer B (20mM Tris, 500mM NaCl, 10% Glycerol, 250mM Imidazole, pH 7.8) was pipetted carefully over the column to not disturb the resin. The high imidazole concentration in Buffer B has a high affinity for the binding regions on the resin and knocks the 6X His-tagged proteins off the resin. From this step, ten 1mL fractions were collected.

To ensure the isolation and purification steps worked, a 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used. The SDS added to the sample denatures the proteins and coats the proteins in a negative charge, allowing the gel electrophoresis to separate by molecular weight only. Using the calculated molecular weight of the protein, it is possible to determine if the protein of interest is present. A 19µL sample was taken from the cell extracts, flow-through samples, wash samples, and each elution fraction (1-

10). To each 19µL sample, 5µL of 10% SDS and 8 µL of 4X sample buffer (0.25M Tris-Cl, pH=6.8, 2% SDS, 10% β-mercaptoethanol, 20% glycerol, 0.01% bromophenol blue) were added. Each sample was boiled for 60 seconds and 20µL of each sample was loaded onto the acrylamide gel. Ten microliters of BioRad Precision Plus unstained protein standard molecular weight ladder were also loaded on the gel. The gel was run at 150V for one hour at room temperature. After running the gel, the stacking gel was cut off, and the gel was rocked in dH_2O three times for ten minutes each to remove the Running Buffer. Thermo Fisher Scientific GelCode™ Blue was used to stain the gel. Imaging was performed with a BioRad ChemDoc™ XRS^+ and a white light conversion tray.

Purification Optimization Tests

After analyzing the N-terminus 6X His-tag variants of INI1(31-125) and INI1(31-144), flow-through samples from these variants were used to complete binding tests to determine if the proteins would bind under different conditions. Only the INI1(31-144) variant was used as it was assumed that both protein variants would react similarly.

For the first binding test, 40µL of INI1(31-144) flow-through sample were placed in each microfuge tube with 10µL of Ni-NTA pre-equilibrated with Buffer A. A control sample was set aside containing only flow-through (suspended in Buffer A from purifications steps) and resin. The second condition halved the concentration of the buffer by adding 40µL of dH_2O. The third condition added 40 µL of 8M urea lysis buffer (100mM Phosphate, 10mM Tris-Cl, 8M Urea, pH 8.0). The fourth condition added 1.2µL of 10% SDS (to a total SDS concentration of 0.3%). The fifth condition added 1µL of Tween 20 (to a total Tween 20 concentration of about 2%). The sixth condition added 1µL of Tween 80 (to a total Tween 80 concentration of about 2%). The

seventh and final condition added 1µL of Triton X-100 (to a total Triton-X concentration of about 2%). The samples rocked overnight at 4°C.

All seven condition tubes were put in a microfuge and spun to 4,000rpm two times. Using a P200 gel-loading tip, the liquid was removed from each resin tube and placed in a fresh, sterile microfuge tube, creating the supernatant tubes. The resin tubes were washed with 50µL of dH_2O by flicking the tube to mix the resin with the water. This step washes away unbound proteins that could have become trapped between the resin beads. The wash was discarded. To each of the resin tubes, 5µL of 10% SDS and 8 µL of 4X sample buffer were added. For each of the supernatant tubes, 19µL of supernatant was transferred to a new microfuge tube and were mixed with 5µL of 10% SDS and 8 µL of 4X sample buffer. All samples were boiled for one minute.

To visualize if the protein bound to the resin or stayed in the supernatant, all samples were run on a 15% SDS-PAGE gel. From the liquid supernatant tubes, 20µL were loaded onto the gel using gel-loading tips. For the resin sample tubes, all liquid in the tube and some resin were loaded onto the gel using a P200 tip. Ten microliters of BioRad Precision Plus unstained protein standard molecular weight ladder were also loaded on the gel. The gel was run at 150V for one hour at room temperature. After running the gel, the stacking gel was cut off, and the gel was rocked in dH_2O three times for ten minutes each to remove the Running Buffer. Thermo Fisher Scientific GelCode™ Blue was used to stain the gel. Imaging was performed with a BioRad ChemDoc™ XRS^+ and a white light conversion tray.

For the second binding test, adjustments were made to see if they made a difference in binding. The procedure was consistent with the first binding test written above; therefore, only procedural changes will be listed below. Only four conditions of the previous conditions were

tested again: control, half-concentration of Buffer A, urea, and 10% SDS. These samples were prepared with the exact same conditions as the previous run. A new lane containing 40µL of N-terminal 6X His-tag INI1(31-125) flow-through and 10µL of resin was added to test if the correct variant was present. The samples were rocked in 4°C for 1.5 hours and were mixed by flicking every 15 minutes to ensure that the solution was well-mixed. Instead of washing with dH_2O, all resin samples were washed with 500µL of Buffer A to remove unbound proteins and the wash was removed with a gel-loading tip. Sample prep did not change. All samples, including resin samples, were loaded onto the 15% SDS-PAGE gel using gel-loading tips to prevent spill over. The gel was run, processed, and visualized using the same methods as above.

Results

Protein Expression and Purification

A 15% SDS-PAGE gel was used to assess the presence and purity of the INI1(31-125) and INI1(31-144) fragments. The first lane of each purification gel contains the cell extract. The cell extract sample includes all the proteins expressed by the *E. coli* and should, therefore, have a wide assortment of bands of varying concentrations and molecular weights. Because the *E. coli* contain a plasmid that codes for the desired truncations of INI1, the cell extract lane should have a highly-concentrated band around the molecular weight of the INI1 truncation if the protein was properly expressed.

The second lane of each gel contains the flow-through sample collected following the loading of the cell extract onto the resin. The positively-charged 6X His-tag on the termini of INI1 binds to the negatively-charged resin while the other cellular components flow through. If the 6X His-tag is positioned on the exterior portion of the INI1 truncation where it is free to bind

to the resin, the flow-through sample should contain all proteins from the cellular extract except for the 6X His-tagged INI1.

The third lane of each gel contains the wash sample obtained after running Buffer A over the column to wash off any unbound protein. This lane is expected to have bands of various molecular weights. Because the wash sample dilutes the protein in 100 mL of Buffer A, faint bands are expected.

Elution fractions 1-10 were taken after running Buffer B over the column and were loaded on the gel. Buffer B contains a high concentration of imidazole which is part of the side chain of the Histidine residues that make up the 6X His-tag. When high concentrations of imidazole interact with the resin, it competes with the 6X His-tag for binding sites and causes the His-tagged protein to fall off the resin and into the elution tube. Therefore, these fractions should contain a strong band around the molecular weight of the protein of interest and little to no other bands. Bands other than the protein of interest indicate that the elution sample is not pure and will need further purification techniques to completely purify the sample.

