2014

Biochemical and Structural studies of AAV-2 Rep68-AAVS1 complex assembly

Clayton Bishop
Virginia Commonwealth University

Biochemical and Structural studies of AAV-2 Rep68-AAVS1 complex assembly

A thesis submitted to the partial fulfillment of the requirements for the degree of
Master of Science at Virginia Commonwealth University

By
Clayton Bishop
Virginia Commonwealth University

Director: Carlos Escalante
Associate Professor, Department of Physiology and Biophysics

Virginia Commonwealth University
Richmond, Virginia
June 4, 2014

ACKNOWLEDGMENTS

I would like to thank my friends and family, especially my parents. Without the support of my parents I would undoubtedly not be in the position I am currently. I would also like to express my sincere gratitude to Dr. Carlos Escalante, as he guided me through my graduate research experience. In addition, I would like to thank Dr. Francisco Zarate-Perez, who directly mentored me in my research and was an invaluable aide to this thesis. His efforts and support cannot be overstated. In addition, I would like to thank Vishaka Santosh and Soumya Remesh for their help and humor in the laboratory. Lastly, I would like to acknowledge the tremendous support of *Coffea arabica*, who spent many long days in the laboratory, and many long nights in the library with me. I owe a deep debt of gratitude to *C. arabica*.

TABLE OF CONTENTS

List of Figures

List of Abbreviations

DNA	Deoxyribonucleic acid
AAV	Adeno-associated Virus
AAVS1	Adeno-associated Virus Site 1
ITR	Inverted Terminal Repeat
RBS	Rep Binding Site
TRS	Terminal Resolution Site
AAV2	Adeno-associated Virus Serotype 2
SF3	Super Family 3 Helicase
OBD	Origin Binding Domain
SAXS	Small Angle X-ray Scattering
IPTG	Isopropyl β-D-1-thiogalactopyranoside
CHAPS	3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate
TCEP	Tris(2-carboxyethyl)phosphine
SDS PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

Abstract

BIOCHEMICAL AND STRUCTURAL STUDIES OF AAV-2 REP68-AAVS1 COMPLEX ASSEMBLY

By Clayton Matthew Bishop, MS

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2014

Major Director: Carlos Escalante, Associate Professor Department of Physiology and Biophysics

Multiple DNA transactions are at the center of almost all processes regulating the AAV life cycle. A common feature shared by all transactions is the binding of the large AAV Rep proteins Rep78/Rep68 onto DNA sites harboring multiple GCTC repeats. AAV mediated site-specific integration is contingent upon the formation of a productive complex between Rep78/Rep68 and the AAVS1 site located at chromosome 19. In order to understand the mechanistic details of the initial assembly process we carried out equilibrium binding experiments of Rep68 and its individual domains with a 42-mer AAVS1 site. Results show that although Rep68 binds AAVS1 with high affinity (69 nM), both the OBD and helicase individual domains bind DNA weakly with affinities of >>60μM and 22μM respectively under our experimental conditions. Mutant Rep68 proteins that have a defective oligomerization interface bind DNA poorly suggesting that productive binding requires both the concerted interaction of the individual domains with DNA and oligomerization. Moreover, we show that a minimal number of two repeats is

required to form a stable complex. In addition, initial studies were done to characterize the interaction between Rep68 and the viral ITR DNA sequences using AUC and electron microscopy.

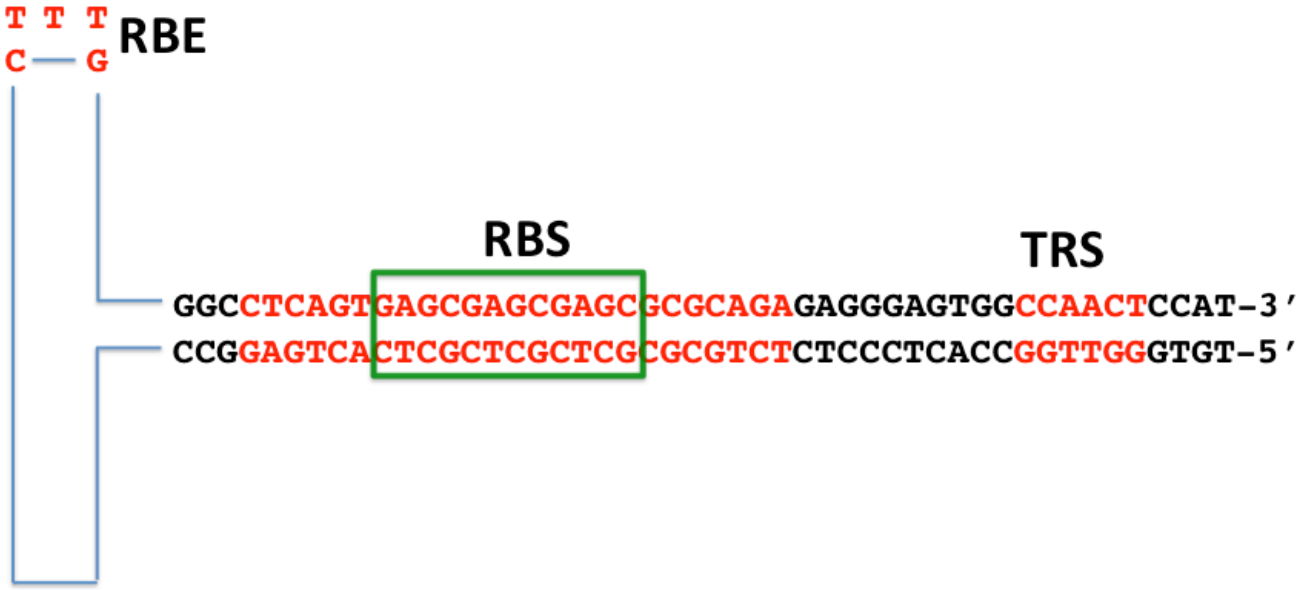
Introduction

1.1 Adeno-Associated Virus

Adeno-associated Virus (AAV), first discovered in 1965 in association with adenovirus particles, is a small (roughly 25 nm), non-enveloped virus. It is a single stranded DNA eukaryotic virus from the family Parvovirus. It cannot replicate itself without co-infection with another "helper" virus, and is thus termed a satellite virus. In addition, it can enter into one of two pathways, a lytic pathway, that depends on the help of a helper virus, in which it replicates itself and releases new virus particles, or a lysogenic pathway, in which it incorporates its genome into the human genome. In the lysogenic pathway AAV integrates its genome into the human genome site-specifically at a site on chromosome 19 known as Adeno-Associated Virus Site 1 (AAVS1). It is, importantly, the only eukaryotic virus known to integrate site-specifically. Adeno-associated virus is a relatively simple virus. Its DNA genome is ~4.7 kilobases and has two inverted terminal repeats (ITR) of 145 base pairs each on each end. The first 125 base pairs of the ITR sequences can fold over on themselves due to their palindromic sequences, and with Watson-Crick base pairing, form T-shaped hairpins (Zhijian Wu et. al, 2009). This folding over enables replication of the viral genome by cellular polymerase by providing a free 3' hydroxyl end. DNA polymerization begins using this free 3' end of the ITR as a primer. Once this is accomplished nicking occurs at the TRS which allows for replication of the ITR sequences (Gonçalves, 2005). Also critical for replication are two sites in the genome known as the Rep Binding Site (RBS) and terminal resolution site (TRS) (Figure 1.1). The AAVS1 site in the human genome also has

both a TRS site and an RBS site in its sequences (Figure 1.4). Between the two inverted terminal repeats in the viral genome are two genes known as the Rep (replication) gene and the Cap (capsid) gene. The capsid gene gives rise to structural elements of the virus. From the Rep gene non-structural proteins, involved in the replication of the viral genome and its integration into the human genome, are produced.

Figure 1.1 : The AAV2 genome with RBS and TRS sites indicated in red. The TRS site is the point at which the AAV2 genome is nicked before being incorporated at the AAVS1 site. The RBS sites are the places at which the Rep proteins bind the viral DNA. The RBS site is characterized by three GCTC repeats, which have been boxed in green.



The AAV2 Inverted Terminal Repeat

1.2 Gene therapy

According to the National Institutes of Health, gene therapy is a technique that "uses genes to treat or prevent disease." Gene therapy may allow for the treatment of diseases by the insertion of a gene into a patient's cells rather than by using drugs, surgery, or other techniques. This may mean replacing a defective gene with a functioning copy, introducing a new gene to help fight disease, or inactivating a mutated gene (Genetic Home Reference, 2014). Because genes that are simply inserted into a cell will not function a vector is generally used to deliver the gene into the cell. In many cases viruses, modified to contain the gene of interest and to be non-pathogenic, are used as vectors. The vector can then be injected or delivered intravenously to a specific tissue where the gene of interest is then delivered. Another option is to remove a piece of tissue from the patient, expose it to the vector, then return it to the patient. Gene therapy has shown some success in treating multiple diseases including hemophilia, severe combined immune deficiency, leukemia, and blindness caused by retinitis pigmentosa. However, several barriers exist to the expanded use of gene therapy as a treatment for illnesses.

1.3 AAV as a Preferred Gene Therapy Vector

Historically little attention has been given to AAV because of its lack of apparent pathogenicity. However, in part of because of this trait AAV has become a preferred vector for gene therapy treatments and trials. For any viral gene therapy vector there are multiple important aspects to consider. These include, but are not limited

to, the virus' ability to enter the cell, move to the nucleus, and a general lack of toxicity. AAV vectors successfully complete all of these tasks (Shyam Daya et. al, 2008).

Of importance is the virus' ability to integrate its genome site-specifically. Because it is the only eukaryotic virus known to have this capability it is an especially attractive potential vector. This is because site-specific integration would ensure long-term transgene expression and would also limit problems with insertional mutagenesis. It is known that the AAVS1 sequence with both the RBS and TRS sites (separated by 8 bp) is necessary and sufficient to target integration (Douglas M. McCarty et. al, 2004) Many other RBS sites have been identified in the human genome but only AAVS1 has a close by TRS site. This suggests that both, in association, are necessary for integration. In addition, the viral genome and the AAVS1 sequence have been tethered *in vitro*. This means that Rep68/78 hold the viral DNA and human DNA in close approximation, which is necessary for integration (Els Henckaerts et. al., 2010). These various phenomena partially explain why AAVS1 is targeted for integration, though the mechanism remains mostly unknown. This research aims to partially elucidate some of the mechanism.

1.4 Rep68

The Rep gene can use two promoters, p5 and p19, and has one intron that can be spliced out or left in, giving rise to four non-structural proteins (Rep78, Rep68, Rep52, and Rep40) which are named for their apparent molecular weights in gels. All of the Rep proteins have helicase activity, and the larger two (Rep68 and Rep78)

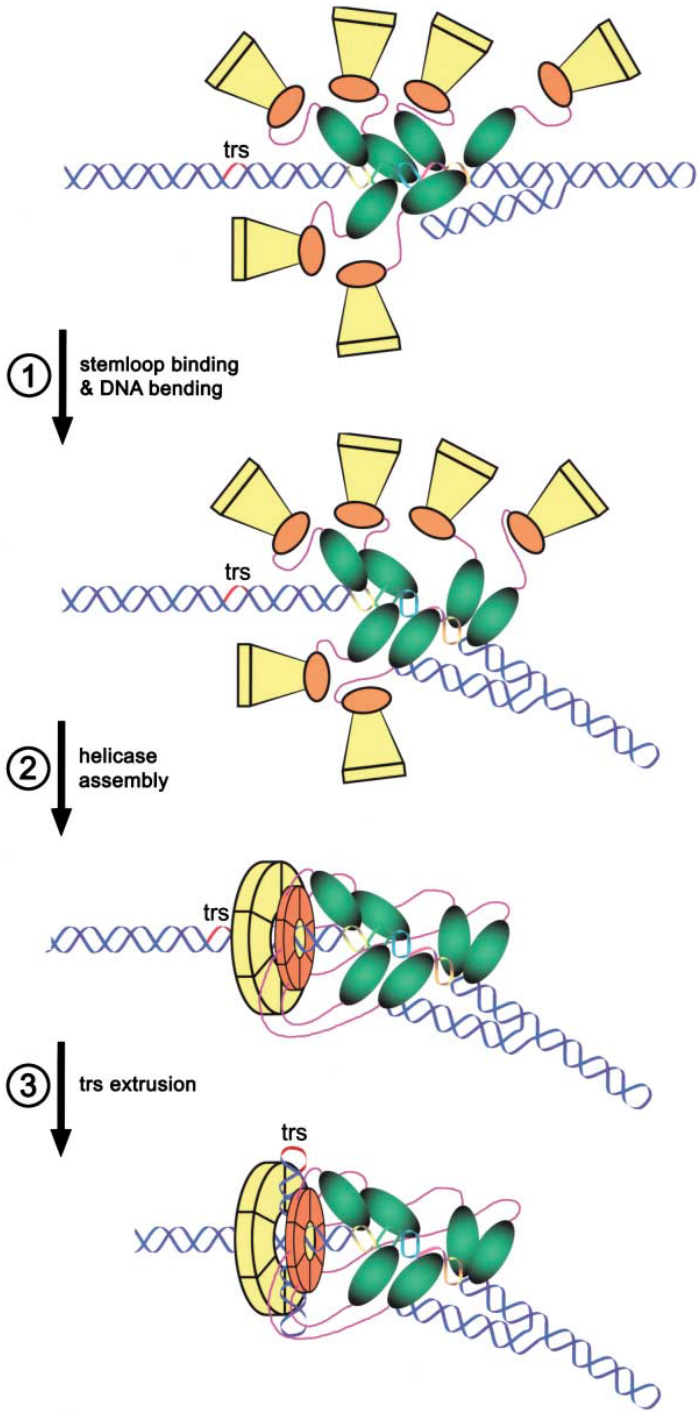
also exhibit nicking ability at the TRS. Rep68 belongs to superfamily 3 of helicases (SF3). As opposed to most helicases whose function is simply to unwind the DNA in front of the advancing replication fork, SF3 helicases also function in the following capacities—as initiators of DNA replication, transcriptional regulators and also help pack DNA into viral particles (Zarate- Perez et. al, 2013). Rep68 has two domains, an origin binding domain (OBD) and a helicase domain, connected by a short 16 amino acid linker. The OBD domain runs from amino acids 1-208 and the helicase (AAA+) domain runs from amino acids 224 to 490. The OBD domain has been shown, in crystal structures, to bind to both the human AAVS1 site and the viral ITR at their respective Rep Binding Sites. These sites consist of multiple GCTC repeats. This binding event is followed by ATP-dependent helicase activity in the region that is just 5' from the RBS. This results in the formation of a stem-loop in which the TRS site is located at the apex of the loop. In this configuration Rep68 can nick the DNA at the TRS site.

