BIOCHEMICAL AND STRUCTURAL ANALYSIS OF THE p58C AND p68N

DOMAINS OF DNA POLYMERASE ALPHA/PRIMASE

By

Brian Edward Weiner

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

August, 2008

Nashville, Tennessee

Approved:

Walter J. Chazin, Ph.D.

Richard N. Armstrong, Ph.D.

David Cortez, Ph.D.

Ellen Fanning, Dr.rer.nat.

Brandt F. Eichman, Ph.D.

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Dr. Walter Chazin. He is a great role model for me and provides an excellent example of how to run an effective research group. His support over my graduate career was tremendously valuable.

I thank Dr. Ellen Fanning for her support as well. The work reported in this dissertation was conducted in close collaboration with her research group. In particular, I thank her student, Hao Huang, who conducted the bulk of the functional studies that will be reported.

I thank the other members of my dissertation committee, Drs. Richard Armstrong, David Cortez, and Brandt Eichman for their input into my dissertation project over the past several years.

I also thank all past and present members of the Chazin research group for providing such a great working environment. I especially thank Susan Meyn for all her hard work in maintaining the lab and for her great attitude she brought to work everyday.

I thank my family and friends for their support, especially my wife Ashley.

Finally, I am grateful to the NIH and the Vanderbilt MSTP for providing funding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAA+   ATPases associated with a variety of cellular activities

ATP   adenosine triphosphate

BME   2-mercaptoethanol

CD   circular dichroism

Cdc   cell division cycle protein

Cdk   cyclin dependent kinase

Cdt1   chromatin licensing and DNA replication factor 1

COSY   correlation spectroscopy

CSPA   chemical shift perturbation assay

Ddk   Dbf-4 dependent kinase

DLB   differential line broadening

dNTP   deoxyribonucleotide

DTT   dithiothreitol

EDTA   ethylenediaminotetraacetic acid

EndoIII   endonuclease III

EPR   electron paramagnetic resonance

FancJ   Fanconi anemia complentation group D protein

GINS   go ichi ni san (5, 1, 2, 3)

GST   glutathione-S-transferase

HD   helicase domain

HiPIP   high-potential iron protein

HR	host-range
Hsc70	heat shock protein cognate 70
HSQC	heteronuclear single quantum coherence
ICP-MS	inductively coupled plasma-mass spectrometry
IPTG	isopropyl thio-beta-D-galactopyranoside
IRP	iron regulatory protein
ITC	isothermal titration calorimetry
k_d	dissociation constant
LC	liquid chromatography
MALDI-TOF	matrix assisted laser desorption ionization – time of flight
Mcm	minichromosome maintenance protein
MS	mass spectrometry
Ni-NTA	nickel-nitrilotriacetic acid
NOESY	nuclear Overhauser effect spectroscopy
OD	optical density
OBD	origin binding domain
OB	oligonucleotide/oligosaccharide binding
ORC	origin recognition complex
p58C	C-terminal domain of p58
p68N	the folded domain of the N-terminal region of p68
PCNA	proliferating cell nuclear antigen
PDB	protein database
Pfu	*Pyrococcus furiosus*

Pho   *Pyrococcus horikoshii*

pol   DNA polymerase

pol-prim   DNA polymerase α/primase

pre-RC   pre-replication complex

PVDF   polyvinylidene fluoride

Rb   retinoblastoma protein

RFC   replication factor C

rMD   restrained molecular dynamics

RMSD   root mean square deviation

RPA   replication protein A

SDS-PAGE   sodium dodecyl sulfate polyacrylamide gel electrophoresis

Sld   synthetically lethal with Dpb11-1

ssDNA   single-stranded deoxyribonucleic acid

Sso   *Sulfolobus solfataricus*

SV40   simian virus 40

Tag   large T antigen

TOCSY   total correlation spectroscopy

TopBP1   topoisomerase II binding protein 1

topo I   topoisomerase I

uv-vis   ultraviolet-visible

VP   viral protein

XPD   xeroderma pigmentosum complentation group D protein

CHAPTER I

INTRODUCTION

DNA Replication Overview

DNA replication is a process that is fundamental to all organisms. The genome must be replicated accurately to ensure that all the genetic material is passed from cell to cell and from parent to offspring. Complexes of several proteins acting in concert are responsible for replicating DNA, and while the endpoint is similar, the mechanism of DNA replication can vary from organism to organism. This dissertation will focus on eukaryotic DNA replication and how a viral model system can be used to facilitate our understanding of this complex process.