Theoretical molecular weights of the protein fragments were determined by the molecular weight calculator provided by ExPASy.org. INI1(31-125) has a molecular weight of 10,990Da and INI1(31-144) has a molecular weight of 13,034Da. The 6X His-tag contains six histidine residues and additional linker amino acids. If the 6X His-tag is added to the N-terminus, it contains 20 linker amino acids and weighs 2,096Da. If the 6X His-tag is added to the C-terminus, it contains 22 linker amino acids and weighs 2,419Da. The exact linker amino acid sequences of the 6X His-tag are shown in Appendix 1, provided by Novagen. Adding the 6X His-tag to the molecular weights of each truncation yields the total molecular weights listed below.

Truncation	6X His-tag Location	Molecular Weight (Da)
INI1(31-125)	N-terminus	13,086
INI1(31-125)	C-terminus	13,409
INI1(31-144)	N-terminus	15,130
INI1(31-144)	C-terminus	15,453

A dark band around these molecular weights on the gel could indicate the presence of the INI1 truncation. It is also important to note the level of uncertainty of SDS-PAGE. On SDS-PAGE gels, proteins drift up to 2kDa from their correct molecular weight due to different factors [28].

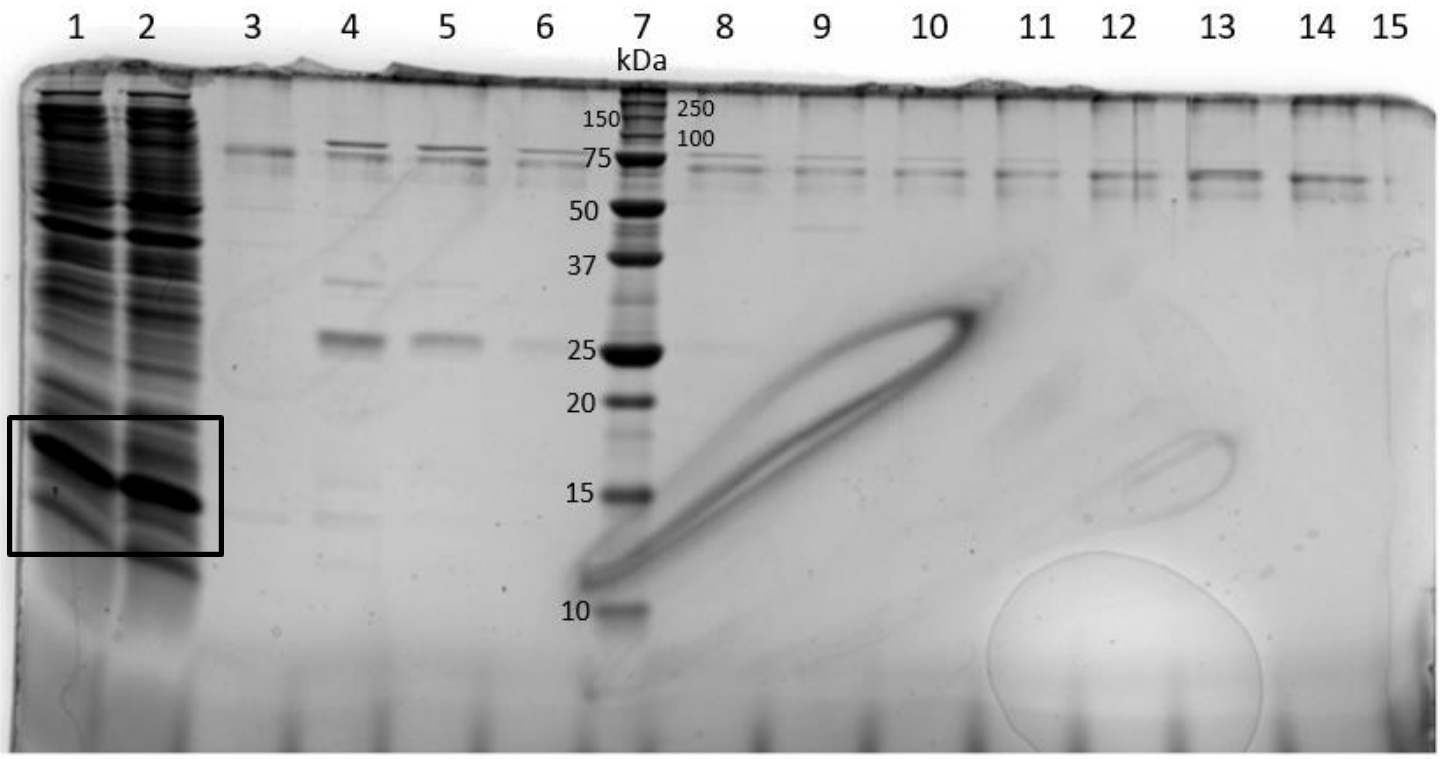


Figure 2: 15% SDS PAGE to ensure expression of N-terminus 6X Histagged INI1 (31-125) using Ni-NTA resin binding. Lane 1: cell extract fraction. Lane 2: flow-through fraction. Lane 3: wash fraction. Lanes 4-6: elution fractions 1-3. Lane 7: molecular weight standard (kDa). Lanes 8-14: elution fractions 4-10.