Rep68, as well as all other SF3 helicases, exhibit 3'=>5' helicase activity. This activity varies directly with protein concentration. The helicase activity occurs at a rate of nearly 350 bp per minute per Rep68 molecule. It requires both $MgCl_2$ and ATP to exhibit helicase activity, with optimal conditions being pH 7.5, and a $MgCl_2$ concentration of 5 mM (Xiaohuai Zhou et. al., 1999). In addition, the endonuclease activity of Rep68 is second order with respect to ATP, and only one of the strands of DNA is nicked (Dong-Soo Im et. al., 1990).

The current proposed model for how Rep68 binds and initially interacts with DNA is as follows. The first step in this model is that five origin binding domains

bind to five tetranucleotide repeats in the RBS, including three GCTC repeats. This is followed by the helicase domains forming a hexameric ring near the TRS site (Hickman et. al., 2004).

Figure 1.2: An illustration of the current proposed model of Rep68 DNA binding from Hickman et. al., 2004.

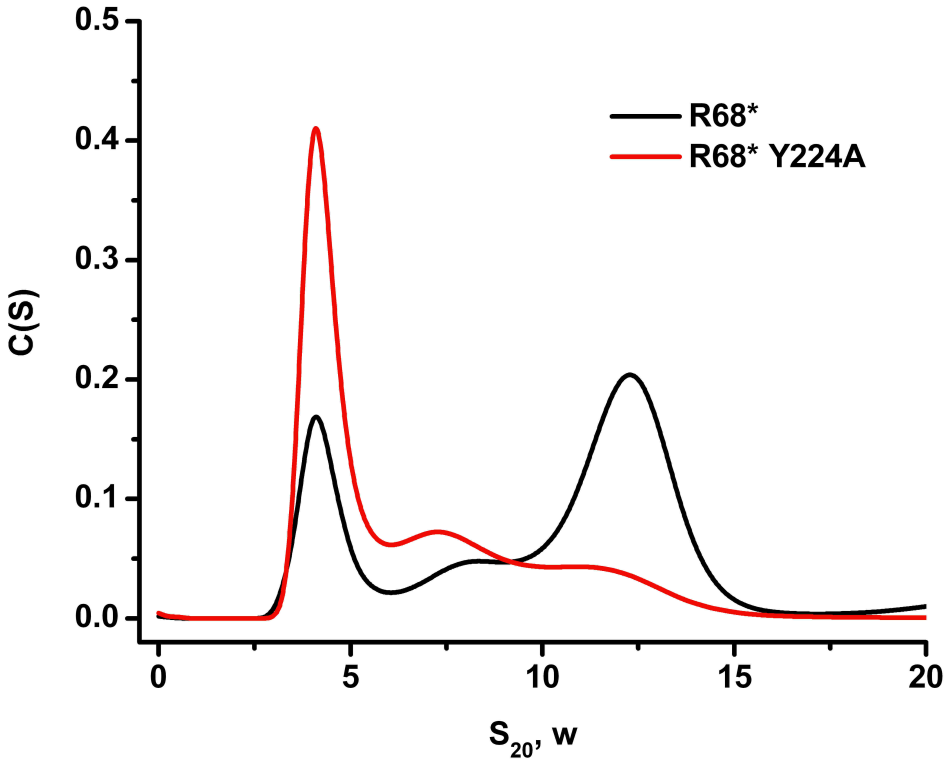


Previously Proposed Model of Rep68 Binding DNA

This model is problematic for several reasons, the first of which is that while five individual OBDs can bind the RBS 5 full length Rep68 proteins cannot. This is due to steric hindrances as the try to lineup along the DNA. This makes this proposed mechanism unlikely. In addition, analytical ultracentrifuge experiments by Francisco Zarate-Perez show that Rep68 and AAVS1 DNA form a heptameric, not a hexameric, complex.

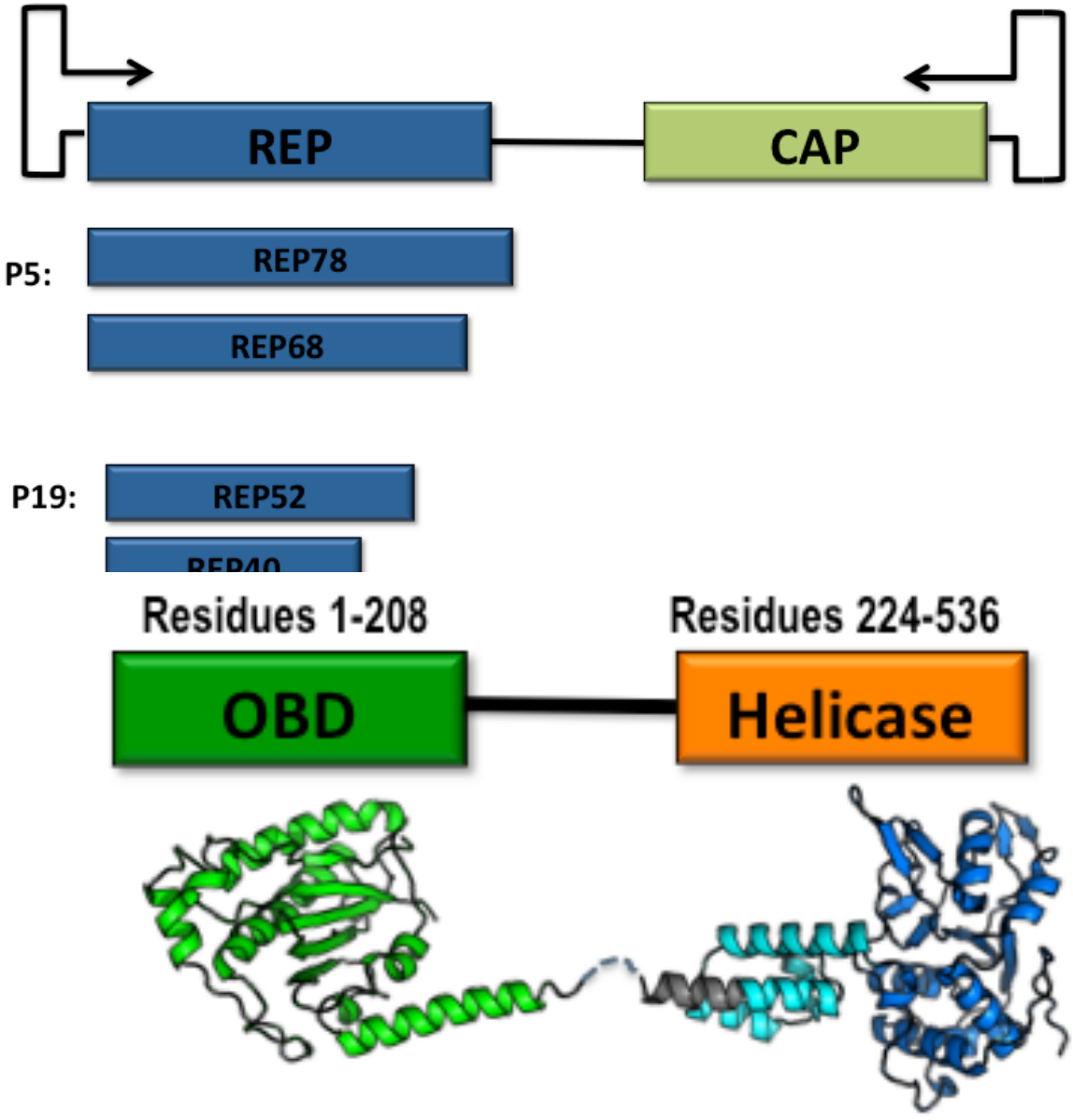
In addition, two residues, Y224 and I251 have been identified as critical for oligomerization. Analytical ultracentrifugation experiments illustrate that Rep68 with a mutation replacing Y224 with an alanine significantly reduces oligomerization (Zarte-Perez et. al., 2012). This can be seen in Figure 1.3. Because these residues can be mutated to reduce oligomerization, it allows for establishing the role that oligomerization plays in the various functions of Rep68, including binding and helicase activity.

Figure 1.3: The results, provided by Francisco Zarate-Perez of an analytical ultracentrifugation experiment that illustrates the difference in oligomeric properties between Rep68* and Rep68* Y224A. Rep68* Y224A fails to form a large oligomeric complex. Results are reported in terms of concentration in terms of s-value against s-value.



AUC Results of Rep68 Mutant

Figure 1.4: An illustration of the AAV2 genome and associated Rep proteins. The four separate Rep proteins, which are non-structural, arise from the use of two separate promoters, and splicing in or out of a single intron. Rep68 consists of two domains, a helicase and an origin binding domain (OBD), joined by a linker.



Schematic of AAV2 Genome with the Rep Proteins and Rep68 with domain structures

1.5 Anisotropic Binding Studies

Determining how well Rep68, as well as its individual domains, and mutants bind the AAVS1 site is important in helping to engineer future chimeric proteins for gene therapy vectors. In order to quantify how well these proteins are able to bind the AAVS1 DNA sequence anisotropic polarization studies will be employed to determine the dissociation constant. The dissociation constant can be quantified by the following equation:

$$K_d = [P][L]/[PL]$$

Where [P] represents molar concentrations of the protein, [L] represents molar concentrations of the ligand (in this case the DNA), and [PL] represents molar concentrations of the protein ligand complex (in this case protein bound to DNA) (Thomas D. Pollard et. al, 2010). Determining the dissociation constant by fluorescence anisotropy is based upon observations of the rotational motion (or tumbling) of fluorescently labeled macromolecules in solution (Pollard et. al, 2010). This fluorescent molecule will prefer to absorb light in a particular plane, known as its absorption dipole. The maximum amount of absorption will occur when the absorption dipole is parallel to the plane of polarized light. The fluorophore will also emit polarized light, which will be slightly depolarized from the excitation plane. Thus the anisotropy of the emitted light can be calculated using the following equation:

$$A = (I_{II} - I_P)/(I_{II} + 2I_P)$$

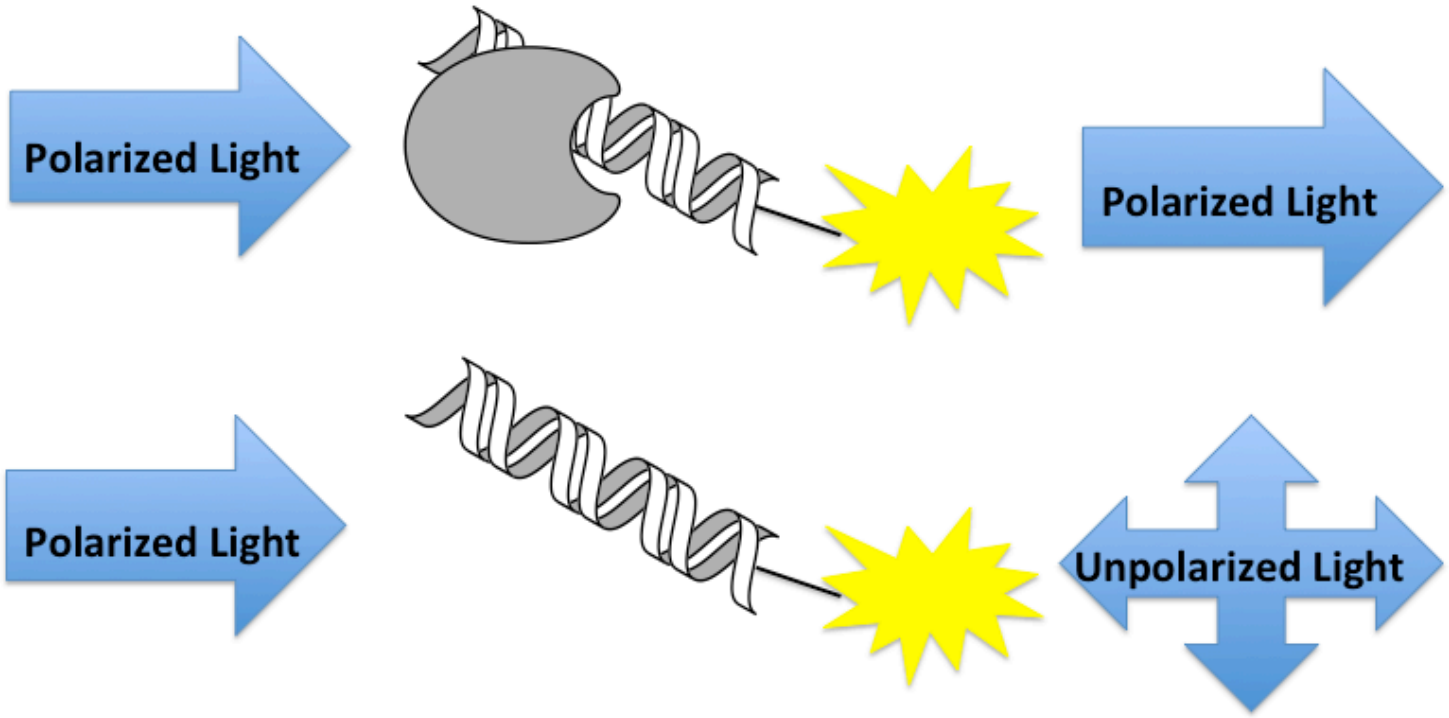
In which A is anisotropy, I_{II} is the intensity of emitted light parallel to excitation and I_P is intensity of emitted light perpendicular to excitation (Wendy A. Lea et. al, 2011). Anisotropy is reduced by rotation of the labeled macromolecule (DNA in this case) in the time between excitation and emission (thus making more of the light polarized in perpendicular to the plane of excitation). That is to say, the rotation of the molecule scatters the light, resulting in less polarization. This reduction in anisotropy can be quantified using the Perrin equation:

$$A = A_0/(1+(\tau/\phi))$$

Where A_0 is the limiting anisotropy (the maximum amount of anisotropy due to the fact that emitted light must be slightly depolarized from excitation), τ is the fluorescence half-life and ϕ is the rotational correlation time. This means is that when a molecule is rotating rapidly its observed anisotropy nears zero. But, when a molecule is rotating slowly the anisotropy approaches the limiting anisotropy. The practical application of this is that when DNA is free (i.e. not bound by a protein) it rotates rapidly resulting in anisotropic values approaching zero. When the DNA has become bound by a protein its rotation slows greatly, resulting in larger values for anisotropy. Because of this relationship a binding curve (and thus dissociation constants) can be calculated from anisotropies when protein is titrated against DNA. Finally anisotropies are converted to percent of DNA bound, which is done using the formula: Percent Bound =

$$((A_x - A_{DNA})/(A_{Max} - A_{DNA}))*100$$

Figure 1.5: Showing the underlying principle of the binding studies. When the protein is bound to the DNA, the DNA strand tumbles more slowly. Because the DNA is tumbling more slowly much of the light remains polarized in the same plane as the excitation. This leads to a larger anisotropy.