There are three major phases of DNA replication: initiation, elongation, and termination. Initiation is characterized by the generation of replication forks at origins of replication. Origins are distinct locations of the genome where protein machinery assembles and begins the process of replicating the genome. The double-stranded DNA (dsDNA) is initially melted and then unwound via the helicase activity of the protein machinery to create the replication bubble, which contains two replication forks. This creates two strands of template single-stranded DNA (ssDNA). The 3' -> 5' strand is known as the leading strand template and the 5' -> 3' strand is termed the lagging strand template. DNA polymerases cannot polymerize deoxyribonucleotides (dNTPs) *de novo*, so a short RNA primer is generated on the leading and lagging strands.

The main components of the elongation phase are the activity of the helicase and DNA polymerases. DNA polymerases create DNA in the 5' -> 3' direction using the template DNA as a guide. Due to the polarity of the lagging strand, the polymerase will have to be repeatedly recruited to the replication fork to generate stretches of DNA termed Okazaki fragments. DNA replication is terminated when a DNA polymerase runs into a region where the DNA has already been replicated. The RNA primers are removed, the resultant gaps are filled in, and the remaining nicks are repaired.

Defects in DNA replication can have dire consequences at both the cellular and organismal level. Incorporation of incorrect bases into the daughter strand may lead to mutations in proteins that interfere with normal cellular processes. Mutations that inactivate tumor suppressors, for example, can lead to tumor formation. Another potential problem during DNA replication is fork stalling due to DNA lesions. In healthy cells, this triggers cell cycle arrest, which allows for the cell to repair the damage or to destroy the cell via apoptosis. The genome must also be replicated faithfully during embryogenesis to ensure that the embryo develops healthily. As such, mutations that cause severe defects in DNA replication do not support life. Given the central role of DNA replication in numerous processes ranging from development to tumorogenesis, understanding how this process is carried out in human cells is an important step in combating disease.

Eukaryotic DNA Replication

DNA replication in eukaryotes is a complex process involving the coordination of many proteins. This process is tightly controlled, allowing for one round of DNA

replication per cell cycle. While origins in lower eukaryotes are well studied, human origins of replication are much more complex (1, 2). This is largely due to the absence of consistent DNA sequences from one origin to the next. The first component to localize to the origin is the origin recognition complex (ORC) in the G1-phase of the cell cycle (3, 4). ORC is a six subunit protein hypothesized to serve as a protein recruiter and scaffold. How ORC identifies origins is not clear. While origins are discrete segments within the chromosomes, ORC binds dsDNA with no apparent sequence specificity (1). Thus other proteins, perhaps those already associated with chromatin, may help guide ORC to origins (5).

A simplified schematic for initiation of eukaryotic DNA replication is shown in Figure 1.1. Once ORC is bound to the origin, it serves as a scaffold for the recruitment of additional proteins required to form the pre-replication (pre-RC) complex. Cell division cycle 6 protein (Cdc6) and chromatin licensing and DNA replication factor 1 (Cdt1) are recruited by ORC to origins (reviewed in (3, 4)). These two proteins are required for the association of minichromisome maintenance 2-7 proteins (Mcm2-7) (6, 7). Cdc6 and three of the subunits of ORC are ATPases, and ATP hydrolysis results in active loading of Mcm2-7 onto origin DNA, whereby the 6 subunits of Mcm2-7 are thought to encircle the dsDNA (4, 8). Several copies of Mcm2-7 are loaded onto each origin via this cycle of ATP hydrolysis (9).
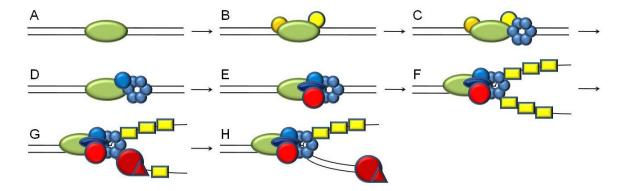
Figure 1.1. Initiation of eukaryotic DNA replication. (A) ORC binds the origin. (B) Cdc6 (orange) and Cdt1 (yellow) are recruited. (C) Mcm2-7 are loaded, licensing the DNA for replication. (D) Mcm10 is loaded as the cell enters S-phase and the pre-RC is phosphorylated by Cdk and Ddk, followed by (E) Cdc45 and GINS. (F) Helicase activity is stimulated by Sld2, Sld3, and Dpb11 allowing RPA (yellow) to bind. (G) Pol-prim is loaded and (H) lays down an RNA-DNA primer.

The pre-RC is a key regulation step in DNA replication. In order to ensure that the genome is not replicated continuously, origins are regulated so that they fire only once per cell cycle via a process termed DNA replication licensing with the key step being the loading of Mcm2-7 (10). Since DNA replication occurs only during S-phase, it is not surprising that licensing is regulated by regulators of the cell cycle. The pre-RC is phosphorylated by cyclin E-cyclin dependent kinase (Cdk) 2, cyclin A-Cdk2, and the DBf-4 dependent kinase (Ddk) which allows DNA replication to continue (8). The availability of Cdt1 is also a key regulatory checkpoint in metazoans. Geminin inhibits Cdt1 via a direct physical interaction. Geminin levels are highly regulated and it is most highly expressed in late S-, G2-, and M-phase (8). This ensures that Cdt1 is only free to associate with the pre-RC during G1-phase.

