# BIOCHEMICAL AND STRUCTURAL ANALYSIS OF SV40 LARGE T ANTIGEN: INSIGHTS INTO CHAPERONE MEDIATED INACTIVATION OF RETINOBLASTOMA TUMOR SUPPRESSOR PROTEIN

By

Christina Kay Williams

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in
Biochemistry
May, 2012
Nashville, Tennessee

Approved by:

Professor Walter J. Chazin, Ph.D.

Professor Richard Armstrong, Ph.D.

Professor Brandt Eichman, Ph.D.

Professor Ellen Fanning, Ph.D.

Professor Scott Hiebert Ph.D.

*To my loving and supportive parents,*
*Austin and Yetta Williams*
*Your prayers brought me through.*

*&*

*In loving memory of my big sister,*
*Kendra Williams-Wood*

## ACKNOWLEDGEMENTS

I am exceedingly grateful to my advisor, Dr. Walter Chazin. Throughout this journey he recognized my personality, skills, and desires and mentored me accordingly. He extracted the scientist from within and endorsed my pursuit of non-research interests. He became my champion when completing this dissertation work seemed highly unlikely. He was understanding and supportive during my treatment for depression. I will always be thankful that he allowed me to be a member of his laboratory. I also thank his assistant, Veronda McClain, for always having a kind and encouraging word for me. Those small moments made a big difference.

I am thankful for my committee members: Drs. Ellen Fanning, Brandt Eichman, Scott Hiebert and Richard Armstrong. My research was full of difficulties and delays but they remained committed in their support, believing that I could get it all done. I know that without their motivating and encouraging word I would have given up.

I and this research project is indebted to the generosity of Dr. Michal Hammel at the SIBYLS beamline. At the very last minute he agreed to perform the critical SAXS experiment that I needed to complete my research. I greatly appreciate that he gave of his time to help me succeed.

I acknowledge the members of the Chazin lab that enhanced my time as a graduate student. I specifically recognize Dr. Siva Vaithiyalingam who really helped me collect and analyze my data as well as helping write my manuscript toward the end of this process. I thank Dr. Steve Damo for his conversation, willingness to answer my questions and to read/edit my documents. I also thank Dr. Dalyir Pretto and Ileana Alers for being great office mates and friends. They opened their life to me and

allowed me to share my life with them. We discussed everything from food, shopping to relationships which provided a calming balance to rollercoaster ride that is scientific research. Even after they finished their degrees, they encouraged and cheered me to the finish line. I wish them all the blessings that life has to offer.

In 2004, I moved to Nashville without family or friends in the vicinity. In the Bioregulation class I met Drs. Robin Bairley, Karen Riggins, Sydika Banks and Kimberly Mulligan. These ladies became my friends and family. Through failed experiments, illness, academic politics, break-ups, road trips, potlucks and restaurants, purchasing houses, marriages and a baby we formed an inseparable bond. I am so happy to say that we defied the statistics which stated that 1 in 5 students would not complete the graduate school process and all persevered through blood, sweat and tear, to earn our Ph.D degree!! Woo Woo!! I have no doubt that I would not be at this point without them talking me out of quitting. I will love them forever.

I am happy to have met Dan Anderson on our Vanderbilt interview weekend. He allowed me to stay with him while I searched for an apartment in Nashville and ever since has let me vent my joys and defeats during our monthly meet-ups at Starbucks.

I am so grateful for my Cornerstone church family. This church embraced, encouraged and nurtured me into a deeper relationship with Christ. My friends Jenny Crick, Sharon Twist, Stephanie Riley, Tineisha Weddle, and Kristi Sabatini did not have any idea what I was studying or understand why it was taking so long but they hugged me, prayed for me and provided me with fun distractions. They truly helped me build the faith and trust in God that was instrumental in getting me to the end of this journey.

In 2007, I met my boyfriend, Stevie Keeton. He has truly been an unexpected blessing in my life. Science is not his background but he read my Science magazines and various websites so that he could better understand what I was doing. He would show me neat scientific facts that helped me not lose my enthusiasm for science when my research was not going well. He has made me a better scientist and woman.

Finally, I acknowledge my family. There are not enough words to describe how thankful I am that Austin and Yetta Williams are my parents. They fully supported my decision to move from New York for the first time to Tennessee. I know that God heard and answered every prayer request they sent. I am so thankful for my sister, Kendra Williams-Wood, whose struggle with breast cancer was a catalyst to the start of my graduate career and memory of her propelled me through. I miss her greatly. I appreciate the continued support and prayers of my sister Beppy Williams. She was the first to get an advance degree in our family which made me determined to complete mine. I thank my brothers, Stephen and Marcus Williams whose courage to join the Marines boosted my courage to return to the lab day after day. To my brother Isaac Williams, your name means laughter and you definitely provided me with much needed laughs over the years. I am grateful for all of my extended family like my Uncle Ivory who checked the website for the announcement of my defense every day for a year and was the first to offer congratulations. I thank the Locketts for locating and welcoming us to the family and encouraging me to the end. I am beyond blessed to have all of you in my life.

*~"For I know the plans I have for you", declares the Lord, "plans to prosper you and not to harm you, plans to give you hope and a future." –Jeremiah 29:11*

TABLE OF CONTENTS

Page

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AAA+ | ATPase Associated domain |
| $\beta$ ME | $\beta$ -mercaptoethanol |
| CD | Circular Dichroism |
| CPH | Cul7, PARC and HERC2 |
| Cul7 | Cullin 7 |
| CRL | Cullin RING ubiquitin Ligase |
| DNA | Deoxyribonucleic Acid |
| DNAJ | J domain containing protein |
| DNAK | Heat shock 70 protein |
| DLS | Dynamic Light Scattering |
| DTT | Dithiothreotol |
| ER | Endoplasmic Reticulum |
| HD | Helicase Domain |
| HPV | Human Papillomavirus |
| HR | Host Range domain |
| Hsp40 | Heat shock protein 40 |
| Hsc70 | Heat shock cognate protein 70 |
| HSQC | Heteronuclear single quantum coherence |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| ITC | Isothermal Calorimetry |
| JCV | JC virus |
| MHC | Major Histocompatibility Complex |
| Ni-NTA | Nickel nitriloacetic acid |
| NMR | Nuclear magnetic resonance spectroscopy |
| OBD | Origin Binding domain |
| PCR | Polymerase chain reaction |
| PE | Early Promoter |

| | |
|---|---|
| PL | Late Promoter |
| P(r) | Pair-distance distribution function |
| Rb | Retinoblastoma tumor suppressor protein |
| Rg | Radius of Gyration |
| RPA | Replication Protein A |
| SAXS | Small-Angle X-ray Scattering |
| SCF | Skp, Cullin, F box |
| SDS-PAGE | Sodium dodecyl sulfate – polyacrylamide gel electrophoresis |
| SEC | Size exclusion chromatography |
| 17KT | 17K T antigen |
| SV40 | Simian virus 40 |
| Tag | Large T antigen |
| Tag | Small t antigen |
| TROSY | Transverse relaxation optimized spectroscopy |
| VP | Viral Protein |
| Zn-D | Zinc binding domain |

# CHAPTER I

# INTRODUCTION

## Simian Virus 40 Overview

### *Simian Virus 40: The History*

Simian Vacuolating Virus 40 (SV40) is one of the first two DNA tumor viruses discovered. In the 1950's there was a severe outbreak of the polio virus. Scientists Dr. Jonas Salk and Albert Sabin had isolated the poliovirus strain and worked to develop vaccines. Dr. Salk determined that if the poliovirus strain was inactivated by formalin, it could then be injected into children. Dr. Sabin would weaken the live poliovirus strain by passaging it through various host cells and then orally administered to children. In 1956, Dr. Sabin was passaging the poliovirus through African Green Monkey kidney cells. In 1960, Dr. Benjamin Sweet and Dr. Maurice Hilleman, of Merck, discovered that viruses are commonly carried by monkeys and will contaminate tissue cell cultures. SV40 was one of those viruses that was able to contaminate monkey kidney cells.[1] In the same year, Dr. Bernice Eddy, of the NIH, determined that SV40 was indeed contaminating the cell cultures used to passage and grow the poliovirus in the production the polio vaccine. Subsequent research isolated SV40 from normal monkey kidney cells, the Sabin poliovirus vaccine stocks, and another adenovirus vaccine that was utilizing the same type of cells. Upon analysis of the Salk poliovirus vaccine, SV40 was discovered there as well. During the Salk vaccine production, the formalin treatment successfully inactivated the poliovirus but was

insufficient to inactivate SV40. From 1955 to 1963 approximately 100 million people in the United States were accidentally exposed to SV40.[2]

In the monkey, SV40 causes a persistent infection in the kidney without really harmful side effects[3]. Reports that SV40 could produce tumors in hamsters, prompted the government to require screening of all the new stocks of the polio vaccine [4–9]. SV40 was also shown to infect human cells, which prompted decades of epidemiogical research across the United States and Europe, following patients who had received contaminated virus to determine whether they developed cancer at a higher rate[10–13]. The results of these studies caused the Centers for Disease Control and Prevention to pronounce that there is no evidence that SV40-contaminated vaccine stocks have caused cancer in the humans[14]. However, technological advances have caused researchers to revisit the relationship of SV40 and human disease including cancer. Some scientists have discovered traces of SV40 in some human tumors like pleural mesothelioma (a form of lung cancer), osteosarcoma, ependymoma and choroid plexus tumors of the brain, and recently non-Hodgkin's lymphoma[15–24]. Still other researchers have presented evidence that SV40 has no relationship with cancer formation [25–30]. This topic is still rigorously researched and has become highly controversial.

The decades of SV40 study have revealed important information about the virus. SV40 has been classified as a polyomavirus. The name Polyomavirus means a virus with the ability to produce multiple tumors. Polyoma is a genus of viruses that have been researched extensively. This genus used to be a member of the now disbanded *Papovaviridae* family along with the rabbit polyomavirus and papillomavirus. One type of papillomavirus is the Human Papillomavirus (HPV). In recent years HPV has been

identified as a causative agent for more than 90% of human cervical cancer[31]. Having this human cancer inducing cousin fuels the idea that SV40 may have the ability to cause cancer as well.

Along with SV40, the members of the genus Polyomavirus members include the JC virus (JCV) and BK virus (BKV). SV40 is a rhesus monkey virus, but BKV and JCV are human polyomaviruses. Both viruses exist commonly in the human population and will form an infection in 90% of humans early on in their life[32]. In fact, approximately 80% of the United States' adult population has antibodies to both BKV and JCV. JCV can infect the respiratory system, kidneys, or brain of humans. In patients with compromised immune systems, JCV can cause the fatal progressive multifocal leukoencephalopathy, a neurodegenerative disease. This virus affects about 5% of all AIDS patients[33]. Polymerase chain reaction (PCR) data has linked JCV to neural cancers[34]. In humans, BKV produces a mild respiratory infection and can affect the kidneys of immunosuppressed transplant patients. BKV has also been found in human tumors[32]. These three members have approximately 70% sequence identity[35] and have similar life cycles. Therefore, it is understandable that pathogenic and possible oncogenic properties of polyomaviruses have led to an increased effort to determine their biology and dynamics.

***Simian Virus 40: The Biology***

SV40 is a small, non-enveloped, icosahedral DNA virus that is approximately 45 nm in diameter. Figure 1.1 shows an X-ray diffraction image of the SV40 particle (pdb id: 1SVA)[36]. The SV40 virus can infect a cell by binding onto the surface of a cell via the

receptors ganglioside GM1, which is located in lipid rafts, and via the major histocompatibility complex (MHC) receptor class 1 molecule. MHC-1 is expressed by most cell types, which suggests that SV40 should be able to bind to many cell surfaces and infect various types of cells. Once attached to MHC-1, SV40 is enclosed into caveolae, which is then released from the membrane into the cell.[37,38] The virus containing caveolar vesicle associates with the caveosomes in the cytoplasm. SV40 can remain in the caveosomes for several hours before it is moved along microtubules to the endoplasmic reticulum (ER). From the ER, SV40 travels through nuclear pore complexes into the nucleus where it stimulates transcription and replication of its genome.[38] A schematic diagram of this process can be seen if Figure 1.2. If the virus has infected a permissive cell, DNA replication will be allowed resulting in cell lysis. However, in a non-permissive cell, DNA replication is not allowed resulting in an aborted viral infection or oncogenesis.

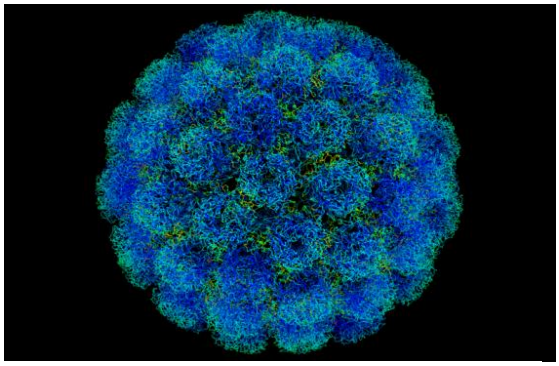


**Figure 1.1: X-ray diffraction image of SV40 virus particle. PDB** id: 1SVA

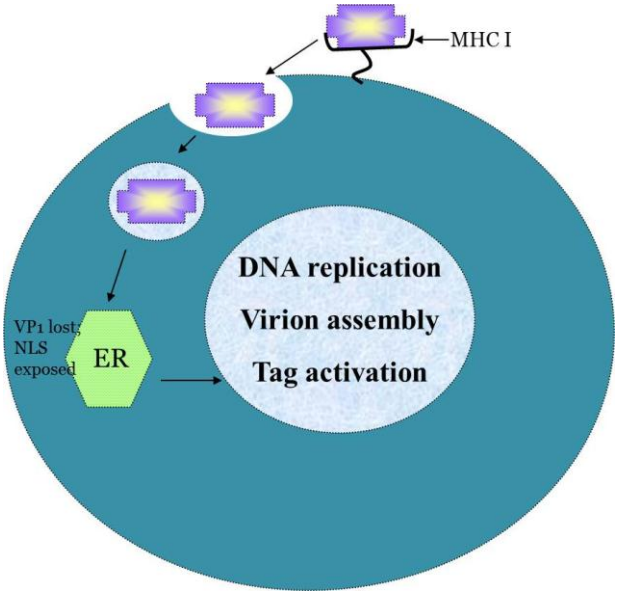



**Figure 1.2: Schematic diagram of SV40 viral infection.** An SV40 virus particle enters the cell via the MHC-1 receptor, travels through the cytoplasm in caveolae to the ER and is then localized to the nucleus.

The genome of SV40 is a single molecule of double-stranded circular DNA that is 5243 bp in length and encodes both early and late genes. The early and late promoters are organized in opposite orientations on either side of the regulatory region, the SV40 origin of replication (ori) (Figure 1.3)[39]. There are seven proteins known to be encoded by the virus. The early promoter is activated by the cellular DNA transcription machinery. The genes encoded from the early promoter (PE) are large T antigen (Tag), small t antigen (tag), and 17K T antigen (17KT).[39] These proteins are expressed soon after viral infection. The molecular weight of Tag is 90 kDa and for tag is 17 kDa. These proteins are translated from the early messenger RNA that is transcribed before the start of viral DNA replication. The large T antigen protein is localized exclusively in the nucleus. The small t antigen exists in all fractions as a soluble form. This activity disrupts the normal pattern of cellular gene expression. The antigen proteins are primary participators in inducing host cell transcription, viral DNA replication, virion assembly and many other functions[40]. Viral DNA replication occurs along with the activation of the late promoter. The genes encoded from the late promoter (PL) are VP1, VP2 and VP3 which are the structural proteins that create the viral capsid. The agno protein is also expressed late and its role may involve assembly and/or disassembly of the viral capsid[40]. Only VP1, VP2 and

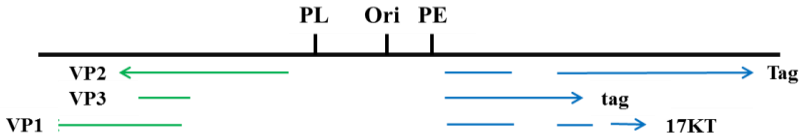



**Figure 1.3. Genome Organization of SV40 virus.** The structural proteins of the virus (green) are encoded from the late promoter (PL). The antigen proteins (blue) are encoded from the early promoter (PE). Ori means the origin. *Adapted from Saenz-Robles et al*, 2001.

VP3 are found in the mature virus particle. The expression of these structural proteins results in the assembly of new virus particles.

As a virus native to kidney cells, which are differentiated cells that don't divide, SV40 had to develop methods for overriding the protective mechanism of cell cycle growth arrest. These methods enable the production of proteins that aid in SV40 genome replication and propagation of viral progeny. It has been determined that T antigen stimulates the majority of these cell seizure activities yielding a cell that has deregulated growth cycles. In fact, T antigen has shown to be necessary and sufficient for cellular transformation[39].

The study of Tag has led to numerous advances in many fields including gene regulation, RNA splicing, nuclear localization, cellular transformation, DNA replication and repair, as well as the functions of tumor suppressor proteins. Study of how T antigen takes over a cell has elucidated the essential cellular proteins and processes. Hence, SV40 virus and its T antigen have proven to be a highly useful model system.

***Simian Virus 40: The Large T antigen***

The SV40 T antigen, a 708 amino acid protein, has the ability to affect many cellular processes. T antigen is the only viral protein required for SV40 DNA replication and cell transformation. T antigen forms oligomeric structures and experiences post-translational modifications. The widespread impact and multiple functions of Tag can be attributed to its multiple domains. The domains interact with one another in a cooperative manner to direct the cellular infection and transformation. Each domain binds to proteins in the host cell and modifies their functions.

T antigen is composed of six structural domains: host range domain (HR), helical domain, AAA+ (ATPases associated with various cellular activities) domain, zinc binding domain (Zn-D), origin binding domain (OBD), and a J domain (DNAJ) (see Figure 1.4)[41,42]. All of the domains have specific biological functions and participate in the coordination of inter and intra-molecular interactions to alter the environment of the host cell to permit and perpetuate viral genome replication and cell transformation.

The HR domain (628-708) occupies the C terminal region of T antigen. While termed a functional domain, neither its exact function nor structural make-up is known. The HR domain is not required for T antigen mediated cellular transformation[43]. It functions to help the Adenovirus replicate in cells (i.e. host range helper function (hr/hf)). The helper function has been defined as the inability of host range mutant viruses to replicate in certain types of African green monkey kidney cells[44]. The HR domain has multiple posttranslational modifications (four phosphorylation sites and one acetylation site),

**Figure 1.4. Structural domains of SV40 T antigen.** J domain (green – 1GH6); origin binding domain (orange – 2TBD); helicase domain (cyan – 1N25); host range domain (red). *Chimera and Microsoft Powerpoint was used to create this representation.*

which suggests that it is highly regulated and therefore must functionally important[45]. In a recent study, mutation of the acetylation site did not affect the replicative abilities of SV40. This same study showed that HR mutant viruses showed defective viral replication and lower T antigen and VP1 mRNA levels. The latter could be alleviated by co-expression with wild type HR domain.[46] Further research is needed to fully define the role of the HR domain and to identify any protein interactions it has.

Connected to the HR domain are the helical (346-414 and 549-627), AAA+ (415-548) and Zn-D (266-345) domains. These domains are all part of the viral replicative helicase apparatus (251-627). The structure of the helicase was determined by X-ray crystallography in the laboratory of Xiaojiang Chen using the T antigen construct 251-627[47]. The components of the helicase domain are responsible for the formation of a double hexamer at the origin of replication (Figure 1.5)[48,49], as well as melting and the unwinding of double stranded DNA. The zinc domain has a globular fold that is stabilized by a zinc atom through a non-classical zinc motif. This domain is believed to be responsible for stabilizing the hexameric structure. The AAA+ domain participates in the ATP binding and hydrolysis. Research has

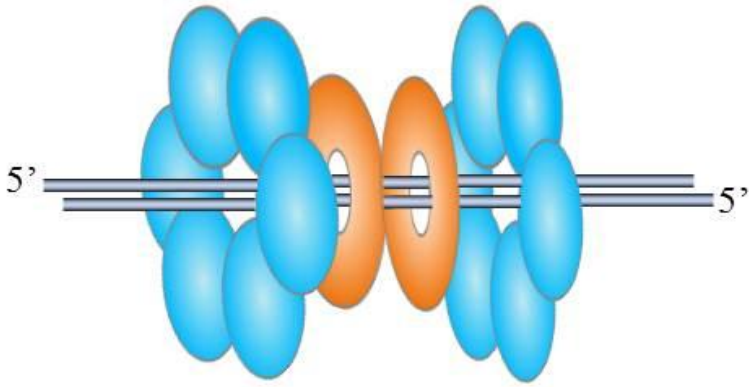



**Figure 1.5. Schematic diagram of the double hexamer at the SV40 origin of replication.** The helicase domain (cyan spheres) forms a hexameric complex. The origin binding domain (orange) assembles on the DNA (gray lines) facing one another T antigen hexamer. Three dimensional structures of the Tag double hexamer assembled at the origin prior to DNA replication were obtained by electron microscopy and negative staining and again by cryoelectron microscopy (Valle et al. 2000; Gomez-Lorenzo et al. 2003).

shown that the presence ATP/ADP promotes the formation of the T antigen hexamer. During replication, T antigen molecules assemble on the viral DNA origin by a sequence specific interaction mediated by the OBD. This results in the formation of a double hexamer via contacts between its functional domains (Figure 1.5). This double hexamer spans approximately seven turns of the double stranded DNA. The helicase domain then uses a series of expansion and constriction motions to melt the origin. The double hexamer then continues to unwind the DNA. As the double stranded DNA enters the double hexamer, the unwound single-stranded DNA exits and is protected by single stranded binding proteins like replication protein A (RPA). This mechanism is believed to be powered by the ATP hydrolysis activity of the AAA+ domain. The helicase domain also interacts with other cellular proteins such as p53 and the polymerase α/primase complex.[42,47,48]

The OBD (131-250) was the first T antigen domain to be structurally characterized[50]. The structure of OBD was determined by solution NMR using the T antigen protein construct containing residues 131-260 (pdb id: 2tbd)[50]. OBD is a compact structure composed of helices and beta sheets that recognizes and correctly positions T antigen on the DNA origin of replication. In the electron microscopy image in Figure 4, the OBDs are the smaller density at the center and the helicase domain is the larger density at the ends. This and other research has provided evidence that OBD and some additional residues may mediate formation of the double hexamer[48,51]. OBD is also involved in the regulation of transcription and replication via the interaction of two of its loops with the viral origin, single or double stranded DNA and with RPA.

OBD also functions in T antigen mediated cell cycle progression by promoting the G1/S transition through interaction with host cell proteins[52].

The J domain (1-102) is found at the N terminus. The X-ray crystal structure of the T antigen J domain (Figure 1.6) was determined using the protein construct contain residues 7-117 (pdb id: 1gh6)[53]. This structure was determined in complex with the retinoblastoma protein pocket domain (378-577/645-772). A J domain is a conserved motif throughout a class of proteins called molecular chaperones. In T antigen, the J domain is attached to the other domains via a thirty residue unstructured region. A cryo-EM study of intact T antigen suggested that the J domain is flexibly tethered to the rest of T antigen because density for only the helicase and OBD domains could be observed[54]. A J domain structure has specific features: four compact, anti-parallel coiled-coil helices; an absolutely conserved histidine, proline, aspartic acid tripeptide (HPD motif) between helix II and III; helix II region that contains basic residues; hydrophobic residues in helices II and III that help stabilize the helix packing[55].

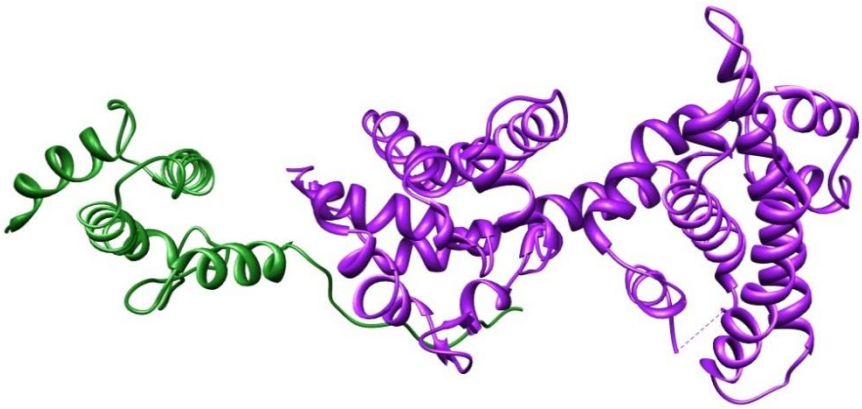


**Figure 1.6. X-ray crystal structures of SV40 T antigen J domain and Retinoblastoma tumor suppressor protein pocket domains**. The J domain of T antighen (green) was crystalized in complex with the Aand B domains of pRb (purple). PDB id: 1GH6.

The J domains are also functionally conserved. Various studies have shown that the J domain on one molecular chaperone can be swapped for that of another and still function[56]. The T antigen J domain was swapped with the J domains from E-coli DNAJ

and yeast Ydj1 and yielded functioning chimeric proteins[57]. Further evidence shows that the J domain of human molecular chaperones Hsj1 and DNAJ2 can be substituted for the J domain of T antigen and still function in DNA replication[58]. The J domain of molecular chaperones must be able to bind a co-chaperone of the DNAK protein family through the conserved HPD motif. T antigen binds to heat shock cognate protein 70 (Hsc70), a mammalian DNAK protein. As with other molecular chaperones, mutation of the T antigen HPD motif abolished the interaction with Hsc70 and its ability to function. A J domain molecular chaperone must stimulate the ATPase activity of their DNAK partner and promote the ATP dependent release of a substrate. A fragment of T antigen (1-136) was able to perform both of these activities[59]. The structural, functional and biochemical characteristics show that the J domain of T antigen is functional and that T antigen can operate as a molecular chaperone. The J domain of T antigen has been shown to be required for multiple cellular processes including DNA replication, cellular transformation, and tumorigenesis[59]. This suggests that the chaperone activity of T antigen may be a key mechanism by which these processes occur.

In addition to the structural and functional domains of T antigen, there is a binding motif that is crucial for the ability of T antigen to transform a cell. This motif is called the LXCXE. This motif is proximal to the J domain and extends from residue 103 to 107 and is required for stable association with many proteins in the pRb protein family. T antigen is able to bind and inactivate pRb proteins. The J domain is required for this activity and it appears that the chaperone function is critical for pRb inactivation.

***SV40 T antigen and molecular chaperones***

Molecular chaperones are expressed in times of cellular stress, like viral infection. The heat shock classes of molecular chaperones were the first to be discovered. Once expressed, their most common function is to assist with proper protein folding. However, their functions have expanded to include preventing protein aggregation, stabilizing unfolded proteins, clathrin uncoating, unfolding proteins for membrane translocation or degradation and others. For example, the chaperonins (Hsp60) are mitochondrial proteins that tend to be responsible for the transportation and refolding of proteins from the cytoplasm into the mitochondrial matrix. The Hsp90 chaperones are the most abundant type in the cytosol and are involved in intracellular transport, maintenance, degradation, cell signaling and tumor repression[60]. The focus of this work is on a member of the Hsp40/Hsc70 chaperone family and its interactions.

The Hsc70 chaperones have a conserved structure which includes an ATPase domain. Hsc70 can hydrolyze ATP resulting in a conformational change and some effect on a bound substrate. This mechanism allows Hsc70 to prevent protein aggregation, refold denatured proteins, transport proteins through membranes, and disrupt multi-protein complexes. However, the intrinsic ATPase activity of Hsc70 is weak. Boosting the ATPase activity is the responsibility of the co-chaperones/Hsp40/J proteins.