Cartoon Showing the Principles of Fluorescence Binding Experiments

Figure 1.6: Cartoon showing the 42 bp DNA strand used in binding experiments. This DNA is a section of human chromosome 19 known as the Adeno-Associated Virus Site 1 (AAVS1) and contains a Rep Binding Site, with three GCTC repeats which have been colored.

Schematic of AAVS1 DNA

1.6 Analytical Ultracentrifugation

Because the functionality of Rep68 is dependent upon the formation of oligomeric complexes, with ATP binding and hydrolysis occurring at the interface of individual proteins, understanding the stoichiometry with which Rep68 binds the ITR sequences at the ends of the viral genome is important. In order to do this analytical ultracentrifugation will be employed.

Ultracentrifugation allows for the characterization of proteins by two different methods: 1) sedimentation velocity—that is to say the speed with which the proteins descend to the bottom of the centrifuge cell and 2) sedimentation equilibrium. In sedimentation velocity, the rotor is spun at high speed, and the macromolecules descend all the way to the bottom. The speed with which they descend to the bottom can be correlated to molecular weight. In sedimentation equilibrium the rotor is spun at low speed, and eventually the descent of the macromolecule due to centrifugal forces will be exactly canceled by the rate of diffusion, resulting in an equilibrium. The shape of this gradient can be a very accurate way of determining molecular weight (James L. Cole et. al, 2008). Determination of sedimentation coefficients has been used in the modeling of the hydrodynamic shape of proteins and protein complexes. Because of this sedimentation velocity can be used to identify the oligomeric state and the stoichiometry of heterogeneous interactions (Jacob Lebowitz et. al, 2002).

The sedimentation coefficient and diffusional coefficient of a macromolecule cane be can be calculated from the Lamm equation:

$$\left(\frac{dC}{dt}\right)r = -\frac{1}{r}\left\{\frac{d}{dr}\left[\omega^2 r^2 sC - Dr\left(\frac{dc}{dr}\right)t\right]\right\}t$$

Where C(x) is the local concentration of solute, r is the radius, t is the time, D is the solute diffusion constant, s is the sedimentation coefficient, and ω is the rotor angular velocity (Igor N. Serdyuk et. al., 2007). SEDFIT is a program that uses Lamm equations to solve for the sedimentation coefficient and the diffusion constant. SEDFIT can then use the Svedberg equation to solve for molecular weight:

$$S = \frac{u}{\omega^{2r}} = \frac{M(1-v\rho)}{NAf} = \frac{MD(1-v\rho)}{RT}$$

where u is the observed radial velocity of the macromolecule, ω is the angular velocity of the rotor, r is the radial position, $\omega^2 r$ is the centrifugal field, M is the molar mass, v is the partial specific volume, ρ is the density of the solvent, N_A is Avogadro's number, f is the frictional coefficient, D is the diffusion coefficient, and R is the gas constant (Lebowitz et. al, 2002). The unit for s-values is the Svedburg which is 10^{-13} seconds. Individual oligomeric configurations will have different S-values, and will sediment to the bottom of the cell at different speeds. This can be seen on an analytical ultracentrifuge.

1.7 Electron Microscopy

Electron microscopy involves the use of a beam of electrons to image a specimen. There are two types of electron microscopy, scanning and transmission. In scanning electron microscopy a highly focused beam of electrons moves over each point of the sample. The electron current leaving the sample is then collected, The electrons interact with the atoms of the sample, and therefore, this collected information can then be used to collect information about the sample's composition and topography (K.C. Smith et. al., 1955). In transmission electron microscopy the beam of electrons passes through a very thin specimen to create an image (Peter J. Goodhew, et. al., 2001). A transmission electron microscope consists of an electron gun, electromagnetic lenses, a viewing screen, and a camera which works within the vacuum of the microscope. Below the electron gun, which emits the beam of electrons, are the condenser lenses which demagnify the beam and control the diameter of the beam as it interacts with the specimen. Below the condenser lenses is the specimen chamber which holds the specimen. The specimen must be able to move laterally as well as to be tilted. Following the specimen chamber are the objective and intermediate lenses which create and then magnify an intermediate image for display (Goodhew et. al., 2001).

In this study transmission electron microscopy will be utilized to study Rep68 and related proteins. In addition, this study will utilize negative staining. Negative staining uses a water soluble heavy metal salt (in this case uranyl formate) to create a thin film that supports biological particles. The staining also creates a contrast in which the biological particle of interest is generally electron transparent,

while the negative stain creates an electron-opaque background (John Kuo, 2007). In this case the sample is dehydrated, and the image is formed by surrounding dense solution created by the heavy metal stain. An advantage to this technique is that it gives a high level of contrast.

TEM will primarily be used in order to perform single particle reconstruction. Single particle reconstruction was developed to increase single to noise ratio and to help deal with intrinsic low contrast (Jonathan Ruprecht et. al., 2001). In single particle reconstruction a 3D image is produced by reconstructing from a large number of images of an object (Joachim Frank, 2010). Programs then are able to use these individual particles to reconstruct the 3-D model.

The first step in utilizing TEM is sample preparation. As was previously stated, this study will utilize negative staining, which has several distinct advantages. It is simple to perform, and provides high contrast as was previously stated. It is also less sensitive to the electron beam, which is of importance because samples are generally very sensitive to electron beam bombardment. However, a major drawback of negative staining is that resolution is limited to about 10 Å. This is because the sample is dehydrated during negative staining, and the grain of the heavy metal salt is about 10Å. Itreveals the distribution of the heavy metal atoms rather than the density of the sample itself (Jonathan Ruprecht et. al., 2001).

Because the electron beam itself is damaging, a “low-dose” technique will also be utilized. In this technique the electron dose is minimized which minimizes damage as well. In this technique three modes, denoted as search, focus and exposure are used. In search, a location on the grid is located at low magnification

(about 3,000x). Using a lower magnification again allows for a minimization of the damage caused by the electron beam itself. Then, the beam is focused at a higher magnification (about 180,000x). Finally in exposure mode the image is taken (Jonathan Ruprecht et. al., 2001).

Because a large data set is needed to reconstruct the 3D structure, in single particle reconstruction , at least several thousand individual particles must be picked out from micrographs. This process involves boxing of individual particles that are not touching other particles (George Harauz et. al., 1988). This step is similar in most reconstruction software. From this point forward reconstruction will be specific to EMAN2. EMAN uses a refinement model to reach a final structure. Because of this a preliminary model must be built (the final model requires a preliminary model which it then refines). An initial model is generated based on selected particles, which have been grouped into distinct classes. This initial model is then subjected to a refinement loop in EMAN2. The preliminary model is iteratively refined against the data. This is done until convergence is achieved. In EMAN2 this means that the Fourier shell correlation stabilizes between iterations. This will mean the noise ratio will have been sufficiently reduced (Steven J. Ludtke et. al., 1999).

Aims of Study

2.1

Rep68 is a multifunctional AAV protein. It is critical to incorporating viral DNA into the human genome site-specifically. Because future gene therapy treatments hope to use AAV as a potential vector it is necessary to understand and characterize the mechanism with which Rep68 integrates the viral genome into the human genome at the human AAVS1 DNA site. The design of any future chimeric Rep68 proteins for vectors will rely on the characterization of this mechanism. This study aims to further the understanding of this mechanism.

The main aim of the study is to characterize how Rep68 interacts with DNA. The first step in this interaction is binding DNA. The first aim looks to characterize this initial interaction by understanding how the individual domains interact with human AAVS1 DNA. Rep68 consists of two domains, OBD and Helicase domain (Rep40), which are joined by a short inter-domain linker. In order to characterize how well these proteins bind the AAVS1 DNA, binding studies will be conducted in which the individual domains (Rep40, OBD and the OBD with the linker) are titrated against DNA in order to determine dissociation constants for the domains. This will help illustrate the roles the domains play in the binding of DNA. Whether one gives affinity, another specificity, or whether the domains interact in order to effectively and specifically bind DNA.

The second aim of this study is to determine what role oligomerization of individual Rep68 molecules plays in the assembly on DNA. It is known that certain amino acid residues play a key role in oligomerization of Rep68 (Y224 and I251).

These residues can be mutated to significantly reduce the ability of Rep68 to oligomerize. These mutant proteins can then be titrated against AAVS1 DNA in order to establish dissociation constants. By characterizing how well these mutant Rep68 proteins bind DNA the role that oligomerization plays in binding DNA can be more fully understood.

A third aim of the project looks to more fully understand the role that the tetranucleotide repeats found in the RBS in both the viral and human DNA play in targeting Rep68 to those specific sequences. By using DNA sequences that have either 1, 2 or 3 of these consensus GCTC sequences, and determining dissociation constants for Rep68 titrated against each of these sequences the project aims to show the minimum number of these repeats necessary for effective binding. This will help to illustrate how Rep68 recognizes and initially binds the RBS in DNA.

A fourth aim was to characterize whether oligomerization had any effect on the helicase activity of Rep68. Because Rep68 is involved in both viral genome replication and integration of the viral genome into the human genome, helicase activity is a critical function. Understanding whether helicase activity requires oligomerization is an important question. Characterizing whether Rep68 functions in a large oligomeric complex when unwinding DNA or in small complexes such as dimers is important to investigate for the future design of chimeric Rep68 proteins. Helicase assays of Rep68* and multiple mutants that lack effective oligomerization will therefore been done in order to quantify helicase activity.

A final aim of this project aims to characterize the interaction of Rep68 with viral ITR DNA. This will include using analytical ultracentrifugation and and electron

microscopy studies, which can help to determine both an initial structure model of Rep68 bound to viral ITR DNA as well as the stoichiometry of how Rep68 binds to the viral ITR DNA.

Materials and Methods

3.1 DNA Binding Experiments

Anisotropic binding experiments were performed using an ISS PC1 fluorimeter (ISS, Champaign, IL). All reactions were performed using a reaction volume of 200 µl. Reactions for Rep68 and all mutants were performed using a buffer of 25 mM HEPES (pH 7.0), 200 mM NaCl, and a final concentration of 10 nM DNA. All reactions for individual domains (OBD, Rep40, OBD plus linker) were performed in the same buffer, except NaCl concentration was lowered to 100 mM to increase binding affinity, as these domains bound very poorly to the AAVS1 site. Protein was added in increasing concentrations until saturation was achieved. Reactions were incubated for 10 minutes prior to measurement. The reactions were excited at a wavelength of 492 nm with an emission wavelength of 520 nm. These wavelengths were chosen because the fluorescein labels absorb at a maximum of 494 nm and emit at 521 nm.

3.2 Transmission Electron Microscopy

A Rep68-ITR 165 complex was purified on a gel filtration Superdex 200 column. Protein samples at 0.1 mg/ml were absorbed directly onto activated carbon-coated grids. Negative staining with 0.75% uranyl formate was used for the prepared grids and then were visualized in a transmission electron microscope Tecnai F20

operated at 200 kV, and images were collected at a magnification of 50,000 X under low dose conditions using a Gatan 4k x 4k charge-coupled device camera. Particle windowing, two-dimensional (2D) alignment, initial model construction and initial model refinement, were carried out with EMAN2 software. The total number of particles used in this reconstruction was 2,557.

3.3 Analytical Ultracentrifugation

Rep68* protein was concentrated to above 3 mg/ml and then added in a 6:1 (protein:DNA) ratio to a 165 bp ITR DNA molecule. 200 µl of this mixture was purified on a superdex 200 10/300 column at 0.5 ml/min. This purification was done using SE buffer. This complex was then concentrated to above 1 OD. Varying concentrations (1 OD, 0.5 OD, 0.25 OD, and 0.125 OD) were then loaded into cells for analysis. Sedimentation velocity experiments were carried out using a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with an eight-hole rotor. Samples were centrifuged in 2-sector carbon-filled Epon centerpieces typically at 25,000 rpm and 20°C. Sectors were loaded with a 420-µl sample volume. Typically, about 200 scans were collected at 5-min intervals at 25,000 rpm. Concentration profiles were collected using both UV absorption (280 nm) and Rayleigh interference optical systems. Results were analyzed using SEDFIT.

3.4 Helicase Assay

A helicase assay was performed to determine whether any of the interface mutations had an effect on the ability of Rep68 to perform one of its obligatory functions—the unwinding of DNA. In this assay a section of double stranded DNA,

tagged with a fluorescent tag, and with a 3' overhang tail, was used. The Rep68 would unwind this DNA. To prevent re-annealing a single stranded "trap" DNA was also added to the mix. To insure that the trap DNA captured the unwound DNA, the trap DNA was present in 5 times the concentration.

Reactions were prepared by adding 1 μM Rep68 protein, 0.5 μM double stranded DNA, 2.5 μM single stranded trap DNA. In addition, because Rep68 needs helicase activity in order to unwind DNA, 5.0 mM ATP, and 5.0 mM $MgCl_2$. Reactions were buffered to 50 μL in 25 mM HEPES, 50 mM NaCl, and pH 7.0. Reactions were allowed to equilibrate for an hour, and then 20 μL of each reaction was loaded onto a 12% polyacrylamide gel. As a negative control a Rep68 mutant, K340H, which has no ATPase activity, and therefore no helicase activity was used.

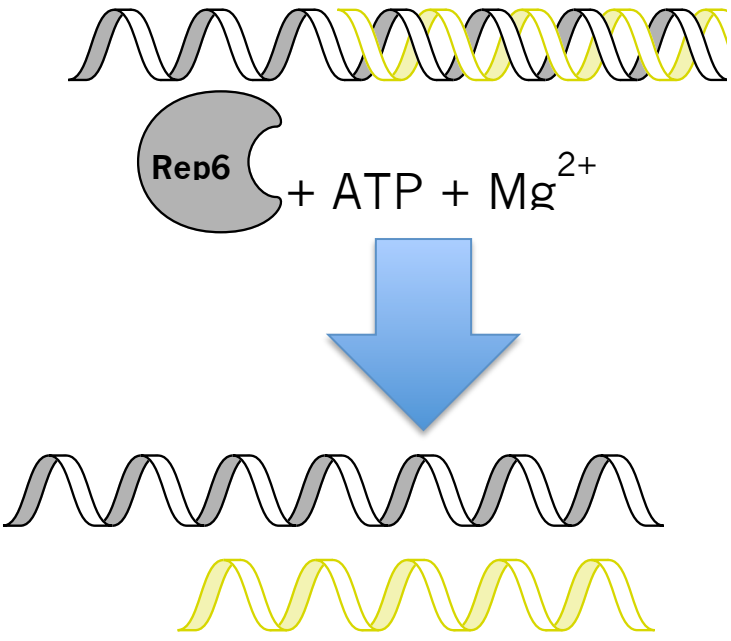
Figure 3.1: Schematic of double stranded, fluorescent tagged, DNA used in helicase assays. In addition, the non-fluorescent trap DNA is also shown. Once unwound, the trap DNA will bind the free DNA, causing a shift in the gel. This shift occurs because the single stranded DNA will travel farther in the gel than the larger double stranded DNA. This process is shown in cartoon as well, with the tagged strand represented in yellow.