As the cell cycle progresses into S-phase, Mcm10 is loaded onto the origin, and serves as a prerequisite for binding of Cdc45 and GINS (11-13). Recently, Cdc45 and GINS have been found to associate with Mcm2-7 to create the replicative helicase

complex (14, 15). Sld2 and Sld3 (synthetically lethal with Dpb11-1) are phosphorylated by Cdk and Ddk, and together with Dpb11 (TopBP1 in humans), they promote origin unwinding by the helicase complex (16-18). The Mcm2-7 proteins belong to the family of AAA+ helicases that will be discussed in more detail later. The helicase utilizes cycles of ATP hydrolysis to continually unwind dsDNA. Melting the origin DNA creates two ssDNA strands that are then bound by replication protein A (RPA) (19). This serves to prevent reannealing of the strands and as protection against nuclease activity.

DNA polymerase α/primase (pol-prim) is then recruited to the newly created replication fork (12) after the addition of DNA polymerase ε (20). DNA polymerases are unable to synthesize DNA *de novo*. Two additional components are required: a template strand, two of which are created when the origin is melted, and a primer, which is created via the primase activity of pol-prim. Pol-prim creates a chimeric RNA-DNA primer of ~30 nt (21). Pol-prim has low processivity and no error-proofing capabilities, making it a poor candidate for the major replicative polymerase. The bulk of DNA is synthesized by DNA polymerase δ (lagging stand) and DNA polymerase ε (leading strand) (22). These polymerases are loaded via a process termed "pol switching". The clamp loader, replication factor C (RFC), loads PCNA, a clamp-like protein that associates with pol δ and ε as they synthesize DNA (22). The loading of pol δ marks the end of the initiation step of DNA replication, and the beginning of the elongation step. It should be noted however, that due to the nature of the lagging strand, the process of primer formation followed by pol switching must be subsequently completed many times.

The complexity of initiation of human DNA replication presents a formidable challenge in understanding this process in detail. While the order in which proteins

associate with the origin has become clearer over the last few years, it differs between species and the molecular details governing these transitions have remained elusive. In order to address the question of how the initiation proteins come together to form a molecular machine, we have decided to study the proteins of SV40 DNA replication. This simplified *in vitro* system serves as a model for eukaryotic DNA replication, as three of the proteins involved in initiation (and 10 out of 11 for complete *in vitro* replication) are required for both processes. Insights into SV40 DNA replication can thus shed light on eukaryotic DNA replication.

Simian Virus 40 DNA Replication

Simian virus 40 (SV40) is a polyomavirus containing a circular dsDNA genome. It was originally discovered as a contaminant in the polio vaccine in the late 1950's. The virus persists as a dormant infection in its natural host, the monkey, *Rhesus macaque*, but has been shown to be tumorigenic in immunocompromised rodent hosts (23). Despite some initial speculation, SV40 has not been proven to be tumorigenic in human hosts (24, 25). Subsequent study of SV40 revealed an interesting DNA replication mechanism, which allows the virus to hijack the host replication machinery.

The SV40 genome is small (~5200 bp), and encodes nine viral proteins (23, 26). Four of these proteins, VP1, VP2, VP3, and VP4 are structural components of the virion. Agnoprotein is involved in both the SV40 life cycle and in host cellular response (27). Small and large T antigen are involved in the induction of cellular transformation (23). Interestingly, large T antigen (Tag) is also a key component of the SV40 DNA replication machinery. Not much is known regarding the two other encoded proteins, 17K T antigen

and small leader protein, other than they are transcribed early, along with Tag (23). Tag was initially shown to support DNA replication of DNA containing an SV40 origin when incubated with a monkey cell extract (28). Subsequent studies identified the human protein components of SV40 DNA replication initiation: topoisomerase I (topo I), RPA, and pol-prim (29). These three proteins, together with the viral Tag, were sufficient to initiate SV40 DNA replication *in vitro*. Further experiments have revealed much about these proteins, and how they function together to initiate DNA replication.