Proteins that contain J domains are highly conserved in nature. All members of the 40 kDa heat shock proteins (Hsp40) family of molecular chaperones contain a J domain. There are three types of J proteins. Type I proteins contain a J domain, a glycine-phenyalanine rich (G/F) region and a zinc finger-like domain. Type II proteins

have a J domain and the G/F region. The C terminal regions of the Type I and Type II Hs40 proteins have some weak homology. Type III proteins only have a J domain.[56,61,62] Figure 1.7 shows a few examples of each type of Hsp40 chaperone[62]. There are many J

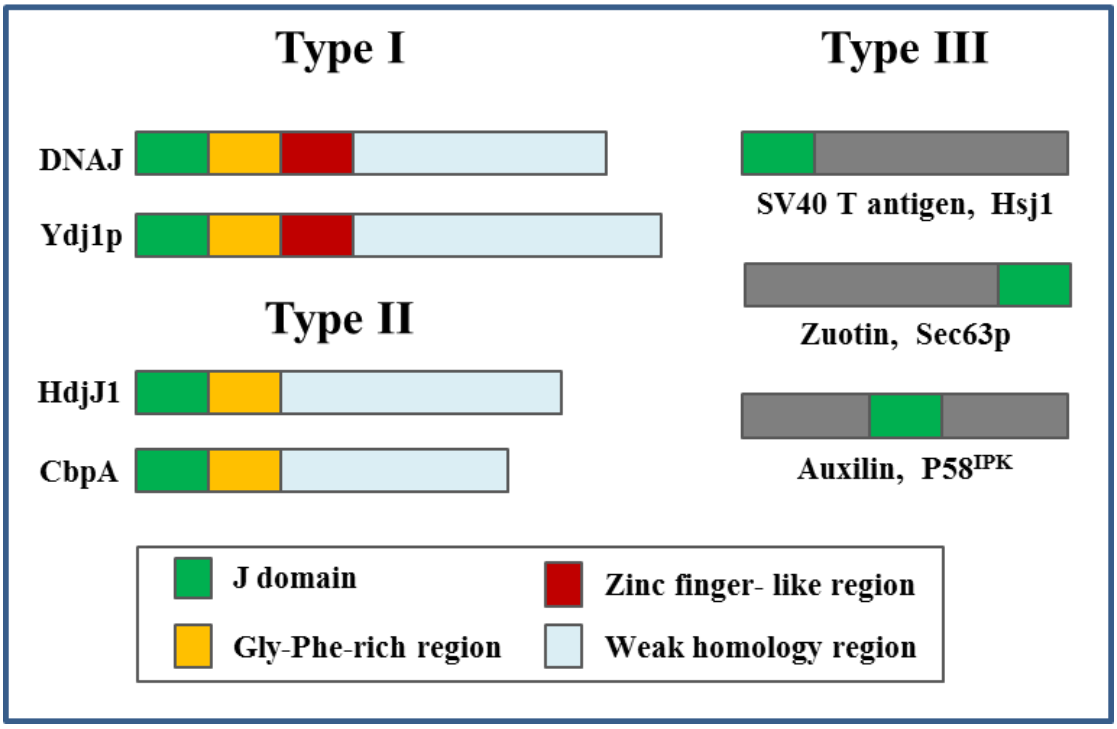



**Figure 1.7. Classification of Hsp40/J proteins.** Type I proteins are the canonical Hsp40s which contain a J domain (green), Gly-Phe (G/F) rich region (gold), and a zinc finger-like region (red). Examples of Type I proteins are DNAJ and Ydj1p, the E-coli and yeast Hsp40 homologs, respectively. Type II proteins contain only the J domain and the G/F region. Examples of Type II are HdjJ1, the human Hsp40 homolog and CbpaA, an E-coli analogue of DNAJ that recognizes curved DNA. Type III, contain only a J domain at varying places in the structure. Examples of Type II include, SV40 T antigen, Hsj1 (neuronal function), zuotin (Z-DNA binding), Sec63p (protein translocation), auxilin (clathrin uncoating), $P58^{IPK}$ (PKR-kinase inhibitor). *Adapted from Kelly 1998.*

proteins with many different functions. Even though their functions vary, they all have the ability to recruit and regulate the activity of the Hsc70[55]. Aside from stimulating ATPase activity of Hsc70, J proteins can also promote substrate interaction, alter the phosphorylation state, and induce degradation of protein substrates.

T antigen is classified as a Type III, Hsp40 molecular chaperone. Research has shown that Hsc70 will bind to T antigen via the universally conserved HPD tripeptide

in the J domain. The ATPase activity of Hsc70 is stimulated by specifically increasing the rate of ATP hydrolysis[56]. It has been noted that maximal ATPase activity occurs when Hsp40 and a substrate are present[63]. In the absence of a protein substrate, the Hsc70/Hsp40 interaction is dynamic and of low affinity. It has been proposed that the Hsp40 is pre-bound with a substrate that it then presents to Hsc70. The energy of the ATPase cycle allows Hsc70 to fold, unfold, disassemble protein complex or perform some other function on the substrate. After the work of Hsc70 is complete, Hsp40 disassociates.[56]

This interaction of T antigen and Hsc70 is essential for cellular transformation. During this process, T antigen binds and modifies the functions of various proteins. The interaction of T antigen with pRb family of tumor suppressor proteins is also essential for cellular transformation. However, instead of assisting pRb in a beneficial manner, T antigen inactivates these proteins. It is probable that T antigen utilizes its ATP hydrolysis-driven chaperone activity to inactivate pRb protein. Inactivation of this protein supports the self-propagation and cellular domination of the SV40 virus.

***Retinoblastoma Protein Family***

The retinoblastoma proteins (Rb) are tumor suppressors that belong to the so-called "pocket protein" family. The members of this family (pRb, p107 and p130) contain a conserved binding region called a pocket through which they bind cellular and viral proteins. The structure contains a small pocket and large pocket (Figure 1.8A). The foundation of the small pocket domain is the two cyclin-folds, A and B, which are separated by a spacer region. This spacer region is different in each family member.

The large pocket contains the small pocket plus the C terminal region.[64] The term pocket implies that the A and B domains form a conserved receptacle or cavity-like structure in which the binding of all proteins would occur. However, various studies have shown that Rb binding proteins associate with one or both of the A and B domains along with other regions of Rb[64]. Therefore, the term pocket is not an accurate description of its structure or mechanism of binding but simply the common term used in the field for the Rb proteins.

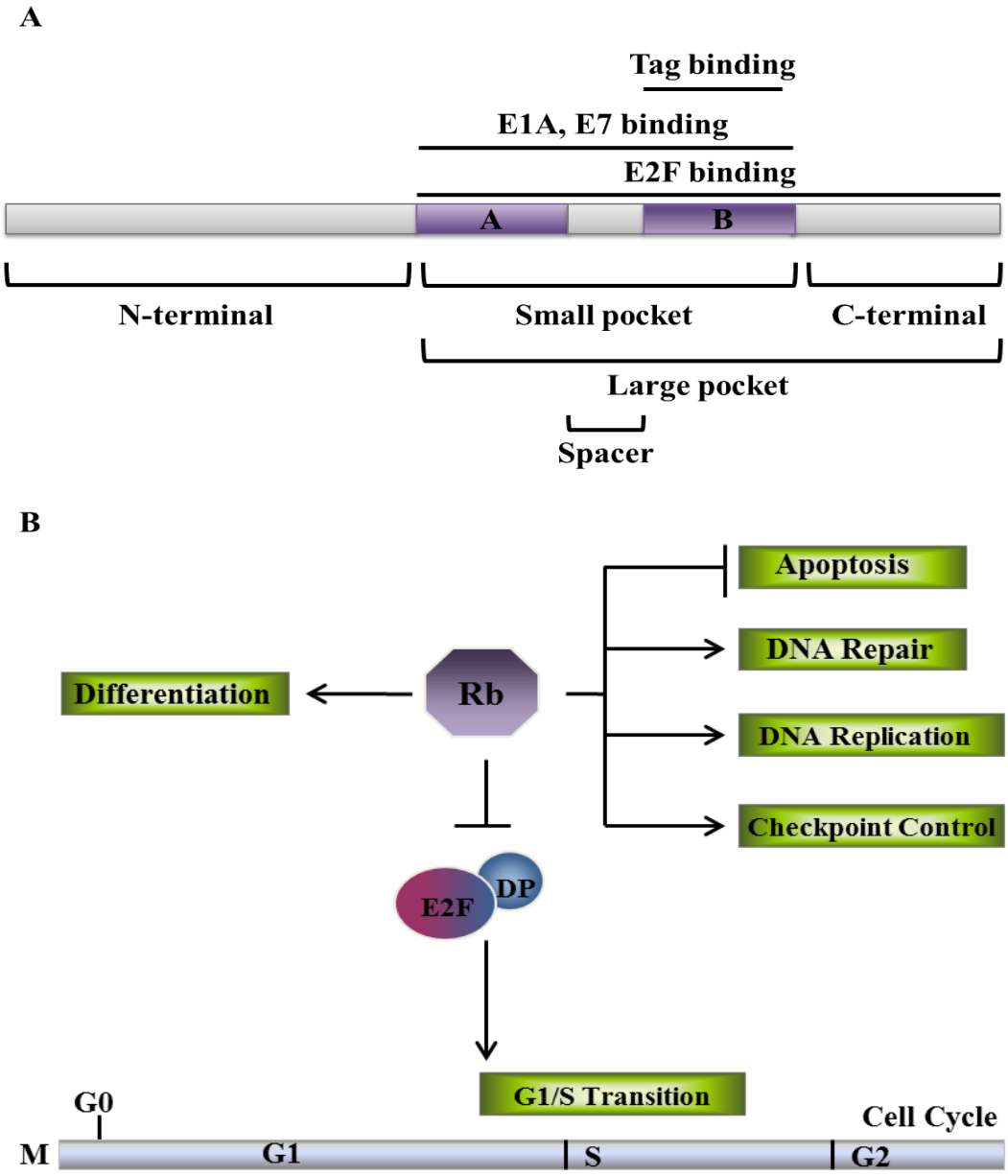



**Figure 1.8. Rb domain structure, binding and function in multiple cellular processes. A.** The domain organization of Rb highlighting the small and large pocket domains as well as the spacer. The Adenovirus oncoprotein E1A, Human Papillomavirus oncoprotein E7, and Simian Virus 40 Tag binds to Rb in the region of the A and B domains resulting in inactivation of the protein. The E2F transcription factor binds within the larger pocket domain. *Adapted from Dick 2007 and Helt et al. 2003.* **B.** Rb proteins are involved in many cellular processes along various points in the cell cycle. *Adapted from Classon et al 2002*

Rb proteins are key regulators at various stages of cell development. Together, these proteins have been implicated in many cellular processes, such as regulation of

the cell cycle, DNA-damage responses, DNA repair, DNA replication, protection against apoptosis, and differentiation, all of which could contribute to its function as a tumor suppressor (Fig 1.8)[65,66]. The pRb protein is the most frequently mutated or inactivated in human cancer.[61] Due to its many roles, it is highly probable that one of the proteins on the pRb path is disrupted in all cancers.

The gene for pRb was the first tumor suppressor gene to be cloned. Since that time great effort has been placed on determining its structure and biochemical function. The pRb protein, which has 928 amino acids, consists of three parts: a structured N terminal region, the above noted pocket domain and a predominately disordered C terminus.[67,68] The pocket domain is the minimum requirement for pRb binding to proteins containing an LXCXE motif. Two crystal structures of pRb pocket domain have been determined. One structure is in complex with an LXCXE peptide from the HPV E7 protein[69] and the other bound to an LXCXE and J domain containing SV40 T antigen fragment[68] (Figure 1.6). Both structures show that pRb is mostly helical. These structures also revealed a well-conserved region on pRb that is the binding cleft for the LXCXE motif.

The primary function of pRb is as a repressor of transcription-associated proteins required for the G1/S phase transition. This repression results in cell cycle growth arrest. The pRb protein interacts with members of the E2F class of transcription factors. There are nine members of the mammalian E2F family: E2F1-2, 3a-b, and 4-8. This family is divided into two functional groups: transcriptional activators (E2F1-3a) and repressors (E2F3b-8).[70,71] The activator E2Fs stimulate cell cycle progression and the repressor E2Fs are thought to be involved in cell cycle exit and cell

differentiation[70]. Both of these groups contain similar domains including a DNA binding domain and the DP1-2 heterodimerization domain (except E2F7-8). This heterodimerization domain allows the E2F proteins to form a complex with their dimerization partner 1 or 2 (DP1, DP2). Binding with DP1 or DP2 provides an additional DNA binding site, increasing E2F binding stability and functional ability as a transcriptional regulator of genes involved in S-phase entry, DNA replication, checkpoint control, repair, mitosis, differentiation and apoptosis[70,72].

Most E2F members (E2F1-5) contain an Rb binding domain as well. The transcriptional activation domain of E2F1-3a and sometimes E2F4 interacts with the pRb protein[70,73]. Under normal circumstances, hypo-phosphorylated pRb protein binds E2F-DP during early G1 phase and inhibits cell cycle progression. In late G1, hyper-phosphorylation of pRb, by both Cdk4/6-cyclin D and Cdk2-cyclin E, releases and thus activates the E2F/DP complex.[67] Disruption of this highly regulated system, yields unrestrained cell growth. When a cell is infected by the SV40 virus, T antigen can form a stable complex with pRb and inactivate it. Inactivation of pRb results in the release of E2F-DP to synthesize DNA, and causes aneuploidy and the promotion of tumor development[74]. The T antigen mediated inactivation of pRb is necessary for cellular transformation. As previously mentioned, the J domain and the T antigen co-chaperone are also necessary for cell transformation. The interaction among the proteins required suggests that T antigen mediated transformation is driven by its chaperone activity.

### *The Chaperone Model*

T antigen binds pRb proteins via a conserved LXCXE motif located on its linker between the J domain and OBD. Upon binding, T antigen can facilitate the release of E2F-DP, which results in aberrant cellular proliferation. This mechanism is dependent upon Hsp70, ATP hydrolysis, and the J domain[61].

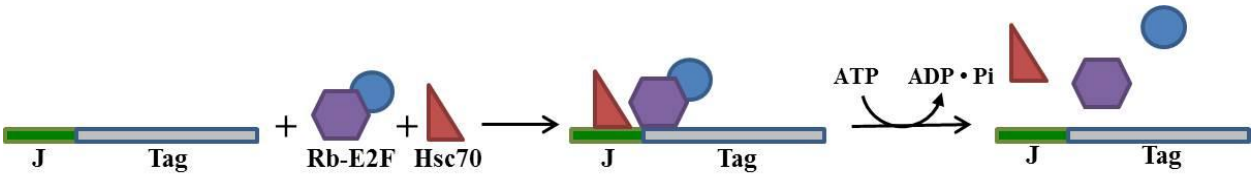



**Figure 1.9. Chaperone model for Rb inactivation.** SV40 Tag binds to the Rb-E2F complex, recruits Hsc70, and using the energy of ATP hydrolysis, releases E2F from interaction with the Rb protein.

A mechanistic model has been proposed whereby T antigen binds the Rb/E2F-DP1 complex, and when the J domain recruits Hsp70 and induces ATP hydrolysis, and the energy is used to liberate E2F-DP (Fig 1.9)[61]. This model suggests that T antigen can alter unique protein complexes, in particular, T antigen either induces disruption of or directly causes a conformational change in the pRb-E2F/DP complex. This suggests that the T antigen-Hsp70 complex must be correctly positioned for action on target proteins. The dynamics governing the coordination of this multi-protein interaction are unknown. Does pRb bind in a manner that allows the J domain to move freely in solution? Does pRb binding completely restrict motion of the J domain? There is a linker attaching the J domain to the rest of T antigen. Does it confer flexibility upon the J domain allowing it to have a wide range of motion? Does the binding of pRb involve the intervening linker and the OBD resulting in a compact complex? My thesis research was designed to investigate these questions and gain insight into the dynamics of the J domain upon binding of the pRb pocket protein.

## Experimental Methods

### *Nuclear Magnetic Resonance spectroscopy*

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful technique by which high resolution information can be obtained on the structure, dynamics and interactions of one or more molecules in solution[75–80]. This technique exploits the NMR phenomenon, which arises when certain atoms are placed in a static magnetic field and exposed to a second oscillating field. The NMR-active nuclei will absorb and re-transmit the electromagnetic radiation at a specific resonance frequency. Different nuclei in a molecule will have different frequencies based upon their local chemical environment. This frequency is proportional to the strength of the magnetic field, and when measured relative to a standard frequency it is termed the chemical shift. Chemical shifts are indicators of the composition and structure of the molecule.

The basic principles of NMR allow this spectroscopic technique to be utilized across several areas of science. NMR spectroscopy is used by chemists to study small organic molecules using one and two-dimensional techniques and by biochemists to study large complex molecules like proteins using multi-dimensional and multi-nuclear techniques. NMR studies of proteins are performed on samples enriched with one or more of the following NMR-active isotopes: $^{2}H$, $^{15}N$, $^{13}C$. These samples are prepared by expressing the protein in a bacterial host with minimal media containing the desired isotopes. It is important that the protein samples are pure and contain a protein concentration of at least 50 μM for two-dimensional experiments and at least 250 μM for three-dimensional experiments.

The most frequently used NMR experiment for the structural study of proteins is the two dimensional $^{15}$N-$^{1}$H Heteronuclear Single Quantum Coherence (HSQC) spectrum. In this experiment, the sample contains a $^{15}$N-enriched protein and the $^{1}$H and $^{15}$N resonance frequencies of each backbone amide are correlated yielding a two-dimensional spectrum. The resonances in the spectrum correspond to each $^{15}$N-$^{1}$H pair, so signals are observed from all amino acids except prolines. The distribution of signals in the HSQC spectrum creates a pattern (sometimes termed fingerprint) that is unique to each protein. The number of peaks in the spectrum should correspond, approximately, to the number of amino acids in the protein (minus prolines plus side

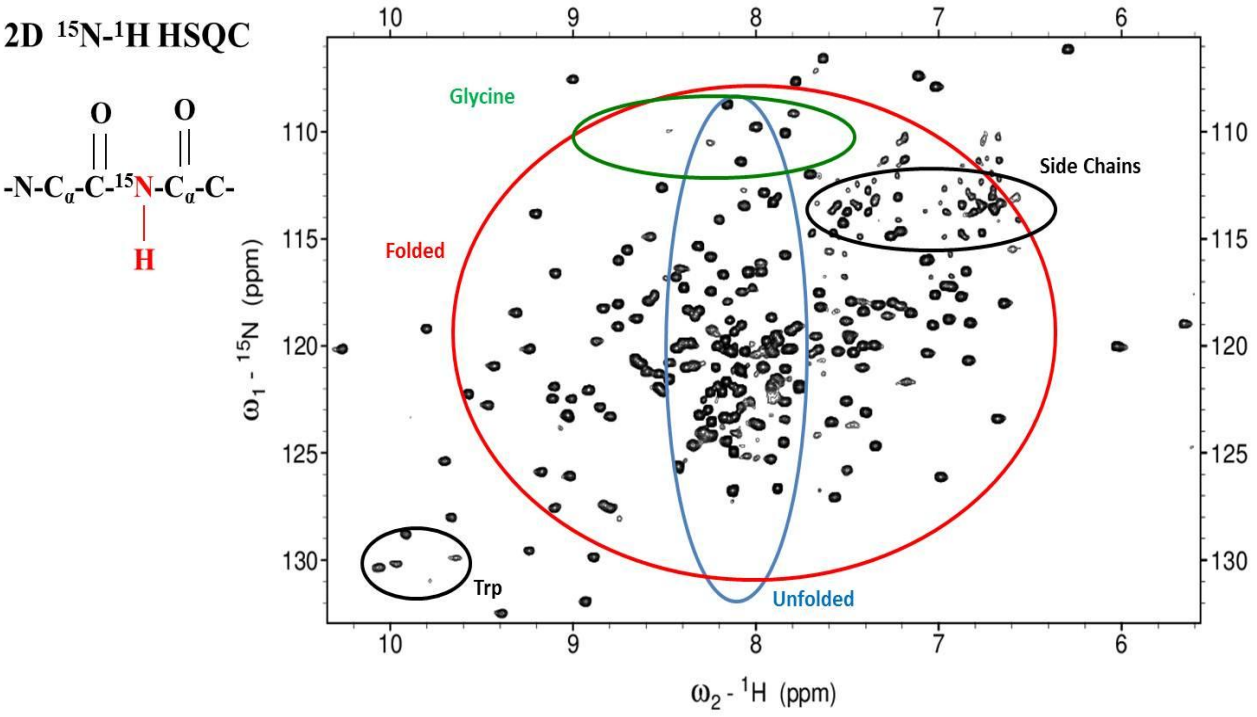



**Figure 1.10. Schematic diagram of a two dimensional HSQC.** By enriching a protein with the 15N-isotope we can perform a 2D NMR experiment that correlates a proton ($^{1}$H) with its covalently bound nitrogen ($^{15}$N). A 2D HSQC experiment of the T antigen construct $N_{260}$ is shown here. There is a NH group for every amino acid (except proline) and every NH group will produce a chemical shift or peak. Typically, NH groups from tryptophan residues appear around 10 ppm (black), asparagine and glutamine side chains around 7 ppm (black oval); and glycines around 8 ppm (green oval). An overview of the tertiary structure of the protein can be obtained: **a**n unfolded protein (blue) will have little proton dispersion whereas the peaks of a folded protein will be spread widely across the spectrum (red).

chains with nitrogen bound protons). If the protein is unfolded, the peaks will be gathered at the center of the $^{1}H$ dimension of the spectrum. If the protein is folded, the peaks are usually well dispersed across the $^{1}H$ dimension of the spectrum and distinguishable from one another. The spectral regions for folded and unfolded proteins as well as typical locations for tryptophans, glycines and side chains can be seen in Figure 1.10. This basic $^{15}N$-$^{1}H$ HSQC experiment is highly effective for studying proteins and protein complexes with molecular masses up to 40 kDa. Larger proteins require the addition of Transverse Relaxation Optimized Spectroscopy (TROSY) techniques to the basic 2D HSQC experiment, to enhance the ability to observe signals.

For these larger proteins and for protein complexes there are more amino acids thus more resonances, which results in spectral overlap. Also, the larger mass of these systems causes slower tumbling of the molecules in solution, faster relaxation rates, broader peaks and poor signal to noise ratios. The $^{15}N$-$^{1}H$ resonances of each peak sample are composed of four components that will relax at different rates. Use of the TROSY–HSQC experiment filters out the fastest relaxing component and selectively detects only the most slowly relaxing component. In the TROSY-HSQC spectrum, the resonances will be narrow and the number of observed signals will be significantly greater than for a standard HSQC spectrum of a large protein. The TROSY experiment is dependent on magnetic field strength. A magnetic field strength greater than 700 MHz is optimal[81,82]. The TROSY-HSQC spectra collected in these studies used an 800 MHz spectrometer.

The unique spectral pattern in the HSQC experiment is a powerful tracking device for monitoring changes to the local electronic environment of the protein.

Changes in the chemical shift or intensity of the peaks can be induced by upon a variation in buffer, pH, temperature, and the introduction of a binding partner. The chemical shift perturbation assay exploits this aspect of the HSQC experiment. In this assay, an unlabeled partner is titrated into the $^{15}$N-enriched protein sample. If an interaction has occurred, perturbations of the peaks (position shift, broadening, or disappearance) will be visible in the spectrum. The kinetics of the interaction determines the type of perturbation that occurs during the titration. Shifting of peak position arises from fast dissociation of the binding complex, or fast exchange. Fast exchange is often observed for weaker binding interactions. The broadening of peaks is a result of intermediate exchange. In slow exchange, peaks will disappear from the spectrum and re-appear at a different position. The primary benefit of this assay is that the protein interaction interface can be identified.

For detailed interpretation of the chemical shift perturbation assay, it is essential that the resonance assignments be determined. Resonance assignments can be obtained using a variety of strategies and these have been reviewed extensively[75]. In these studies, the backbone resonances of the J domain of SV40 Tag were assigned using two- and three-dimensional NMR experiments on $^{13}$C, $^{15}$N enriched protein. The NMR experiments collected were: 2D $^{15}$N-$^{1}$H HSQC and $^{13}$C-HSQC, as well as 3D HNCACB, CBCA(CO)NH, HN(CO)CA, HNCA, HNCO, $^{13}$C-NOESY, 3D $^{15}$N-NOESY, and $^{15}$N-TOCSY. Once the resonances have been assigned, the amino acids that are perturbed upon a binding interaction can be identified and mapped onto the structural surface of the protein. Utilized in this manner, NMR spectroscopy allows a highly efficient and

direct approach to study the structural and functional significance of protein-protein interactions.

***Small Angle X-Ray Scattering (SAXS)***

SAXS belongs to a family of scattering techniques used to characterize biological materials such as proteins, and nucleic acids. The initial principles of SAXS were developed in the 1930s by Guinier and Fournet, who showed that this technique provided information on the size, shape and internal structure of ordered and disordered systems. Technological and scientific advances have enhanced the importance and advantages of SAXS and increased its popularity. Unlike X-ray crystallography, SAXS is done is solution and does not require a crystalline sample. SAXS also does not have molecular mass limitations like that of NMR spectroscopy and can be routinely performed on protein samples ranging from ~10 kDa to 300 kDa in size[83]. However, due to the spatial averaging of the random orientation of the molecules in solution, SAXS does not provide high resolution information although it can provide structural

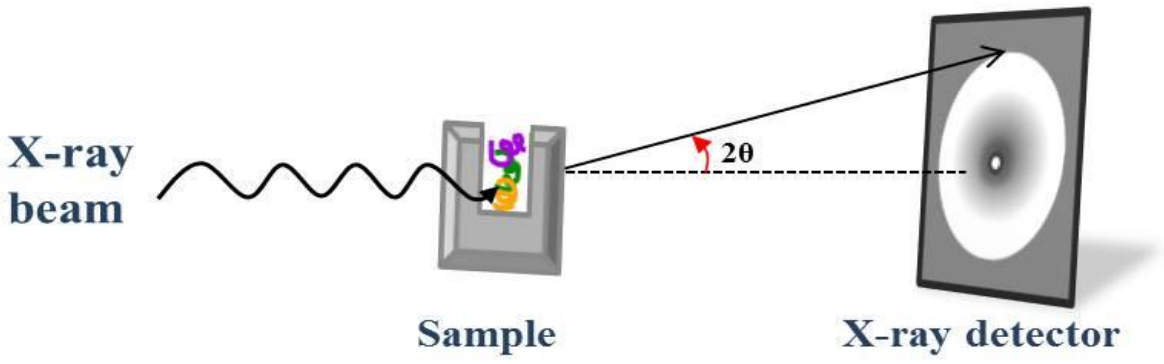



**Figure 1.11. Schematic diagram of SAXS.** A monochromatic beam of X-rays is shot at a protein sample. Some of the X-rays from the sample scatter, but most simply pass through the sample without any interaction with the protein. The scattered X-rays yield a scattering pattern that is detected by a 2-dimensional flat X-ray detector, typically. The detector is positioned perpendicular to the direction of the X-ray beam. The resultant scattering pattern contains information on the structure of the protein. *Adapted from Putnam et al. 2007*.

insight for challenging systems. SAXS is especially valuable as a complement to structural techniques like Electron Microscopy, X-ray crystallography and NMR spectroscopy. As such, SAXS has proven highly significant to the generation of descriptions for the structural assembly and large-scale dynamics of macromolecules and their complexes[84,85].

The SAXS experiment involves the elastic scattering of X-rays by a solution containing 2-10 mg/mL of protein or complex, which are recorded at very low angles (typically 0.1 - 10°). The scattered radiation is registered by a detector producing a unique two-dimensional scattering pattern (Figure 1.11). The random positions and orientations of the molecules, due to tumbling in solution, yield an intensity that is isotropic and proportional to the scattering from a single molecule averaged over all orientations[83]. From the total scattering intensity of the solution, the intensities contributed by the buffer are subtracted leaving only the intensities of the protein. The scattering intensities in the solution are dominated by the abundant solvent so appropriate concentrations are necessary to obtain good intensity from the protein. Also, it is imperative that protein samples for SAXS experiments be pure and monodisperse. Separation of

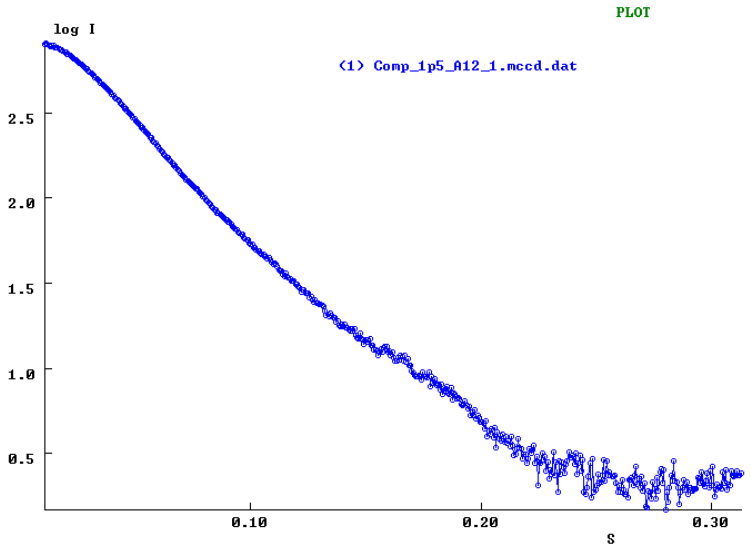



**Figure 1.12. Scattering curve from $N_{260}$ bound to pRbA/B.** This plot is of data that has been integrated and normalized. On the y-axis is intensity and the scattering angle is on the x-axis.

the scattering intensities contributed by the impurities, aggregates, or polydispersity creates a significant challenge for SAXS analysis.