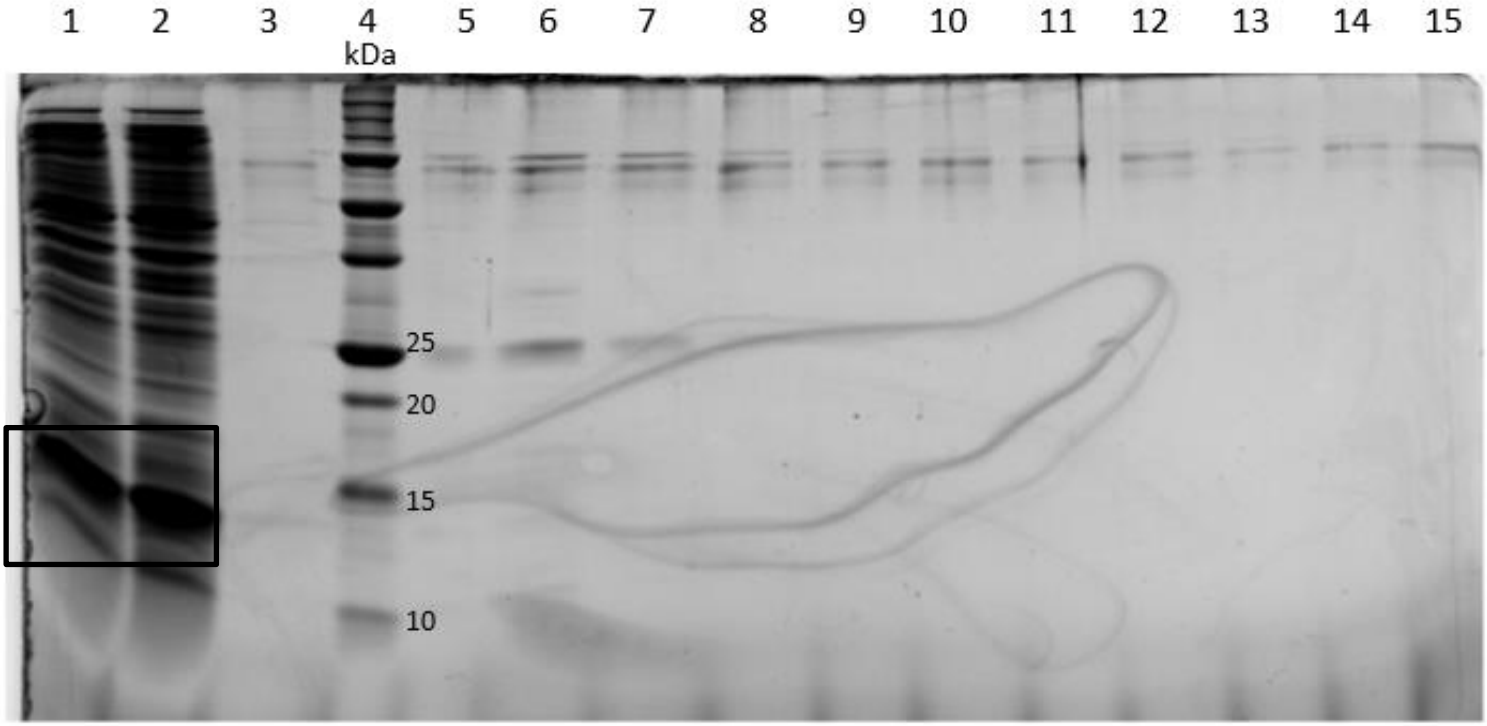


Figure 3: 15% SDS PAGE to ensure expression of N-terminus 6X Histagged INI1 (31-144) using Ni-NTA resin binding. Lane 1: cell extract fraction. Lane 2: flow-through fraction. Lane 3: wash fraction. Lane 4: molecular weight standard (kDa). Lane 5: empty. Lanes 6-15: elution fractions 1-10.

Figure 2 and Figure 3 show the expression and purity of the N-terminal 6X His-tagged INI1(31-125) and INI1(31-144) truncations, respectively. Figure 2 shows no strong bands around 13kDa in the elution fractions. In both Elution 1 lanes on both gels, there are very light bands around the correct molecular weights of the proteins that could be the protein of interest that are not observed in other lanes. There are dark bands around 13kDa in the cell extract fraction lane and the flow-through fraction lane (indicated by the box). Figure 3 shows similar results as no prominent bands are present around 15kDa in the elution fractions. Dark bands are present in the cell extract fraction lane (indicated by the box) and the flow-through fraction lane around 15kDa. The elliptical marks on both gels are due to dye bubbles becoming trapped under the gel after the rocker broke.

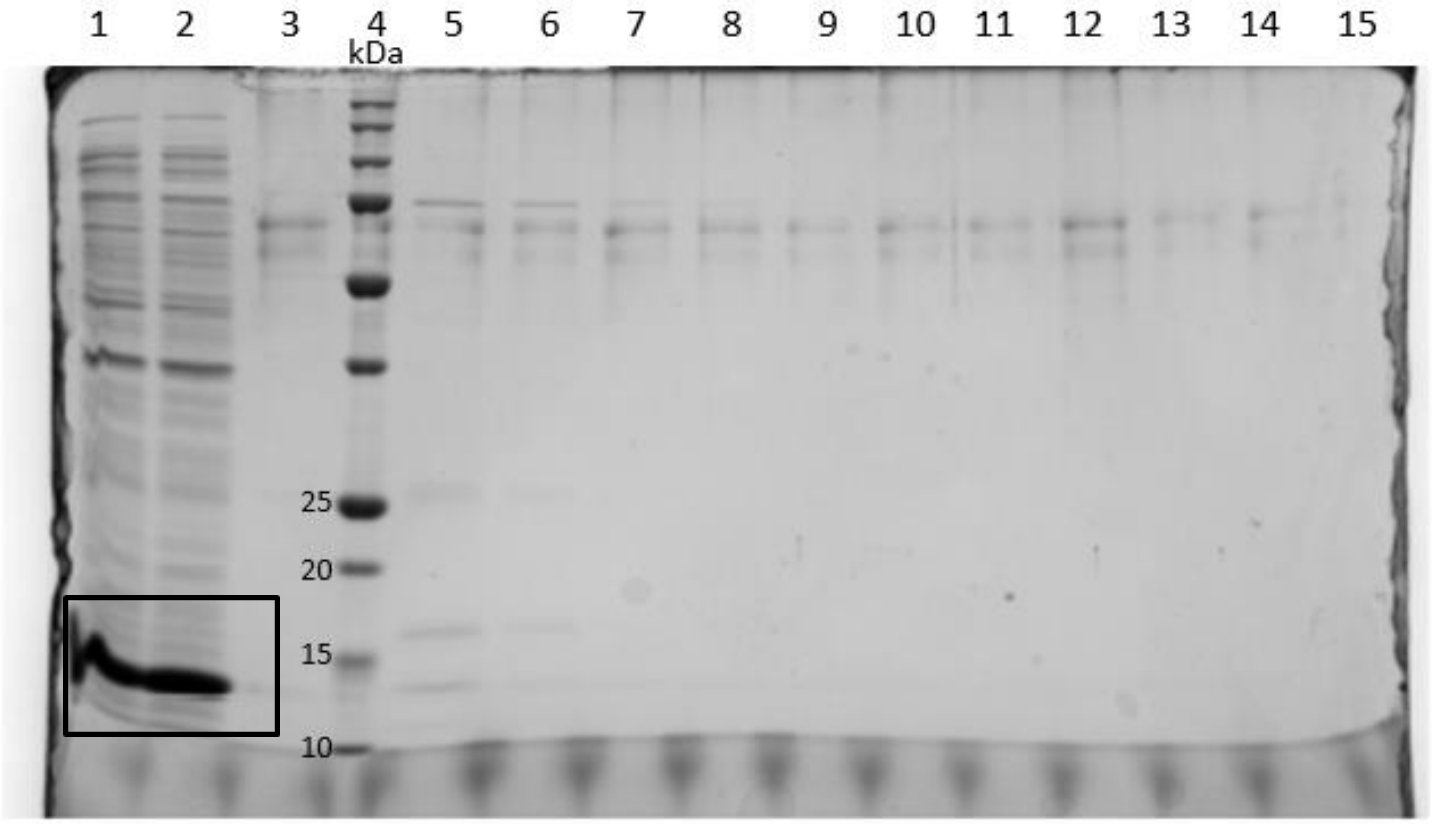


Figure 4: 15% SDS PAGE to ensure expression of C-terminus 6X Histagged INI1 (31-125) using Ni-NTA resin binding. Lane 1: cell extract fraction. Lane 2: flow-through fraction. Lane 3: wash fraction. Lane 4: molecular weight standard (kDa). Lane 5-14: elution fractions 1-10.