SV40 DNA Replication: Initiation Proteins

SV40 Large T Antigen (Tag)

Tag is the major early gene product encoded by the SV40 chromosome (23). It plays critical roles in both cellular transformation and viral DNA replication. It is composed of four functional domains: an N-terminal J-domain (a.a. 1-102), an origin binding domain (OBD) (a.a. 131-259), a helicase domain (a.a. 260-627), and a C-terminal host-range (HR) domain (a.a. 628-708). Structures have been determined for the J-domain (30), OBD (31), and helicase domain (32). The HR domain was originally predicted to be unstructured, however limited proteolysis experiments indicated the possibility of some structure in that region (33, 34).
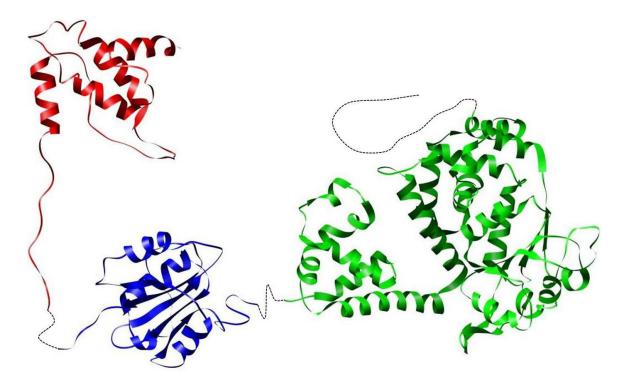
Figure 1.2. Tag structure. Tag is composed of three independent domains per subunit: the J-domain (red, PDB code 1GHG (30)), the OBD (blue, PDB code 1TBD (31)) and the helicase domain (green, PDB code 1N25 (32)). Structures for each domain were determined independently. Figure layout adapted from (34).

The J-domain and HR domain are not directly involved in SV40 DNA replication. The J-domain is primarily involved in cellular transformation. It interacts with Hsc70 to activate the E2F transcription factor which promotes cellular proliferation (35). Prior to S-phase, E2F is inactivated via a physical interaction with the tumor suppressor, Rb. Tag physically associates with Rb via an LXCXE motif, just C-terminal to the structured region of the J-domain (30). ATP hydrolysis of the J-domain-bound Hsc70 causes a release of E2F from the RB/Tag complex, thus activating it and promoting cell cycle progression (35). The HR domain is involved in host-range determination as well as in assembly of viral particles (36). It is dispensable for cellular transformation, however (37).

As their names imply, the OBD and helicase domain play crucial roles in SV40 DNA replication. The OBD is a small globular domain that is involved in origin recognition via binding the pentanucleotide GAGGC DNA sequence (31). Both the OBD and the helicase domains are involved in critical protein-protein interactions that will be discussed further in this introduction.

The helicase domain is composed of three subdomains: D1, a Zn-binding domain; D2, a AAA+ domain; and D3, an α-helical, globular domain (32). Tag is a hexamer in solution, and x-ray crystal structures have been determined for various nucleotide binding modes (38). The Zn domain is globular, and does not have a traditional Zn-finger motif that is often involved in protein-DNA interactions. Thus the Zn domain is most likely not involved in DNA binding, and indeed, it has been shown to be required for Tag hexamerization (32). The AAA+ domain is also involved in hexamerization. AAA+ proteins share common sequence and structural properties, such as: a Walker A motif, involved in ATP binding; a Walker B motif, involved ATP hydrolysis; and the motif C, which detects ATP/ADP in the nucleotide-binding pocket (39). All of these motifs are found in D2.

Tag assembles at origins as a double hexamer, with the origin binding domains on the interior, and the helicase domains on the exterior (40-42). Hydrolysis of ATP results in a large conformational change in Tag, allowing it to unwind dsDNA. As ATP is hydrolyzed, D2 swings toward D1, using the fifth α-helix as a hinge (38). This results in movement of the β hairpins in the central channel, which allows for the translocation along dsDNA.

Topoisomerase I

Topoisomerases are enzymes that relieve torsional stress created during DNA processing (reviewed in (43)). As dsDNA is unwound, for example during initiation of DNA replication, DNA downstream of the helicase becomes supercoiled, creating superhelical tension. Topoisomerases relieve this tension via one of two mechanisms. Type I topoisomerases, which as the name implies, includes the human Topo I, relax supercoiled DNA by cleaving one strand, and passing the intact strand through the gap. The gap is then religated and the topoisomerase dissociates. Type II topoisomerases cleave both strands of dsDNA, and pass the dsDNA through this gap. The double strand break is then repaired.

Topoisomerase activity is essential for human and SV40 DNA replication. Topo I contains 765 amino acids, and is a monomer composed of four domains. The N-terminal domain (a.a. 1-214) is poorly conserved and dispensable for Topo I activity *in vitro* (44). It is thought to be involved in subcellular localization (45). The core domain (a.a. 215-635), linker domain (a.a. 636-712), and C-terminal domain (a.a. 713-765) make up the rest of the protein, for which a crystal structure has been determined (46). The core and C-terminal domains contain the active site residues including Tyr723, which is the residue that creates an intermediate phosphotyrosine bond with the cleaved DNA strand (43).