Double Stranded DNA

5′–TCTGTACTGCTCCATATGATGCCTATCC–3′
F–AGACATGACGAGGTATAC–5′

Trap ssDNA

3′–AGACATCAGGAGGTATAC–5′



Schematic of Helicase Assay DNA and Cartoon of Assay

Results

4.1 Expression and Purification of Proteins

All constructs of His_6-PreScission protease (engineered from human rhinovirus 3C protease which cleaves the Gln and Gly residues of the recognition sequence of LeuGluValLeuPheGln/GlyPro) cleavage site-Rep68 fusion protein (including all mutants) were transformed into BL21(DE3)pLysS competent *eschericia Coli* cells. Individual colonies were then picked and transferred into 100 mL pre-cultures of LB Miller Broth that grew overnight at 37°C. 1 L cultures were then inoculated using 15 mL of the pre-culture. Once reaching an optical density of 0.5 OD the cultures were then induced using 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). After a 3-4 hour induction period cells were then pelleted and collected. Cells were then resuspended in 25 mL of Buffer A (20 mM Tris-HCl [pH 7.9 at 4°C], 500 mM NaCl, 5 mM imidazole, 10% glycerol, 0.2% CHAPS {3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}, 1 mM TCEP [Tris(2-carboxyethyl)phosphine]) per 1 L of culture. The cells were then lysed using eighteen 10-s cycles of sonication, with 20-second pauses between each cycle. The lysate was then subjected to centrifugation using a Beckman-Coulter J-series centrifuge at 16,000 rpm for 30 minutes.

Following centrifugation the cell extract was then purified using a 5 mL Nickel-NTA column that was pre-equilibrated with Buffer A. The protein was then eluted using Buffer B (which is the same as Buffer A but with 1 M imidazole). The His-PP tag was then cleaved using PreScission protease treatment using 150 μg PP/mg His-PP-Rep68. The protein was incubated overnight at 4°C with the

protease. After confirmation of cleavage, using a 10% polyacrylamide SDS PAGE gel comparing protein that was not incubated with protease to the protein that was incubated, the protein was then desalted into Buffer A using a HiPrep™ 26/10 desalting column (GE Healthcare). The protein was then purified again on a 5 mL Nickel-NTA column to separate the His-PP tag and uncleaved protein from the cleaved Rep68. Finally the protein was run on a HiLoad Superdex 200 16/60 column (GE Healthcare) using size exclusion (SE) (25 mM Tris-HCl [pH 8.0], 200 mM NaCl, 2 mM TCEP). Proteins were concentrated and then stored at -80°C. Individual proteins include Rep68* (a mutant where C151S was mutated to prevent aggregation). This mutant is used as a stand-in for the wild-type. Further mutants include Rep68* I251A, where an isoleucine at residue 251 is mutated to an alanine, Rep68* Y224A, where a tyrosine at residue 224 is mutated to an alanine, and a combination of those two mutants called Rep68* Y224AI251A.

The individual domains of Rep68 were also expressed as OBD208 (OBD without the linker domain), OBD224 (OBD with the linker domain), and Rep40 (the helicase domain). These proteins were respectively His_6-Thrombin protease, His_6-Thrombin protease, and His_6-TEV protease fusion proteins. OBD208 and OBD224 were cleaved with 3 units of Thrombin protease per 1 mg of protein for 3 hours at 25^{o}C. Rep40 was cleaved at 4^{o}C overnight with TEV at 1 mg TEV per 64 mg of Rep40. Purification for the individual domains was otherwise the same as for Rep68*.

Figure 4.1: Chromatograph showing the result of running Rep68* cell lysate over a Nickel-NTA column. The three peaks correspond to the flowthrough, elution with 5% buffer B, and elution with 30% buffer B. And a corresponding 10% SDS-Polyacrylamide gel. Once eluted with 30% of buffer B Rep68 is almost pure. The flow through, and the portion eluted with 5% buffer B contain large amounts of unwanted proteins.

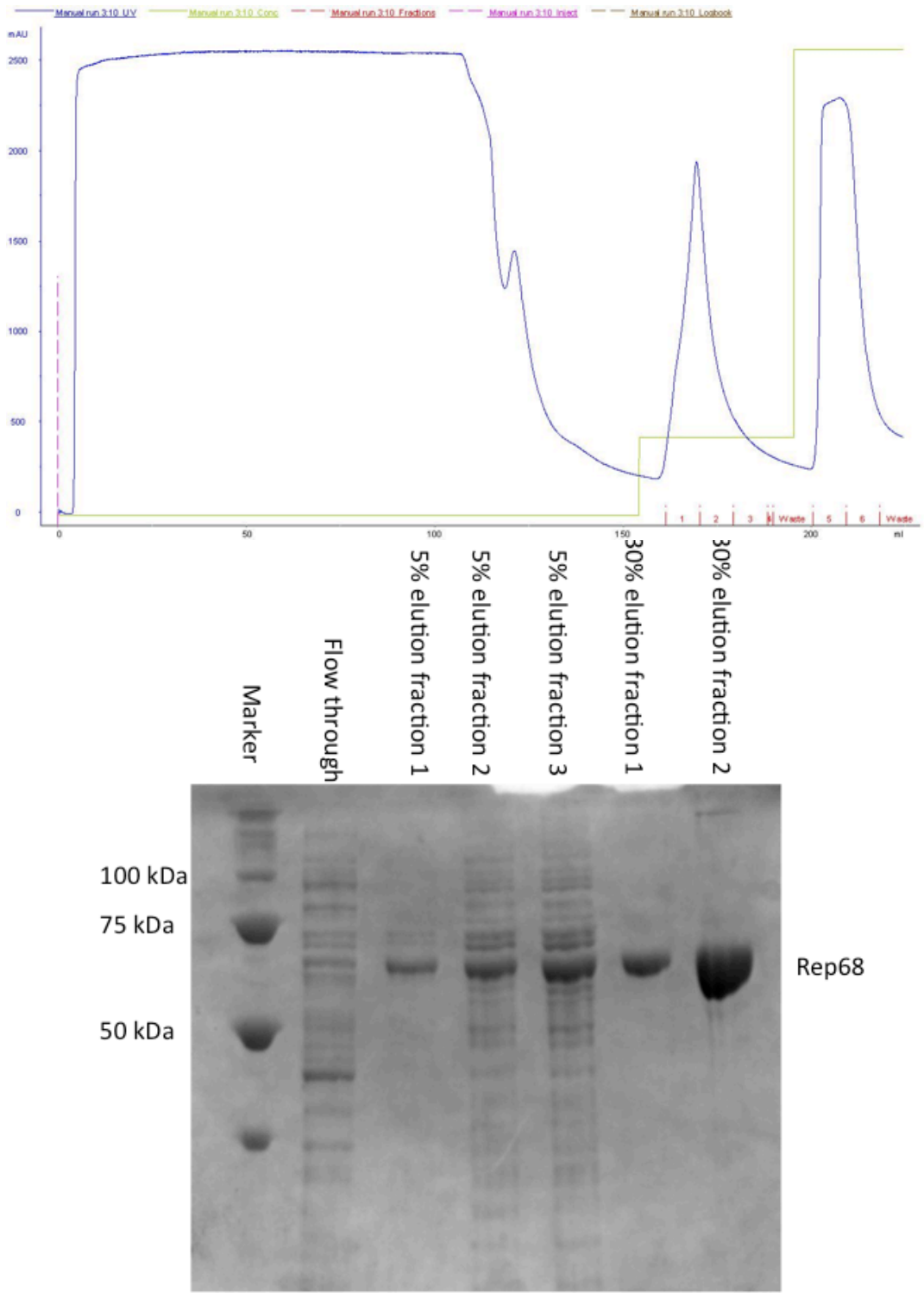
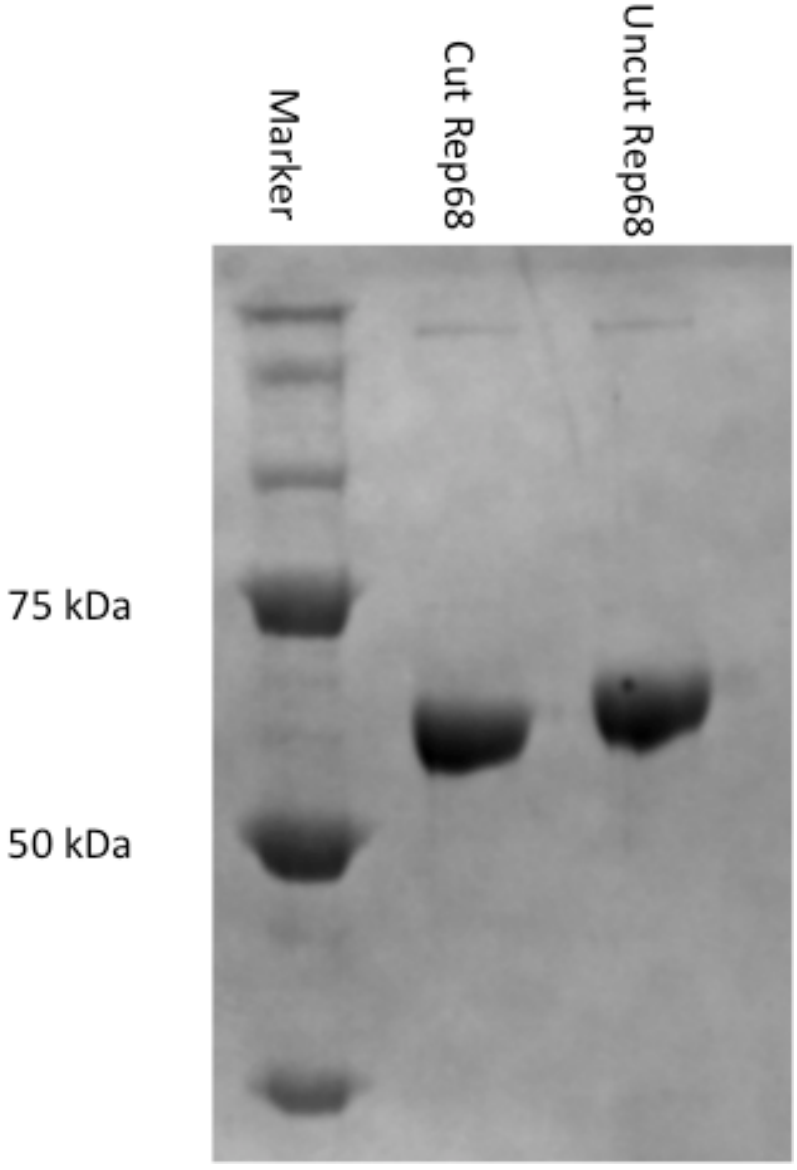


Figure 4.1: Chromatograph and Polyacrylamide Gel of Cell Lysate Nickel-NTA Column Purification

Figure 4.2: 10% SDS-polyacrylamide gel showing Rep68 that has been cut overnight at 4°C with Precission protease. This cleavage removes the Histidine tag so the protein will no longer bind Nickel-NTA columns. This gel compares the cut protein to the uncut protein. The cut protein is slightly smaller and therefore runs slightly lower in the gel.



Polyacrylamide Gel of Rep68 His-Tag Removal

Figure 4.3: Chromatograph showing the result of running cut Rep68* over a Nickel-NTA column. The protein mostly comes out in the flowthrough, seen as the first peak in the UV. A smaller peak is seen in the 5% buffer B elution, and virtually no peak is seen with the 30% buffer B elution, meaning most of the protein was successfully cut. A corresponding 10% SDS-Polyacrylamide gel, because it now lacks a His-tag most of the Rep68* elutes in the flowthrough or 5% elution. Almost no Rep68* is seen in the 30% buffer B elution.

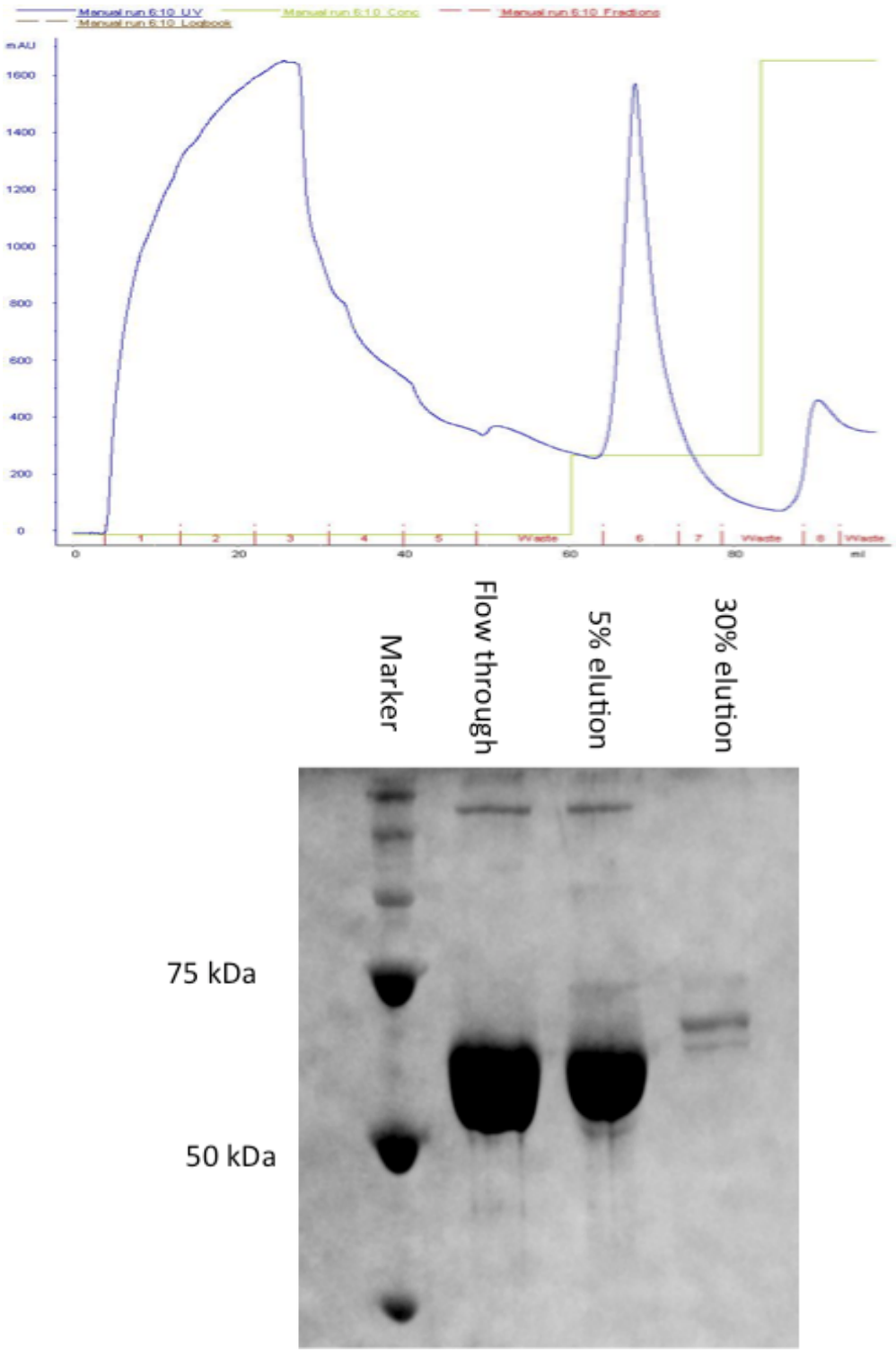


Figure 4.3: Chromatograph and Polyacrylamide Gel of Second Nickel-NTA Column Purification

Figure 4.4: Chromatograph showing the purification of Rep68* on a HiLoad Superdex 200 16/60 column. The column separates proteins based on size. A corresponding 10% SDS-Polyacrylamide gel shows the result of the filtration. The gel shows that the protein is essentially clean besides some aggregation bands seen at the top of the gel. The previous chromatograph corresponds to this gel.