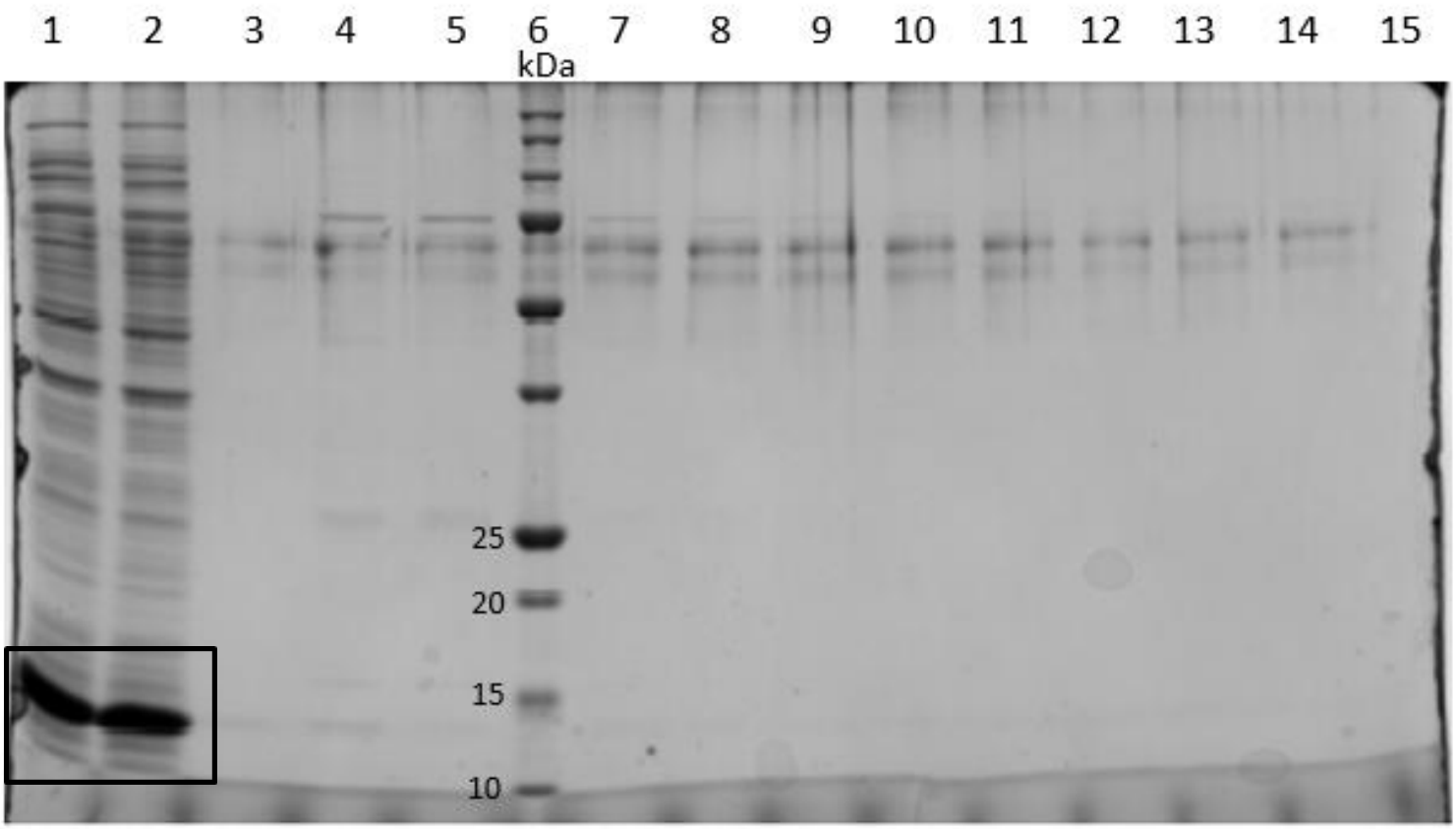


Figure 5: 15% SDS PAGE to ensure expression of C-terminus 6X Histagged INI1 (31-144) using Ni-NTA resin binding. Lane 1: cell extract fraction. Lane 2: flow-through fraction. Lane 3: wash fraction. Lanes 4-5: elution fractions 1-2. Lane 6: molecular weight standard (kDa). Lanes 7-14: elution fractions 1-10.

Figure 4 and Figure 5 show the expression and purity of C-terminal 6X His-tagged INI1(31-125) and INI1(31-144) truncations, respectively. The results for the N-terminal 6X His-tagged INI1 gels are consistent with the results from the C-terminal 6X His-tag INI1 gels. In both C-terminal 6X His-tagged INI1 gels, a very light band at the correct molecular weight is present in the Elution 1 lanes only. Both cell extracts and the flow-through samples show a dark band slightly under 15kDa on both gels (shown in boxes).

Purification Optimization Tests

To determine if the INI1 truncations would bind to the Ni-NTA resin under different conditions, binding tests were conducted. The flow-through samples from protein expression methods were mixed with the pre-equilibrated resin and the variable component. After allowing time for binding, the supernatant was removed from each tube and the remaining resin beads were boiled. When Ni-NTA resin is boiled for one minute after adding the sample prep materials, the resin-bound proteins unbind and solubilize into solution. If the protein did not bind to the resin, a strong band should appear in the supernatant sample. If the protein did bind to the resin, a strong band should appear in the resin sample.

For the binding tests, only INI1(31-144) was used. It was assumed that if one condition worked for INI1(31-144), it would also work for INI1(31-125) because they are similar in size, sequence, and molecular weight.