To obtain the scattering profile, scattering intensities of only the protein are spherically averaged across all molecular orientations and the intensities, I($q$), are plotted as a function of the scattering angle, $q$. This one-dimensional scattering curve represents the protein in reciprocal space (Figure 1.12). Analysis of low-q region, the area closest to the beam stop (also known as the Guinier region), is important for determining whether any radiation damage or aggregation has occurred. The high-q region is further from the beam stop and can reveal information about the shape of the protein.[84]

Initial analysis of the SAXS scattering curve involves Guinier analysis, which refers to the assessment of the scattering curve at very small scattering angles, and should be performed in the region where the product $q$ x $R_g \leq 1.3$. The plotted data should fit to a straight line. If the residuals for the fit show curvature (i.e. 'smiling'), that is evidence of aggregation in the protein sample (Figure 1.13). Guinier analysis

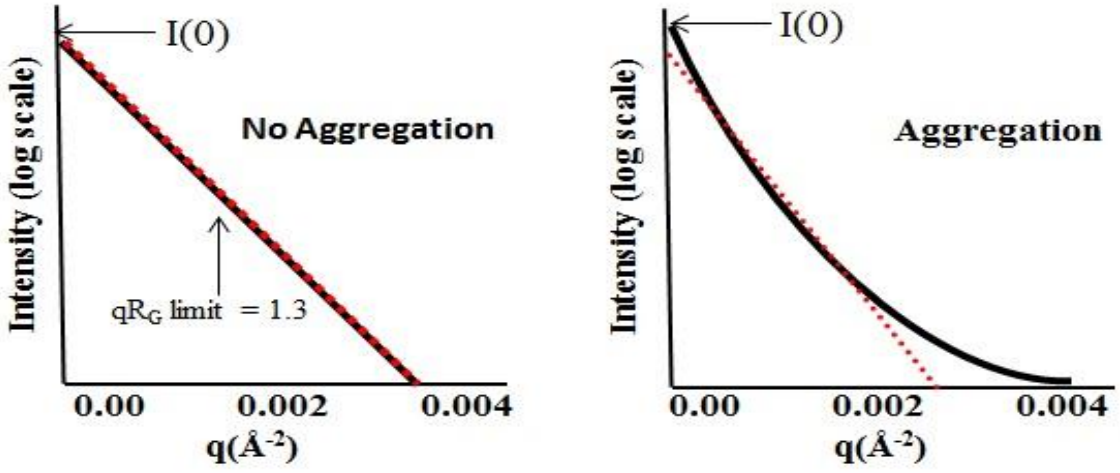



**Figure 1.13. Schematic diagram of a Guinier plot.** A non-linear log vs $q^2$ indicates the presence of protein aggregation in the sample. *Adapted from Putnam et al. 2007.*

also allows estimating the radius of gyration ($R_g$) of the protein directly from the scattering data. $R_g$ represents the mass distribution of the molecule around its center of gravity[84]. The Guinier derived $R_g$ is also known as the reciprocal space $R_g$. Comparison of the $R_g$ values for a protein alone and with a binding partner, can provide insight on whether the interaction induced a conformational change.

Qualitative assessment of the internal flexibility, structural disorder and conformational changes of proteins/complexes in solution can be obtained by applying the Kratky analysis to the scattering curve[85]. Kratky analysis is the plot of $q^2*I(q)$ as a function of $q$. For unstructured molecules, the Kratky curve sharply increases and then reaches a plateau in the high $q$ regions. For highly structured, well folded molecules, the curve is parabolic and converges toward the baseline in the high $q$ region[84]. Molecules with a combination of structured and unstructured portions will have a Kratky curve with a parabolic peak in the low $q$ that turns upward well above the baseline in the high $q$ region (Figure 1.14)[84,86].

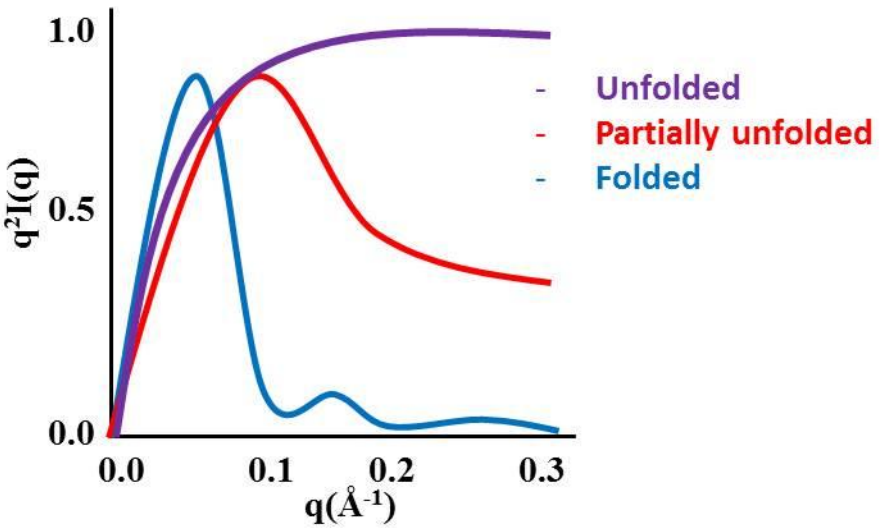



**Figure 1.14. Schematic diagram of a Kratky plot.** This plot uses the merged data to assess the folded nature of the protein. A globular protein will have a distinct peak in the low q region. A protein of random coil has no peak but a curve that will plateau. A protein with both globular and random coil characteristics will have a plot representing this. *Adapted from Putnam et al. 2007*.

Fourier transformation of the scattering curve into real space yields the pair-distance distribution function, P(r). The P(r) function represents the frequency of paired distances between all the electrons in a protein structure[84,87]. The P(r) value is plotted against real space, r (Å) and yields a bell shaped curve that is sensitive to the shape of the protein and to any changes that may occur upon an interaction. The curve for an unstructured protein would be narrower then broad for a structured compact protein. A molecule containing unstructured regions and multiple structured domains, will have a curve containing various features, like shoulders, secondary peaks and an elongated tail that would reflect the shape or domain organization of the protein (Figure 1.15)[84,88]. From the P(r) plot, the maximum dimension (D*max*) of the protein can be determined. The Dmax is the r value at which the curve for the P(r) function returns to zero. An $R_g$ can also be determined from this data transformation by integrating the P(r) function with $r^2$ over all values of r. Unlike the $R_g$ obtained from the Guinier analysis, this value utilizes all the data and is calculated in real space. The $R_g$ values from both methods should be comparable, if not that could indicate aggregation in the sample[84]. If the structure of the protein of interest is available, theoretical Rg and Dmax values can be calculated from the coordinates and used for direct comparison with the values derived from the SAXS data. Overall, the P(r)

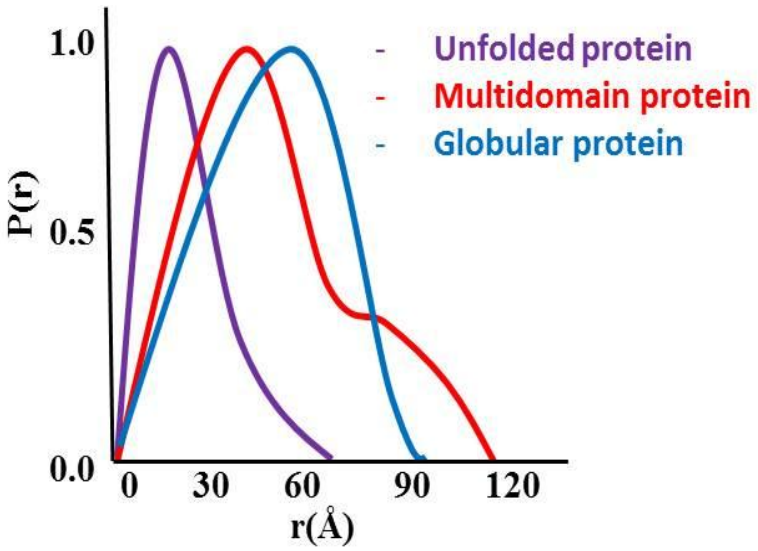



**Figure 1.15. Correlation of the P(r) function with various molecular shapes.** *Adapted from Putnam et al. 2007*

function provides important information on the structure of the molecule and may be useful for investigating conformational changes that arise by varying conditions.

The Porod–Debye law can be applied to the scattering data to more accurately assess the presence of flexibility in proteins and protein complexes[89]. The Porod-Debye region represents the scattering intensity decay by capturing the pertinent information for identifying folded and flexible proteins. The Porod-Debye law describes a fourth power approximation to the relationship between q and the observed intensities, I(q)[89,90]. The Porod-Debye parameter is the plot of $q^4 \bullet I(q)$ as a function of $q^4$. The plot should display a curve asymptotically approaching a constant value as q becomes larger. The curve for a highly compact structured protein will plateau rapidly at low q values. The curve for a fully unfolded protein would be devoid of any discernible plateau. A protein with a mixture of folded and unfolded regions will have a curve at some point in between the two mentioned above.

## Research Overview

This dissertation focuses on the relationship between SV40 T antigen and its binding partners pRb and Cullin7. Of particular interest is how the structure and dynamics of the J domain of T antigen is affected upon protein binding. These studies are aimed at understanding how the J domains participate in the functional activities such as cellular transformation of T antigen. In Chapter II we address the dynamic nature of the interaction between the J domain of T antigen and the pocket domain of pRb (pRbA/B). I have approached this question using NMR and SAXS to investigate the flexibility of the J domain in the context of the larger T antigen construct, $N_{260}$,

which contains the J and origin binding domains with a ~30 residues linker between them. Experiments were performed in the absence and presence of pRbA/B. Protein expression, purification and suitable buffer conditions were optimized for the T antigen and pRbA/B constructs. The binding characteristics of T antigen and pRb were analyzed. All NMR data was collected at the Vanderbilt Biomolecular NMR facility. All SAXS data were collected at the SYBILS beam line located at Berkley National Laboratory by Dr. Michal Hammel. These studies showed that the dynamic nature of the J domain is retained upon binding pRbA/B.

Studies described in Chapter III identify the residues in the CPH domain of Cullin 7 that are significant for the interaction with the J domain of T antigen. Cullin 7 is required for T antigen mediated cellular transformation. This work is aimed at determining whether the conserved CPH domain is the binding location for T antigen. The binding of the J domain and CPH domain were tested using NMR chemical shift perturbation assays. A weak interaction was found to occur between the two proteins. The perturbed chemical shifts were analyzed for significance and the residues identified in the interaction were mapped onto the NMR solution structure of the $Cul7_{CPH}$ domain. These data suggested a protein surface upon which T antigen could bind.

Finally, the data presented in this thesis are discussed in Chapter IV. Overall, my research advances our understanding on how SV40 T antigen interacts with proteins of a host cell and provides a foundation for further research on T antigen mediated cellular transformation.

# CHAPTER II

## BINDING TO RETINOBLASTOMA POCKET DOMAIN DOES NOT ALTER THE INTER-DOMAIN FLEXIBILITY OF THE J DOMAIN OF SV40 LARGE T ANTIGEN

### Introduction

Simian Virus 40 (SV40), a double stranded DNA polyomavirus, encodes the multi-functional multi-domain large T antigen (Tag) protein. Tag is required at several stages of viral productive infection, and is necessary and sometimes sufficient, to induce cellular transformation[39]. During viral infection, Tag is required for the initiation and elongation steps of viral DNA replication, transcriptional repression of the late viral promoter, regulating cellular transcription, and virion assembly. SV40 transforms cells by evasion, inactivation, and deregulation of the growth pathways of the host[91]. It performs these actions primarily through the action of Tag.

The amino-terminus of Tag is a J domain (Fig. 2.1A), and genetic and biochemical studies have established that Tag is a DnaJ molecular chaperone[61]. DnaJ co-chaperones work in concert with DnaK family chaperones. Tag binds the mammalian DnaK homologue Hsc70, and like other DnaJ-DnaK interactions, the binding of Tag to Hsc70 leads to activation of the Hsc70 ATPase activity. The J domain and its interaction with Hsc70 are essential for Tag mediated cellular transformation[62].†

† This chapter is published: Williams CK, Vaithiyalingam S, Hammel M, Pipas J, Chazin WJ. *Binding to retinoblastoma pocket domain does not alter the inter-domain flexibility of the J domain of SV40 large T antigen.* Arch Biochem Biophys. 518(2):111-118 (2012).

One well defined role for the J domain in transformation is the disruption of the interaction between retinoblastoma (Rb) tumor suppressor proteins and transcriptional activator E2F proteins[92]. The Rb family contains pRb, p107, and p130 proteins.[i] These

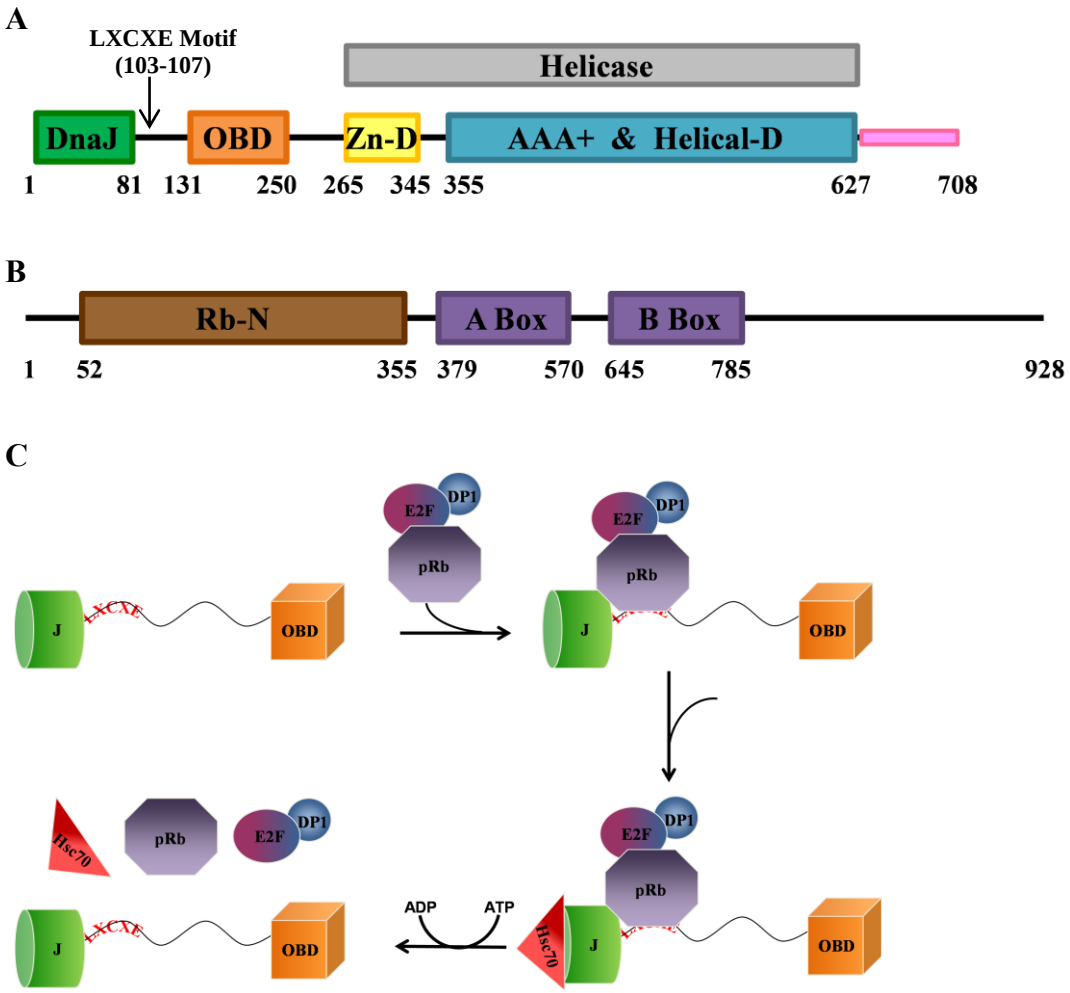



**Figure 2.1. Domain structure of SV40 Tag (A) and pRb (B), and an overview of the chaperone model (C).** Tag binds Rb/E2F complexes primarily via the LXCXE motif, recruits its co-chaperone Hsc70, ATP hydrolysis occurs and the complex is disassembled.

multi-domain proteins (Fig. 2.1B) have been implicated in transcriptional activation and several other cellular processes[93]. One function of Rb proteins is to regulate the cell cycle by inhibiting proliferation. Rb is bound to the E2F/DP1 in G1 phase and represses its activity[94]. In late G1 phase, Rb is inactivated via phosphorylation by

cyclin dependent kinases, which in turn releases E2F/DP1 and enables entry into the S phase and progression of the cell cycle.[95] Inactivation of Rb causes deregulated E2F activity, inappropriate cellular proliferation and may potentiate progression to tumorigenesis.[96]

During both productive infection and cell transformation Tag induces the release of E2F transcription factors from the three retinoblastoma proteins, pRb, p107, and p130.[64,92] The release of E2F from Rb is dependent on a functional J domain and Hsc70-mediated ATP hydrolysis.[61] These observations led to a "chaperone" model in which Tag first binds to Rb-E2F complexes, and then energy derived from ATP hydrolysis by Hsc70, recruited by the J domain, is used to liberate E2F from Rb (Fig. 2.1C)[97].

The chaperone model involves Tag recruiting Hsc70 and multi-protein complexes so that the chaperone machinery can alter specific protein-protein interactions. In this model, Tag can alter the target complex by: (1) dislodging specific proteins from the complex; (2) altering the conformation and thus the activity of proteins in the complex; (3) targeting specific proteins in the complex for post-translational modification and/or degradation. This model is consistent with studies of Tag action on pRb/E2F complexes, in which it has been shown that energy derived from Hsc70-mediated ATP hydrolysis is used to release E2F4 from its association with p130.[92] In the absence of ATP hydrolysis, Tag associates with both Hsc70 and p130/E2F4 complexes.[92] Hence, the assembly of Tag and Rb is a biologically important intermediate of the chaperone reaction.

If this model is correct, Hsc70 must have the ability to act on different cellular complexes, each containing a unique set of proteins. For example, genetic studies indicate that, in addition to targeting Rb-E2F complexes bound near the amino-terminus, the J domain acts on targets bound in the carboxy-terminal portion of Tag[98]. Thus, in each case the J domain must be in position so that Hsc70 is correctly oriented relative to the target complex. This suggests that the J domain must be able to adopt multiple orientations relative to the other Tag domains (Figure 1). This hypothesis is supported by cryo-EM studies of Tag[54], in which the absence of the J domain in the structure is presumed to arise from it having different orientations in different molecules on the grid.

Tag binds Rb proteins primarily through a consensus LXCXE binding motif (Tag 103-107) that resides in the linker between the OBD and the J domain (Fig. 1)[99]. However, the details of how Tag coordinates the protein interactions involved in inactivation of Rb remains poorly understood. A crystal structure has been determined of a complex of the N-terminal region of Tag containing the J domain and the LXCXE motif ($N_{117}$), in complex with the pocket domain of Rb (pRbA/B)[68]. The interaction interface of Tag in this structure primarily involves the LXCXE motif, but a few contacts between the J domain and pRbA/B were also observed. The chaperone model requires that the J domain be able to adopt different orientations in order to correctly position Hsc70 for action on its multiple targets. In contrast, the crystal structure suggests the J domain and pRb are in a fixed orientation. We propose that flexibility must exist in the linker between the J and OB domains of Tag to provide the variation in positioning required for function. To test this hypothesis we used a combination of

NMR and small angle X-ray scattering (SAXS) experiments to characterize the relative flexibility of the linker between J and OB domains, and to determine if this flexibility is restricted upon binding to the Rb substrate.

## Materials And Methods

### *DNA constructs*

A plasmid expressing SV40 Tag construct $N_{260}$ (4-260) obtained from Professor Xiaojiang Chen (University of Southern California) was subcloned into an in-house kanamycin resistant pBG101 vector (Dr. Laura Mizoue, Center for Structural Biology) containing an H3C cleavable N-terminal 6xHis-GST tag. The Tag constructs $N_{117}$ (4-117) and $N_{102}$ (4-102) were subcloned from $N_{260}$ into the ampicillin resistant pET15b vector containing an N-terminal 6xHis tag. A pRbA/B (379-570/645-772) construct in an ampicillin resistant pGEX-KG vector containing a thrombin cleavable GST tag was used in these studies because it corresponds to that used in the crystal structure of the pRb-Tag complex.[68]

### *Protein Expression*

The constructs of SV40 Tag ($N_{260,}$ $N_{117}$, $N_{102}$) were expressed in BL21 (DE3) Escherichia coli (*E. coli*) competent cells grown in LB or M9 minimal media containing $^{15}NH_4Cl$ and/or $^{13}C$ glucose as needed. The cells containing $N_{260}$ were grown at 37 ºC and induced with 0.25 mM isopropyl β-D-thiogalactopyranoside (IPTG) overnight at 18 ºC. Expression of $N_{117}$ or $N_{102}$ was induced with 1 mM IPTG at 37 ºC for 5 hours. The pRbA/B protein was expressed in *E. coli* XA90 cells. The cells were grown in Terrific

Broth at 37 ºC. Protein was induced with 0.4 mM IPTG and expressed overnight at 18 ºC. All cells were harvested using a JLA 8.1 Beckman rotor at 7000 rpm and 4 ºC for 20 minutes. Pellets were stored at -20 ºC.

***Protein Purification***

The Tag protein constructs were purified using Nickel affinity chromatography (Ni-NTA) in 25 mM Tris (pH 8.0) and 300 mM NaCl. Cleavage of the His tag on $N_{102}$ and $N_{117}$ was performed with thrombin at room temperature for 3 hrs while the protein remained bound to the resin. The cleaved protein was eluted by washing the resin with the purification buffer and followed by size exclusion chromatography (SEC) using a Superdex S75 column equilibrated with 20 mM sodium phosphate ($NaH_2PO_4$) (pH 7.2), 100 mM NaCl, and 5 mM dithiothreotol (DTT). Cleavage of the His-GST tag from $N_{260}$ with H3C protease was performed via overnight dialysis in buffer containing 25 mM Tris (pH 8.0), 200 mM NaCl and 2 mM DTT. A second Ni-NTA purification step was used to remove the His-GST tag and the H3C protease. An elution gradient of 5-20 mM imidazole was used to elute the cleaved $N_{260}$ protein, followed by SEC over a Superdex S200 column equilibrated with 20 mM $NaH_2PO_4$ (pH 7.2), 100 mM NaCl, and 5 mM DTT.

The pRbA/B protein was purified using glutathione affinity chromatography in 1X phosphate buffered saline (PBS), 2 mM DTT, pH 7.2 and 10 mM imidazole for the elution. The fusion protein was dialyzed into 20 mM $NaH_2PO_4$ (pH 6.0), 50 mM NaCl, and 5 mM b-mercapto ethanol (BME). The GST tag was cleaved at room temperature without agitation in the presence of 10% glycerol via addition of thrombin (100 units)

every two hours for eight hours. Then the cleavage reaction remained at room temperature overnight. The pRbA/B was further purified with ion exchange chromatography (SourceS) using 20 mM $NaH_2PO_4$ (pH 6.0), 5 mM DTT with a 50 mM - 1 M NaCl elution gradient. The final purification step was SEC using a Superdex S200 column equilibrated with 20 mM $NaH_2PO_4$ (pH 7.2), 100 mM NaCl, and 5 mM DTT.

***Nuclear Magnetic Resonance Spectroscopy***

NMR experiments were conducted using Bruker DRX 600 and 800 MHz spectrometers equipped with z-axis gradient TXI cryoprobes. All experiments were collected at 25 ºC. To obtain backbone resonance assignments, 800 μM $^{13}C$,$^{15}N$-enriched $N_{117}$ was prepared in 1X PBS (pH 6.5), 2 mM DTT and 10% $D_2O$. The experiments acquired were 2D $^{15}N$-$^{1}H$ HSQC and $^{13}C$-HSQC, and 3D HNCACB, CBCA(CO)NH, HN(CO)CA, HNCA, HNCO, $^{13}C$-NOESY, 3D $^{15}N$-NOESY, and $^{15}N$-TOCSY. Chemical shift perturbations induced by binding of pRb/AB were elucidated by comparing spectra of free $^{15}N$-enriched $N_{117}$ and $N_{260}$ with the corresponding 1:1 complex with unlabeled pRb/AB in a buffer containing 20 mM $NaH_2PO_4$ (pH 7.2), 100 mM NaCl, 5 mM DTT, and 5% $D_2O$. $^{15}N$-$^{1}H$ HSQC spectra were acquired for $N_{117}$ and $^{15}N$-$^{1}H$ TROSY HSQC for $N_{260}$. All NMR data were processed using Bruker software TopSpin[100] and analyzed using Sparky[101].

***Small angle x-ray scattering***

SAXS data were collected at the ALS beamline 12.3.1 (SIBYLS) LBNL Berkeley, California[102]. The wavelength of the incident beam λ was 1.0 Å and the

sample-to-detector distances set to 1.5 m resulting in scattering vectors, q, ranging from 0.001 $Å^{-1}$ to 0.32 $Å^{-1}$. The scattering vector is defined as $q = 4\pi \sin\theta/\lambda$, where $2\theta$ is the scattering angle. All experiments were performed at 20 °C and data were processed as described[102]. The experimental SAXS data for different protein concentrations were examined for evidence of aggregation using Guinier plots[103]. The radius of gyration $R_g$ was derived by the Guinier approximation $I(q) = I(0) \exp(-q^2R_g^2/3)$ with the limits $qR_g<1.3$. The program GNOM[83] was used to compute the pair-distance distribution functions, P(r). This approach also provided the maximum dimension of the macromolecule, $D_{max}$. In order to determine $R_g$ and $D_{max}$ for the OBD and J domain, the FoXS server[104] was used to back-calculate SAXS from coordinates for OBD (PDBid: 2tbd) and for J domain ($N_{102}$) (PDBid: 1gh6:A) and this data was then used for Guinier and GNOM analyses.

The recently reported approach based on the Porod–Debye law was used to qualitatively assess the presence of structural disorder in the proteins[89]. The Porod-Debye law describes a fourth power approximation to the relationship between q and the observed intensities, I(q)[90,105]. Transformation of scattering data as $q^4 \cdot I(q)$ vs. $q^4$ (Porod-Debye plot) should display a curve asymptotically approaching a constant value as q approaches infinity. In practice, the plot for a highly globular protein will plateau rapidly at low $q^4$ values, whereas the plot for a fully unfolded protein would be devoid of any discernible plateau[89].

### *Computational modeling*

Ab initio shape envelopes were calculated with the program GASBOR[83]. Ten GASBOR runs were merged using DAMAVER. Overlays with PDB coordinates were performed using SUPCOMB using the atomic coordinates from the crystal structures of the complex of pRbA/B and $N_{117}$ (PDBid: 1gh6) and the OBD domain (PDBid: 2tbd). Molecular graphics images were generated using PYMOL (DeLano Scientific, Palo Alto, CA, USA).

### *Molecular graphics*

All images were generated using Chimera 1.5.2 (Regents of the University of California, San Francisco, CA, USA).

## Results

SV40 Tag is a modular protein containing three globular domains (J, origin binding, helicase) and a disordered C-terminal 'host-range' domain (Fig. 1). A segment of 29 residues links the J domain to the origin binding domain (OBD). Previous studies have shown the OBD aligns with the helicase domain (HD)[54]. A crystal structure of the J domain and a portion of the linker bound to pRb was reported and identified the residues involved in this interaction[53]. Potential models for the effect of pRb binding on the structural organization of Tag domains range from interaction exclusively with the LXCXE motif to formation of a tri-partate complex with the J domain and the OBD fully engaged. Differences in these models impact the Tag mechanism of action on pRb function, in particular, how the J domain can adjust its orientation to correctly position

Hsc70 for action on its targets. Our experiments were directed toward investigating changes in the relative flexibility of the J domain upon pRb binding. Since the OBD is structurally aligned with the core of Tag[54], we have selected for our studies a construct containing the residues spanning the J domain through the OBD ($N_{260}$).