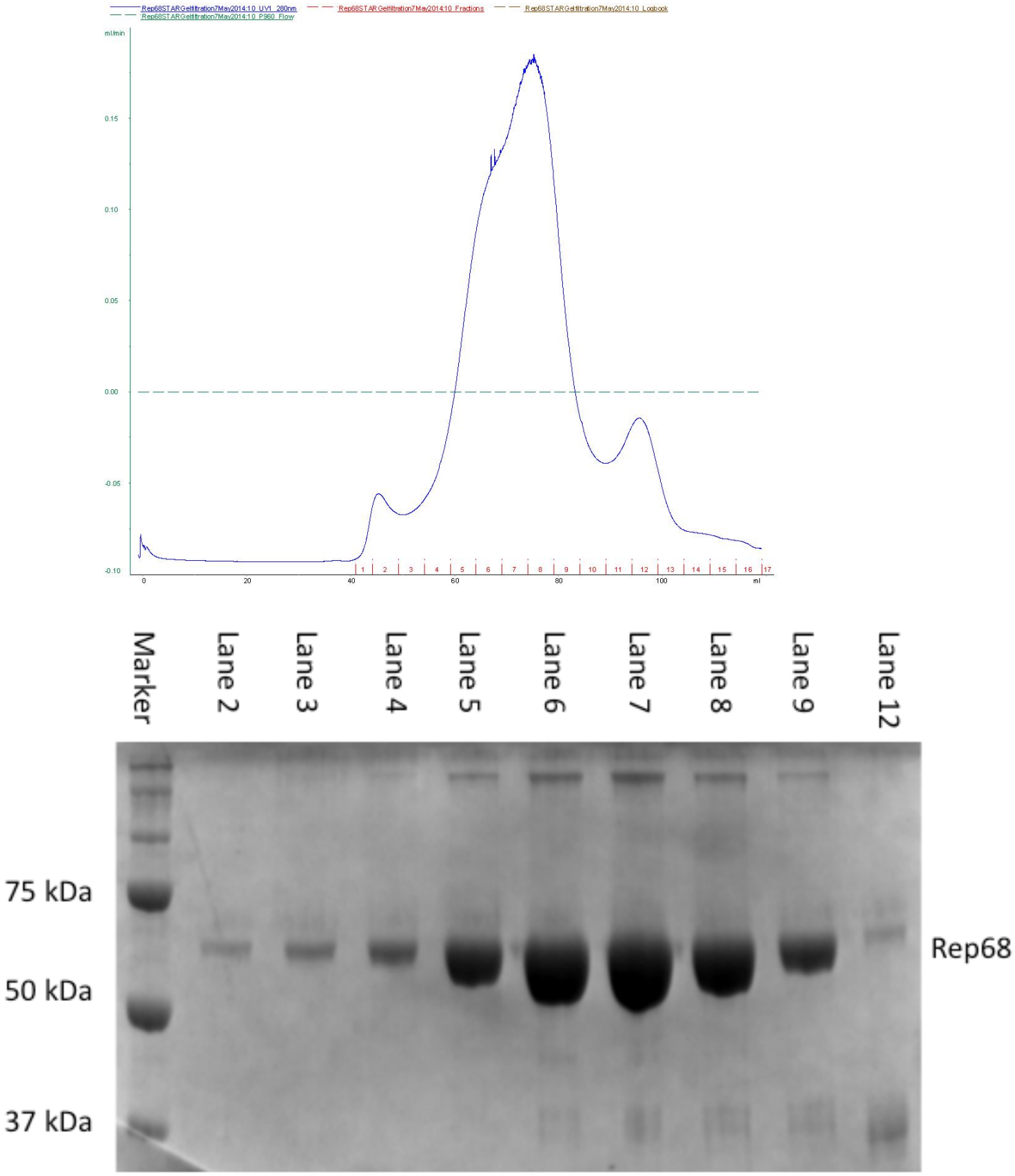
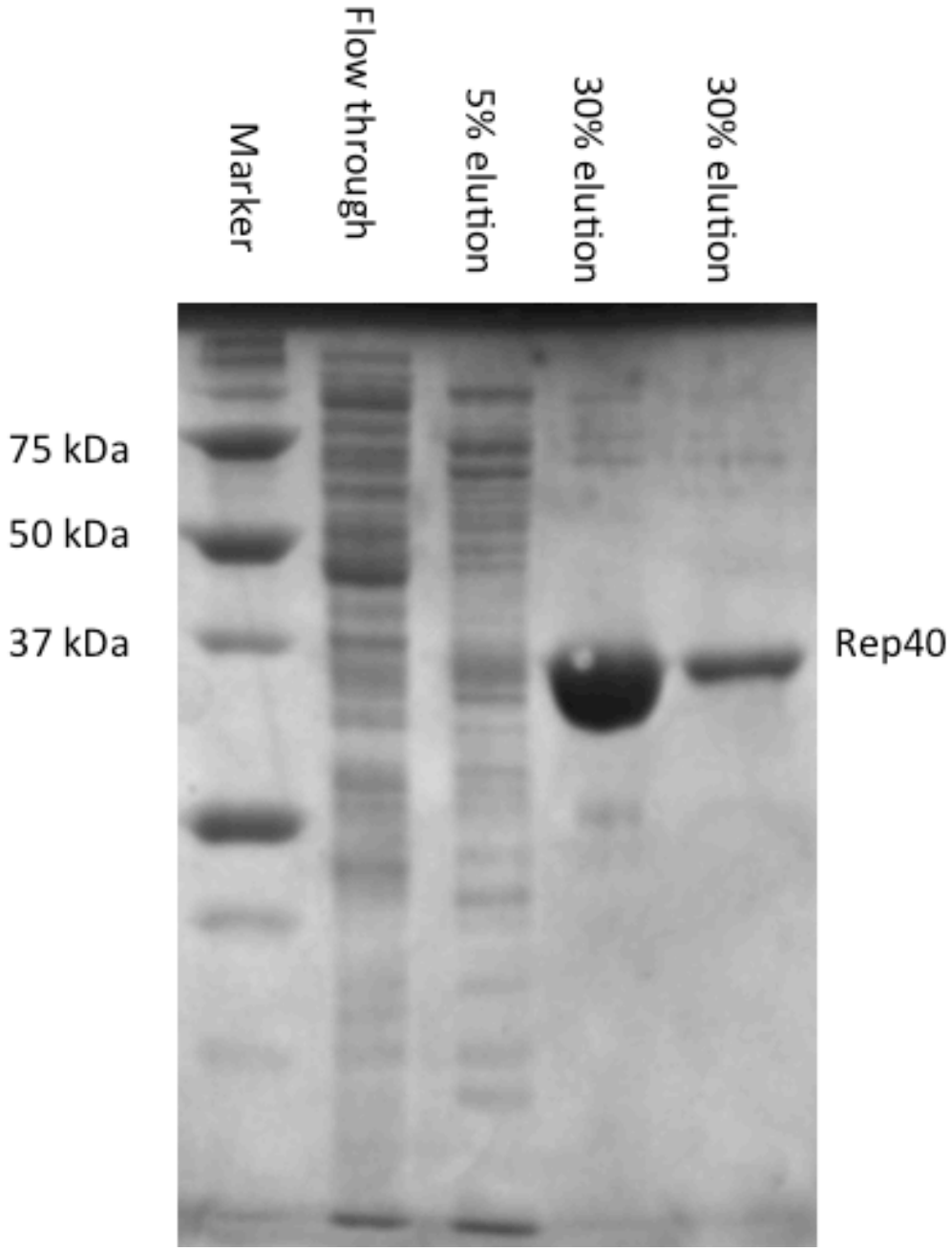


Figure 4.4: Chromatograph and Polyacrylamide Gel of HiLoad Superdex 200 16/60 Column Purification

Figure 4.5: 10% SDS-polyacrylamide gel showing the purification of Rep40 from cell lysate. The elution of the column using 30% buffer B results in an almost pure protein solution.



Polyacrylamide Gel of Rep40 Purification on Nickel-NTA Column

4.2 Binding Curves of Rep68 Individual Domains with Human AAVS1 DNA

For anisotropic binding curves a 41 base pair section of the AAVS1 site was used. One strand of this double stranded DNA was tagged with a fluorescein tag at the 5' prime end. All measurements were taken in triplicate, and then averaged to arrive at the final curve. Then data was then entered into OriginLab and analyzed using a non-linear curve fit using a one-site binding formula. The dissociation constant was determined for the Rep68* mutant using a Hill co-efficient of 1.3 due to presumed cooperativity between individual proteins. This coefficient was calculated in OriginLab. The dissociation constants of the individual domains, including Rep40 (helicase), OBD (OBD208), and OBD and linker (OBD224), were determined using a single binding site. All values were converted to percent of DNA bound, except for OBD208 for which no dissociation constant could be attained due to poor binding. The results illustrate that the individual domains have very poor binding as compared to the whole Rep68 protein. Rep68* had a dissociation constant of 69.01 nM, while Rep40 and OBD224 had dissociation constants of 22.05 μM and 50.33 μM respectively. A dissociation constant could not be determined for OBD208 because binding was so poor.

Figure 4.6: Binding curves for Rep68*and the individual domains for Rep68. Above each curve is a cartoon schematic showing which domains were a part of the construct. Results were reporte in percent of AAVS1 DNA bound against protein concentration. The dissociation constant for the whole Rep68* protein as well as the individual domains is displayed in a table.

Protein	Dissociation Constant
Rep68*	69 nM
Rep40	22 μM
OBD208	>>60 μM
OBD224	50 μM

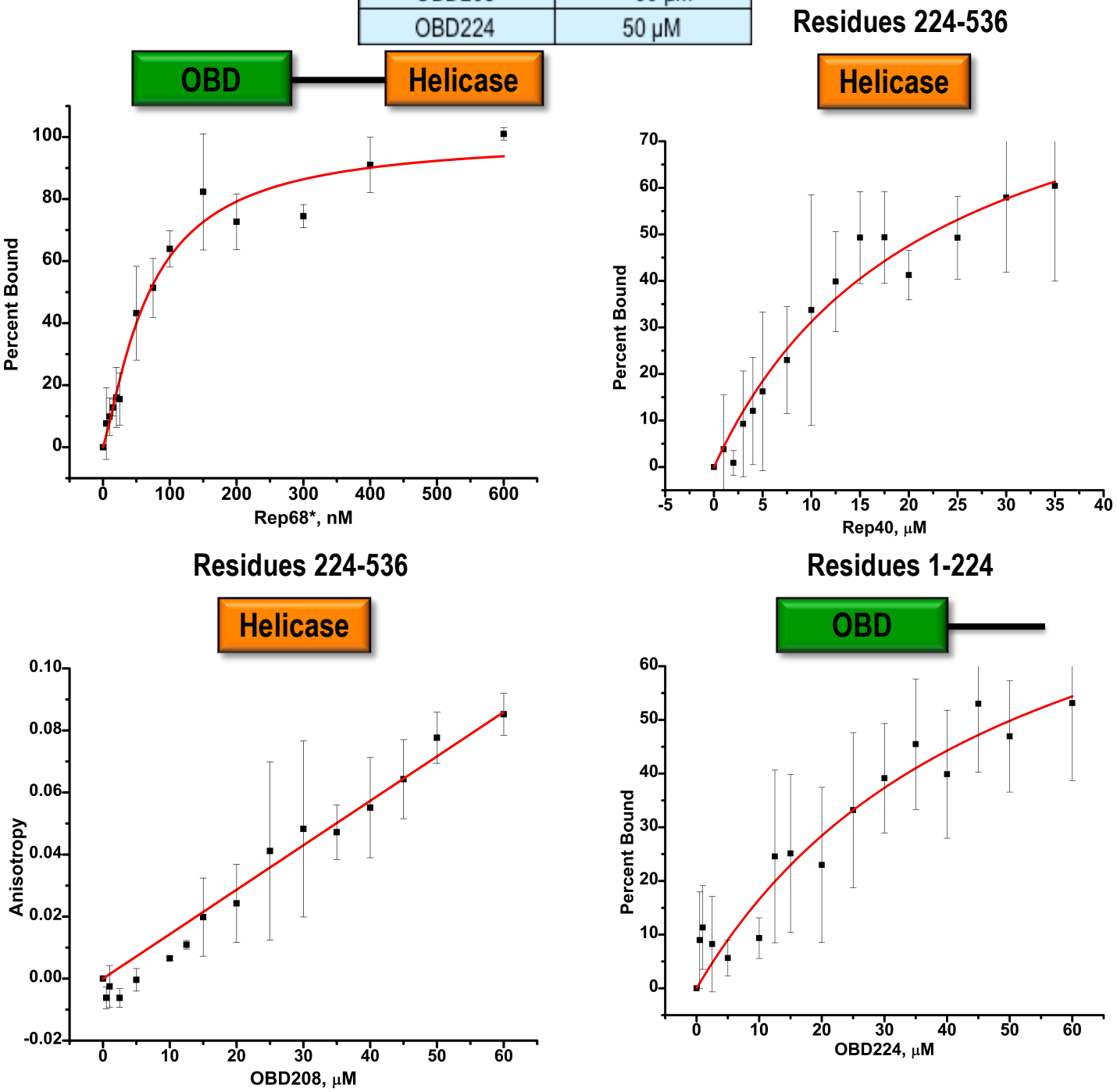
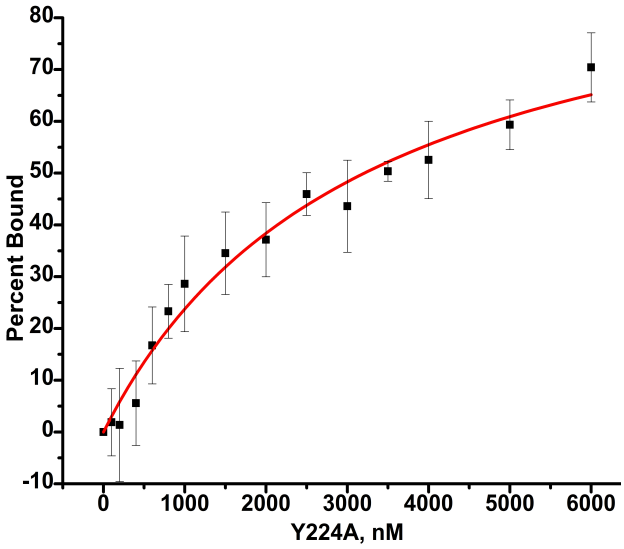