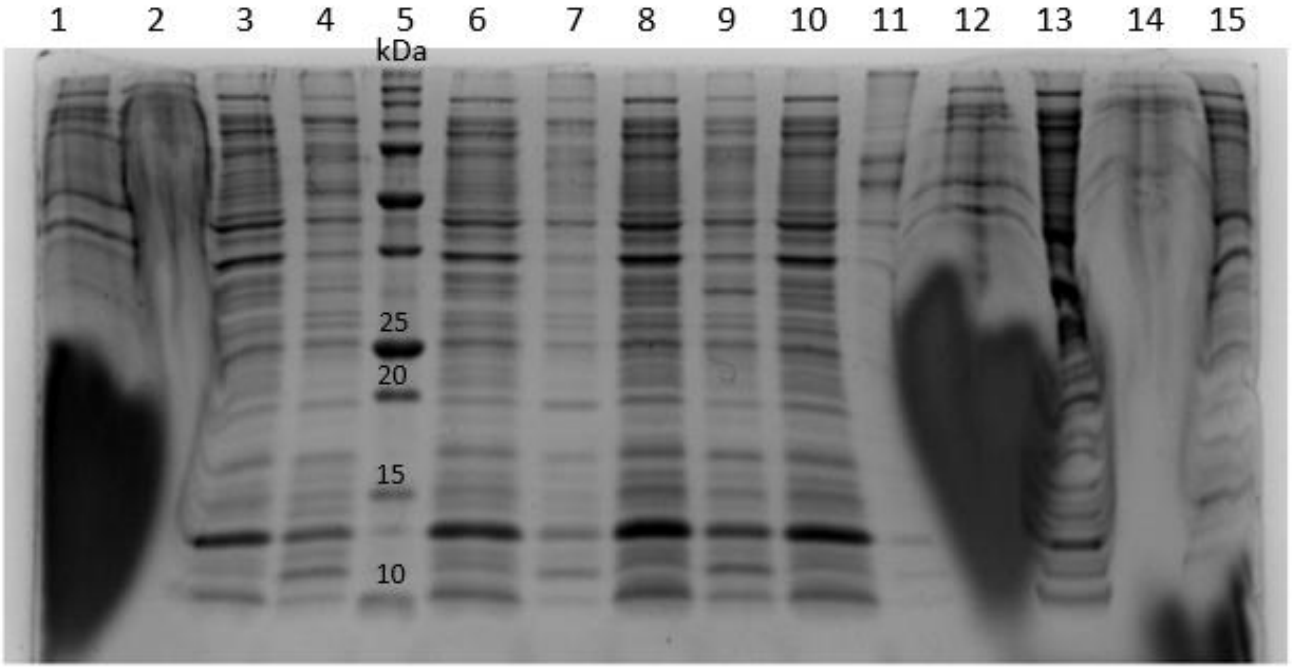


Figure 6: 15% SDS-PAGE of first binding test of N-terminus 6X Histagged INI1(31-144). Lane 1: Triton X-100 supernatant. Lane 2: Triton X-100 resin. Lane 3: half buffer A concentration supernatant. Lane 4: half buffer A concentration resin. Lane 5: molecular weight standard (kDa). Lane 6: urea supernatant. Lane 7: urea resin. Lane 8: control supernatant. Lane 9: control resin. Lane 10: 10% SDS supernatant. Lane 11: 10% SDS resin. Lane 12: Tween 20 supernatant. Lane 13: Tween 20 resin. Lane 14: Tween 80 supernatant. Lane 15: Tween 80 resin.

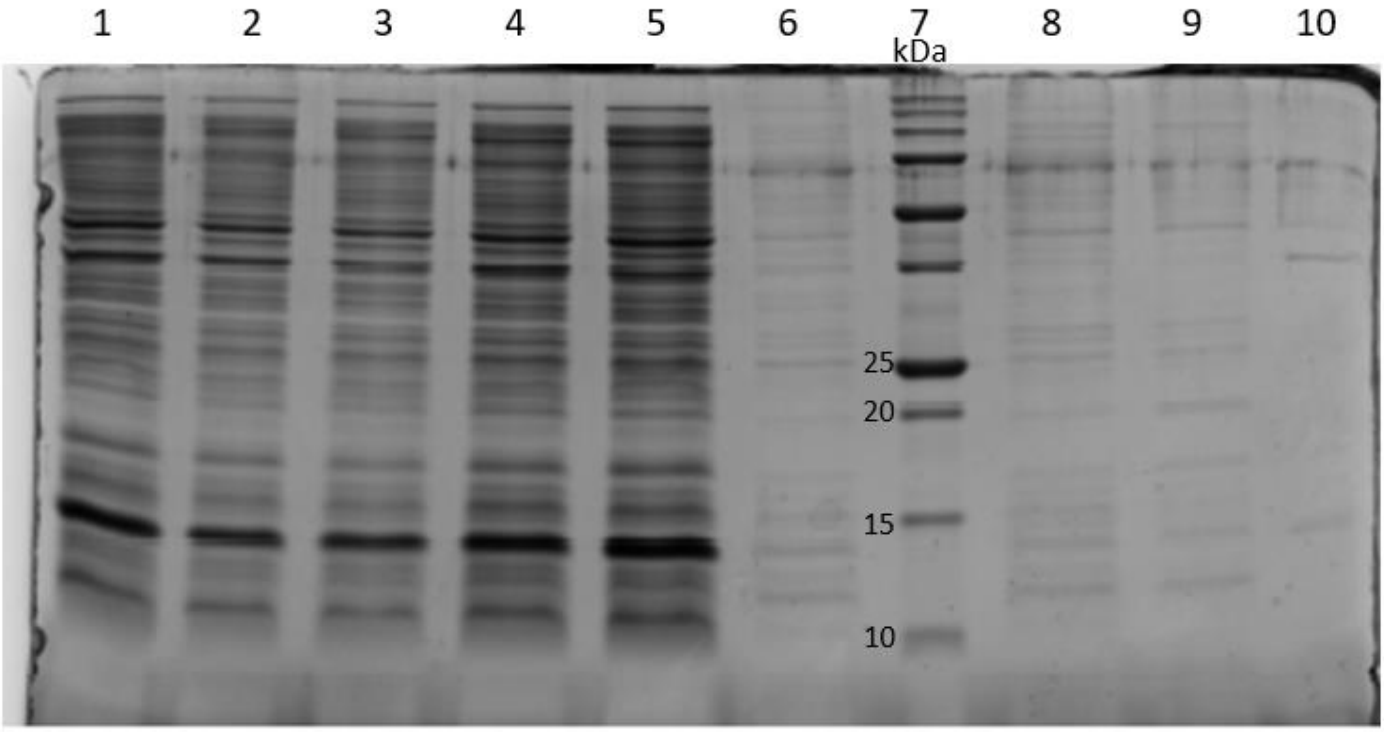


Figure 7: 15% SDS-PAGE of second binding test of N-terminus 6X Histagged INI1(31-144). Lane 1: control supernatant. Lane 2: half-buffer A concentration supernatant. Lane 3: urea supernatant. Lane 4: 10% SDS supernatant. Lane 5: INI1(31-125) flow-through sample test lane. Lane 6: control resin. Lane 7: molecular weight standard (kDa). Lane 8: half-buffer A concentration resin. Lane 9: urea resin. Lane 10: 10% SDS resin.