***The J and OB domains of Tag are structurally independent***

Our approach involved characterizing the structural and dynamic properties of the Tag $N_{260}$ construct first using heteronuclear NMR spectroscopy (Figure 2.2). Note that TROSY-HSQC, as opposed to the standard HSQC technique was required for the larger $N_{260}$ construct to obtain sufficient resolution of the peaks. The spectra for all

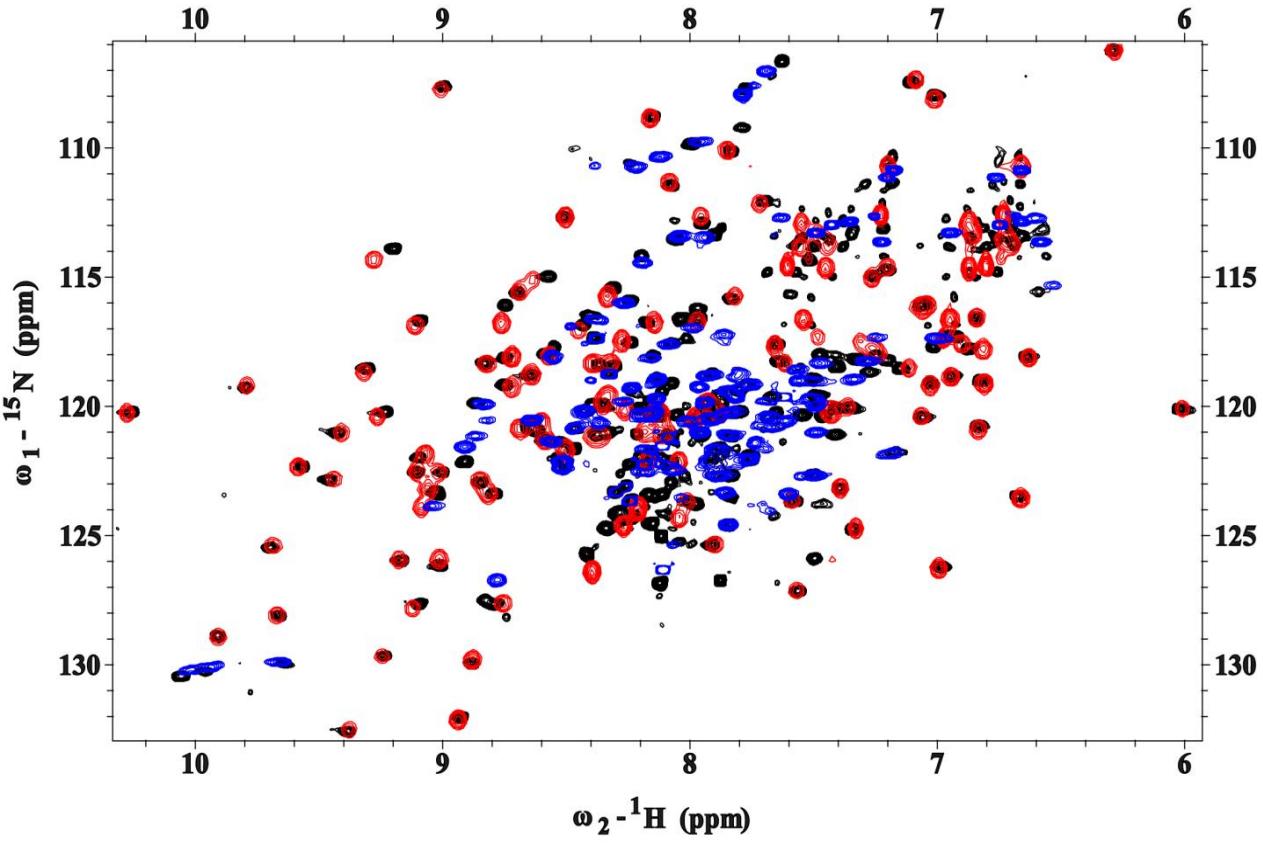



**Figure 2.2. The J and OB domains of SV40 Large T antigen are structurally independent.** Overlays of 15N-1H HSQC NMR spectra of $N_{117}$ (red) and OBD (blue) on the 15N-1H TROSY-HSQC spectrum of $N_{260}$ (black). Resonances arising from each domain correspond closely to those observed in the larger $N_{260}$ construct.

three constructs show well dispersed signals, which is indicative of the presence of folded globular domains. Inspection of these spectra reveals that $N_{260}$ has many signals that reflect the presence of the J domain and OBD. In addition, a number of signals in the center of the spectrum are observed, which were attributed to unstructured regions including the 29 residue linker between the two folded domains.

To accurately identify the signals observed in the spectrum of $N_{260}$, chemical shift assignments for the J domain and OBD are required. The corresponding NMR signal assignments have been reported for OBD[50] but not for the J domain. Consequently, $^{15}N$- and $^{13}C$, $^{15N}$-enriched samples of the J domain $N_{102}$ and $N_{117}$

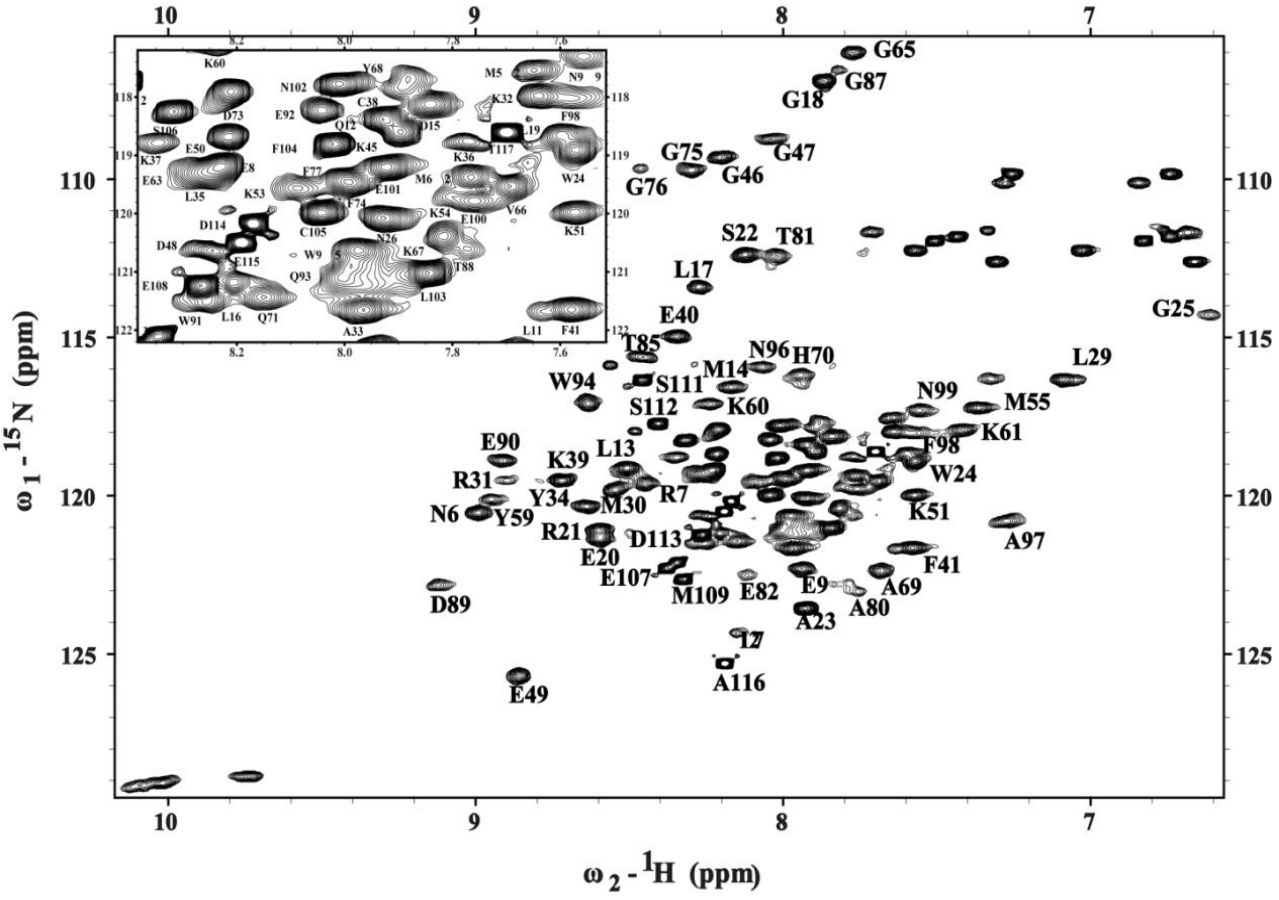



**Figure 2.3. HSQC spectrum of Tag construct, $N_{117}$.** The amino acid assignments of the $N_{117}$ resonances are labeled. The insert shows the crowded region around 8 ppm in the $^{1}H$ dimension and 120 ppm in the $^{15}N$ dimension.

constructs were prepared and a series of standard multi-dimensional heteronuclear NMR experiments were recorded (See Materials and Methods). The assignments for the backbone amide resonances of $N_{117}$ are provided in Figure 2.3. Note that these include the resonances of residues in the linker, including the key LXCXE motif spanning residues 103-107.

Overlays of the individual J domain and OBD spectra onto the $N_{260}$ spectrum are shown in Figure 2.2. It is clear from these comparisons that many distinct resonances corresponding to the J domain and OBD appear in identical locations in the spectrum of $N_{260}$. The observation that the J domain and OBD have the same NMR chemical shifts, whether alone or tethered together in the context of the $N_{260}$ construct, indicates that the J domain and OBD are structurally independent of each other and implies that the linker between them is flexible.

***The J domain is flexibly tethered to the OBD***

To test if the tether between the two domains of $N_{260}$ is flexible, we turned to small angle X-ray scattering (SAXS) experiments. SAXS is well suited for characterizing the structural architecture of multi-domain proteins in solution and determining whether or not the domains are arranged in a fixed orientation[102]. Analysis of the SAXS data for $N_{260}$ shows it has a maximal protein dimension ($D_{max}$) and a radius of gyration ($R_g$) (from a Guinier plot, Figure 2.4) of 124 Å and 30.6 Å, respectively. These values are significantly larger than the sum of $R_g$ (14.75 + 14.58 = 29.33 Å) and $D_{max}$ (44.0 + 42.5 = 86.5 Å) values for the two globular domains, calculated from the x-ray crystal structure of the J domain[53] and the representative NMR structure of the

OBD[50]. More significantly, the broadened P(r) function and elongation of the P(r) tail (Fig. 2.5B) are characteristic of population of highly extended conformations where the J domain and OBD are far from each other. In addition, the upward slope in the Kratky plot at high q values (Fig. 2.5A) indicates the presence of disordered polypeptide in the protein. In summary, the ensemble of the SAXS data indicated that the J domain of $N_{260}$ is flexibly attached to the OBD.

### *pRbA/B interacts with the Tag LXCXE motif and the J domain but not the OBD*

Binding of pRb to consensus LXCXE motifs, including that of SV40 Tag, is well established[106–108]. The crystal structure of the complex of the Tag J domain extended to the

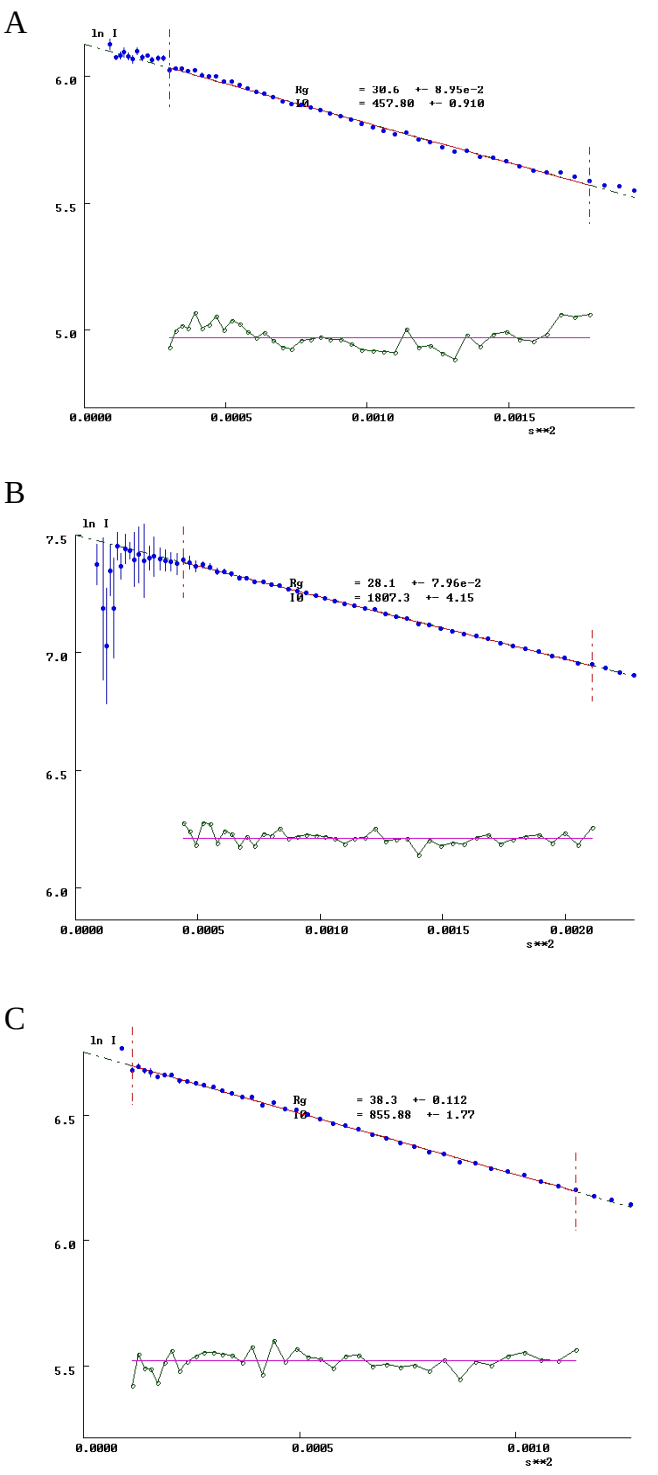



**Figure 2.4. Guinier analysis of SAXS data for N260, pRbA/B and the $N_{260}$-pRb/AB complex.** Guinier plots of (A) $N_{260}$, (B) pRbA/B, and (C) $N_{260}$-pRbA/B complex.

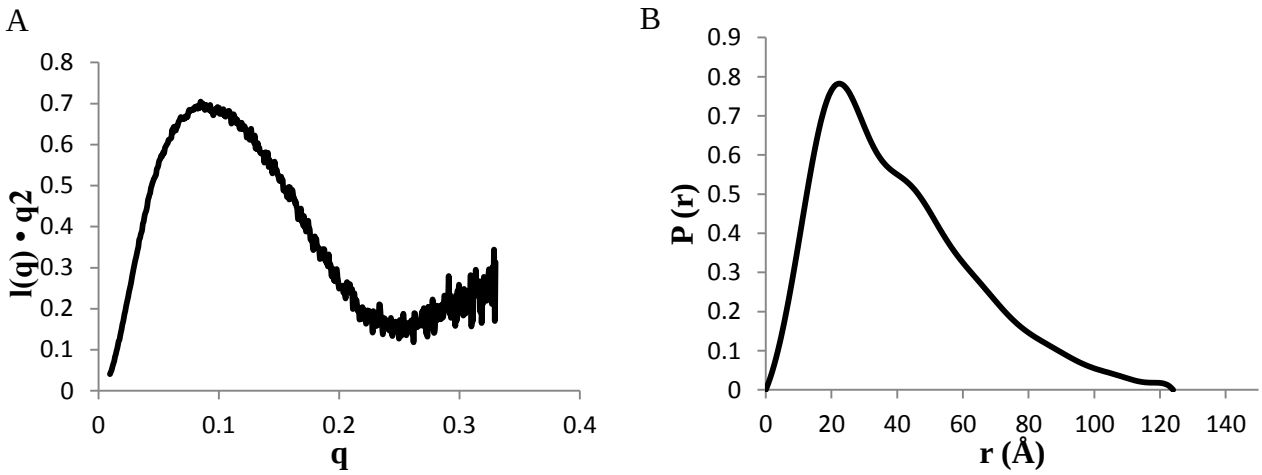



**Figure 2.5. SAXS analysis for SV40 Tag $N_{260}$.** (A) Kratky plot and (B) P(r) function.

LXCXE motif ($N_{117}$) with pRbA/B revealed a few additional contacts to the J domain[53], although the importance of these contacts to the interaction of Tag and pRb has not been directly investigated. Of importance here, if the contacts with the J domain are significant, the binding of pRb may fix the orientation of pRb and J, which has implications for the chaperone model. Following the crystal structure and previous studies of Rb binding to LXCXE motif proteins[108], we utilized the pRbA/B construct that contains the core A and B structured domains with a truncated linker.

An NMR approach was used first to characterize contacts between Tag and pRb. Figure 2.6A shows an overlay of the $^{15}N$-$^{1}H$ HSQC spectrum of $^{15}N$-enriched $N_{117}$ alone (black) and with a 1:1 ratio of pRbA/B (red). The comparison of these data revealed that binding of pRbA/B causes a combination of resonance disappearance, line broadening, perturbation of chemical shifts, and some unchanged peaks. These observations are indicative of a direct interaction with pRbA/B that extends beyond the LXCXE motif to residues in the J domain. The perturbation of some but not all signals in the J domain is consistent with a specific interaction. The pronounced broadening and disappearance of J domain peaks in the presence of Rb might be interpreted as due

to the close proximity of the J domain to pRbA/B. However, given the large size of the $N_{117}$-pRbA/B complex (52 kDa) and the fact that a substantial number of J domain signals are present in a basic $^{15}N$-$^{1}H$ HSQC spectrum, the most likely explanation is that the J domain does not tumble in synchrony with pRbA/B, which implies that the interaction with the J domain is transient.

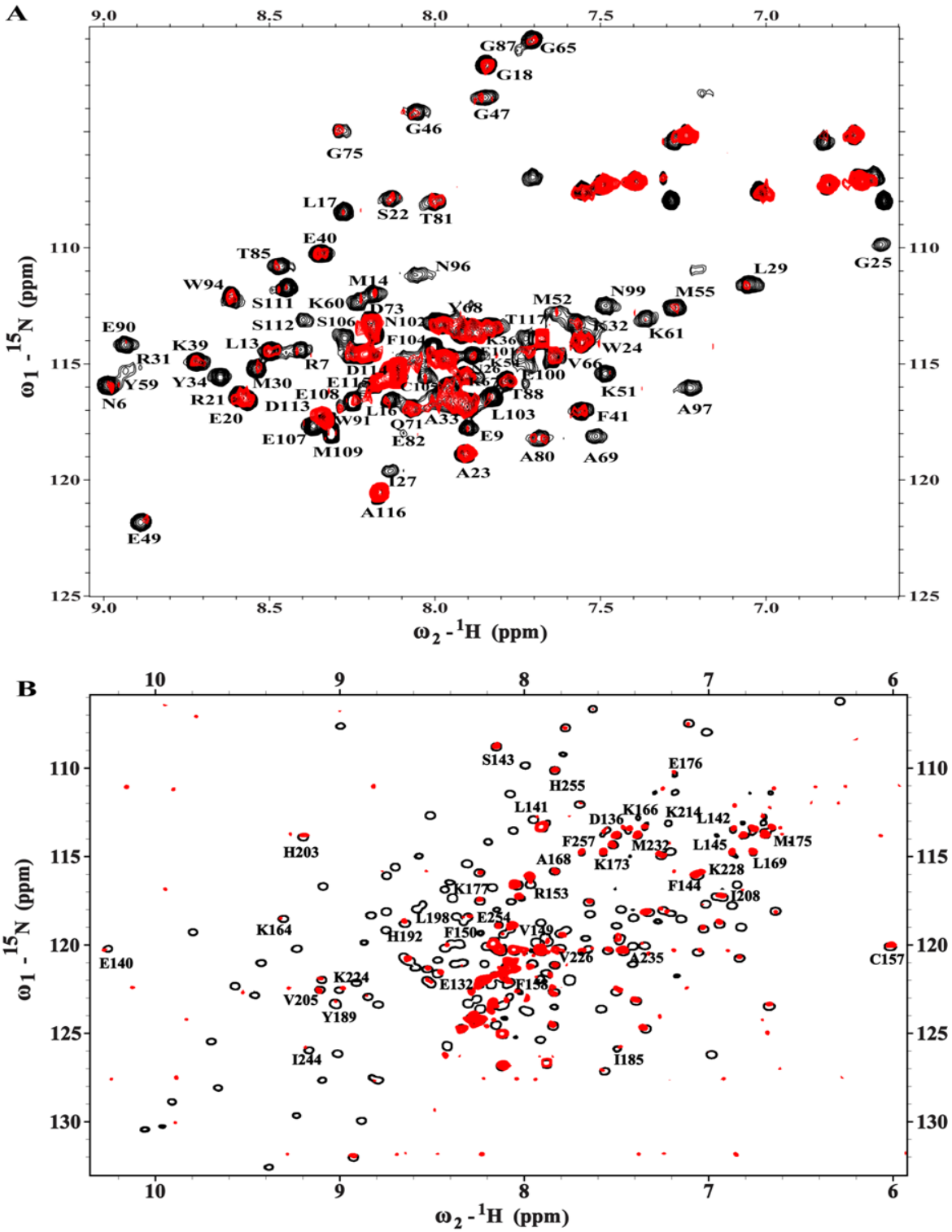



**Figure 2.6. Interaction of pRbA/B with Tag.** (A) Region from the 600 MHz 15N-1H HSQC spectra for 15N-enriched N117 acquired in the absence (black) and presence (red) of a 1.0 molar equivalent of pRbA/B. Sequence-specific assignments are provided for the resolved peaks in the spectrum. (B) 800-MHz 15N-1H TROSY-HSQC spectra for 15N-enriched $N_{260}$ acquired in the absence (black) and presence (red) of pRbA/B. Sequence-specific assignments are provided for resolved peaks that are present in the spectrum of the complex, all of which arise from OBD residues.

To obtain further insights into the interaction with pRbA/B, we took advantage of our NMR chemical shift assignments for the J domain and identified the residues whose resonances are perturbed by the addition of pRbA/B (Table 2.1).

The disappearance of all of the resonances from the LXCXE motif is consistent with it serving as an essential component of the binding interface and with the high affinity ($K_d$ 0.44 mM) binding of the isolated motif to pRbA/B[50]. We note that the residues whose resonances are perturbed include all of the J domain residues that are reported to be in contact with pRbA/B in the crystal structure[53] as well as a number of additional residues. Figure 2.7 shows a ribbon diagram of J domain with these perturbed residues mapped onto the structure. Since residues whose signals are perturbed are scattered throughout the J domain and are not clustered near the C-terminus close to the LXCXE motif, then it is possible to rule out that the perturbations are due to simple proximity effects. Chemical shift perturbations may arise from direct inter-molecular contacts or from structural changes induced upon binding of the ligand. Hence, it is not possible to determine if the binding site in solution is identical to that observed in the crystal structure, but the available evidence suggests this is likely to be so. Regardless, as noted above, the very observation of NMR signals from the J domain in this 52 kDa complex strongly implies the interaction between pRbA/B and the J domain is more dynamic than implied by the crystal structure.

| **T antigen Residues affected by RbA/B interaction** |
|---|
| N6; R7; E9; L13; M14; L17; S22; I27; L29; M30; R31; **K32** Y34; **K39**; E40; G46; G47; E49; K51; M55; Y59; K60; K61; G65; G75; G76; T81; E82; T85; G87; E90; Y94; **N96**; A97; **N99**; **E100**; E101; **L103**; **F104**; **C105**; S106; **E107**; **E108**; **M109**; S111; **S112** |

**Table 2.1.** T antigen residues perturbed upon interaction with pRbA/B as observed by NMR (black) and X-ray crystallography (bold).

To determine whether the OBD is also involved in the interaction with pRbA/B, an NMR chemical shift perturbation assay was performed using the $N_{260}$ construct. Figure 2.6B shows an overlay of the $^{15}N$-$^{1}H$ TROSY-HSQC spectrum of $N_{260}$ alone (black) and with a 1:1 ratio of pRbA/B (red), along with labels that identify the origin of the resolved resonances. As described above for the experiment with $N_{117}$, the extensive degree of peak disappearance and increases in line widths are clear evidence of formation of a large complex (69 kDa). Of critical importance, all of the assigned signals arise from the OBD but none from the J domain.

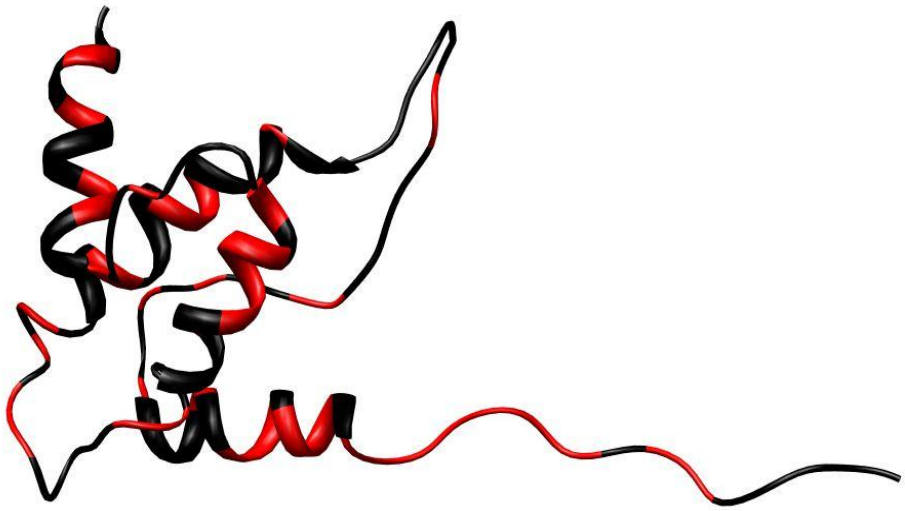


**Figure 2.7. Map of residues on the J domain whose resonances are altered upon binding of pRbA/B.** Ribbon diagram of the J domain (PDBid: 1GH6) colored red for residues whose resonances disappear or are broadened in the complex (Table 1).

This observation does not rule out that the more pronounced broadening and disappearance of J domain versus OBD signals in the presence of Rb is simply due to it being closer to the Rb than the OB domain. However, as for the $N_{260}$-pRbA/B complex, the fact that we see signals in a basic $^{15}N$-$^{1}H$ TROSY-HSQC spectrum of this 69 kDa complex indicates that the OBD does not tumble in synchrony with pRbA/B. The most likely interpretation of these observations is that the OBD does not interact with pRbA/B, but given the limitations of the NMR analysis, it is not possible to rule out

alternate explanations such as a highly transient weak interaction with a small OBD interface.

***The Tag J domain remains flexibly attached to the OBD after pRbA/B is bound.***

To test the effect of pRbA/B binding on the flexibility of $N_{260}$, we performed additional SAXS experiments on the purified pRbA/B and the complex of pRbA/B with

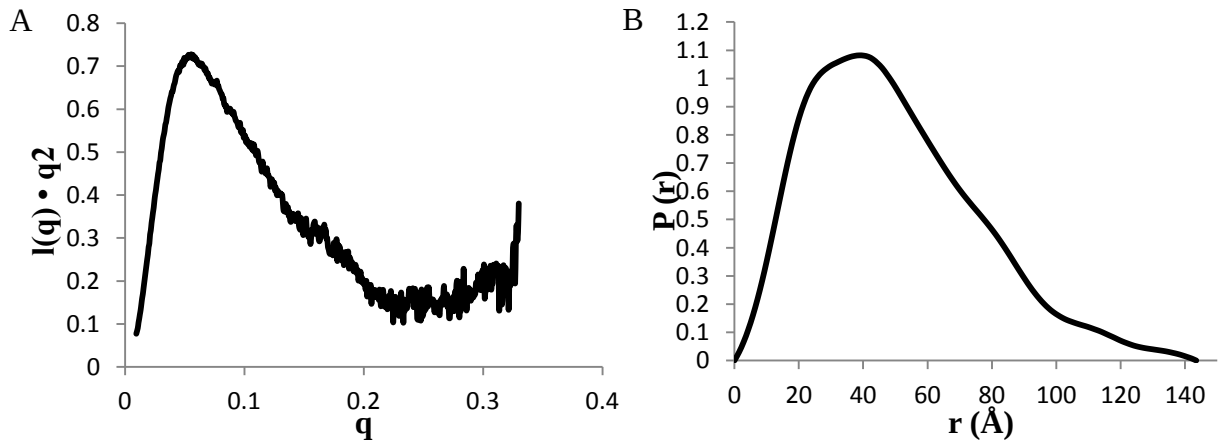



**Figure 2.8. SAXS data for the complex of pRbA/B and $N_{260}$.** (A) Kratky plot and (B) P(r) function.

$N_{260}$ (Figure 2.8). The data for pRbA/B matched well to that previously published[109]. The calculated P(r) function of the complex had a $D_{max}$ of ~145 Å (Table 2.2) about 20 Å greater than for $N_{260}$ alone. In addition, a new feature in the curve was observed in the P(r) function at r ~80 Å, which corresponds to the average distance between the J-domain and pRbA/B domains (Fig. 2.8B).