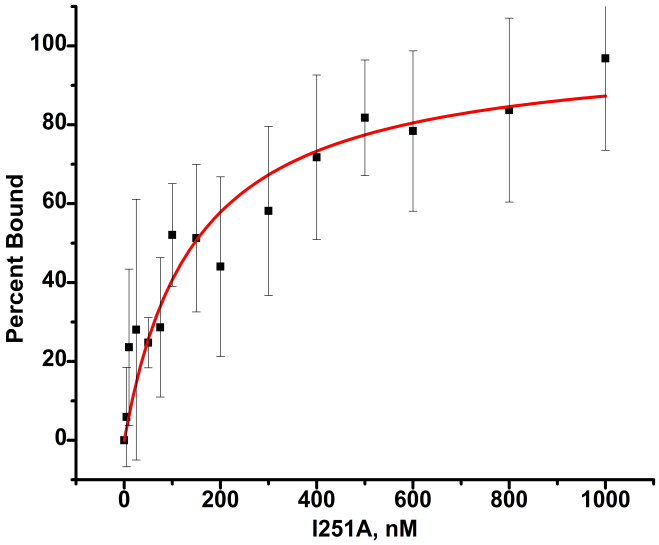
Figure 4.6: Binding Curves and Dissociation Constants for Rep68* and Individual Domains

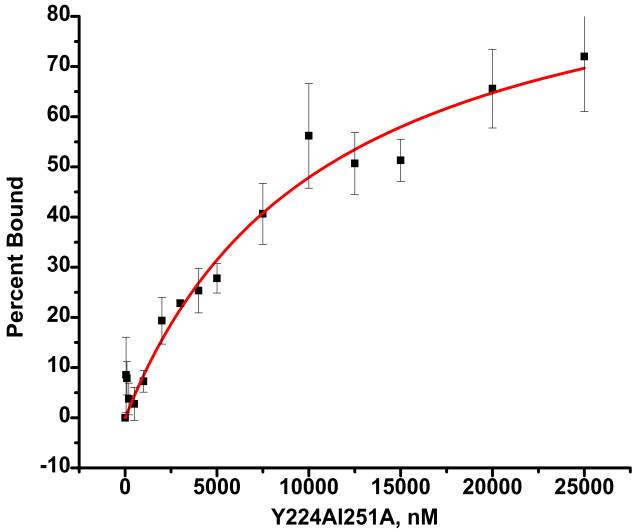
4.3 Binding Curves for Rep68* and Mutants with Human AAVS1 DNA

Rep68 has been shown, by Francisco Zarate-Perez, to oligomerize into a heptameric complex when bound to human AAVS1 DNA. The second aim of this study was to characterize how oligomerization of Rep68 protein molecules was involved in binding of and assembly on human AAVS1 DNA. The results from this section clearly illustrate that oligomerization is an important step in maintaining and stabilizing binding of human AAVS1 DNA. Binding curves were also obtained for the mutants of Rep68*, including Rep68* I251A, Rep68* Y224A, and Rep68*Y224AI251A. It has been demonstrated that Rep68 normally forms large oligomers (Zarate-Perez et. al, 2013). These mutations were targeted to interfere with oligomerization of the protein. Previous studies have shown that these mutations interfere with the ability of the protein to form oligomers(Zarate-Perez et. al, 2012). The results showed that mutants with reduced oligomerization showed greatly reduced binding ability. The dissociation constants for Rep68*I251A, Rep68*Y224A and Rep68*Y224AI251A were 145.69 nM, 3.21 μM, and 10.88 μM respectively.

Figure 4.7: Binding curves for Rep68*Y224A mutant, Rep68*I251A mutant, and Rep68*Y224AI251A mutant with AAVS1 DNA. Results are reported in percent of AAVS1 DNA bound against protein concentration. Also displayed is a table comparing the dissociation constants of these three mutants with AAVS1 DNA with the dissociation constant for Rep68*.







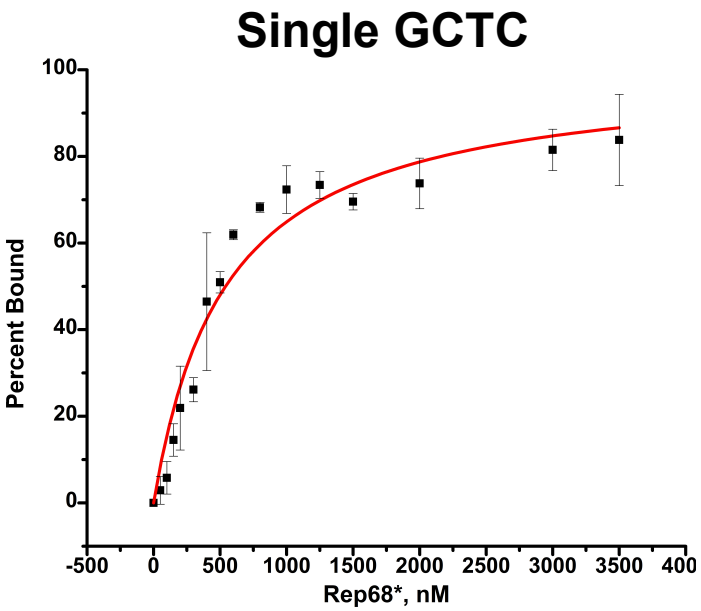
Protein	Dissociation Constant
Rep68*	69 nM
Rep68* Y224A	3.21 μM
Rep68* I251A	145 nM
Rep68* Y224AI251A	10 μM

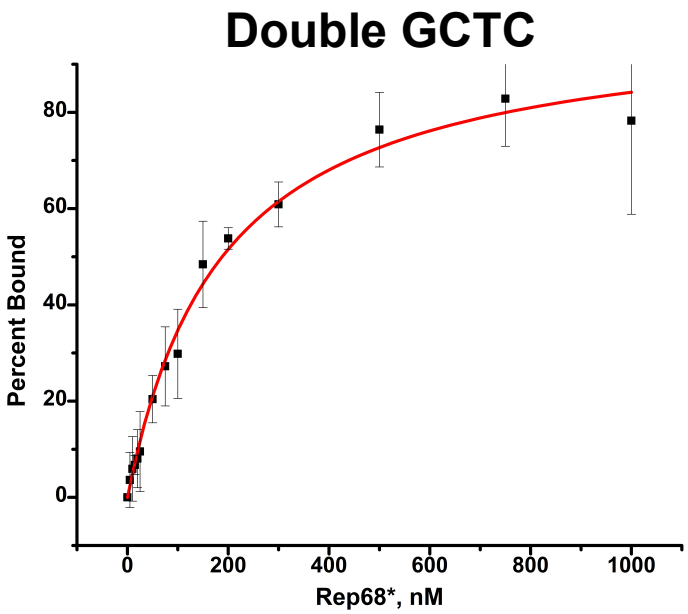
Binding Curves and Dissociation Constants for Rep68* Mutants

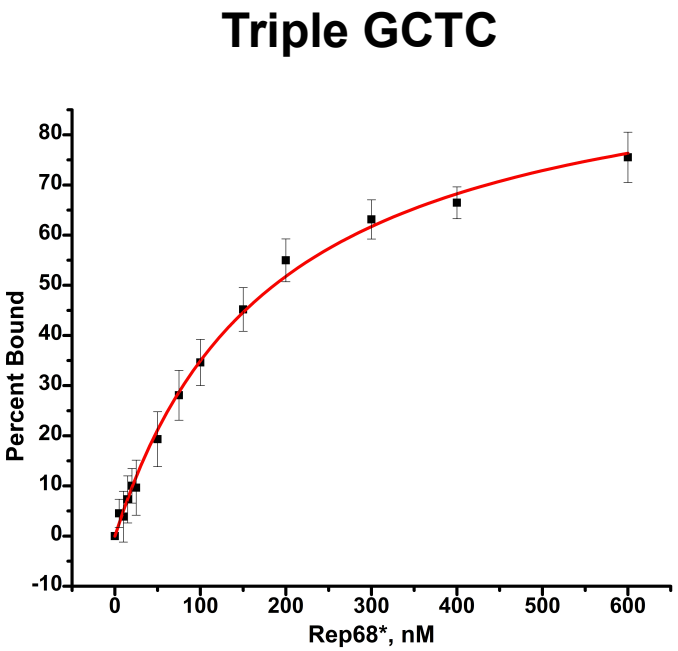
4.4 Binding Curves for Rep68* with Mutant DNA

Rep68 specifically targets a sequence in the human AAVS1 DNA site known as the RBS, which is characterized by three GCTC repeats. Understanding how Rep68 targets this sequence is an important research topic for designing chimeric proteins. Therefore the third aim of this study was to determine the minimum number of GCTC repeats that are necessary for effective Rep68 targeting and binding of DNA. 41 base pair DNA strands were prepared that had random sequences, but contained either one, two or three GCTC repeats. The Rep Binding Site (RBS) contains three GCTC repeats that help Rep proteins specifically bind DNA. By seeing how well Rep68* binds random DNA sequences with varied numbers of GCTC repeats it was hoped that a minimal number of GCTC repeats for effective binding of Rep proteins to DNA.

Figure 4.8: Binding curves for Rep68* with a 41 base pair DNA sequences that contains one GCTC, two GCTC repeats, or three GCTC repeats. The curves show results in percent of DNA bound against protein concentration. A table also displays







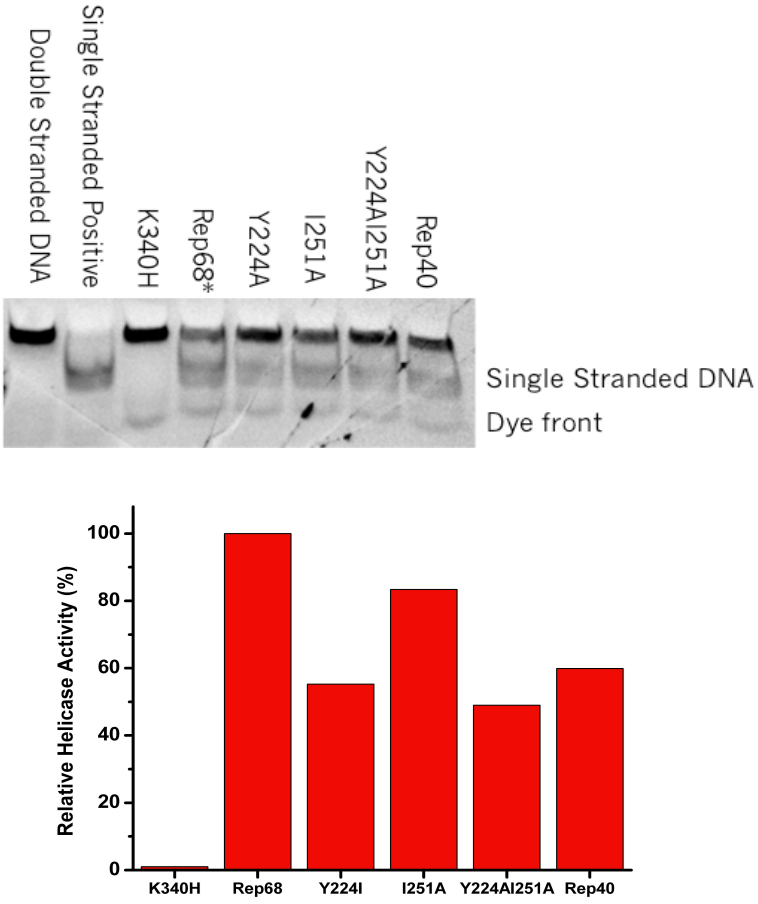
DNA	Dissociation Constant
Single Repeat	540 nM
Double Repeat	187 nM
Triple Repeat	186 nM

Binding Curves and Dissociation Constants for Rep68* with Mutant DNA Sequences

4.5 Helicase Assay

The fourth aim of the study was to show whether ability to oligomerize had any effect on Rep68's helicase activity. The helicase assay shows that mutants that are incapable of oligomerization, and bind DNA poorly are still capable of helicase activity. When the double stranded DNA is unwound by the Rep68 protein, the shorter fluorescent strand is prevented from re-annealing by the trap DNA, which competes to bind the complementary strand. Because the trap DNA is present in 5x the concentration, it out-competes the fluorescent strand. This fluorescent strand then travels farther in the gel. So, a lower band is seen as a positive result. Even though a lower band is present in Rep40 and the mutants with reduced oligomerization, the helicase activity is reduced. This was determined by quantifying the pixel density in the lower bands using the program Fiji, which measures pixel intesity. This was the converted to a relative activity based on the intensity of the lower band of Rep68*. Rep68*K340H, a mutant with no helicase activity, was used as a baseline control. In addition, the results showed that as oligomerization ability was reduced helicase acitivy was also somewhat diminished. The results are shown in the following polyacrylamide gel.

Figure 4.9: Polyacrylamide gel showing the result of a helicase assay. The lower band corresponds to fluorescent DNA that has been unwound and trapped by the single stranded "trap" DNA. The upper band corresponds to intact double stranded DNA. A positive control of tagged single stranded DNA was used in the second lane. This result clearly illustrates that mutants that fail to oligomerize still retain helicase activity. In addition, a bar graph of the relative helicase activities based on the pixel density of the single stranded DNA bands is shown. Rep68* is considered to have a relative activity of 100%.