Topo I clamps around dsDNA with two lobes on either side of the DNA duplex (43). The upper lobe which contains core subdomains I and II is also termed the cap. The lower lobe is composed of core subdomain III and the C-terminal domain. The tight

clamp structure of DNA-bound Topo I allows for efficient enzymatic activity and helps to ensure that the enzyme does not prematurely dissociate leaving nicked dsDNA.

Replication Protein A

RPA is the major eukaryotic ssDNA binding protein (19). It is essential for many DNA processing events including replication, recombination and repair (47). RPA binds ssDNA with high affinity and also interacts with the relevant DNA processing machinery.

RPA is composed of three subunits: RPA70, RPA32, and RPA14, named according to their molecular weights. Similar to Tag, isolated domain structures have been determined for each RPA domain (Figure 1.3). RPA has six oligonucleotide/oligosaccharide (OB-fold) domains, and a winged-helix-turn-helix RPA32C domain. OB-folds are characterized by a pair of three-stranded antiparallel β sheets, with an α helix between the third and fourth strands (48). The OB-fold domains in RPA are RPA70N, RPA70A, RPA70B, RPA70C, RPA32D, and RPA14.

RPA70A, RPA70B, RPA70C, and RPA32D bind ssDNA and orient RPA with a 5' -> 3' polarity with respect to the DNA (49). The polarity of binding to DNA arises because RPA70A binds ssDNA with the highest affinity, RPA70B with the second highest affinity, RPA70C with the third highest, and RPA32D binds the weakest. RPA has been shown to interact with DNA using three distinct binding modes. In the compact mode, RPA70A and RPA70B bind 8-10 nt. RPA70C can then bind ssDNA to cover ~20 nt in total. In the fully extended binding mode, all four OB-folds bind ssDNA covering 28-30 nt (47).

OB-folds are also capable of interacting with proteins. In RPA, RPA70N, RPA70A, and RPA70B, have been shown to be involved in protein-protein interactions (50-53). Additionally, RPA32C is a major protein-protein interaction domain, and has been shown to be directly involved in several interactions (54, 55).
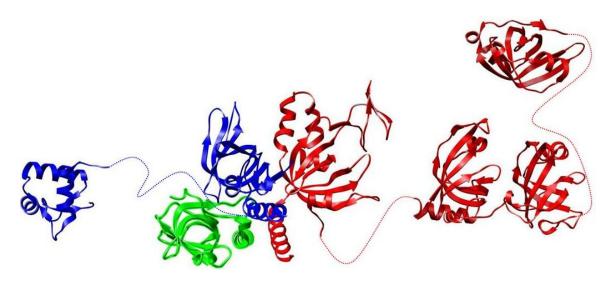
Figure 1.3. Structure of RPA. RPA70 (red) contains four OB-fold domains: 70N (PDB 2B3G, (50)), 70A, 70B (PDB 1JMC contains RPA70AB, (56)), and 70C. RPA32 (blue) contains one OB-fold domain, 32D, and a winged-helix-turn-helix domain, 32C (PDB code 1DPU, (54)). RPA14 (green) contains a single OB-fold (PDB 1L1O contains the trimerization core, which consists of RPA70C, RPA32D, and RPA14, (57)). Figure layout adapted from (58).

DNA Polymerase α/Primase

Pol-prim is the only DNA polymerase capable of synthesizing DNA *de novo*, due to its associated primase activity (21). Thus pol-prim serves as the initiator polymerase during DNA replication. Pol-prim is composed of four subunits, named according to their molecular weights: p180, p68, p58, and p48. Figure 1.4 contains a schematic figure of the subunit organization. Unlike the other three SV40 DNA replication proteins, there are no available structures for any pol-prim domain. Even the domain architecture itself

is unclear. Several functional regions of p180 and p68 have been identified, but whether these form structured domains remains to be determined (59, 60). Due to their discrete functions, the subunits will be discussed in more detail separately.
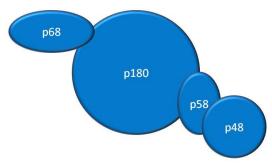
Figure 1.4. Subunit organization of pol-prim.

p180 - The p180 subunit contains the DNA polymerase α (pol α) activity. Studies of truncation mutants of the mouse pol α identified three functional domains: an N-terminal domain (a.a. 1-329), a core domain (a.a. 330-1279), and a subunit assembly domain (a.a. 1235-1465) (60).   The core domain is necessary and sufficient for polymerase activity.   The subunit assembly domain interacts with the p68 and p58 subunits.