We performed a Porod-Debye analysis for pRbA/B, $N_{260}$ and their complex to obtain direct insights into their relative flexibility [89]. In plots of the normalized $q^4 \cdot I(q)$ vs. $q^4$, we find the globular pRbA/B reaches a plateau at lower q than the flexible $N_{260}$, and the complex is in between (Fig. 2.9). Table 2.3 lists the Porod volumes

| Proteins | $N_{260}$ | pRbA/B | $N_{260}$ - pRbA/B Complex |
|---|---|---|---|
| **Rg (Å), Guinier Analysis** | 30.6 ($0.95e^{-2}$) | 28.1 ($8.0e^{-2}$) | 41.6 (0.1) |
| **Rg (Å), P(r) analysis** | 32.9 | 27.1 | 42.2 |
| **Dmax (Å), P(r) analysis** | 124 | 96.5 | 143.5 |

**Table 2.2. Comparison of SAXS-derived parameters for $N_{260}$, pRbA/B, and the $N_{260}$-pRb/AB complex.** *The error for the Rg* from the *Guinier analysis (in Angstroms) is below the value.*

| | Porod Vol. ($Å^3$) | $d_{protein}$ ($g•cm^3$) |
|---|---|---|
| **pRbA/B** | 62000 | 1.07 |
| **$N_{260}$** | 64000 | 0.78 |
| **$N_{260}$ - pRbA/B** | 126000 | 0.92 |

**Table 2.3. Porod-Debye Parameters derived from the SAXS data.** These values are for $N_{260}$, pRb/AB, and the $N_{260}$-pRb/AB complex.

determined from the three SAXS data sets, along with the experimental packing densities. The ideal value for a compact globular protein is 1.37 g-cm$^3$. The value of 0.8 g-cm$^3$ for $N_{260}$ is consistent with extensive inter-domain flexibility. Although a higher value of 0.9 g-cm$^3$ is found for the complex, this is still not close to the canonical value for a compact protein.

*Ab initio* molecular envelopes were calculated directly from the scattering data using GASBOR for the free proteins and the complex to which the corresponding structures were manually fitted (Fig. 2.10). A well-defined conformational envelope that fits well to the coordinates from the x-ray crystal structure was obtained for the

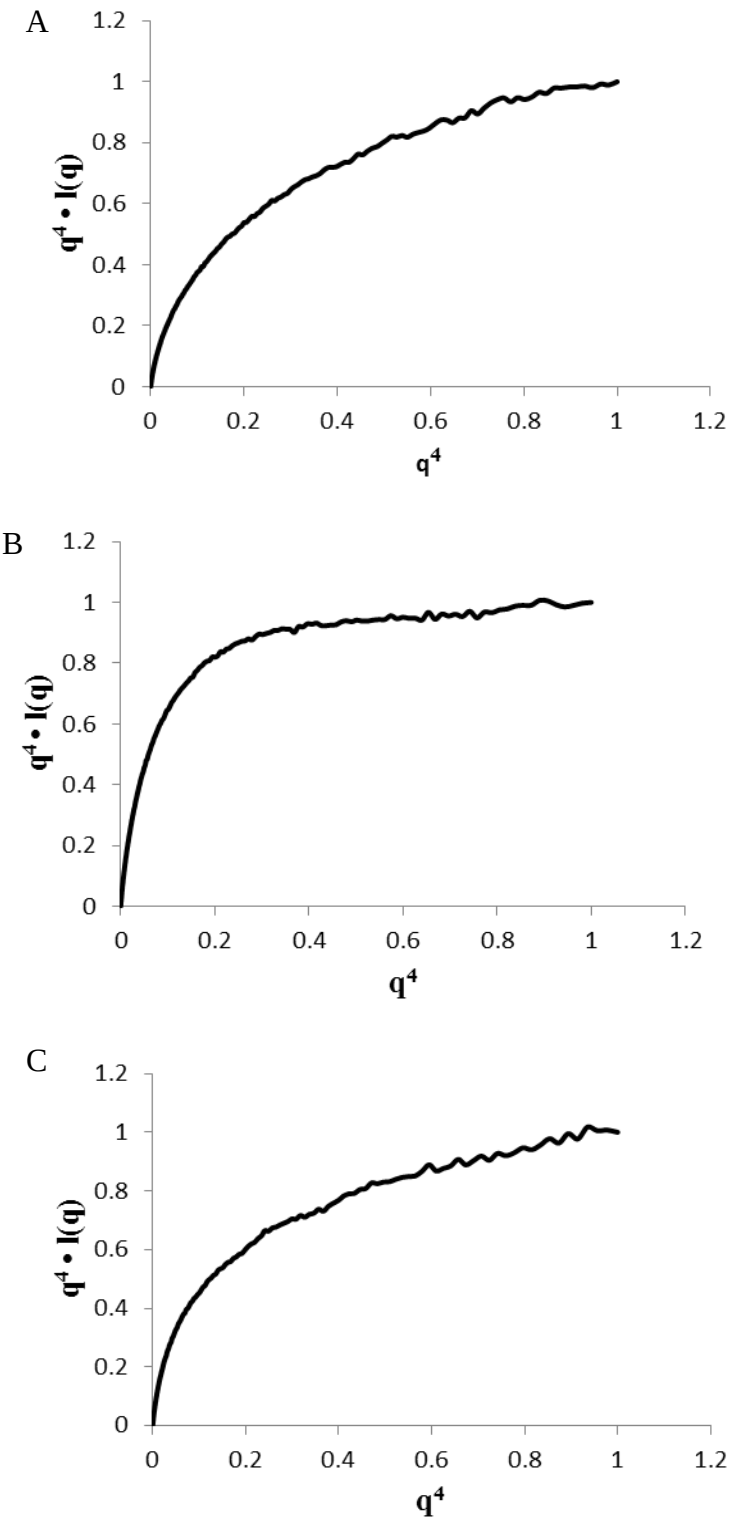



**Figure 2.9. Porod-Debye analysis of SAXS data for the $N_{260}$-pRb/AB complex.** Plots of q4*I(q) vs. q4 plots for (A) $N_{260}$, (B) pRbA/B, and (C) $N_{260}$-pRbA/B complex.

globular, single domain pRbA/B. In contrast, the scattering data of $N_{260}$ alone yielded an elongated shape, which reflects the highly flexible nature of the protein. For the complex, the coordinates for pRbA/B and the J domain fit well into the molecular envelope, but the OBD is difficult to position, which suggests that it remains flexibly tethered to the rest of the complex. This observation suggests there is little if any global compaction of $N_{260}$ upon binding of pRb and that the OBD remains flexible, consistent with the NMR results and the upward slope in the Kratky plot at high q values for the complex (Fig. 2.8A), indicating the presence of at

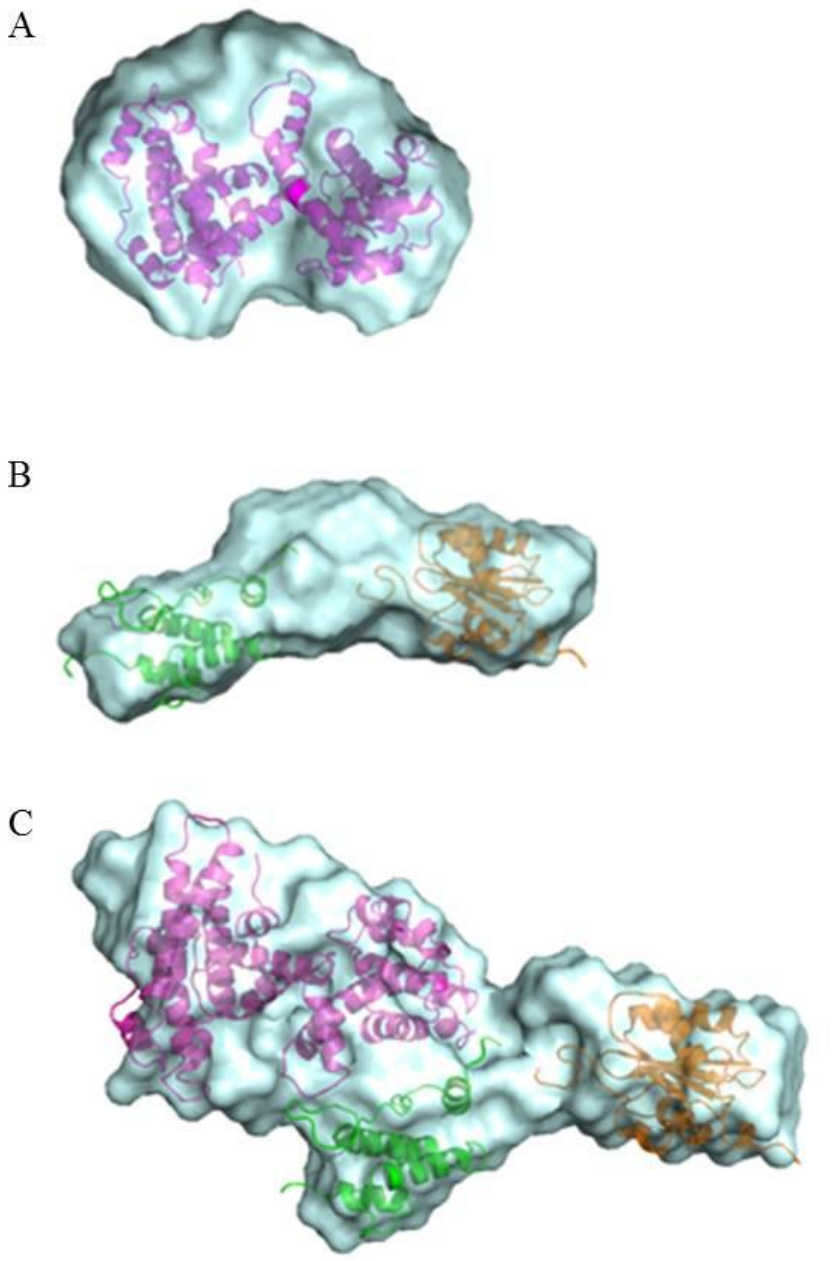



**Figure 2.10. Molecular envelopes calculated from the SAXS scattering data.** Envelope of (A) pRbA/B alone fitted with the crystal structure from 1GH6, (B) $N_{260}$ alone fitted with the structures of the J domain (green - 1GH6) and OBD (orange - 2TBD), (C) $N_{260}$-pRbA/B complex.

least some disordered polypeptide. Together, the data indicate that the Tag J domain remains flexibly attached to the OBD when pRbA/B is bound.

## Discussion

SV40 Tag is a modular protein with multiple domains linked by flexible tethers. The dynamic nature of modular proteins presents challenges for characterization by traditional structural biology approaches. Here we applied a combination of NMR and SAXS to investigate how binding of Rb to Tag affects the structural dynamics of the Tag architecture and to determine implications of these data for the evolving chaperone model.

The chaperone model involves Tag recruitment of Hsc70 and multi-protein complexes involved in transcription and DNA replication that lead to alterations in specific protein-protein interactions. If this model is correct, the chaperone must have the ability to act on different cellular complexes, each containing a unique set of proteins. For example, genetic studies indicate that, in addition to targeting Rb-E2F complexes bound near the amino-terminus, the J domain acts on targets bound to the C-terminus of Tag[98]. In each case the J domain must be in position so that Hsc70 is correctly oriented relative to the target complex. This suggests that the J domain must be able to adopt multiple orientations relative to the other Tag domains. This hypothesis is supported by cryo-EM studies of Tag[54], in which the absence of the J domain in the structure is presumed to arise from it having different orientations in different molecules on the grid. One prediction of the chaperone model is that Hsc70 directly contacts Rb, E2F or both. Another prediction of the chaperone model is that to

effect release of E2F from Rb, Hsc70 must be positioned to influence the Rb-E2F interaction.

A mechanism for release of E2F/DP1 from Rb has recently been proposed[67]. Rb has structured N terminal and central (pocket) domains as well as an unstructured C terminal domain (Fig. 1). The pocket domain is composed of two structured domains (A and B box) connected by a 64 residue unstructured linker. The transactivation domain of E2F binds to the unphosphorylated pocket domain, making critical interactions with both the A and B sub-domains. When phosphorylated, the Rb linker binds to pocket domain at the E2F binding site. The linker can be displaced from the pocket in the presence of the transactivation domain of E2F. This suggests that there is a binding competition between E2F and the pocket linker, which could explain the mechanism of Rb inactivation and E2F release.

Recently, evidence for the direct contact of Hsc70 with E2F has been reported[110]. Hsc70 is known to bind unfolded regions of proteins in a substrate binding domain. An unstructured region was observed in E2F and shown to increase stimulation of ATP hydrolysis by Hsc70, but only in the presence of a J protein[110]. This suggests that the J domain is required to attain proper positioning of Hsc70 for contact with the E2F substrate.

These observations and our results support the chaperone model and highlight the intrinsically dynamic nature of the protein complexes. Our data has shown that the J domain is structurally and dynamically independent of the other Tag domains. This provides a required degree of flexibility for altering the spatial localization of the Hsc70-Tag-Rb complex to drive the dissociation of E2F transcription factors. In

particular, despite binding of both Hsc70 and Rb, the residual dynamics can allow positioning Hsc70 for direct contact with E2F yielding an increased rate of ATP hydrolysis and disassembly of the entire complex, which would lead to improper progression of the cell cycle.

# CHAPTER III

## TOWARD STRUCTURAL CHARACTERIZATION OF SV40 T ANTIGEN J DOMAIN WITH THE CULLIN 7 CPH DOMAIN

### Introduction

The SV40 T antigen protein is capable of transforming a variety of cells and inducing the formation of tumors in animal model. T antigen is a multi-domain, mulit-functional protein that depends upon binding and strategically inactivating proteins of the host cell that regulate the growth cycle. It inactivates the tumor suppressor proteins of the pRb family as well as p53. Over the years, T antigen research has focused on determining the mechanisms by which it transforms a cell. This has led to the discovery of other proteins that are bound by T antigen. The mammalian cullin family protein, Cul7, is one of those reported to be associated with T antigen.

The cullin RING ubiquitin ligase (CRL) protein family is the largest class of ubiquitin ligases, which are critical for the regulation of protein degradation. Proper degradation of proteins is important for many cellular processes, including transcription and signal transduction. CRL proteins are involved in the regulation of these proceses[111,112]. In mammalian cells, there are at least six cullin proteins that have been identified (Cul1, 2, 3, 4a, 4b, 5 and 7). They all contain the conserved cullin homology domain. Cullin forms a scaffold for mulit-protein complexes called SCF (Skp, Cullin, F box) that carry out ubiquitination of target substrates. Cullins are activated by the ubiquitin like molecule Nedd8, which leads to the binding of a RING containing protein

such as Rbx1/Roc1 and recruitment of a ubiquitin conjugating enzyme to the SCF complex[113]. The cullin of interest in this work is Cul7.

In 1992, D. Kohrman and M. Imperiale reported that SV40 T antigen formed a stable complex with a 185 kDa host cell protein[114]. While using co-immuno-precipitation experiments to characterize T antigen-pRb complexes in rodent cells, this protein (p185) was pulled out of solution with the complex. The association of this mystery protein with T antigen was independent of pRb or p53 binding. While working with truncated version of the T antigen protein, the binding of p185 was localized to residues 1-121, which contains the J domain. Additionally, the deletion of residues 2-108 from T antigen abrogated binding to the 185 kDa protein. This suggests that the J domain is important for the binding of this protein and is likely to be functionally important, as well[114].

In 2000, Tsai et al. discovered a 193 kDa protein binding to T antigen via the J domain[115]. Mutation of T antigen residues 107 and 108 did not affect the binding implying that this interaction is independent of pRb binding. The size and binding similarities hinted that this may be the same as the 185 kDa protein. Sequence analysis revealed that the protein contains an LXCXE motif, the consensus binding sequence for pRb proteins, at residues 1052-1056. This protein also contained a Bcl-2 homology 3 (BH3) domain at residues 1556-1572, which is associated with proteins that promote apoptosis. Fluorescence studies found that this protein promoted cell death during the $G_1$ phase before the onset of DNA synthesis. Co-expression of the 193 kDa protein with T antigen inhibited its ability to induce cell death.[115]

In 2004, Ali et al. identified that the 185 kDa and 193 kDa proteins were all the recently characterized cullin protein, Cul7[113,116]. Their research recapitulated the previous findings that Cul7 binds to T antigen via the J domain. T antigen residues 69-83 are particularly important for the binding of Cul7[117]. Cul7 binds T antigen in a manner independent of pRb, and p53. The binding of Cul7 also did not disrupt T antigen mediated viral DNA replication. However, T antigen protein that is defective for binding of pRb or Cul7 could not support cellular transformation. This suggests that Cul7 has a role in cellular transformation. Cul7 also bound to the traditional cullin protein, Rbx1/Roc1 and F box proteins, Fbx29 and Fbw6[113,116].

At this time, the targets of Cul7 were unknown. Subsequent research has shown that Cul7 targets cyclin D1, a regulator of the G1/S phase cell cycle transition, and insulin receptor substrate-1, a mediator of the insulin/insulin-like growth factor 1 signaling system[118]. Dysregulated Cul7 is directly linked to the autosomal recessive diseases 3-M and Yaktus because Cul7 mutations were identified in families with these conditions.[118,119] There are 25 Cul7 mutations associated with 3-M patients alone[119]. The phenotype of both diseases is pre- and post-natal growth retardation[118]. The association of Cul7 with these proteins and diseases suggests that it may play a role in cell cycle progression, senescence and growth control.

When Cul7 is compared to its other family members, some distinctive features are seen. Cul7 has been shown to be present only in vertebrates. The human Cul7 is 1698 amino acids which makes it larger than its sister cullins[118]. Besides the conserved cullin domain, Cul7 has a DOC domain that is involved in ligand binding and a CPH (conserved in Cul7, PARC and HERC2 proteins) domain. A schematic diagram of the

Cul7 domains can be seen in Figure 3.1. Other cullins tend to only interact with proteins within the context of the SCF complex; however Cul7 stably interacts with proteins like p53 and T antigen. The interaction of Cul7 with these proteins has garnered much interest because these associations imply that Cul7 is involved in cellular growth processes and is therefore a possible therapeutic target.

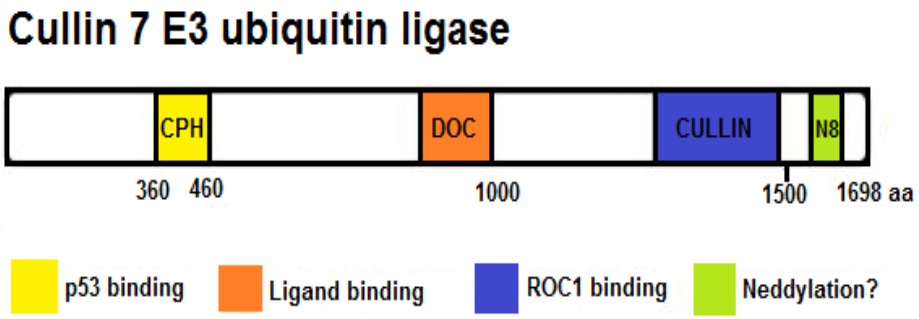



**Figure 3.1. Domain organization of Cullin 7.** This E3 ligase contains a CPH domain (yellow) through which it binds p53; a DOC domain (orange) involved in ligand binding, the conserved cullin domain (blue), and possibly a region through which the post-translational modification, neddylation, may occur (green). *Adapted from Sarikas et al.*

The p53 tumor suppressor protein is inactivated in many cancers, so detailed insight into the mechanisms of its regulation has been a research focus for many years. Cullin E3 ligases typically participate in the degradation of a protein target. Cul7 does support the mono- and di-ubiquitination of p53 but there is no evidence that it promotes p53 degradation.[118,120,121] Reports show that p53 is required after DNA damage to upregulate the Cul7 mRNA and protein levels, which may suggest this interaction has a role in transcriptional regulation[120]. In a functional screen, Cul7 prevented apoptosis and promoted the transformation of neuroblastoma cells in a p53 dependent manner[122]. More research must be done to confirm that Cul7 regulates apoptosis. To better understand the structure, interaction and function of the Cul7-p53 binding, the Arrowsmith laboratory turned to solution NMR spectroscopy studies[123]. NMR

experiments revealed that the CPH domain (360-460) of Cul7 binds to the tetramerization domain of the p53 protein. They also determined a structure of the CPH domain (Figure 3.2). The CPH domain has a β-barrel tertiary structure consisting of three loops, five antiparallel β-strands, and an α-helix. The CPH domain was deemed necessary and sufficient for p53 binding[124]. As mentioned previously, T antigen binds and inactivates p53. As the SV40 virus takes over a cell, the various roles of Cul7 make it an attractive protein for T antigen to target and usurp.

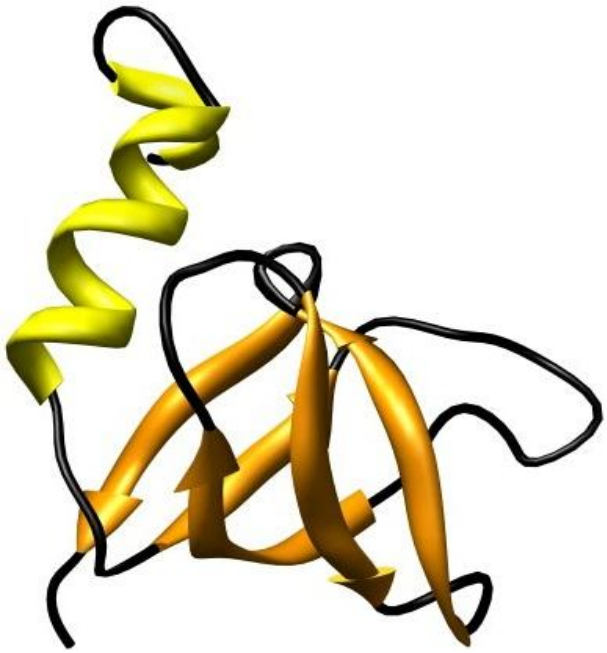


**Figure 3.2. Structure of the Cul7 CPH domain.** The CPH domain contains and alpha helix (yellow) and four beta sheets (orange). *PDBid:2JNG.*

The Cul7 protein binds to the J domain of the T antigen protein. The critical T antigen residues (69-83) for that interaction have been identified. The goal of this research study was to determine the location on Cul7 to which the J domain binds. The Fanning laboratory performed yeast two hybrid experiments that implicated the CPH domain as the binding site for T antigen (E. Fanning, personal communication). The molecular weights of the J and CPH domains (11.7 kDa and 11.5 kDa, respectively) make them readily amenable to NMR spectroscopy. Hence, NMR chemical shift perturbation assays were used to verify the interaction between the J and CPH doamins and to determine which Cul7 residues are involved in the binding interface.

## Materials And Methods

### *Cloning*

The pET15b plasmid containing $Cul7_{CPH}$ was received from the Fanning laboratory. Primers were designed to clone T antigen construct $N_{102}$ containing residues (4-102) which span the functional J domain. For the $N_{102}$ fragment, the forward primer (5'-GGGAATTCCATATGGTTTTAAACAGAGAG-3') contains a Nde1 restriction site and ATG start codon. The reverse primer is (5'-CCGCTCGAGTCAGTTTTCCTCATTAAAGGC-3') contains an Xho1 restriction site (underlined) and either TTA or TAA stop codon. For the $N_{117}$ construct (4-117), the forward primer is the same as the $N_{102}$ forward primer. The reverse primer is (5'-CCGCTCGAGTCAAGTAGCCTCATCATC-3'). The restriction enzymes used were the same as for $N_{102}$. Both constructs were cloned into the pET15b vector which contains a thrombin cleavable 6X-His tag.

### *Protein Expression*

The pET15b plasmid containing $N_{102}$ or $Cul7_{CPH}$ was transformed into BL21(DE3) competent cells. The unlabeled protein was prepared using LB medium containing ampicillin (100 mg/mL) and grown at 37 °C with shaking until $OD_{600}$ reaches between 0.5-0.9. Protein expression was induced with IPTG to a final concentration of 1 mM and continued for 3-5 hrs at 37 °C with shaking. The expression profile of $Cul7_{CPH}$ can be seen in Figure 3.5. The culture was centrifuged at 7 K rpm for 20 minutes and the resulting pellet was stored at -20 $^{o}$C. Uniform $^{15}N$- labeled proteins were produced similarly, except for using M9 media enriched with $^{15}NH_4Cl$.

***Protein Purification***

The purification of $N_{102}$ and $Cul7_{CPH}$ proceeded in the same manner. The cell pellet was resuspended with a protease inhibitor tablet in 25 mL of the following buffer: 25 mM Tris at pH 7.5, 250 mM NaCl, 2 mM DTT. The lysate was sonicated for 5 minutes (5 sec on; 5 sec off) and then centrifuged at 20K rpm for 20 minutes. The lysate was filtered and incubated with 5 mL column volume of NiNTA resin for 20 minutes. The column was then drained and washed with 20x column volume of the above buffer. The 6xHis tag was cleaved by mixing 10 units (1unit/uL) of thrombin with one column volume of buffer and adding it to the column. The slurry of resin bound protein and thrombin are then incubated with rocking at 4$^{o}$C for 4 to 8 hours. The cleaved protein was then drained and the column washed with 5 column volumes of buffer to remove any remaining cleaved protein.

***Isothermal Calorimetry***

Isothermal Calorimetry (ITC) responds to changes in heat evolution as molecules interact. Titration of a binding partner into the sample cell containing the protein of interest allows for a binding curve to be generated. ITC was performed to investigate the interaction between the T antigen construct $N_{102}$ and $Cul7_{CPH}$. Both samples are purified in a buffer containing DTT. A buffer containing DTT cannot be used during ITC experiments, however BME is allowed. So, the protein samples were dialyzed into 25mM Tris at pH 7.5 and 250mM NaCl and then into 25mM Tris at pH 7.5, 250mM NaCl, and 1mM BME. The $N_{102}$ was concentrated to ~750 μM and $Cul7_{CPH}$ to ~30 μM. The $Cul7_{CPH}$ was placed in the sample cell and $N_{102}$ was injected during the experiment.

The data was collected at 25 ºC using a MicroCal VP-ITC and analyzed using the manufacturer's software.

***Dynamic Light Scattering***

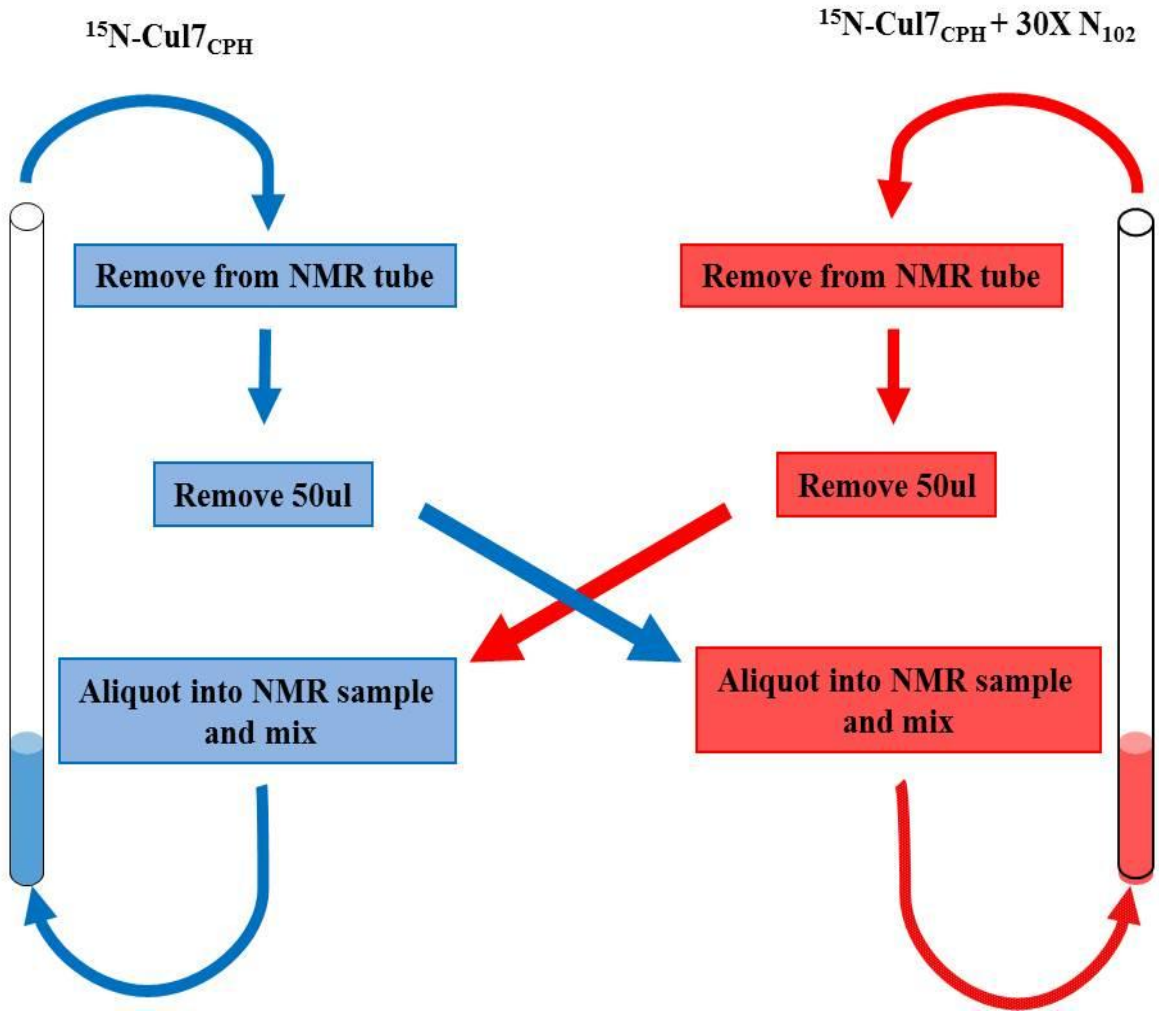
Dynamic Light Scattering (DLS) measures the intensity fluctuations of scattered light arising from the random motions of a particle in solution (ie. Brownian motion). The Brownian motion is inversely proportional to size therefore the size of the molecule can be deduced. This technique was used to determine whether $N_{102}$ was monomeric or forming some higher order oligomer. The buffer used for this experiment was 25mM Tris at pH 7.5, 250mM NaCl and 1mM BME. The concentration of $N_{102}$ used for the DLS experiment was 75µM.