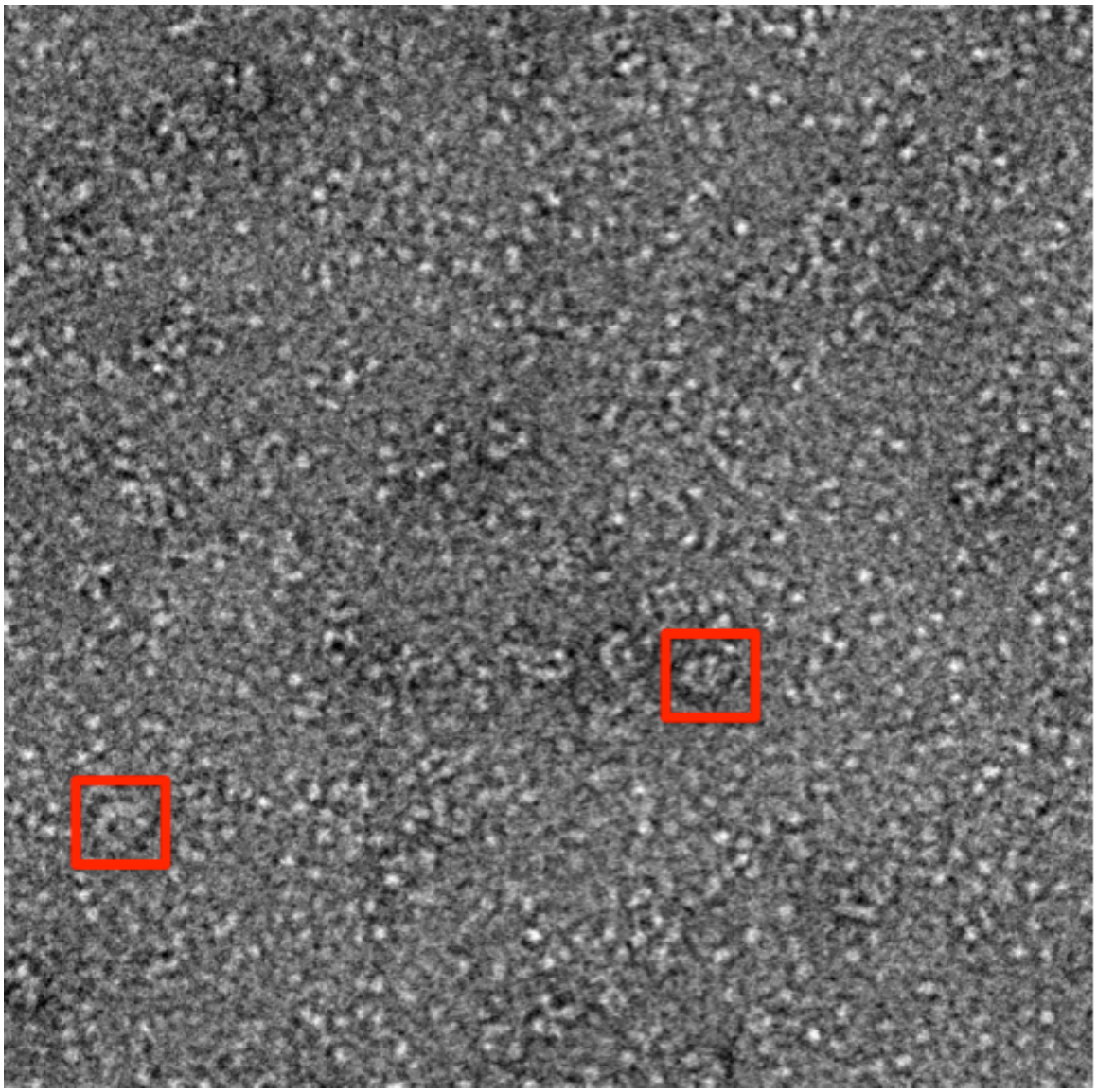


Polyacrylamide Gel of Helicase Assay and Graph of Relative Intensities

4.6 Transmission Electron Microscopy Analysis

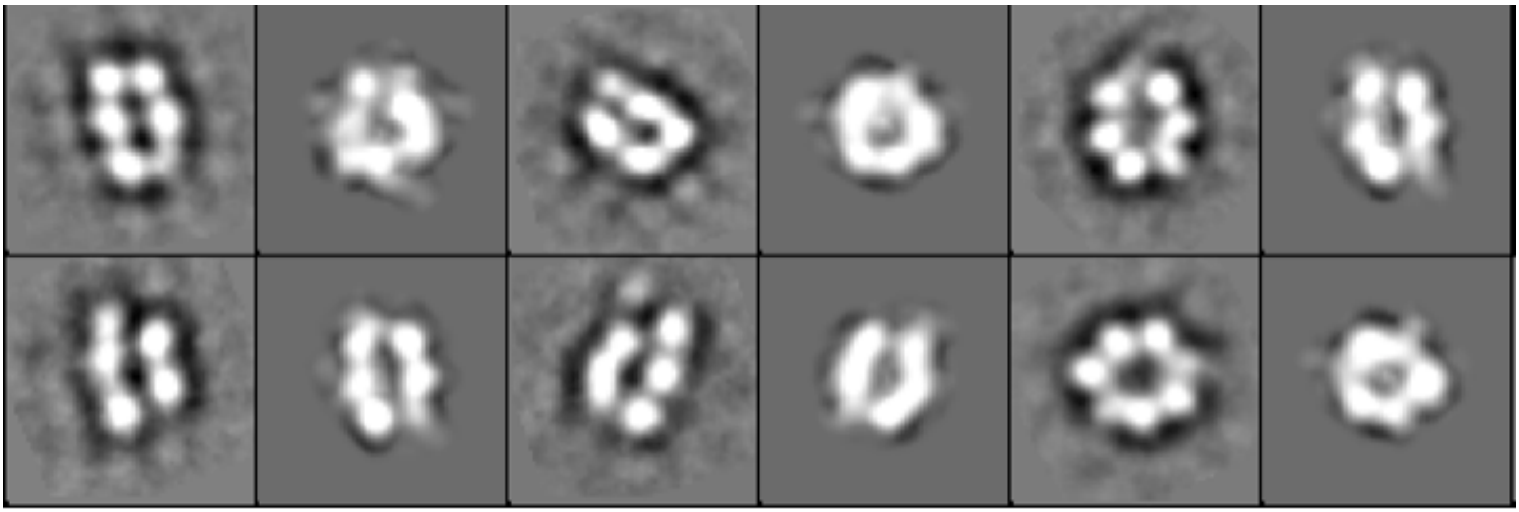
The transmission electron microscopy analysis combined 2,557 particles. Individual particles were picked from 200 individual micrographs. A sample micrograph is seen in figure 3.12. By combining a large number of particles, a degree of accuracy can be obtained. “Bad” particles need to be removed because having a bad particle does more damage than having a single additional particle will add to the reconstruction. These particles are then analyzed using EMAN2 to group them into discrete classes. Essentially, particles that looked similar were used to construct individual classes. Similarly to removing bad particles, bad classes are now removed. An initial model must then be constructed. The 3-D re-projection of the initial model can then be displayed next to the class averages that helped to reconstruct the model. This can be seen in figure 3.11 with the classes on the left and the 3D re-projection on the right. This initial model must then be refined. This model refinement produces a refined, or final model. The initial and refined models were visualized and analyzed ChimeraUCSF. The refined model had dimensions of 148.9Å by 103Å. The model showed what appeared to be individual Rep68 proteins. However, this model was of low quality in part, because symmetry was not imposed. According to the authors of the EMAN software without symmetry imposition during reconstruction “symmetry-breaking artifacts will occur due to high noise levels present in the individual particle images,” (Ludtke et. al.,1999). Future studies will need to help determine the stoichiometry and whether symmetry is seen in order for future models to be reconstructed with symmetry imposition. This will allow for a higher degree of accuracy.

Figure 4.10: An individual micrograph of Rep68ITR165 complex. Particles of the complex are boxed in red.



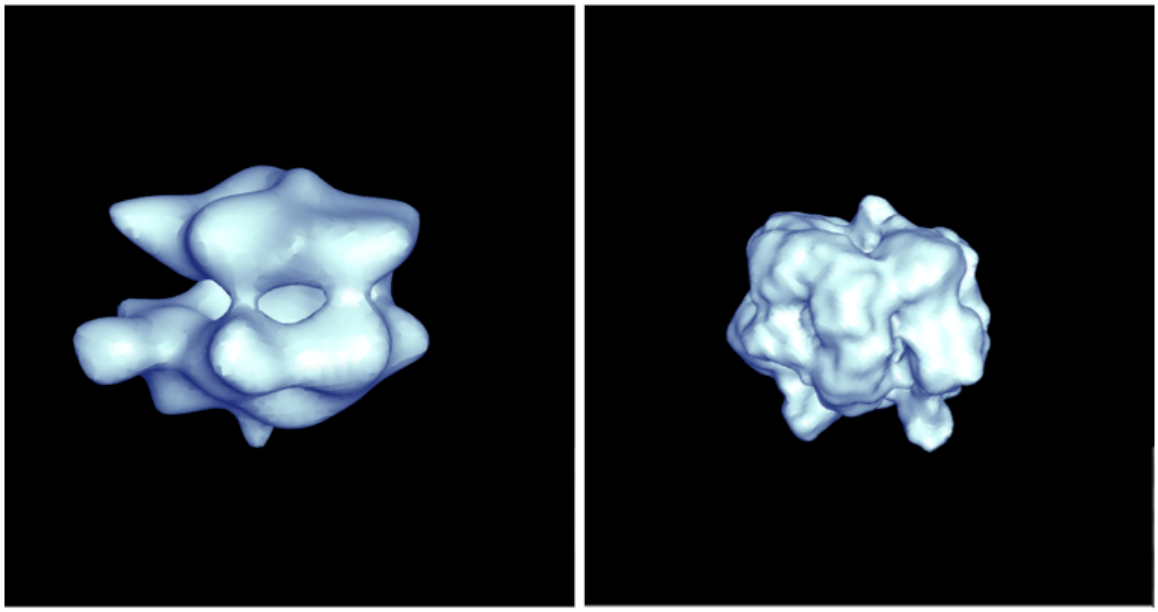
Micrograph of Rep68ITR165 Complexes

Figure 4.11: Shows a selection of class averages used for 3D reconstruction, side-by-side with re-projections of the 3D map.



Classes and 2-D Projections of Rep68ITR165 Complexes in EMAN2

Figure 4.12: Shows the result of EMAN2 reconstruction from the 2-D projections. An initial model is seen on the left, and a refined final model is seen on the right. The refined model has several bean shaped segments, which correspond to individual Rep68 proteins.

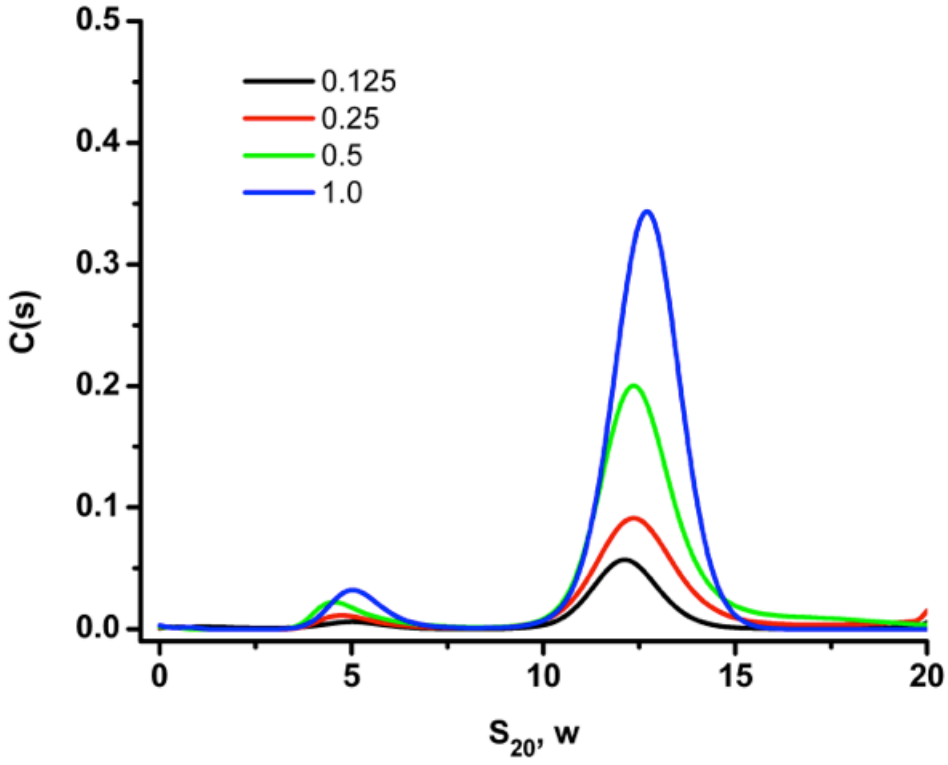


Initial and Refined Models of Rep68ITR Complex

4.7 Analytical Ultracentrifugation

Results of analytical ultracentrifugation of a Rep68*-ITR DNA complex resulted in the following plot of sedimentation coefficients (Figure 3.15). Using SEDFIT the molecular weight of the complex was calculated to be `403kDa. This gives a preliminary stoichiometry of ~6 protein molecules per ITR DNA molecule. However, further analysis needs to be done to confirm this preliminary result, and therefore conclusions about stoichiometry cannot yet be drawn. However, what can be said is that it appears that Rep68*-ITR DNA forms a large, stable oligomeric complex.

Figure 4.13: Plot of sedimentation coefficients for Rep68*-ITR complexes at various optical densities. The center of the peaks appears to give an S-value 12.6. The vertical axis corresponds to concentration, while the horizontal axis corresponds to S-value.