Pol α, along with pol ε and pol δ, belongs to the member of family B polymerases, based on sequence and structural similarities (21).   The structure of the family B polymerase, the gp43 protein from the bacteriophage RB69, has been determined, and it offers insights into how the pol α core domain might be structured (61). Family B polymerases fold into a structure that resembles a right hand, containing a palm, thumb, and fingers (21). The palm domain contains two active site aspartic acid residues that coordinate two magnesium ions involved in catalysis via the two-metal ion mechanism

13

common among polymerases (including primases). The fingers orient the template and dNTPs, and the thumb is involved in stimulating processivity via DNA-binding. Multiple sequence alignments comparing pol α with other family B polymerases have identified six highly conserved regions that are involved in structural integrity as well as catalysis (61, 62).

p68 - The second largest subunit of pol-prim is an accessory subunit containing no known enzymatic activity. p68, also referred to as the B subunit, is required for initiation and elongation in SV40 DNA replication (63). p68 physically interacts with Tag (59) during SV40 DNA replication, and with the eukaryotic DNA replication proteins, ORC and Cdc45 (64, 65). Thus p68 is thought to play a role in recruiting pol-prim to replication forks. p68 has also been shown to be required for pol-prim import into the nucleus (60, 66).

Another important property of p68 is that it is phosphorylated in a cell cycle dependent manner. Both cyclin A-Cdk2 and cyclin E-Cdk2 are capable of phosphorylating several residues between residues 141 and 160. This serves as another regulatory step in eukaryotic DNA replication, as the phosphorylation state of p68 can influence the both the catalytic activity of p180 (67) and its protein-protein interactions (68).

p48 - p48 and p58 constitute the DNA primase portion of pol-prim. The active site residues are found in p48. Based on multiple sequence alignments and alanine scanning experiments, Asp109, Asp111, and Asp306 have been proposed as the metal-binding residues in the active site (69, 70). p48 shares sequence homology with members

of the family X polymerases, such as DNA polymerase β, and is thought to catalyze primer formation using the two-metal ion mechanism.

The initial step in primer formation is the rate-limiting dinucleotide formation (71). Primase exhibits some sequence specificity at this stage, as only purines may be incorporated as the 5' nucleotide (72). Primase is capable of dissociating at this stage *in vitro* (71), but likely remains bound to the template *in vivo*. Next, primase extends the dinucleotide to create a unit length primer of 7-10 ribonucleotides. Finally, the primer-template complex is then transferred to the p180 active site for addition of dNTPs to the primer (73).

One of the roadblocks limiting our insight in eukaryotic DNA primase function is that eukaryotic and bacterial primases are structurally and functionally different (74). Bacterial DnaG-type primase are monomers and do not associate with a pol-prim complex. Additionally, while pol-prim creates a chimeric RNA-DNA primer prior to pol switching, DnaG-type primases only synthesize short RNA primers. Thus despite the structural information available for DnaG-type primases, they are not very informative in elucidating eukaryotic DNA primase function. However, eukaryotic and archaeal primases do share sequence, and presumably structural, similarities.

Structures of primases from *Pyrococcus furiosis* (75), *Pyrococcus horikoshii* (76, 77), and *Sulfolobus solfataricus* (78) have been determined (Figure 1.5). The structures of these proteins are actually quite different from DNA polymerase structures as they are lacking the familiar hand-like structure with the palm, thumb, and fingers (74). The archaeal primases contain two closely associated globular domains. The larger domain contains the putative active site residues, and the smaller domain is variable between the

organisms. Acidic residues that align with Asp109 and Asp111 are part of an active site that is actually quite similar to the active site of the pol X family polymerases. Thus despite the overall difference in fold, the mechanism of catalysis may be quite similar.
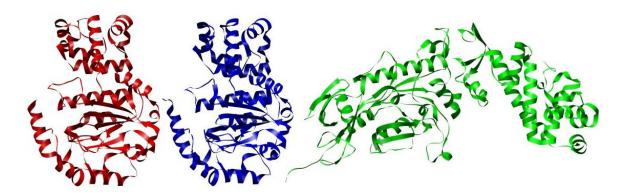
Figure 1.5. Archaeal primase structures. p48 homologs from *P. furiosis* (red, PDB code 1G71 (75)) and *P. horikoshii* (blue, PDB code 1v33 (76)) are extremely similar due to high sequence identity. The primase dimer lacking the C-terminal portion of the p58 homolog from *S. solfataricus* is shown in green (p48 on left, and p58 on right, PDB code 1ZT2 (78)).

p58 - Like p68, p58 is not known to possess any enzymatic activity. However, p58 stimulates p48 activity *in vitro* (79). p58 has also been shown to be important for initiation, elongation, and counting (creating a unit-length primer of 7-10 ribonucleotides) (80). Additionally, p58 interacts with both primer and template, suggesting a role during the termination step of primase activity when the primer is passed to the p180 subunit (81).