***NMR Chemical Shift Perturbation Assays***

All NMR titration experiments were performed on the 500 MHz Bruker spectrometer equipped with a cryoprobe and operating at 25 °C. Two-dimensional $^{15}$N-$^{1}$H-HSQC spectra were acquired on $^{15}$N-enriched samples of $Cul7_{CPH}$ in 90% $H_2O$/10% $D_2O$, and in the presence of unlabeled $N_{102}$ titrated to assess the interaction and to map the binding surface. The buffer used for the NMR titration experiments was 100 mM NaCl, 25 mM Tris and 2 mM DTT at pH 7.5. The ratio of unlabeled $N_{102}$ to $^{15}$N-$Cul7_{CPH}$ used in this experiment was 1:30 with $^{15}$N-$Cul7_{CPH}$ at 100 uM and $N_{102}$ at 3 mM. Because of the high ratio, a cross titration method was used in this experiment so that the volume of the sample would not significantly change and thus dilute the $^{15}$N-$Cul7_{CPH}$ concentration. Two NMR samples were prepared. One contained only $^{15}$N-

$Cul7_{CPH}$ at 100 uM and the other contained the 1:30 concentration of $^{15}N$-$Cul7_{CPH}$:$N_{102}$. A two-dimensional $^{15}N$-$^{1}H$-HSQC spectrum was collected for each sample. To obtain the points between the zero and high concentrations of $N_{102}$, 50 μL was removed from each sample, put into the other, vortexed to fully mix and then spun down to release any bubbles. Then $^{15}N$-$^{1}H$-HSQC spectra were collected for the new samples. The exchange was performed multiple times and the concentration of these middle points was calculated. Figure 3.3 provides a schematic diagram of this technique and a listing of all the titration points in this experiment. All NMR data were processed using TOPSPIN (Bruker Biospin Ltd.) and analyzed using Sparky (University of California, San Francisco).

A




B

**Titration Points**

| | |
|---|---|
| 1:0 | 1:16.9 |
| 1:3 | 1:17.3 |
| 1:7.8 | 1:17.8 |
| 1:9.2 | 1:18.5 |
| 1:11.6 | 1:20.8 |
| 1:12.3 | 1:22.2 |
| 1:12.9 | 1:27 |
| 1:13.3 | 1:30 |

**Figure 3.3. NMR Titration of $N_{102}$ into $^{15}Cul7_{CPH}$.** A. Schematic diagram of the titration method. B. List of the titration points collected

## Results

### *Production of $N_{102}$ and $Cul7_{CPH}$ protein*

In order to characterize the CPH domain of Cul7 and the putative changes that occur upon binding of the T antigen J domain, production of pure protein for NMR spectroscopy studies was necessary. The T antigen construct $N_{102}$ was successfully subcloned, expressed and purified using NiNTA affinity chromatography. Figure 3.4 shows the high yield of unlabeled $N_{102}$ protein from one liter of cell culture directly after the NiNTA affinity chromatography step. The expression of unlabeled $Cul7_{CPH}$ was low (Figure 3.5), which leads to even lower expression of $^{15}N$-enriched protein (Figure 3.6). From 2 liters of cell culture, the protein yield is ~2 mg/mL. While producing enough protein was the rate limiting step, once produced the protein is stable at 4 ºC. However, the poor yield only allows for only a uni-directional NMR titration where $Cul7_{CPH}$ is the protein kept at a constant concentration. This purified protein was used for ITC and NMR experiments.

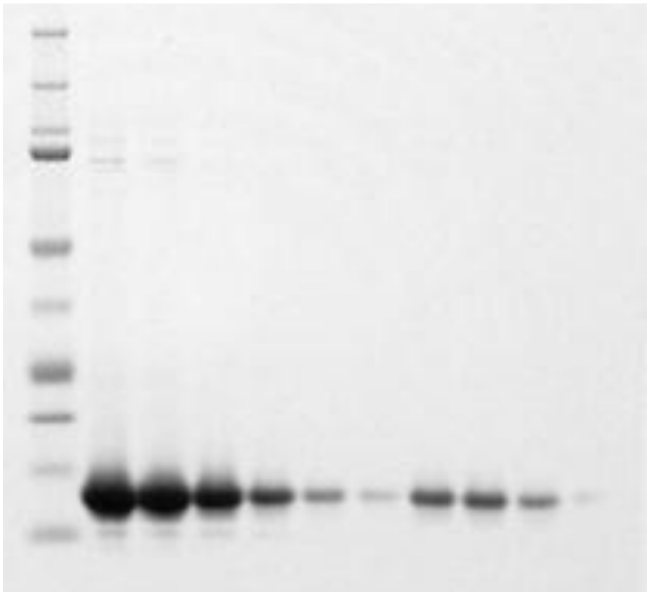


**Figure 3.4. Purified $N_{102}$ protein.** This SDS gel is of 1L of unlabeled protein. Lane 1 is the molecular marker. The remaining lanes contain purified protein.

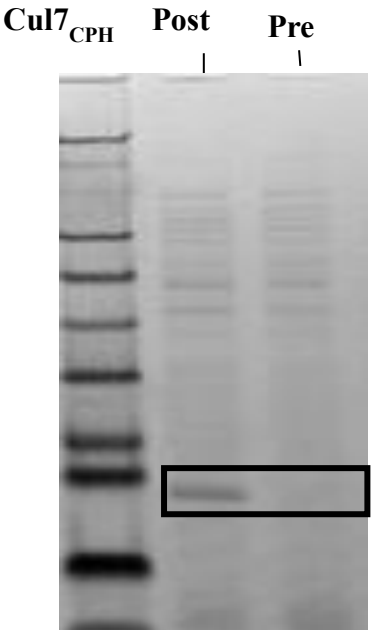



**Figure 3.5. E-coli expression of $Cu7_{CPH}$.** Lane 1 of this SDS PAGE gel is the molecular marker. Lane 2 is the protein expression post IPTG induction. Lane 3 is pre-induction protein.

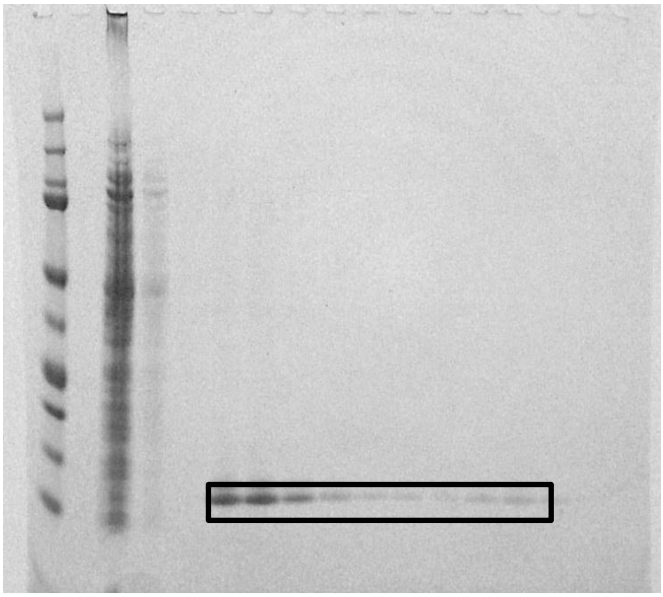


**Figure 3.6. SDS Purification gel of 2L of $^{15}Cul7_{CPH}$.** The first lane contains the molecular weight marker. The second and third lanes are the beads prior to protein load and the flow through, respectively. Lanes four and five are washes done to remove unbound proteins. Elutions containing purified $Cul7_{CPH}$ are highlighted in the black box.

***The J domain of T antigen is monomeric and folded in solution***

Dynamic light scattering (DLS) measurements were used to characterize the oligermeric state of the J domain construct $N_{102}$. The DLS sample contained $N_{102}$ at a concentration of 74 uM. Analysis of the spectrum revealed a single peak at ~12 kDa which corresponds to the molecular weight of $N_{102}$ of 11.4 kDa (Figure 3.7). NMR spectroscopy was used for further characterization of the J domain construct. Optimization of NMR spectral conditions, which involved testing various buffers, temperature and pH conditions, was primarily performed on the longer J domain construct ($N_{117}$) but was also applied to $N_{102}$. Circular dichroism (CD) spectroscopy, (data not shown), used to evaluate the secondary structure of the J domain, was performed on the $N_{117}$ construct, as well. The buffer considered best for collection of J

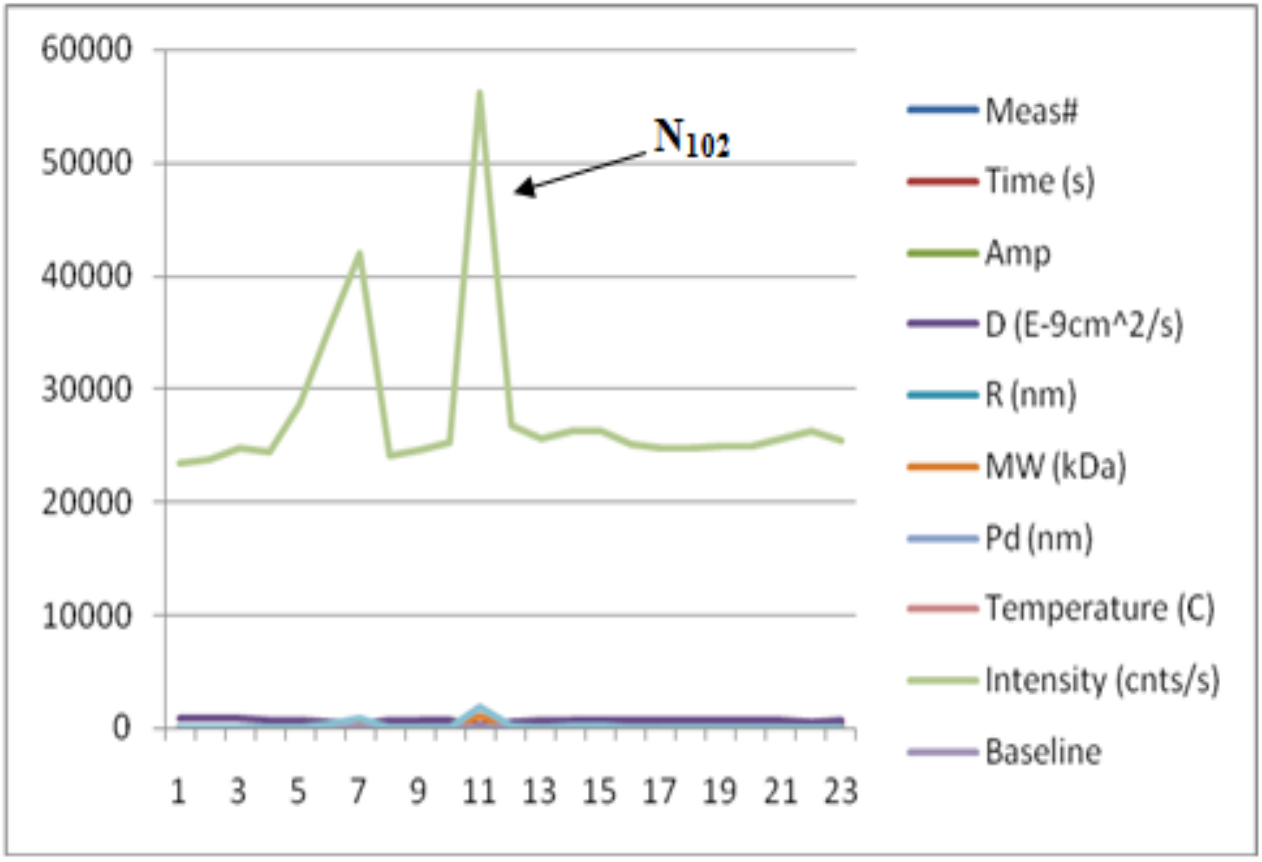



**Figure 3.7. Dynamic Light Scattering of $N_{102}$.** The DLS results showed two peaks. The smaller peak at 7 kDa corresponds to the signal from the solvent. The larger peak at 11 kDa is that of protein. The presence of a single protein peak shows that $N_{102}$ is monodisperse in solution.

domain NMR spectrum consisted of 50mM $Na_2HPO_4$ at pH 6.5, 100mM NaCl and 2mM DTT. A protein concentration of 200uM was used to collect two-dimensional HSQC spectra for each construct (Figure 3.8). A total of 87 peaks can be distinguished for the $N_{102}$ protein and 97 peaks for $N_{117}$. These peaks were well dispersed across the spectrum which is an indication of a well-folded protein. The spectrum of $N_{102}$ could perfectly overlay on the spectrum of the longer $N_{117}$ protein, which confirms that the solution conditions for one can be used for the other. The peaks belonging to the extra residues in $N_{117}$ could be clearly seen in the central region of the HSQC suggesting that they are unstructured and flexible.

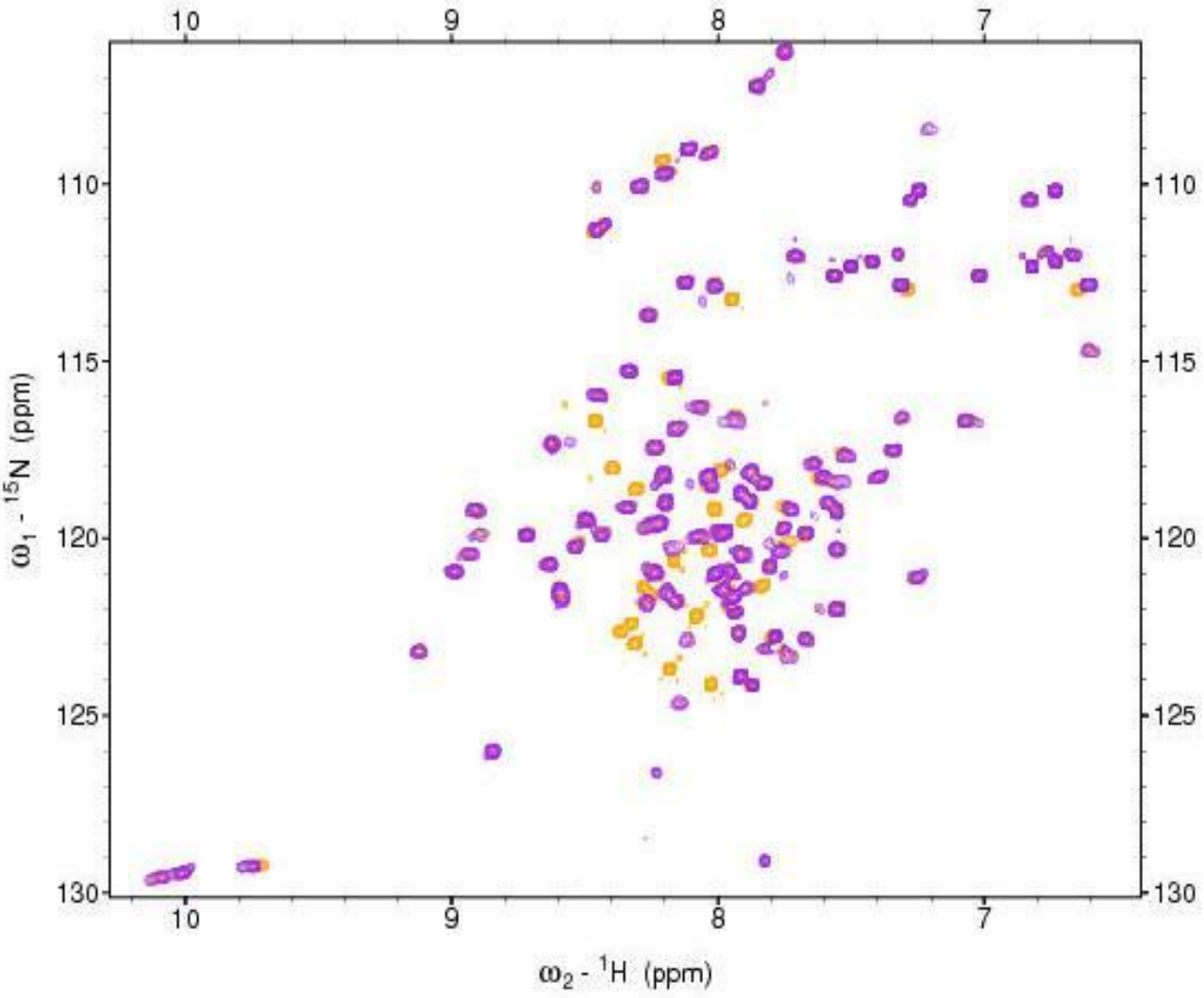



**Figure 3.8. Overlay of HSQC spectra for the J domain constructs.** Orange: HSQC spectrum of $N_{117}$; Purple: HSQC of $N_{102}$

***Toward the biophysical characterization of T antigen/Cullin7 interaction***

As previously mentioned, unpublished yeast two-hybrid experiments were performed in the Fanning laboratory. The yeast two-hybrid technique measures protein-protein interactions by measuring transcription of a reporter gene. If protein X and protein Y interact, then their DNA-binding domain and activation domain will combine to form a functional transcriptional activator (TA). The TA will then proceed to transcribe the reporter gene that is paired with its promoter yielding colony formation.

The Fanning lab tested the binding of $N_{102}$ construct with $Cul7_{CPH}$ along with full length T antigen with p53 and laminin as positive and negative controls, respectively (Table 3.1). Their research indicated that the J domain of T antigen may bind Cul7 via its CPH domain.

| **Binding Domain** | **Activation Domain** | **Colony Growth** |
|---|---|---|
| p53 (positive) | Tag (full length) | Yes |
| Laminin (negative) | Tag (full length) | No |
| Cul7(360-433) | Tag (4-102) | Yes |

**Table 3.1. Yeast two-hybrid of T antigen ($N_{102}$) with Cullin7 (CPH domain).**
The results indicate that there is an interaction between the CPH domain of Cul7 and the J domain of T antigen.
**Results from the Fanning laboratory at Vanderbilt University.*

The region on T antigen that Cul7 binds has been structurally characterized. This data prompted the structural investigation of the location on Cul7 at which T antigen interacts.

The first step towards attaining a more detailed understanding of the Cul7 interface for T antigen binding was to perform ITC experiments. Due to the low protein yield of $Cul7_{CPH}$, 50 μM $Cul7_{CPH}$ (360-460) was placed in the ITC sample cell at 25 ºC and 1 mM of $N_{102}$ was titrated into the cell. Both proteins were in an identical buffer, which contained 25 mM Tris (pH 7.5), 250 mM NaCl, and 1 mM BME. Unfortunately, we did not observe any heat evolution that would allow fitting of the data to a binding model and calculating the dissociation constant (Figure 3.9). While the results of ITC showed no observable heat evolved during the titration of $N_{102}$ into $Cul7_{CPH}$, this does

not necessarily rule out binding, but is simply indicative of a low ΔH of the reaction. A more sensitive technique, like NMR spectroscopy, was needed to continue the characterization of this interaction.

### *Mapping the Cul7$_{CPH}$ binding surface utilized by the J domain of T antigen*

The solution structure of Cul7$_{CPH}$ was previously determined using NMR methods by the Arrowmith laboratory and the assignments were deposited in the BMRB[123]. This structure and the assignments were used in combination with NMR titration data to map the Cul7$_{CPH}$ surface involved in the interaction with the J domain of T antigen. NMR is a powerful structural technique that has proven useful in the study of protein-protein interactions[125]. Performing an NMR chemical shift perturbation assay is the most common method to study protein interactions. In this assay, a two dimensional HSQC is collected on an $^{15}$N-labeled protein. An unlabeled binding partner is then added to the sample containing the $^{15}$N-labeled protein at predetermined concentrations. Additional HSQC spectra are recorded after each addition of the unlabeled protein. The

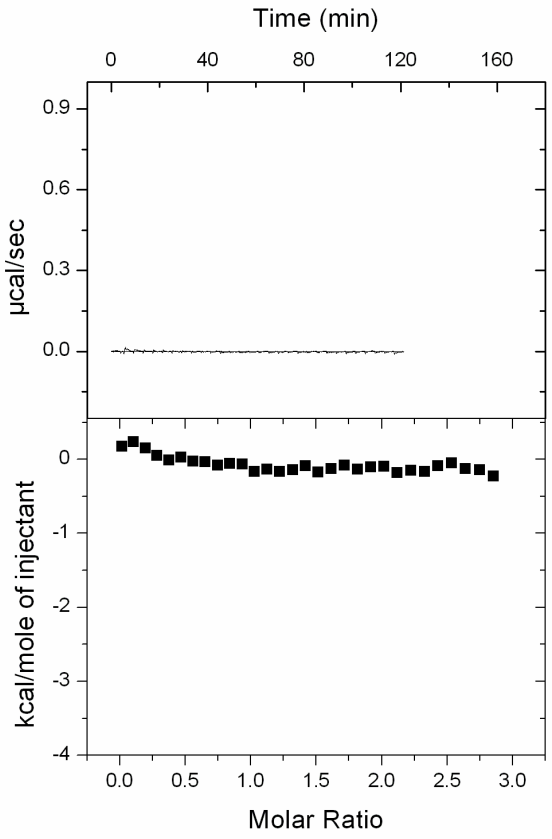



**Figure 3.9. ITC experiment of Cul7$_{CPH}$ with N$_{102}$.** The ITC curve was linear which shows that an interaction was not observable by ITC.

chemical shift is a parameter that is very sensitive to changes in environment. If an interaction has occurred, the peaks corresponding to residues at the binding interface will be perturbed. Peak perturbations can either be a change in the location of the peak or a decrease in its intensity. Changes in the structure of the protein upon an interaction will also produce chemical shift perturbations. When a high resolution structure and the corresponding resonance assignments are known, the binding interface can be accurately identified and mapped onto the structure.

As previously mentioned, the initial step is to obtain a spectrum of the $^{15}$N-

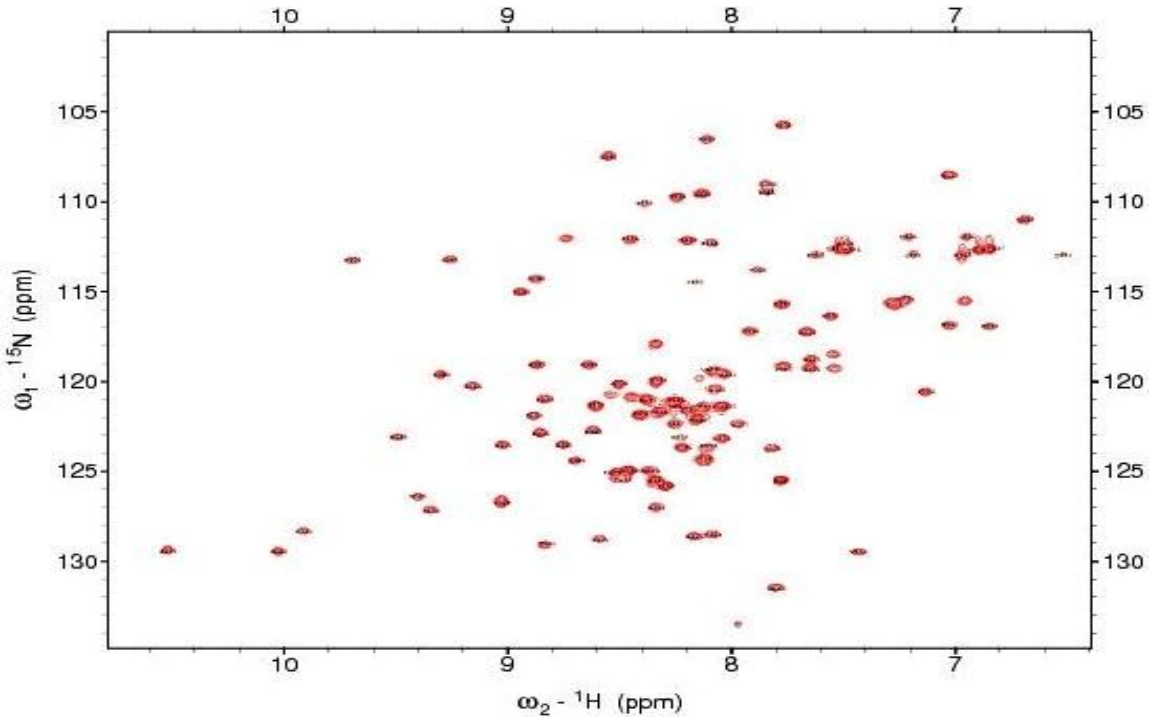



**Figure 3.10. HSQC spectrum of Cul7$_{CPH}$.** The peaks are dispersed across the spectrum which confirms that this is a well folded protein.

labeled protein alone in solution, in this case Cul7$_{CPH}$. The HSQC spectrum of Cul7$_{CPH}$ was collected in 25 mM Tris (pH 7.5), 250 mM NaCl, and 2 mM DTT (Figure 3.10). This spectrum showed well dispersed signals, which indicates that Cul7$_{CPH}$ is a well folded domain and suitable for further NMR experiments. The chemical shifts in the

spectrum were quite similar to the previously assigned $Cul7_{CPH}$ resonances and could be transferred onto the peaks in this spectrum[123]. A total of 95 of the 101 peaks of the $Cul7_{CPH}$ (360-360) were unambiguously assigned.

Collection of the initial spectrum was followed by establishing the titration conditions. The first NMR titration performed at points along the concentration range of 0:1 to 6:1 ratio of $N_{102}$:$^{15}Cul7_{CPH}$. This experiment did not reveal significant chemical shifts so collecting higher titration points was necessary. Hence, the NMR titration experiment was re-structured to contain just three points: 0:1, 5:1 and 20:1. The buffer was modified to contain 100 mM NaCl, 25 mM Tris at pH 7.5, and 2 mM DTT from a buffer that contained 250 mM NaCl. The final concentrations of the protein stock solutions were 11.5 mM for $N_{102}$ and 250 µM for $^{15}Cul7_{CPH}$. From the $^{15}Cul7_{CPH}$ stock, a 100 µM NMR sample was prepared. After the NMR spectra were collected and processed, the 20:1 spectrum was overlaid upon the 0:1 spectrum. There were some obvious chemical shifts and intensity changes in several of the resonances. These resonance perturbations were matched to the chemical shift assignments (Table 3.2). A few of the residues perturbed are also changed upon the binding of Cul7 to p53[123]. The results of this analysis suggest that J domain may occupy the same or similar binding site as p53.

| **$Cul7_{CPH}$ residues perturbed upon $N_{102}$ binding** |
|---|
| Y372; D375; L377; Q378; E391; R401; S403; N404; Q411; E429; W422; H424 |

**Table 3.2. Residues perturbed during 20:1 $N_{102}$:$^{15}Cul7_{CPH}$ NMR titration.** Residues in red were perturbed upon binding p53 as reported in Kaustov 2007.

To obtain detailed structural information regarding the T antigen/Cullin7 binding interface, the NMR titration experiment was restructured and performed once again. As increasing amounts of a binding partner is added, the volume of the sample is increased and the concentration of the $^{15}$N-labeled protein is diluted. The titration experiment designed to combat this problem has been described in the Experimental Methods section. Briefly, the final concentration for each protein stock used in this experiment was: [$N_{102}$] = 6 mM and [$^{15}Cul7_{CPH}$] =338.8 µM. Two NMR samples were prepared.