The first binding test in Figure 6 has significant distortion of lanes 1-3 and 11-15, rendering them unreadable. The remaining lanes show a multitude of bands in each lane as well as a strong band slightly under 15kDa. Some of the resin sample lanes look identical to their supernatant samples but are slightly fainter. Due to these results, the second binding test illustrated in Figure 7 was conducted. Figure 7 shows highly-concentrated samples in lanes 1-5. Lane 5, which contains the N-terminal 6X His-tagged INI1(31-125) sample has a dark band under 15kDa that is a smaller molecular weight than the heaviest band in lanes 1-4. Additionally, lanes 6, 8, 9, and 10 show no distinct bands.

Discussion

Protein Purification and Expression

The purpose of these four gels was to determine the presence and purity of the INI1 truncations. In Figure 2 and Figure 3, only very faint bands with the correct molecular weight of the N-terminal 6X His-tagged INI1 truncations were found in the elution 1 sample lanes. However, there are strong bands slightly under the 15kDa marker in both the cell extract and flow-through samples of both the INI1(31-125) and INI1(31-144) gels (indicated by boxes). This finding suggests that although the protein was highly expressed by the *E. coli*, the protein only slightly bound to the resin and a predominate portion of the protein flowed-through the column with the rest of the cellular proteins.

Another method can be used to determine the approximate molecular weight of a protein band on the gel. A standard curve can be made by measuring how far each of the molecular weight ladder bands migrated from the top of the gel. These measurements create a standard curve and a trendline equation. To approximate the molecular weight of the proteins on the gel, the migratory distance of the band is measured in centimeters and plugged in for the X value of the equation. A sample standard curve and trendline equation for N-terminal 6X His-tagged INI1(31-144) from the gel in Figure 3 are shown in Figure 8. In this example, the strong bands in the cell extract and flow-through samples in Figure 3 are 3.85cm from the top of the gel. Plugging 3.85 into the trendline equation yields a molecular weight of 14,885Da which is about 250Da smaller than the theoretically calculated 15,130Da molecular weight of N-terminus 6X His-tagged INI1(31-144). Although this mathematical approximation method is often close to the theoretical value, it does have its limitations. If the R^2 value is less than 0.985, molecular weights begin shifting from theoretical values by up to 1,000Da. In addition, the accuracy of the standard

curve is limited by the uncertainty of measuring with a ruler that does not have smaller dimensions.

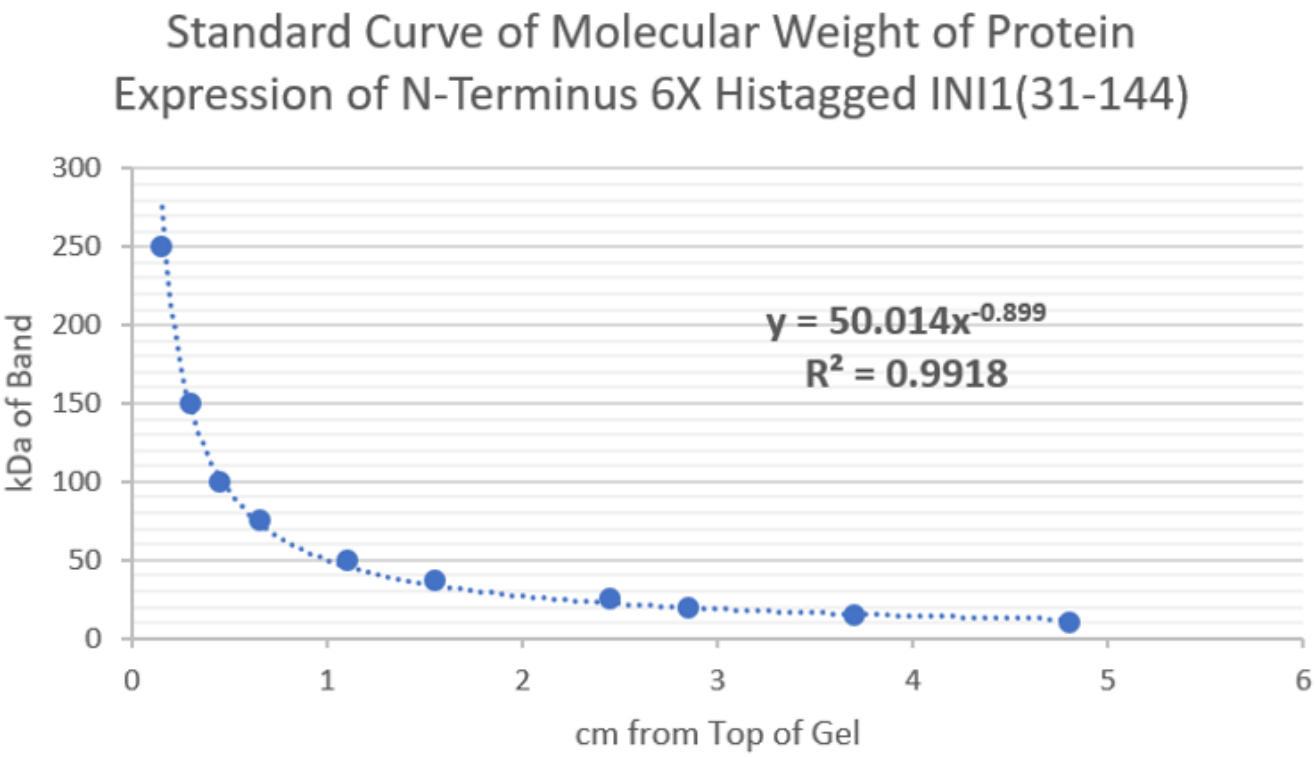


Figure 8: Standard Curve of Molecular Weight of Protein Expression of N-Terminus 6X Histagged INI1(31-144).

For the INI1(31-125) gel in Figure 2, the highly-concentrated bands in the cell extract and flow-through lanes and the light band in elution sample 1 are slightly below 15,000Da. The standard curve for this gel is poor as the R^2 value is 0.982 and the approximation for the molecular weight is 15,004Da. Because the N-terminal 6X His-tagged INI1(31-125) protein is approximately 13,086Da, it is possible that this prominent band is our protein of interest. This is consistent with the light band observed in Elution sample 1. For the N-terminal 6X His-tagged INI1(31-144) gel in Figure 3, the highly concentrated bands in the cell extract and the flow-

through and the light band in the elution 1 lane are below 15,000Da. The N-terminal 6X His-tagged INI1(31-144) protein should be around 15,130Da which leads to three possible explanations for the difference in molecular weights. The first possible explanation is that this is not the protein of interest. Although the darkest band appears to be around the same molecular weight as the desired protein, it is possible that this is not the desired protein. The second possible explanation is that the protein might not be made correctly. The INI1(31-144) protein weighs 13,034Da before the 6X His-tag is added. The dark bands of the cell extract and flow-through samples in Figure 3 lie around roughly 14,000Da; therefore, this could indicate that this is the protein of interest but the 6X His-tag is not present. The third possible explanation is that the shift is caused by the uncertainty of SDS-PAGE gels. As mentioned before, bands on gels can shift up to 2,000Da from their correct molecular weights depending on the properties of the gel and materials [28].