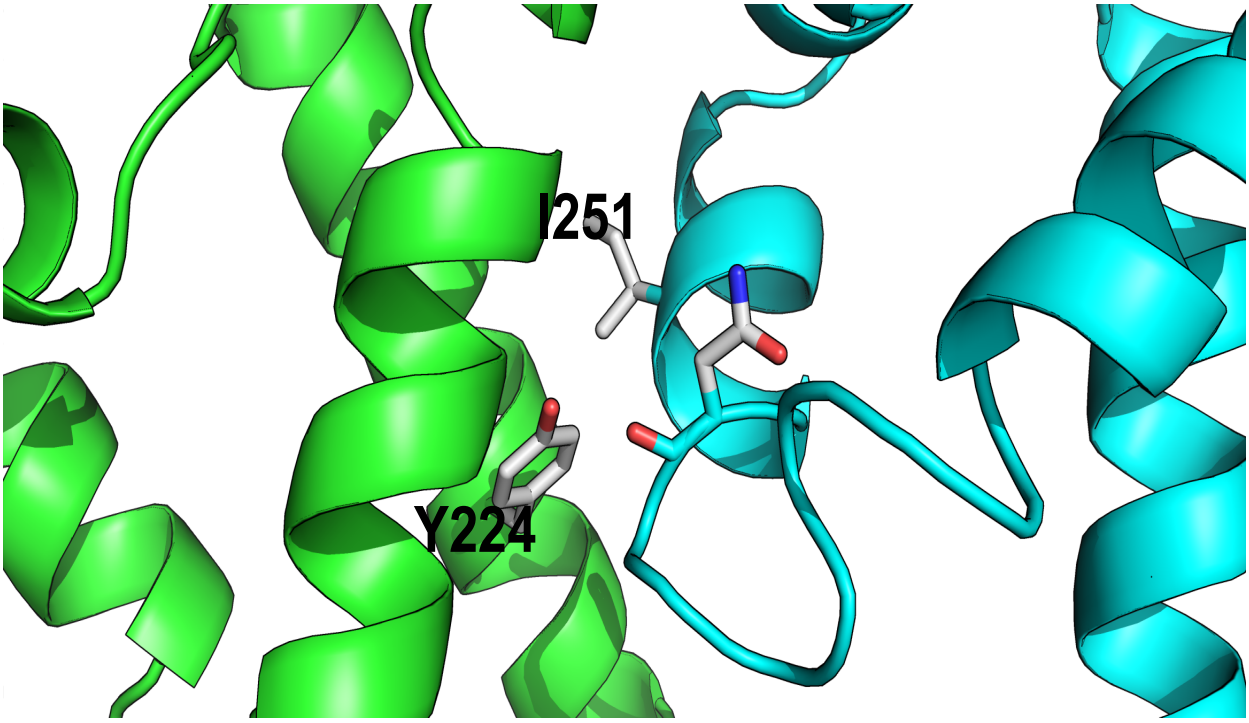
AUC Sedimentation Velocity Results of Rep68ITR Complex

Discussion

These studies characterized the binding of Rep68 protein to AAVS1 DNA, as well as to characterize the stoichiometry of the binding of Rep68 to the ITR section of the AAV genome. The first section of these studies characterized the binding of the individual domains of Rep68, the OBD and the helicase domain, to AAVS1 DNA. The results illustrated, vividly, that the individual domains bind extremely poorly. This supports a view that the individual domains of the protein must bind in concert, simultaneously, in order to effectively bind DNA. Alone they cannot bind DNA in any substantive fashion. The helicase domain also bound with significantly greater affinity than the OBD or the OBD plus the linker domain. The helicase domain had a dissociation constant of 22.05 μM, much smaller than the dissociation constants of the OBD or the OBD with the linker. In fact, no dissociation constant could be determined for the OBD by itself. This illustrates that the helicase domain likely provides much of the affinity for binding, while the OBD likely provides much of the specificity for the interaction with the RBS.

The next section of the study illustrated, using mutants that lacked the ability to oligomerize, that oligeromerization of Rep68 is critical to sustaining binding AAVS1 DNA. Two residues, Y224 and I251, have been shown to be critical to oligomerization. They interact in the interface of two Rep68 proteins and help sustain oligomerization. When these residues are mutated to alanines, this interaction is significantly reduced or abolished, resulting in an inability to oligomerize into a large heptameric complex as is normally seen.

Figure 5.1: A structure showing the interface between two Rep68 proteins. I251 and Y224, residues important in oligomerization, are labeled. It is postulated that these residues interact, either through hydrophobic interactions or some other interaction, to help stabilize oligomerization. This structure comes from a crystal structure of Rep68 from Carlos Escalante, PhD and Faik Musayev PhD.



Structure Illustrating Rep68 Oligomer Interface

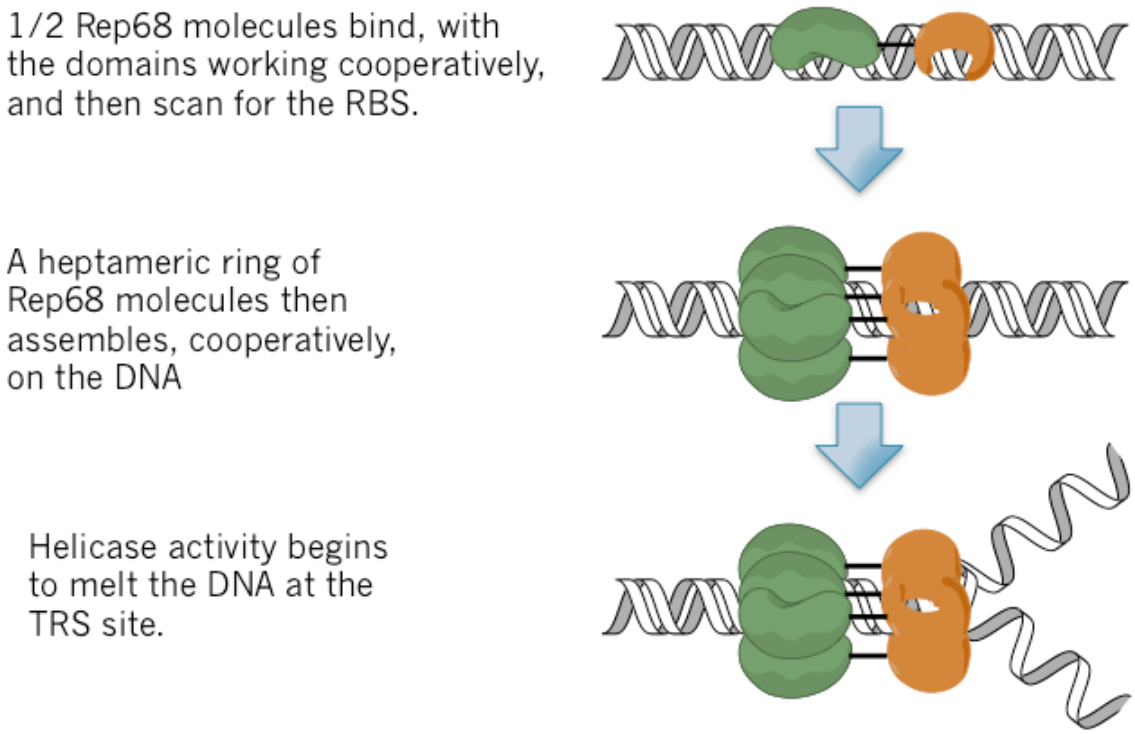
Three mutants were used to demonstrate the effect that oligomerization has on binding. These mutants, Y224A, I251A, and Y224AI251A, showed that oligomerization is critical for sustaining binding. The dissociation constants for these mutants were all more than double the dissociation constant of Rep68* and the double mutant had a dissociation constant, 10.88 μM, several orders of magnitude larger than Rep68*. Because it has been demonstrated that Rep68 assembles on AAVS1 DNA as a ring, it therefore cannot bind pre-oligomerized. Therefore it must assemble as a oligomeric complex on the DNA itself. This means that once one Rep68 protein binds AAVS1 DNA, that this binding is sustained, likely through cooperativity, by further binding events. The binding of further Rep68 proteins as a heptameric ring prevents a high "off" rate, and helps stabilize the interaction with DNA.

Lastly, by using mutant DNA sequences with one, two, or all three of the GCTC repeats that characterize the RBS, it was determined that two repeats are sufficient for effective binding. What the results showed was that DNA with a single repeat resulted in relatively poor binding with Rep68*, with a dissociation constant of 540.32 nM. Adding a second repeat illustrated a marked improvement in the binding. A second repeat lower the dissociation constant to 187.90 nM. However, adding a third repeat showed virtually no added benefit and resulted in a dissociation constant of 186.13 nM. This illustrates that two GCTC repeats are sufficient for effective binding, and that added repeats have marginal or no added benefit.

These results, taken collectively, illustrate that in order to bind AAVS1 DNA, the domains must work in concert, that in order to stabilize the interactions with DNA the individual proteins must oligomerize to form a heptameric ring, and that there must be, at minimum, two GCTC repeats for the protein to be able to target a sequence.

The overall conclusion of these studies resulted in an updated model of how Rep68 binds AAVS1 DNA. In a previous model, it was assumed that five individual OBD domains bound five tetranucleotide repeats in the RBS (these include the three GCTC repeats). The next step in the previous model was that a hexameric ring of helicase domains assembled near the TRS site. These new findings allow for a newer, more accurate view of the binding process to be proposed. First, we know that individual OBD domains cannot maintain binding on their own. So, we know that the first step of the binding interaction must be a whole Rep68 protein binding. Secondly, because the data clearly show that only two GCTC repeats are necessary for effective binding and that added repeats do not enhance binding, it is likely that only one to two Rep68 proteins need to bind initially. The next step is that a heptameric ring of Rep68 proteins assembles onto the site, exhibiting cooperativity through oligomerization. The final step in this initial process is the unwinding of DNA through helicase activity.

Figure 5.2: A cartoon of a new proposed model for initial Rep68-AAVS1 DNA interactions.



Cartoon of New Proposed Model for Rep68 Binding to AAVS1 DNA

As the only non-structural proteins in AAV the Rep proteins must fulfill a wide variety of tasks. Because of this, chimeric Rep proteins are essential to any future AAV gene therapy vector. Engineering these future chimeric proteins depends upon understanding the many functions and mechanisms of these proteins, including their interactions with both viral and human DNA. These studies helped illuminate some of these interactions, including the mechanism of cooperative assembly of Rep68 protein on AAVS1 DNA.

Literature cited

Daya, S., & Berns, K.I. (2008). Gene Therapy Using Adeno-Associated Virus Vectors. *Clinical Microbiology Review,* 21(4): 583–593.

Zarate-Pereze, F. ,Mansilla-Soto, J. Bardelli, M., Burgner, J.W., Villamil-Jarauta, M., Kekilli, D., Samso, M., Linden, R.M., Escalante, C.R. (2013). Oligomeric Properties of Adeno-Associated Virus Rep68 Reflects Its Multifunctionality. *Journal of Virology*, 87(2): 1232–1241.

Lebowitz, J., Lewis, M.S., Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Science*, 11(9): 2067–2079.

Frank, J. (2009). Single-Particle Reconstruction of Biological Macromolecules in electron microscopy – 30 years. *Quarterly Review of Biophysics,* 42(3): 139-158.

Linden, M.R., Winocour, E., Berns, K.I. (1996). The recombination signals for adeno-associated virus site-specific integration. *Proceedings of the National Academy of Science*, 93: 7966-7972

Wu, Z., Yang, H., Colosi, P. (2010). Effect of Genome Size on AAV Vector Packaging. *Molecular Therapy*, 18(1): 80-86.

MacCarty, D.M., Young, S.M. Jr., Samulski, R.J. (2004). Intregration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annual Review of Genetics*, 38: 819-45.

Lea, W.A., Simeonov, A. (2011). Fluorescence Polarization Assays in Small Molecule Screening. *Expert Opinion on drug discovery*, 6(1): 17-32.

Pollard, T.D. (2010). A Guide to Simple and Informative Binding Assays. *Molecular Biology of the Cell*, 21(23): 4061-4067.

Harvard Medical School. Basic Science Partnership: How Does an Electron Microscope Basically Work? http://bsp.med.harvard.edu/node/221 (accessed 19 May, 2014).

Zarate-Perez, F., Bardelli, M., Burgner J.W., Villamil-Jarauta, M., Das, K., Kekilli, D. Mansilla-Soto, J., Linden, R.M., Escalante, C.R. (2012). The Interdomain Linker of AAV-2 Rep68 Is an Integral Part of Its Oligomerization Domain: Role of a Conserved SF3 Helicase Residue in Oligomerization. *PLOSP Pathogens*, http://www.plospathogens.org/article/info%3Adoi%2F10.1371%2Fjournal.ppat.1002764#s2 (accessed 19 May, 2014).

The National Insititutes of Health. Genetic Home Reference: What is Gene Therapy? http://ghr.nlm.nih.gov/handbook/therapy/genetherapy (accessed 19 May, 2014).

Cole, J.L., Lary, J.W., Moody, T.P., Laue, T.M. (2008). Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods in Cell Biology*, 84: 143-179.

Kuo, J. *Electron Microscopy: Methods and Protocols*, 2nd ed.; Humana Press Inc.: Totawa, New Jersey, 2007.

Gonçalves, M.A.F.V. (2005). Adeno-associated virus: from defective virus to effective vector. *Virology Journal*, 2:43.

Henckaerts, E., Linden, R.M. (2010). Adeno-associated virus: a key to the human genome? *Future Viology*, 5(5): 555-574.

Zhou, X., Zolotukhin, I., Im, D.S., Muzyczka, N. (1999). Biochemical Characterization of Aden-associated Virus Rep68 DNA Helicase and ATPase Activities. *Journal of Viriology*, 73(2): 1580-1590.

Im, D.S., Muzyczka, N. (1990). The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell*, 61(3): 447-457.

Castro-Hartmann, P., Heck, G., Eltit, J.M., Fawcett, P., Samso, M. (2013). The ArrayGrid: A methodology for applying multiple samples to a single TEM specimen grid. *Ultramicroscopy*, 135: 105-112.

Harauz, G., Boekema, E., van Heel, M. (1988). Statistical Image Analysis of Electron Micrographs of Ribosomal Subunits. *Methods in Enzymology*, 164: 35-49.

Ludtke, S.J., Baldwin, P.R., Chiu, W. (1999). EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. *Journal of Structural Biology*, 128(1): 82-97.

Hickman, A.N., Ronning, D.R., Perez, Z.N., Kotin, R.M., Dyda, F. The Nuclease Domain of Adeno-Associated Virus Rep Coordinates Replication Initiation Using Two Distinct DNA Recognition Interfaces. (2004). *Molecular Cell,* 13: 403-414.

Serdyuk, I.N., Zaccai, R.N., Zaccai, J. *Methods in Molecular Biophysics: Structure,*

Dynamics, Function; Cambridge University Press: New York, 2007.

Ludtke, S.J. (2010). 3-D structures of macromolecules using single particle analysis in EMAN. *Methods in Molecular Biology*, 673: 157-173.

Smith, K.C.A., Oatley, M.A. (1955). The scanning electron microscope and its field of applications. *British Journal of Applied Physics*, 6: 391-399.

Ruprecht, J., Nield, J. (2001). Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction. *Progress in Biophysics and Molecular Biology*, 27(3): 121-164.

Schuck, P. (2000). Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal*, 78:1606-1619.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 13: 1605-12.