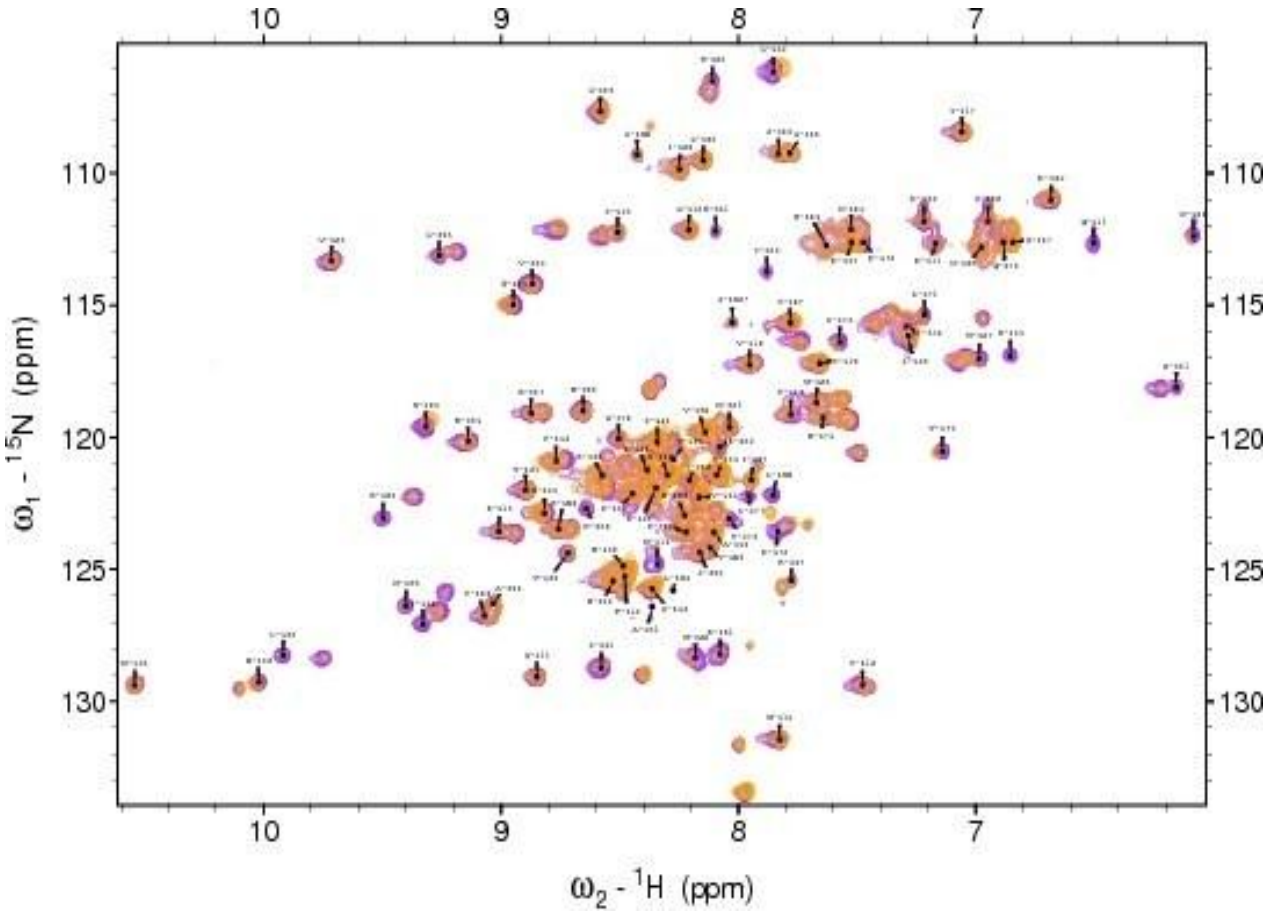



**Figure 3.11. Overlay of NMR titration spectra.** HSQC of titration point 0:1 (purple); HSQC of titration point 30:1 (orange)

One contained just 100 µM $^{15}Cul7_{CPH}$ (ratio 0:1) while the other contained 3 mM $N_{102}$ and 100 µM $^{15}Cul7_{CPH}$ (ratio 30:1). A one-dimensional NMR spectrum was collected on the 30:1 ($N_{102}$:$^{15}Cul7_{CPH}$) sample and on a 50 µM NMR sample from the unlabeled

$N_{102}$ protein stock to ensure that protein was present and properly folded. The experiment continued with the collection of two-dimensional HSQC spectra for each titration point. An overlay of the 30:1 spectrum onto the 0:1 spectrum can be seen in Figure 3.11. This overlay shows that there were selective changes in intensity of certain peaks as well as definite perturbations of chemical shifts. This data highlights that a binding interaction has occurred. However, the large amount of $N_{102}$ required to elicit the modest peak shift indicates that the interaction is weak.

The significance of the observed residue perturbations were evaluated and measured. The average change in chemical shift of all the residues was calculated

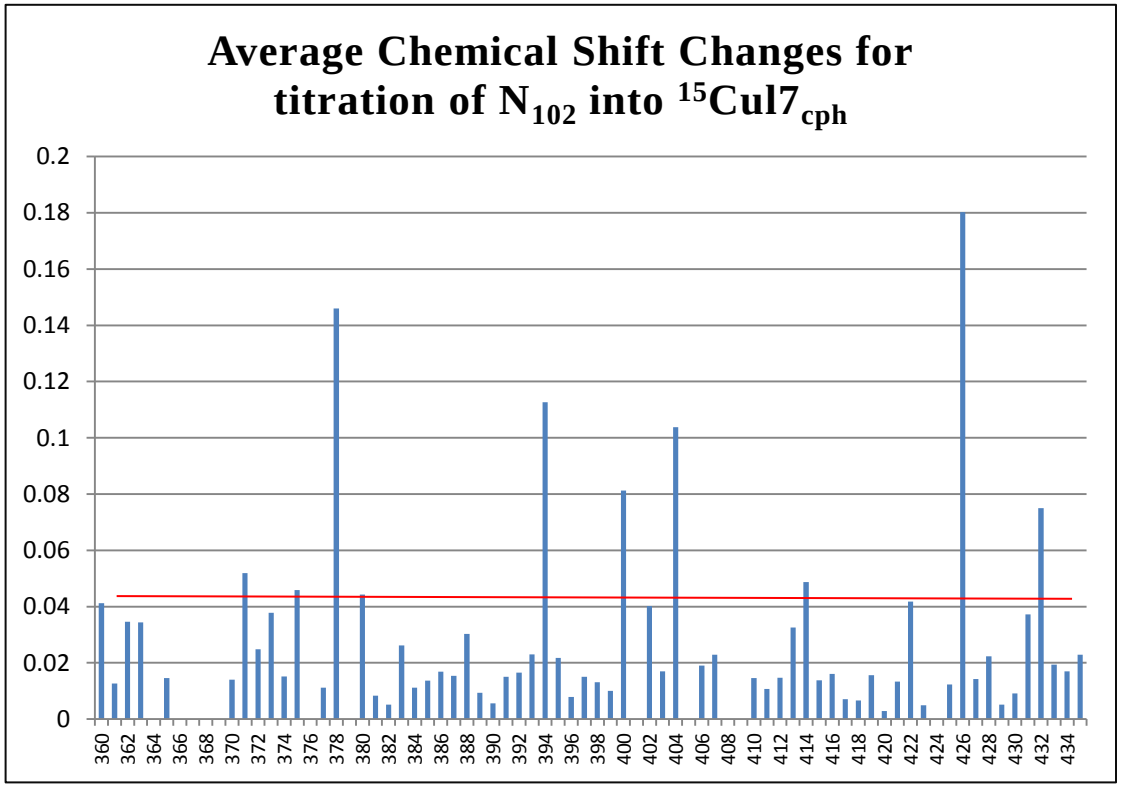



**Figure 3.12. Plot of the average change in chemical shift.** The red line shows approximately where the standard deviation, 0.0338 is located.

using the following formula: $\Delta\delta = [(0.17\Delta\delta N)2 + (\Delta\delta HN)2]1/2$[126]. The results of this calculation can be seen in Figure 3.12. Residues that have a chemical shift change

greater than the standard deviation (std. dev. = 0.0338) may be significant in the binding interaction. The calculation was applied to all of the Cul7CPH residues (360-460), however only residues until residue 435 were plotted. The PDB structure of $Cul7_{CPH}$ includes only residues 360-435 because residues 436-460 were determined to be disordered[123]. The residues with significant perturbations are listed in Table 3.3. Only the residues above the standard deviation (those in the yellow and green columns) were considered for further analysis.

| Red | Yellow | Green | **Average change in chemical shift:** 0.028119 |
|---|---|---|---|
| E362 | R360 | Q378 | **Standard Deviation:** 0.033818 |
| F363 | V373 | A394 | |
| D388 | L371 | F400 | **Residue Range** |
| *D436* | D375 | N404 | ***Red***: 0.028119 - 0.033818 |
| *M440* | G380 | F413 | ***Yellow***: 0.033818 - 0.08 |
| *V450* | Q402 | H426 | ***Green***: 0.08 and above |
| *V454* | W414 | *Q447* | Significant residues after structure ends |
| | W422 | | Residues that may not be significant |
| | L431 | | |
| | G432 | | |
| | *E438* | | |
| | *D439* | | |
| | *Y441* | | |
| | *E442* | | |
| | *A443* | | |

**Table 3.3. The list of the residues with significant perturbations.** Only residues greater than standard deviation were considered in the rest of the analysis.

The change in chemical shifts, in both the $^{15}N$ and $^{1}H$ dimensions, was then plotted against the different concentrations of $N_{102}$ for each of the significant residues,. This leads to the identification of residues that showed erratic chemical shifts that are not due to the binding event. The charts for the analysis of the amide and proton chemical shifts are in Figure 3.13. For example, the residue, Q378 showed an

inconsistent chemical shift pattern which is not likely due to the binding event so it was withheld from further analysis. The other residues showed a consistent change in chemical shift as the concentration of $N_{102}$ was increased. This data confirmed that some of the residues involved in binding p53 (ie. N404 and W422) are also involved in binding the J domain of T antigen. However, there are many differences, which suggest that T antigen may not bind in the exact same manner as p53 but rather it occupies a similar surface. The residues of the $Cul7_{CPH}$ domain that displayed significant perturbation upon binding to $N_{102}$ were mapped onto the NMR solution structure of the domain (Figure 3.14). While there are not a lot of residues that were significantly perturbed, they do appear to occupy a common region on the CPH domain. The perturbed residues reside on either side of a central channel formed by a flexible region and a beta sheet. These studies have provided insight into the T antigen binding site on Cul7. However, further experiments are required to gain additional insights into the precise mechanism of binding.

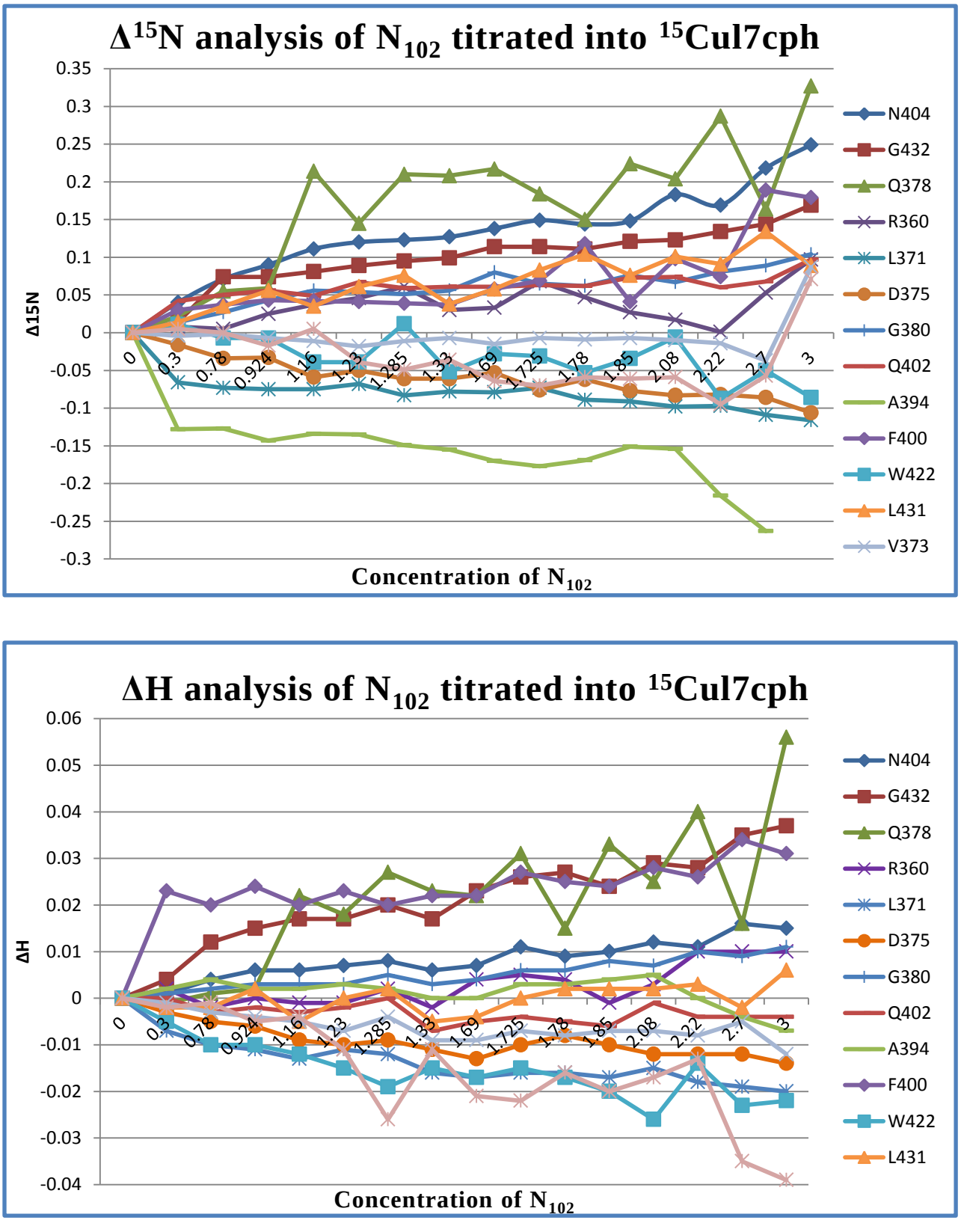



**Figure 3.13. Plot of the change in chemical shift.** The changes in the amide (top) and proton (bottom) dimensions, plotted against the residues identified as significant, against the concentrations of $N_{102}$. Q378 was eliminated as a significant residue.

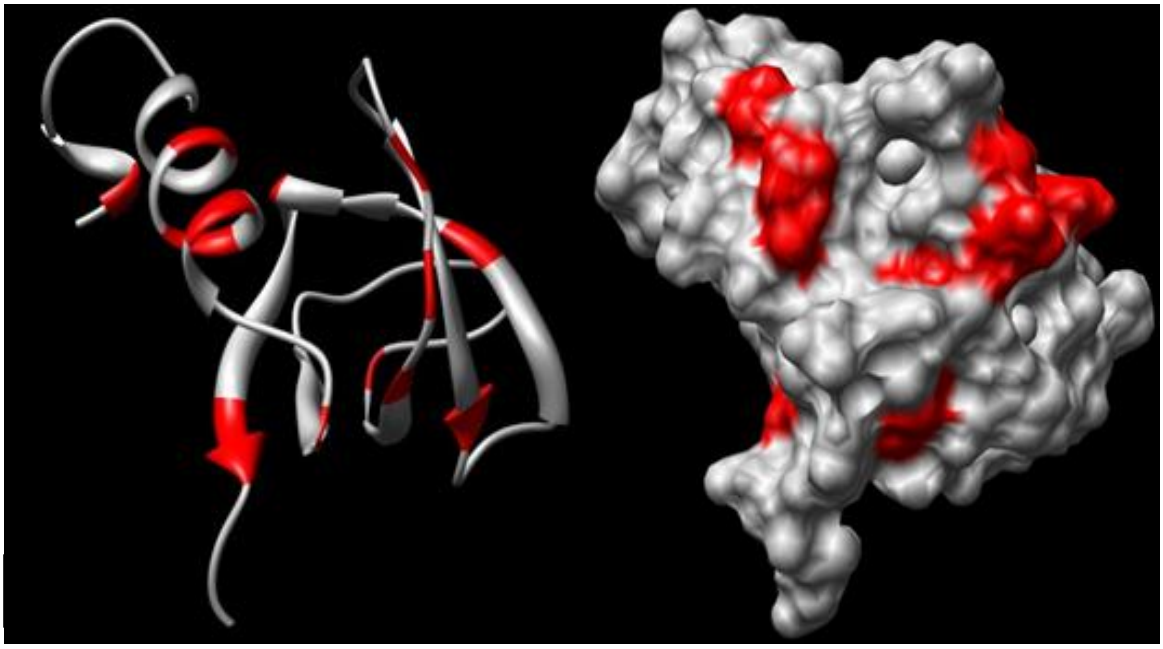


**Figure 3.14. Structure of Cul7$_{CPH}$.** A. Red- the residues significant in the binding of J domain to Cul7. B. Solid surface representation. Note: Figures are in the same orientation. PDBid: 2JNG

## Discussion

The role of SV40 T antigen is to utilize or usurp proteins in a host cell for the intention of producing a deregulated and transformed cell. Cullin7 is an E3 ligase that can mediate protein degradation. The binding interaction of T antigen to Cullin7 has been established in the literature[113,124]. It has been demonstrated that this interaction is essential for cellular transformation[113]. It is believed that T antigen induces cellular transformation by utilizing its function as a chaperone protein. This need for Cullin7 suggests it may have a role in the T antigen chaperone mediated protein inactivation for the propagation of the SV40 virus and cellular transformation.

Mutational studies identified T antigen residues 69-83 as the binding surface for Cul7. These residues are located in the J domain of T antigen and their deletion abrogates its ability to bind Cul7 but not its ability to association with p53 and pRb[124].

The deletion of these T antigen residues also reduces its cellular transformation abilities. Our study was designed to identify the putative binding region on Cul7 that is required to bind T antigen.

The literature has identified the CPH domain in Cul7 as sufficient for binding to the tumor suppressor protein p53. This domain is conserved, which suggests that it has an important role. Since the CPH domain is already known to be a protein binding domain, it was proposed that T antigen may bind there as well. The role of T antigen is to disrupt the regular functions of cellular proteins. Therefore occupying a binding surface that belongs to a host protein could definitely cause a disruption in the normal cellular processes.

A biochemical and structural approach was used to investigate the J and CPH domain interaction. The results indicated that a weak interaction had formed between the two domains. The weak binding may speak to the mechanism of the function or indicate that some additional residues outside the CPH domain and/or other host cell proteins may be involved to strengthen the interaction. In the study of the p53-$Cul7_{CPH}$ domain interaction the affinity was also weak (~530 μM)[123]. However, when compared, the affinity of p53 was greater for the full length protein than for the CPH domain alone[123]. This may hold true for the binding of T antigen as well.

In mapping all of the residues involved in this interaction onto the structures of the J and CPH domains, a possible mode on binding is revealed (Figure 3.15). The T antigen J domain residues involved in binding Cul7 are on the loop between the third

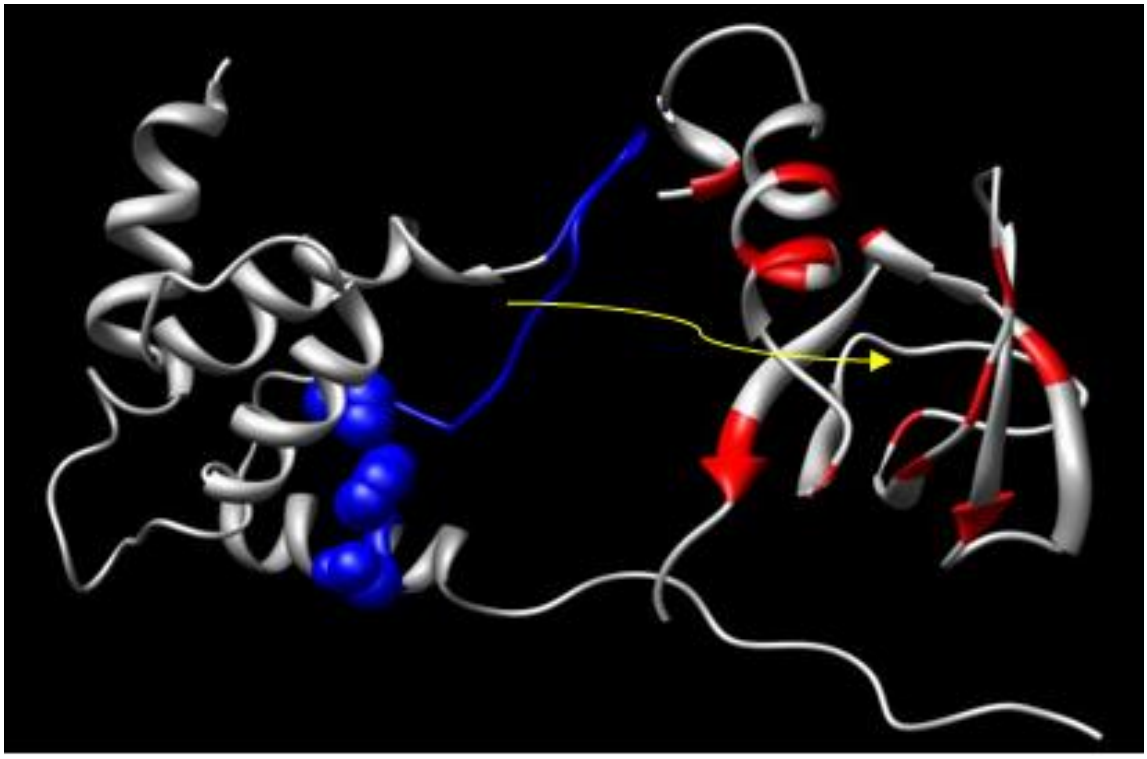


**Figure 3.15. Mode of T antigen and Cullin 7 binding.** The structures of the T antigen J domain (left, 1GH6) and Cul7CPH (right, 2JNG) have the known interacting residues highlighted in blue and red, respectively. The yellow arrow signifies the loop of the J domain could insert into the cavity in Cul7CPH.

and fourth helices. The CPH domain residues demonstrated to be involved in the interaction surround a central channel in the structure. It is possible that the T antigen/Cullin7 mechanism of interaction occurs via the insertion of the J domain loop into the central region of the CPH domain. It is probable that this is a highly dynamic interaction involving other proteins of the host cell. As mentioned previously, the binding of Cul7 to T antigen does not disrupt pRb or p53 binding. These proteins, along with the co-chaperone of T antigen Hsc70 may be present. It would be useful if a

stable complex of T antigen, Hsc70 and pRb or p53 could be formed and viewed by SAXS to determine their orientation in solution. On a smaller note, performing the reverse NMR $N_{102}$/$Cul7_{CPH}$ titration would allow structural verification of the involved residues on T antigen. It would also reveal whether the Hsc70 binding site, the HPD tripeptide, is perturbed upon binding. The protein expression and purification of Cul7CPH would have to be optimized to accommodate a reverse titration. Cullin 7 also has an LXCXE sequence, which is the Rb family consensus binding motif. NMR titration experiments could be used to investigate whether Cul7 is able to bind Rb proteins. If Rb does bind, how its interaction with Cul7 changes in the presence of T antigen would be interesting. Further structure based studies coupled functional studies, involving mutating the amino acids at the binding interface and growth assays would provide more insight into the mechanisms by which T antigen transforms cells and participates in the regulation and growth of cells.

# CHAPTER IV

# DISCUSSION AND FUTURE DIRECTIONS

## Research Summation

### *The role of the Tag in cell transformation*

SV40 T antigen is a complex multi-functional protein that binds to various cellular proteins. The complexity of Tag compensates for the simplicity of the SV40 genome, which does not contain the necessary components to self-replicate. The virus depends heavily on the host cell but the genome does contain regulatory signals that direct Tag production. As the concentration of Tag grows, the protein binds to the origin region of the virus genome, which initiates DNA replication. In order for viral replication to occur, it is important that the host cell enters S phase. In S-phase, the viral genome is replicated along with the host cellular genome. Binding host proteins, such as DNA polymerase-α and inactivating tumor suppressor proteins, such as pRb, allows for the cells to enter S-phase, promoting DNA replication and viral propagation. Tag mediated inactivation of tumor suppressor proteins, supersedes the inhibition of cell growth yielding more opportunities for viral production[127]. While, the primary goal of Tag is to induce production of additional SV40 particles, cellular transformation is simply a side effect due to its inactivation of tumor suppressor proteins. However, these proteins can be inactivated under non-viral conditions, as seen in many cancers. Using the SV40 system to research this Tag-mediated inactivation mechanism may yield valuable insight.

Initial research suggested that T antigen induced cell transformation by sequestering all the pools of tumor suppressor proteins[61]. Further research has discovered that the J domain is necessary for cellular transformation. More specifically, the interaction of the J domain with host cell proteins is the crucial aspect of the transformation ability. The presence of the J domain in T antigen, classifies it as a chaperone protein. This chaperone functionality conferred by the J domain, suggests that a more dynamic mechanism for cellular transformation occurs.

Efficient cell transformation requires a functional J domain, the co-chaperone Hsc70, an intact LXCXE motif for pRb binding and a bound Cullin7. It also requires a carboxy-terminal domain that is involved in the inactivation of the p53 protein. There are conflicting reports on whether the J domain is needed in the inactivation of p53.[61] All of these proteins are involved in regulating cell growth in some manner.

The binding of pRb is the first step in the T antigen mediated inactivation. This research demonstrated that this interaction is strong as well as residue specific. While bound to pRbA/B, J domain recruits and binds Hsc70. The next step in the model is ATP hydrolysis and the separation of the Rb-E2F complex. Further research must been carried out to determine how Rb and E2F are separated. Under normal cellular conditions, the phosphorylation state of Rb changes yielding the release of E2F. As a protein recruiting and binding machine, perhaps T antigen recruits a kinase to phosphorylate Rb. Although it could simply be a conformational change that induces the release. After E2F is released, the fate of the Rb protein needs to be explored. Cullin7 contains an LXCXE motif, so perhaps Rb becomes as substrate to Cullin7 mediated poly-ubiquitination and degradation.

If Cullin7 is truly part of this mechanism, the point at which it binds to T antigen should be investigated and added to the model. It is known that the binding of Cullin7 does not disrupt bound pRb but the relationship with Hsc70 is unknown. Whether these proteins form stable or transient complex should be investigated. My research provided a SAXS profile of T antigen bound to pRbA/B. SAXS performed on a complex of the T antigen construct containing the J domain and LXCXE motif ($N_{117}$), pRbA/B and $Cul7_{CPH}$ domain could shed light on the organization of the proteins in this putative complex.

***Influence of pRb binding on T antigen dynamics***

The studies presented in Chapter II explore the dynamics of the J domain of SV40 T antigen and the effect that binding of pRb protein has upon its motional freedom. The J domain was found to be structurally and functionally independent of the other T antigen domains. The NMR HSQC spectrum of the T antigen construct containing the J domain and OBD ($N_{260}$) clearly demonstrated that these domains tumble independently in solution (Figure 2.2). The SAXS derived P(r) functions showed interatomic distances corresponding to flexibly attached domains. The NMR results revealed that upon binding of pRb, the OBD remained independent of the interaction and the J domain retained some motional freedom. The SAXS data generated Dmax and Rg values reflecting no compaction of T antigen upon pRb binding. These results support a model in which the J domain remains dynamic and therefore able to recruit additional proteins to a complex that inactivates tumor suppressor proteins. This model is supported by studies showing the J domain of T

antigen recruits Hsc70[92,128], and that binding to Cullin7 does not disrupt the interaction with pRb[124]. The dynamic nature of the J domain and T antigen in general, permits the formation of multiple protein complexes that deregulate cellular processes yielding the production of a transformed cell.

***The dynamic J domain***

SV40 mediated cell transformation requires the dynamic activity of the T antigen protein to disrupt normal cell functions. Disruption of these functions is made possible by the multiple domains of T antigen, in particular the J domain. The J domain is flexibly attached to the remainder of the protein. In the exploration of replication protein A, flexibly linked domains were deemed an important foundation for dynamic architecture as well as for the recruitment and binding of proteins and ultimately function[129,130]. These same observations can be applied to the J domain of T antigen and provide insight on its ability to transform cells.