The gels in Figure 4 and Figure 5 show the presence and purity of the C-terminal 6X His-tagged INI1 truncations. These gels show identical results to the gels from the N-terminal 6X His-tagged INI1 truncations. Again, there are highly-concentrated bands in the cell extract and flow-through samples but only very faint bands near the molecular weight of the protein of interest in the elution 1 sample lanes. Figure 4 of C-terminus 6X His-tagged INI1(31-125) shows dark bands in the cell extract and flow-through that are slightly under 15kDa (std curve = 15,469Da), indicating that the dark bands are most likely the target protein but are not strongly binding like in the N-terminal 6X His-tagged INI1(31-144) gel. Figure 5 of C-terminus 6X His-tagged INI1(31-144) also has dark bands below 15,000Da (std. curve = 15,445Da) in the cell extract and flow-through samples. The calculated molecular weight is very close to the theoretical weight of the C-terminus 6X His-tagged INI1(31-144) protein, suggesting that the 6X

His-tag could be present but hidden, resulting in it not binding to the column. These results indicate that only a small amount of protein bound to the resin, suggesting that there is either a problem with the His-tag, the protein folding, the resin, or the conditions in which the protein bound.

It is worth noting that a doublet appears around 100kDa on every gel. These are common contaminant proteins that are present when performing Ni-NTA affinity chromatography with pSKB3 vectors in *E. coli* cells. These proteins probably contain a stretch of Histidine residues or positively-charged residues that mimic a 6X His-tag.

Overall, these four gels show that the location of the 6X His-tag on the INI1 protein does not matter, as both termini 6X His-tags in both INI1 truncations did not bind to the resin. Because of this finding, a series of binding tests were performed to determine if the protein did not bind to the resin because of the binding conditions of the solution or if there is another issue present.

Purification Optimization Tests

Because binding relies on the presence of the 6X His-tag, there are two possibilities that could explain why the INI1 truncations did not strongly bind. The first is that the 6X His-tag is not on the protein. Although the presence of the 6X His-tag was verified when the plasmid was sequenced, it is possible that the *E. coli* expressed the INI1 truncations with the 6X His-tag attached but cleaved the 6X His-tag off during post-translational processing. The uncertainty of the SDS-PAGE, the uncertainty of the mathematical molecular weight calculation, and the small size of the 6X His-tag makes it difficult to declare that the 6X His-tag is missing based on molecular weight approximations alone. If the 6X His-tag is not present, the INI1 truncations

will not bind to the 6X His-tag specific resin. The other possibility is that the 6X His-tag is folded into the interior of the protein. Full-length INI1 labeled with 6X His-tag can be successfully purified using the Ni-NTA resin, suggesting that the 6X His-tag is present on the exterior of the full-length protein and able to bind to the resin. However, because both INI1 truncations are shorter forms of the protein, it is possible that they fold differently due to the absence of additional hydrophobic and hydrophilic residues.

To test if either of these possibilities is the reason for the lack of binding, different types of denaturants/detergents as well as different buffer conditions were used. The first binding test shown in Figure 6 tested binding in unchanged conditions (control), half the buffer A concentration, urea lysis buffer, 10% SDS, Triton X-100, Tween 20, and Tween 80 to see if it affected the binding. There were several issues with this gel. The most obvious is that the Triton X-100, Tween 20, and Tween 80 lanes caused severe distortion of their lanes and surrounding lanes. This could be because Triton X-100, Tween 20, and Tween 80 are all non-ionic, hydrophobic detergents, which could explain the diffusion of the sample through the hydrophilic gel instead of towards the positive pole. This drift could also have been caused by too much Triton X-100, Tween 20, and Tween 80 in the samples themselves. These three detergents were too viscus to pipette normally so it is possible that too much Triton X-100, Tween 20, and Tween 80 could have been the issue. In addition, when comparing the supernatant lane and the resin lane of the same condition, the proteins present are nearly identical. It is expected that the proteins missing from the supernatant lane should be the only proteins found in the resin lane; however, this is not the case with this gel. This is likely caused by loading the gel with standard P20 tips as their wide tip caused spill over into other wells. Lastly, when looking at the urea condition, it looks like the urea supernatant sample contains most of the desired protein. This is

worrisome because the urea binding condition result is the most reliable as a high urea concentration should completely denature the protein, thus revealing the 6X His-tag for binding. If the 6X His-tag is present, it is expected to cause the protein to strongly bind to the resin and have a dark band in the urea resin sample lane. However, this pattern was not observed. Because of these issues, the binding tests were repeated with tweaks to the procedure.

Figure 7 shows the results of the second binding attempt. Three changes were made to the procedure: 1) the Triton X-100, Tween 20, and Tween 80 conditions were removed to eliminate the risk of another distorted gel, 2) gel loading tips were used to load all samples onto the gel to minimize spill over, and 3) the resin tubes were flicked every 15 minutes during the binding step to ensure that the solution was properly mixed. A test lane (Lane 5) with non-resin-bound flow-through INI1(31-125) was run to ensure that the samples are not the same. When comparing the prominent dark line in Lane 5 to the dark lines in Lanes 1-4, the band in Lane 5 is observed at a slightly lower molecular weight, indicating that the two samples do contain different INI1 truncations.

The gel in Figure 7 is much neater than the gel in Figure 6, and spillover was kept to a minimum during loading of the gel. The Lanes 2-4 on the left of the gel are the supernatant samples while the right side of the gel contains the resin samples. There is no distinguishable band in any of the resin sample lanes, suggesting that the truncated INI1 protein did not bind to the resin under any of the conditions. Because the conditions contained denaturants, the 6X His-tag should have been exposed, allowing the protein to bind. This data as well as the varying molecular weight discrepancies of both truncations suggest that the 6X His-tag is not present on the protein or is not functioning properly. If the 6X His-tag was present and fully-functioning,

there should have been more 6X His-tagged INI1 bound to the resin and in the resin lane of one of the detergent conditions in the second binding test.

Future Studies

The most imminent goal is to find a way to isolate INI1(31-125) and INI1(31-144). From the findings, either the 6X His-tag is not present or the 6X His-tag is not functioning as intended. Because of this, a new purification approach must be used such as ion exchange chromatography or size-exclusion chromatography. The project cannot continue until an efficient purification procedure for these truncations is determined. A purification optimization test was not performed on the C-terminal 6X His-tagged INI1(31-144) or either INI1(31-125) sample. It is possible that these truncations might bind more to the resin than the N-terminal 6X His-tagged INI1(31-144).