Very little is known about the domain architecture of p58. Both the N- and C-terminal halves of the subunit interact with p48 (79). The N-terminal region of p58 has homology to the piece of the primase large subunit of *S. solfataricus* that was co-crystallized with the small (p48 homolog) subunit (Figure 1.5, (78)). Additionally,

multiple sequence alignments revealed homology between a region of p58 and a small domain of DNA pol β (80). Given the unique roles of p58 during eukaryotic DNA replication, additional structural studies are warranted to fully understand its activity.

The SV40 DNA Replication Initiation Machine

Tag, Topo I, pol-prim, and RPA interact to coordinate the early stages of DNA replication. How these four proteins come together to initiate SV40 DNA replication is of considerable interest. DNA processing enzymes are often modular, containing flexibly linked domains (82). This property facilitates a handoff mechanism, whereby specific proteins can be recruited when needed and displaced when they have completed their task (82). This is especially apparent in SV40 DNA replication, as both Tag and RPA are highly modular, and preliminary studies of pol-prim suggest a modular architecture as well. Previous studies have taken advantage of this property by studying independent domain-domain interactions, and then testing the functional consequence of those interactions in the context of the intact full-length proteins. The data generated from these studies (summarized in Table 1.1) have provided insight into how the DNA initiation proteins associate to form a molecular machine.

Table 1.1. Protein-protein interactions involved in SV40 DNA replication initiation. Adapted from (47).

Interaction	Domains	Role of interaction	Binding site mapped?	Ref
Tag – Topo I	HD - Cap	Topo I loading	Y	(83, 84)
Tag – Pol-prim	HD – $p68_{1-240}$	Pol-prim loading?	N	(59)
	? – $p180_{195-313}$	?	N	(85)
	? – p48/p58	Pol-prim loading on lagging strand?	N	(86)
Tag – RPA	OBD – 70AB	RPA loading onto ssDNA	Y	(53, 87)
	OBD – 32C	Pol-prim loading onto RPA coated ssDNA	Y	(55)
RPA – Pol-prim	70NAB – p48/p58	?	N	(88-90)
	32C? – p48/p58	?	N	(90)

The current model of SV40 DNA replication is shown schematically in Figure 1.6. The first step in this process is Tag binding to the origin, mediated through the OBD (91). Tag recognizes the specific DNA sequence and assembles as a double hexamer, encircling the dsDNA with the OBDs on the interior, and the helicase domains on the exterior. Pol-prim is then recruited by Tag, presumably via the p68-helicase domain interaction (91).   Next, Topo I is recruited, also via an interaction with the helicase domain of Tag. Two Topo I molecules are recruited, per double hexamer, presumably one on each side of Tag. At this stage, Tag can begin melting the dsDNA via the mechanism described previously. Topo I is positioned to relieve the torsional strain created by this process. Tag actively loads RPA onto the ssDNA as it is extruded via the interaction between the OBD and RPA70AB (53). RPA is initially loaded in the compact mode, only interacting with DNA via RPA70AB. As more ssDNA is extruded, RPA then binds in its extended binding mode, utilizing all four DNA-binding domains. Pol-prim then interacts with the ssDNA template, which is also mediated by Tag. In addition to interacting with RPA70AB, Tag OBD also interacts with RPA32C. This interaction promotes primer synthesis, and is likely the result of the dissociation of RPA at the

primer initiation site (55). RPA is likely shifted from the high affinity, extended binding mode, to the lower affinity, compact binding mode, which gives pol-prim access to the ssDNA.   Pol-prim then begins synthesizing the RNA-DNA primer, completing the initiation stage of SV40 DNA replication. Pol-prim and RPA also interact, and pol-prim is presumably able to dissociate RPA from ssDNA via these interactions.
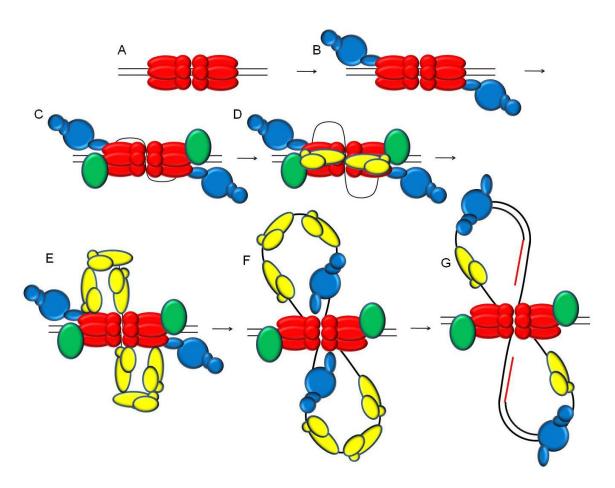
Figure 1.6. Initiation of SV40 DNA replication. (A) Tag (red) assembles at the origin as a double hexamer. (B) Pol-prim (blue) is recruited by Tag. (C) Topo I (green) binds and the origin is melted. (D) RPA (yellow) binds to the OBD of Tag, and (E) is actively loaded onto the emerging ssDNA. (F) Pol-prim is loaded onto the ssDNA, and (G) creates a chimeric RNA-DNA primer.