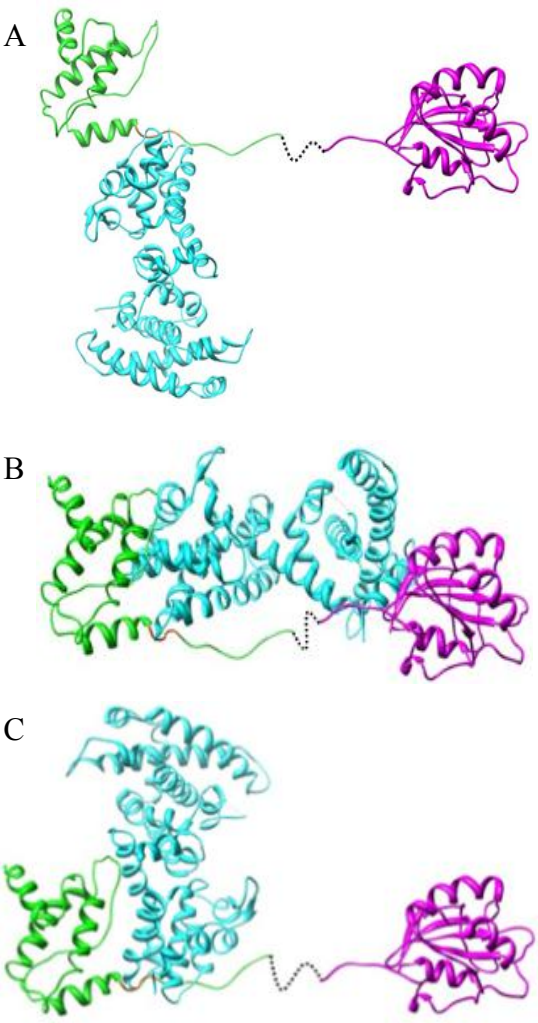



**Figure 4.1. Mode of T antigen and pRbA/B binding.** The structures of the T antigen J domain (green, 1GH6), connected with OBD (purple) bound to pRbA/B (cyan) in three conformations.

The questions posed at the beginning of this research were directed to the mode in

which pRb bound to T antigen. Previous research demonstrated that a peptide containing the LXCXE motif is sufficient to bind the pRb protein[106]. Could pRb bind in a manner that does not affect the domains of T antigen as represented in Figure 4.1A? The crystal structure of J domain and pRbA/B showed that two thirds of the interaction surface is through the LXCXE motif[68]. NMR studies described in Chapter II confirm that there are J domain residues that are affected by pRb binding. While pRb interacts with the J domain, does it bind to the OBD as well forming a compact complex as seen in Figure 4.1B? The SAXS data ruled out this mode of binding. The NMR evidence shown in Chapter II supports an interaction in which some J domain residues are involved but many remain unaffected. These unchanged residues reflect retention of conformational freedom or flexibility in the structure. Flexibility in a domain provides an opportunity for binding of additional proteins. As the mechanism of T antigen mediated transformation is researched, the necessity for a dynamic J domain becomes more apparent. An examination of how restricting the motion of the J domain affects protein binding could be performed by deleting and/or mutating the inter-helical unstructured regions or the linker between the J and OBD. A static J domain may have difficulty recruiting proteins and thus transforming cells.

***Chaperones and Cancer***

Research has shown that T antigen uses its chaperone ability to inactivate tumor suppressor proteins, transform cells and produce tumors. It is not clear that SV40 T antigen does this in human cells. However, some protein or protein complex is inactivating Rb and p53 proteins. It is possible that some non-viral chaperone is

inactivating these proteins in a manner similar to T antigen. One may believe that this human Rb inactivating protein would have been found, with all the technology and scientific advances that have occurred since the discovery of SV40. However, there are new proteins discovered every day and the connection between chaperones and cancer is gaining renewed interest. In fact, Hsc70 and Hsp75 has been shown to bind to pRb[131,132]. The Hsc70 protein, the co-chaperone of T antigen, is mutated in some breast cancers[61]. The Hsp70 protein, a chaperone that is expressed under cell stress conditions, along with its co-chaperone Hsp40, has also been shown to inhibit apoptosis[133,134]. This same protein is the co-chaperone of Hsp90. Hsp90 functions in multi-protein complexes through which it regulates the stability and activity of client proteins such as HER-2 and SRC, which are important in breast cancer progression. The overexpression of Hsp90 has been observed in a variety of cancers[135]. Recently medical researchers have begun targeting heat shock proteins as sources of therapeutic intervention in a number of diseases[136].

On the other hand, there are gaps in our knowledge of Tag mediated inactivation of pRb so there may not be a Tag-like chaperone. In the current mechanism, it is not clear how Tag releases pRb from E2F. It is known that in a normal environment, alteration of the pRb phosphorylation state releases the bound E2F. It is also known that pRb is inactivated by CDK4/6 phosphorylation[137]. Could Tag recruit a kinase to act upon pRb or mimic the change via induced structural conformation? Revealing and properly placing the remaining pieces of the Tag chaperone mediated inactivation puzzle and comparing it with known data may yield a complete answer as to how pRb is being inactivated.

***Significance***

The focus of the work presented in this dissertation highlights the influence of cellular proteins on the SV40 T antigen with specific focus on the J domain. This research contributes observations on the dynamics of T antigen and its protein interactions. These findings will contribute to fully identifying the mechanism of T antigen in protein recruitment and cellular transformations. SV40 T antigen has been extensively researched. However the independent nature of the domains, in particular the J domain and host range domain, suggests that there is still more insight to be gained by studying domain fragments and their binding partners. This research demonstrated that protein structural dynamics is influenced, whether modestly or dramatically, by the proteins to which they interact. Understanding the interactions on at higher resolutions will allow for understanding the function and activity of the multi-protein machines.

The proteins that T antigen target are critical components in cellular growth and their deregulation results in diseases like cancer. My research has expanded the knowledge of the dynamic nature of the J domain of T antigen and surveyed the effects of pRbA/B and $Cul7_{CPH}$ protein binding. Elucidating the mechanism by which T antigen inactivates proteins like pRb, is critical for understanding how a non-viral protein could perform the same function. The majority of cancers have a defect in the pRb protein, so determining the mode of pRb inactivation is fundamental to its treatment. When that non-viral, pRb inactivating protein is discovered, this may open the door to the development of new anticancer therapeutics.

## REFERENCES


1. Sweet, B. H. & Hilleman, M. R. The vacuolating virus, S.V. 40. *Proc. Soc. Exp. Biol. Med.* **105**, 420–427 (1960).
2. Shah, K. & Nathanson, N. Human exposure to SV40: review and comment. *Am. J. Epidemiol.* **103**, 1–12 (1976).
3. Newman, J. S., Baskin, G. B. & Frisque, R. J. Identification of SV40 in brain, kidney and urine of healthy and SIV-infected rhesus monkeys. *J. Neurovirol.* **4**, 394–406 (1998).
4. Eddy, B. E., Borman, G. S., Grubbs, G. E. & Young, R. D. Identification of the oncogenic substance in rhesus monkey kidney cell culture as simian virus 40. *Virology* **17**, 65–75 (1962).
5. Girardi, A. J., Sweet, B. H., Slotnick, V. B. & Hilleman, M. R. Development of tumors in hamsters inoculated in the neonatal period with vacuolating virus, SV-40. *Proc. Soc. Exp. Biol. Med.* **109**, 649–660 (1962).
6. Kirschstein, R. L. & Gerber, P. Ependymomas produced after intracerebral inoculation of SV40 into new-born hamsters. *Nature* **195**, 299–300 (1962).
7. Fraumeni, J. F., Jr, Ederer, F. & Miller, R. W. An evaluation of the carcinogenicity of simian virus 40 in man. *JAMA* **185**, 713–718 (1963).
8. Eddy, B. E., Borman, G. S., Berkeley, W. H. & Young, R. D. Tumors induced in hamsters by injection of rhesus monkey kidney cell extracts. *Proc. Soc. Exp. Biol. Med.* **107**, 191–197 (1961).
9. Diamandopoulos, G. T. Leukemia, lymphoma, and osteosarcoma induced in the Syrian golden hamster by simian virus 40. *Science* **176**, 173–175 (1972).
10. Olin, P. & Giesecke, J. Potential exposure to SV40 in polio vaccines used in Sweden during 1957: no impact on cancer incidence rates 1960 to 1993. *Dev. Biol. Stand.* **94**, 227–233 (1998).
11. Geissler, E. SV40 and human brain tumors. *Prog. Med. Virol.* **37**, 211–222 (1990).
12. Mortimer, E. A., Jr *et al.* Long-term follow-up of persons inadvertently inoculated with SV40 as neonates. *N. Engl. J. Med.* **305**, 1517–1518 (1981).
13. Strickler, H. D. *et al.* Contamination of poliovirus vaccines with simian virus 40 (1955-1963) and subsequent cancer rates. *JAMA* **279**, 292–295 (1998).
14. CDC - Concerns - Cancer, Simian Virus 40 (SV40), and Polio Vaccine Fact Sheet - Vaccine Safety. at <http://www.cdc.gov/vaccinesafety/updates/archive/polio_and_cancer_factsheet.htm>
15. Carbone, M. *et al.* Simian virus 40-like DNA sequences in human pleural mesothelioma. *Oncogene* **9**, 1781–1790 (1994).
16. Testa, J. R. *et al.* A multi-institutional study confirms the presence and expression of simian virus 40 in human malignant mesotheliomas. *Cancer Res.* **58**, 4505–4509 (1998).

17. Galateau-Salle, F. *et al.* SV40-like DNA sequences in pleural mesothelioma, bronchopulmonary carcinoma, and non-malignant pulmonary diseases. *J. Pathol.* **184**, 252–257 (1998).
18. Lednicky, J. A., Garcea, R. L., Bergsagel, D. J. & Butel, J. S. Natural simian virus 40 strains are present in human choroid plexus and ependymoma tumors. *Virology* **212**, 710–717 (1995).
19. Martini, F. *et al.* SV40 early region and large T antigen in human brain tumors, peripheral blood cells, and sperm fluids from healthy individuals. *Cancer Res.* **56**, 4820–4825 (1996).
20. Bergsagel, D. J., Finegold, M. J., Butel, J. S., Kupsky, W. J. & Garcea, R. L. DNA sequences similar to those of simian virus 40 in ependymomas and choroid plexus tumors of childhood. *N. Engl. J. Med.* **326**, 988–993 (1992).
21. Mendoza, S. M., Konishi, T. & Miller, C. W. Integration of SV40 in human osteosarcoma DNA. *Oncogene* **17**, 2457–2462 (1998).
22. Carbone, M. *et al.* SV40-like sequences in human bone tumors. *Oncogene* **13**, 527–535 (1996).
23. Lednicky, J. A., Stewart, A. R., Jenkins, J. J., 3rd, Finegold, M. J. & Butel, J. S. SV40 DNA in human osteosarcomas shows sequence variation among T-antigen genes. *Int. J. Cancer* **72**, 791–800 (1997).
24. Vilchez, R. A. *et al.* Association between simian virus 40 and non-Hodgkin lymphoma. *Lancet* **359**, 817–823 (2002).
25. Strickler, H. D. *et al.* Trends in U.S. pleural mesothelioma incidence rates following simian virus 40 contamination of early poliovirus vaccines. *J. Natl. Cancer Inst.* **95**, 38–45 (2003).
26. Engels, E. A. *et al.* Absence of simian virus 40 in human brain tumors from northern India. *Int. J. Cancer* **101**, 348–352 (2002).
27. Gordon, G. J. *et al.* Detection and quantification of SV40 large T-antigen DNA in mesothelioma tissues and cell lines. *Oncol. Rep.* **9**, 631–634 (2002).
28. Hübner, R. & Van Marck, E. Reappraisal of the strong association between simian virus 40 and human malignant mesothelioma of the pleura (Belgium). *Cancer Causes Control* **13**, 121–129 (2002).
29. de Sanjose, S. *et al.* Lack of serological evidence for an association between simian virus 40 and lymphoma. *Int. J. Cancer* **104**, 522–524 (2003).
30. Brenner, A. V. *et al.* Polio vaccination and risk of brain tumors in adults: no apparent association. *Cancer Epidemiol. Biomarkers Prev.* **12**, 177–178 (2003).
31. Liu, X., Clements, A., Zhao, K. & Marmorstein, R. Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. *J. Biol. Chem.* **281**, 578–586 (2006).
32. Imperiale, M. J. The human polyomaviruses, BKV and JCV: molecular pathogenesis of acute disease and potential role in cancer. *Virology* **267**, 1–7 (2000).
33. Berger, J. R. & Major, E. O. Progressive multifocal leukoencephalopathy. *Semin Neurol* **19**, 193–200 (1999).
34. Caldarelli-Stefano, R. *et al.* JC virus in human glial-derived tumors. *Hum. Pathol.* **31**, 394–395 (2000).
35. Frisque, R. J., Bream, G. L. & Cannella, M. T. Human polyomavirus JC virus genome. *J. Virol.* **51**, 458–469 (1984).

36. Stehle, T., Gamblin, S. J., Yan, Y. & Harrison, S. C. The structure of simian virus 40 refined at 3.1 A resolution. *Structure* **4**, 165–182 (1996).
37. Stang, E., Kartenbeck, J. & Parton, R. G. Major histocompatibility complex class I molecules mediate association of SV40 with caveolae. *Mol. Biol. Cell* **8**, 47–57 (1997).
38. Pietiäinen, V. M., Marjomäki, V., Heino, J. & Hyypiä, T. Viral entry, lipid rafts and caveosomes. *Ann. Med.* **37**, 394–403 (2005).
39. Sáenz-Robles, M. T., Sullivan, C. S. & Pipas, J. M. Transforming functions of Simian Virus 40. *Oncogene* **20**, 7899–7907 (2001).
40. C. N Cole Polyomaviridae: the viruses and their replication. *Fields Virology* 1997–2043 (1996).
41. Fanning, E. & Zhao, K. SV40 DNA replication: from the A gene to a nanomachine. *Virology* **384**, 352–359 (2009).
42. Gai, D. *et al.* Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen. *J. Biol. Chem.* **279**, 38952–38959 (2004).
43. Tevethia, M. J., Pipas, J. M., Kierstead, T. & Cole, C. Requirements for immortalization of primary mouse embryo fibroblasts probed with mutants bearing deletions in the 3' end of SV40 gene A. *Virology* **162**, 76–89 (1988).
44. Pipas, J. M. Mutations near the carboxyl terminus of the simian virus 40 large tumor antigen alter viral host range. *J. Virol.* **54**, 569–575 (1985).
45. Spence, S. L. & Pipas, J. M. Simian virus 40 large T antigen host range domain functions in virion assembly. *J. Virol.* **68**, 4227–4240 (1994).
46. Poulin, D. L. & DeCaprio, J. A. The carboxyl-terminal domain of large T antigen rescues SV40 host range activity in trans independent of acetylation. *Virology* **349**, 212–221 (2006).
47. Li, D. *et al.* Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* **423**, 512–518 (2003).
48. Valle, M., Gruss, C., Halmer, L., Carazo, J. M. & Donate, L. E. Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. *Mol. Cell. Biol.* **20**, 34–41 (2000).
49. Gomez-Lorenzo, M. G. *et al.* Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. *EMBO J* **22**, 6205–6213 (2003).
50. Luo, X., Sanford, D. G., Bullock, P. A. & Bachovchin, W. W. Solution structure of the origin DNA-binding domain of SV40 T-antigen. *Nat. Struct. Biol.* **3**, 1034–1039 (1996).
51. Weisshart, K. *et al.* Two regions of simian virus 40 T antigen determine cooperativity of double-hexamer assembly on the viral origin of DNA replication and promote hexamer interactions during bidirectional origin DNA unwinding. *J. Virol.* **73**, 2201–2211 (1999).
52. Fanning, E., Xiaorong, Z. & Xiaohua, J. Polyomavirus Life Cycle. *DNA Tumor viruses book* (2008).
53. Kim, H. Y., Ahn, B. Y. & Cho, Y. Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *EMBO J.* **20**, 295–304 (2001).

54. Gomez-Lorenzo, M. G. *et al.* Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. *EMBO J.* **22**, 6205–6213 (2003).
55. Walsh, P., Bursać, D., Law, Y. C., Cyr, D. & Lithgow, T. The J-protein family: modulating protein assembly, disassembly and translocation. *EMBO Rep.* **5**, 567–571 (2004).
56. Hennessy, F., Nicoll, W. S., Zimmermann, R., Cheetham, M. E. & Blatch, G. L. Not all J domains are created equal: implications for the specificity of Hsp40-Hsp70 interactions. *Protein Sci.* **14**, 1697–1709 (2005).
57. Kelley, W. L. & Georgopoulos, C. The T/t common exon of simian virus 40, JC, and BK polyomavirus T antigens can functionally replace the J-domain of the Escherichia coli DnaJ molecular chaperone. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3679–3684 (1997).
58. Campbell, K. S. *et al.* DnaJ/hsp40 chaperone domain of SV40 large T antigen promotes efficient viral DNA replication. *Genes Dev.* **11**, 1098–1110 (1997).
59. Srinivasan, A. *et al.* The amino-terminal transforming region of simian virus 40 large T and small t antigens functions as a J domain. *Mol. Cell. Biol.* **17**, 4761–4773 (1997).
60. Stebbins, C. E. *et al.* Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell* **89**, 239–250 (1997).
61. Sullivan, C. S. & Pipas, J. M. T antigens of simian virus 40: molecular chaperones for viral replication and tumorigenesis. *Microbiol. Mol. Biol. Rev.* **66**, 179–202 (2002).
62. Kelley, W. L. The J-domain family and the recruitment of chaperone power. *Trends Biochem. Sci.* **23**, 222–227 (1998).
63. Kelley, W. L. Molecular chaperones: How J domains turn on Hsp70s. *Curr. Biol.* **9**, R305–308 (1999).
64. Helt, A.-M. & Galloway, D. A. Mechanisms by which DNA tumor virus oncoproteins target the Rb family of pocket proteins. *Carcinogenesis* **24**, 159–169 (2003).
65. Dick, F. A. Structure-function analysis of the retinoblastoma tumor suppressor protein - is the whole a sum of its parts? *Cell Div* **2**, 26 (2007).
66. Classon, M. & Harlow, E. The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* **2**, 910–917 (2002).
67. Burke, J. R., Deshong, A. J., Pelton, J. G. & Rubin, S. M. Phosphorylation-induced conformational changes in the retinoblastoma protein inhibit E2F transactivation domain binding. *J. Biol. Chem.* **285**, 16286–16293 (2010).
68. Kim, H. Y., Ahn, B. Y. & Cho, Y. Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *EMBO J.* **20**, 295–304 (2001).
69. Lee, J. O., Russo, A. A. & Pavletich, N. P. Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature* **391**, 859–865 (1998).
70. Dimova, D. K. & Dyson, N. J. The E2F transcriptional network: old acquaintances with new faces. **24**, 2810–2826

71. Maiti, B. *et al.* Cloning and characterization of mouse E2F8, a novel mammalian E2F family member capable of blocking cellular proliferation. *J. Biol. Chem.* **280**, 18211–18220 (2005).
72. Sozzani, R. *et al.* Interplay between Arabidopsis Activating Factors E2Fb and E2Fa in Cell Cycle Progression and Development. *Plant Physiol* **140**, 1355–1366 (2006).
73. Chen, H.-Z., Tsai, S.-Y. & Leone, G. Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat. Rev. Cancer* **9**, 785–797 (2009).
74. Hernando, E. *et al.* Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control. *Nature* **430**, 797–802 (2004).
75. Wüthrich, K. Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* **265**, 22059–22062 (1990).
76. Evans, J. N. . *Biomolecular NMR Spectroscopy*. (Oxford University Press: 1995).
77. Rule, G. & Hitchens, T. K. *Fundamentals of NRM Spectroscopy*. (Springer: The Netherlands, 2006).
78. Teng, Q. *Handbook of Structural Biology: Practical NMR Applications*. (Academic/Plenum Publishers: 2004).
79. Wüthrich, K. *NMR of Proteins and Nucleic Acids*. (Wiley-Interscience: 1986).
80. Cavanagh, J., Fairbrother, W. ., Palmer, A. . & Skelton, H. . *Protein NMR Spectroscopy Principles and Practice.* (Academic Press: 2006).
81. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12366–12371 (1997).
82. TROSY_large_protein_training_Feb2704.pdf. at <http://www.nmr.sinica.edu.tw/Cours/Course20040227/TROSY_large_protein_training_Feb2704.pdf>
83. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
84. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
85. Rambo, R. P. & Tainer, J. A. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr. Opin. Struct. Biol.* **20**, 128–137 (2010).
86. Pretto, D. I. *et al.* Structural dynamics and single-stranded DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering. *Biochemistry* **49**, 2880–2889 (2010).
87. Tsutakawa, S. E., Hura, G. L., Frankel, K. A., Cooper, P. K. & Tainer, J. A. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. *J. Struct. Biol.* **158**, 214–223 (2007).
88. Bernadó, P. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. Biophys. J.* **39**, 769–780 (2010).

89. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–571 (2011).
90. Debye, P., Anderson, H. R. & Brumberger, H. Scattering by an Inhomogeneous Solid. II. The Correlation Function and Its Application. *Journal of Applied Physics* **28**, 679 (1957).
91. Pipas, J. M. SV40: Cell transformation and tumorigenesis. *Virology* **384**, 294–303 (2009).
92. Sullivan, C. S., Cantalupo, P. & Pipas, J. M. The molecular chaperone activity of simian virus 40 large T antigen is required to disrupt Rb-E2F family complexes by an ATP-dependent mechanism. *Mol. Cell. Biol.* **20**, 6233–6243 (2000).
93. Giacinti, C. & Giordano, A. RB and cell cycle progression. *Oncogene* **25**, 5220–5227 (2006).
94. Hiebert, S. W., Chellappan, S. P., Horowitz, J. M. & Nevins, J. R. The interaction of RB with E2F coincides with an inhibition of the transcriptional activity of E2F. *Genes Dev.* **6**, 177–185 (1992).
95. Lundberg, A. S. & Weinberg, R. A. Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes. *Mol. Cell. Biol.* **18**, 753–761 (1998).
96. Hernando, E. *et al.* Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control. *Nature* **430**, 797–802 (2004).
97. Brodsky, J. L. & Pipas, J. M. Polyomavirus T antigens: molecular chaperones for multiprotein complexes. *J. Virol.* **72**, 5329–5334 (1998).
98. Rushton, J. J., Jiang, D., Srinivasan, A., Pipas, J. M. & Robbins, P. D. Simian virus 40 T antigen can regulate p53-mediated transcription independent of binding p53. *J. Virol.* **71**, 5620–5623 (1997).
99. Zalvide, J., Stubdal, H. & DeCaprio, J. A. The J domain of simian virus 40 large T antigen is required to functionally inactivate RB family proteins. *Mol. Cell. Biol.* **18**, 1408–1415 (1998).
100. TopSpin. at <http://www.bruker-biospin.com/topspin-overview.html>
101. T.D. Goddard & D.G. Kneller *SPARKY 3*. (University of California, San Francisco., 2006).
102. Hura, G. L. *et al.* Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612 (2009).
103. A. Guinier & G. Fournet *Small angle scattering of X-rays*. (Wiley-Interscience: New York, 1955).
104. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–544 (2010).
105. Porod, G. Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen I. *Teil. Kolloid Zeitschrift* **124**, 83–115 (1951).
106. Singh, M., Krajewski, M., Mikolajka, A. & Holak, T. A. Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences. *J. Biol. Chem.* **280**, 37868–37876 (2005).
107. Dong, W. L., Caldeira, S., Sehr, P., Pawlita, M. & Tommasino, M. Determination of the binding affinity of different human papillomavirus E7 proteins for the tumour suppressor pRb by a plate-binding assay. *J. Virol. Methods* **98**, 91–98 (2001).

108. Lee, C. & Cho, Y. Interactions of SV40 large T antigen and other viral proteins with retinoblastoma tumour suppressor. *Rev. Med. Virol.* **12**, 81–92 (2002).
109. Balog, E. R. M., Burke, J. R., Hura, G. L. & Rubin, S. M. Crystal structure of the unliganded retinoblastoma protein pocket domain. *Proteins* **79**, 2010–2014 (2011).
110. Garimella, R. *et al.* Hsc70 contacts helix III of the J domain from polyomavirus T antigens: addressing a dilemma in the chaperone hypothesis of how they release E2F from pRb. *Biochemistry* **45**, 6917–6929 (2006).
111. Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **6**, 9–20 (2005).
112. Bosu, D. R. & Kipreos, E. T. Cullin-RING ubiquitin ligases: global regulation and activation cycles. *Cell Div* **3**, 7 (2008).
113. Ali, S. H., Kasper, J. S., Arai, T. & DeCaprio, J. A. Cul7/p185/p193 binding to simian virus 40 large T antigen has a role in cellular transformation. *J. Virol.* **78**, 2749–2757 (2004).
114. Kohrman, D. C. & Imperiale, M. J. Simian virus 40 large T antigen stably complexes with a 185-kilodalton host protein. *J. Virol.* **66**, 1752–1760 (1992).
115. Tsai, S. C. *et al.* Simian virus 40 large T antigen binds a novel Bcl-2 homology domain 3-containing proapoptosis protein in the cytoplasm. *J. Biol. Chem.* **275**, 3239–3246 (2000).
116. Dias, D. C., Dolios, G., Wang, R. & Pan, Z.-Q. CUL7: A DOC domain-containing cullin selectively binds Skp1.Fbx29 to form an SCF-like complex. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16601–16606 (2002).
117. Kasper, J. S., Kuwabara, H., Arai, T., Ali, S. H. & DeCaprio, J. A. Simian virus 40 large T antigen's association with the CUL7 SCF complex contributes to cellular transformation. *J. Virol.* **79**, 11685–11692 (2005).
118. Sarikas, A., Xu, X., Field, L. J. & Pan, Z.-Q. The cullin7 E3 ubiquitin ligase: a novel player in growth control. *Cell Cycle* **7**, 3154–3161 (2008).
119. Huber, C. *et al.* Identification of mutations in CUL7 in 3-M syndrome. *Nat. Genet.* **37**, 1119–1124 (2005).
120. Jung, P. *et al.* Induction of cullin 7 by DNA damage attenuates p53 function. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11388–11393 (2007).
121. Andrews, P., He, Y. J. & Xiong, Y. Cytoplasmic localized ubiquitin ligase cullin 7 binds to p53 and promotes cell growth by antagonizing p53 function. *Oncogene* **25**, 4534–4548 (2006).
122. Kim, S. S. *et al.* CUL7 is a novel antiapoptotic oncogene. *Cancer Res.* **67**, 9616–9622 (2007).
123. Kaustov, L. *et al.* The conserved CPH domains of Cul7 and PARC are protein-protein interaction modules that bind the tetramerization domain of p53. *J. Biol. Chem.* **282**, 11300–11307 (2007).
124. Kasper, J. S., Arai, T. & DeCaprio, J. A. A novel p53-binding domain in CUL7. *Biochem. Biophys. Res. Commun.* **348**, 132–138 (2006).
125. Takeuchi, K. & Wagner, G. NMR studies of protein interactions. *Curr. Opin. Struct. Biol.* **16**, 109–117 (2006).
126. Chitayat, S., Kanelis, V., Koschinsky, M. L. & Smith, S. P. Nuclear magnetic resonance (NMR) solution structure, dynamics, and binding properties of the kringle IV type 8 module of apolipoprotein(a). *Biochemistry* **46**, 1732–1742 (2007).

127. Ali, S. H. & DeCaprio, J. A. Cellular transformation by SV40 large T antigen: interaction with host proteins. *Semin. Cancer Biol.* **11**, 15–23 (2001).
128. Sullivan, C. S., Gilbert, S. P. & Pipas, J. M. ATP-dependent simian virus 40 T-antigen-Hsc70 complex formation. *J. Virol.* **75**, 1601–1610 (2001).
129. Pretto, D. I. *et al.* Structural dynamics and single-stranded DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering. *Biochemistry* **49**, 2880–2889 (2010).
130. Brosey, C. A. *et al.* NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA. *J. Am. Chem. Soc.* **131**, 6346–6347 (2009).
131. Inoue, A. *et al.* 70-kDa heat shock cognate protein interacts directly with the N-terminal region of the retinoblastoma gene product pRb. Identification of a novel region of pRb-mediating protein interaction. *J. Biol. Chem.* **270**, 22571–22576 (1995).
132. Chen, C. F. *et al.* A new member of the hsp90 family of molecular chaperones interacts with the retinoblastoma protein during mitosis and after heat shock. *Mol. Cell. Biol.* **16**, 4691–4699 (1996).
133. Nylandsted, J. *et al.* Selective depletion of heat shock protein 70 (Hsp70) activates a tumor-specific death program that is independent of caspases and bypasses Bcl-2. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7871–7876 (2000).
134. Seo, J. S. *et al.* T cell lymphoma in transgenic mice expressing the human Hsp70 gene. *Biochem. Biophys. Res. Commun.* **218**, 582–587 (1996).
135. Kim, L. S. & Kim, J. H. Heat Shock Protein as Molecular Targets for Breast Cancer Therapeutics. *J Breast Cancer* **14**, 167–174 (2011).
136. Rousaki, A. *et al.* Allosteric drugs: the interaction of antitumor compound MKT-077 with human Hsp70 chaperones. *J. Mol. Biol.* **411**, 614–632 (2011).
137. Sherr, C. J. & McCormick, F. The RB and p53 pathways in cancer. *Cancer Cell* **2**, 103–112 (2002).