The eventual goal of this project is to isolate INI1(31-125) and INI1(31-144), potential DNA binding regions, and perform gel shift assays to determine if these domains bind DNA. Although this project did not reach this step, this is still the goal of the project and these tests will be completed in the future if effective purification techniques are discovered.

The goal of studying INI1 is to determine its function and understand its mutation's role in causing cancer. By determining its binding regions, researchers can better understand how INI1 functions and this information can be used to develop a treatment or cure.

Works Cited

[1] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. Science. 2001;291:1304-51.

[2] Mirabella AC, Foster BM, Bartke T. Chromatin deregulation in disease. Chromosoma. 2016;125:75-93.

[3] Zhou B-R, Feng H, Kato H, Dai L, Yang Y, Zhou Y, et al. Structural insights into the histone H1-nucleosome complex. Proceedings of the National Academy of Sciences. 2013;110:19390.

[4] International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931-45.

[5] Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell research. 2011;21:381-95.

[6] Fan J, Baeza J, Denu JM. Chapter Six - Investigating Histone Acetylation Stoichiometry and Turnover Rate. In: Marmorstein R, editor. Methods in Enzymology: Academic Press; 2016. p. 125-48.

[7] Saksouk N, Simboeck E, Déjardin J. Constitutive heterochromatin formation and transcription in mammals. Epigenetics & Chromatin. 2015;8:3.

[8] Teif VB, Rippe K. Predicting nucleosome positions on the DNA: combining intrinsic sequence preferences and remodeler activities. Nucleic acids research. 2009;37:5641-55.

[9] Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, part II: ATP-dependent chromatin remodeling. Trends in Molecular Medicine. 2007;13:373-80.

[10] Roberts CWM, Orkin SH. The SWI/SNF complex — chromatin and cancer. Nature Reviews Cancer. 2004;4:133-42.

[11] Tang L, Nogales E, Ciferri C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. Progress in biophysics and molecular biology. 2010;102:122-8.

[12] Quinn J, Fyrberg AM, Ganster RW, Schmidt MC, Peterson CL. DNA-binding properties of the yeast SWI/SNF complex. Nature. 1996;379:844-7.

[13] Mohrmann L, Verrijzer CP. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2005;1681:59-73.

[14] Versteege I. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature. 1998;394:203-6.

[15] Kim KH, Roberts CWM. Mechanisms by which SMARCB1 loss drives rhabdoid tumor growth. Cancer genetics. 2014;207:365-72.

[16] Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. Molecular & Cellular Proteomics. 2014;13:397.

[17] Steele D. Characterizations of Intergrase Interactor 1 Binding Chromatin. Oklahoma State University. 2016.

[18] Stojanova A, Penn LZ. The role of INI1/hSNF5 in gene regulation and cancer. Biochemistry and Cell Biology. 2009;87:163-77.

[19] Craig E, Zhang Z-K, Davies KP, Kalpana GV. A masked NES in INI1/hSNF5 mediates hCRM1-dependent nuclear export: implications for tumorigenesis. The EMBO journal. 2002;21:31-42.

[20] Klochendler-Yeivin A, Fiette L, Barra J, Muchardt C, Babinet C, Yaniv M. The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. EMBO reports. 2000;1:500.

[21] Biegel JA. Alterations of the hSNF5/INI1 gene in central nervous system atypical teratoid/rhabdoid tumors and renal and extrarenal rhabdoid tumors. Clinical cancer research. 2002;8:3461-7.

[22] Sevenet N, Lellouch-Tubiana A, Schofield D, Hoang-Xuan K, Gessler M, Birnbaum D, et al. Spectrum of hSNF5/INI1 Somatic Mutations in Human Cancer and Genotype-Phenotype Correlations. Human Molecular Genetics. 1999;8:2359-68.

[23] Eaton KW, Tooke LS, Wainwright LM, Judkins AR, Biegel JA. Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. Pediatric blood & cancer. 2011;56:7-15.

[24] von Hoff K, Hinkes B, Dannenmann-Stern E, von Bueren AO, Warmuth-Metz M, Soerensen N, et al. Frequency, Risk-Factors and Survival of Children With Atypical Teratoid Rhabdoid Tumors (AT/RT) of the CNS Diagnosed between 1988 and 2004, and Registered to the German HIT Database. Pediatric Blood & Cancer. 2011;57:978-85.

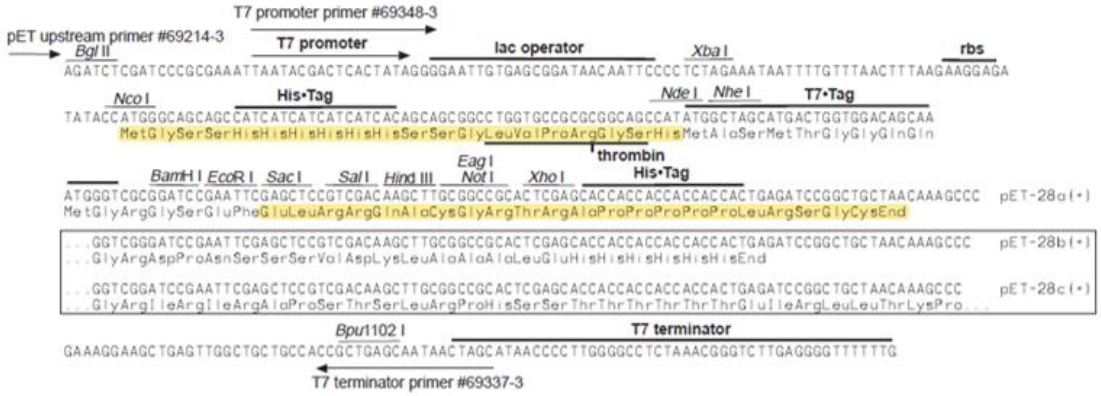
[25] Kohashi K, Oda Y. Oncogenic roles of SMARCB1/INI1 and its deficient tumors. Cancer science. 2017;108:547-52.

[26] Harrison SC. A structural taxonomy of DNA-binding domains. Nature. 1991;353:715-9.

[27] Steele D. Unpublished Data. 2016.

[28] Shirai A, Matsuyama A, Yashiroda Y, Hashimoto A, Kawamura Y, Arai R, et al. Global Analysis of Gel Mobility of Proteins and Its Use in Target Identification2008.

Appendix



Appendix 1: The highlighted regions indicated the amino acid sequence of the 6X Histag added to the N-terminus (top) and the C-terminus (bottom).