Experimental Overview

The studies described in this dissertation are aimed at furthering our knowledge of how pol-prim is involved in the initiation of SV40 DNA replication. The following experiments were conducted in close collaboration with Dr. Ellen Fanning's research group at Vanderbilt University. All of the functional data originates from her laboratory, and the person responsible for generating the data will be identified in the appropriate figure legends. Together with the Fanning group, the similar strategy of studying domain-domain interacts in the context of the initiation complex was pursued. However, since structural information regarding pol-prim is lacking, identifying stable, structured domains was a necessary first step. Using a variety of techniques structured domains were identified in the N-terminus of p68 (p68N) and the C-terminus of p58 (p58C).

The p58C domain was initially identified using limited proteolysis coupled with mass spectrometry (Chapter II). Circular dichroism (CD), which can detect protein secondary structure was used to verify that the domain was structure. This finding was confirmed by nuclear magnetic resonance (NMR) experiments. Interestingly, p58C was found to contain an essential iron-sulfur cluster. The protein solution was golden brown, and the cluster was identified using ultraviolet-visible (uv-vis) spectrophotometry. Electron paramagnetic resonance spectroscopy (EPR) was used to directly probe the cluster, and we were able to classify the clusters as a high-potential [4Fe-4S] type. Functional experiments using mutant proteins lacking the iron-sulfur cluster revealed that this cluster was required for p58's stimulation of p48's activity.

The p68N domain was found to interact with the Tag helicase domain, and it was found to be essential for SV40 DNA replication. The p68N-helicase domain interaction

was studied biophysically using isothermal titration calorimetry (ITC) and NMR to calculate the strength of the interaction and to map the binding interface (Chapter III). ITC responds to heat changes as molecules interact, and titrating a binding partner into the sample cell allows for a binding curve to be created. NMR is a powerful tool for studying protein-protein interactions, as residues at a binding interface are often perturbed when bound to the interacting protein. These biophysical experiments were used to create p68N mutants deficient in Tag-binding for use in functional assays.

The p68N solution structure was determined using NMR techniques (Chapter IV). Briefly, backbone and side chain resonances were assigned using standard three-dimensional experiments. Initial structures were generated using NMR-derived distance restraints as well as dihedral restraints. These structures were then further refined using restrained molecular dynamics.

Finally, these results are discussed, and the model for SV40 initiation is revised to incorporate these new findings (Chapter V). This work advances our knowledge of DNA replication by identifying a novel cofactor amongst DNA replication proteins, and it offers the first high resolution structure of a pol-prim domain.

CHAPTER II

AN IRON-SULFUR CLUSTER IN THE C-TERMINAL DOMAIN OF THE p58
SUBUNIT OF HUMAN DNA PRIMASE[*]

Introduction

DNA polymerase alpha-primase (pol-prim) associates with eukaryotic replication

forks in S-phase during the initiation of DNA replication (21, 22). Pol-prim synthesizes a

chimeric RNA-DNA primer of ~30 nucleotides that is then extended by more processive

DNA polymerases that synthesize the leading and lagging strands. Pol-prim is composed

of four subunits (p180, p68, p58, and p48). The p180 subunit has the DNA polymerase

catalytic activity and binds to both the p68 and p58 subunits. The p68 subunit has a

regulatory function that is not completely understood. It is required for initiation of yeast

chromosomal replication (92, 93) and cell-free SV40 DNA replication (63). In addition,

phosphorylation of p68 alters the activity of pol-prim in SV40 replication (64, 67, 94,

95).

The two smallest subunits, p48 and p58, together function as the DNA primase by

creating an RNA primer of 7-10 nucleotides (96, 97). The p48 subunit contains the

catalytic site (98).   The p58 subunit stabilizes p48 and participates in initiation,

elongation, and "counting" the ribonucleotides polymerized (80). Interestingly, p58 is

also involved in transferring the RNA strand directly into the active site of the associated

p180 subunit, which extends the growing nucleotide with dNTPs to complete the

formation of the RNA-DNA primer (21, 72, 81). Knowledge of the molecular basis for

regulation of the length of RNA portion of the primer and internal transfer to the p180 subunit is very limited.

Despite the fundamental importance of primase in DNA replication, the only structural information available for a heterodimeric primase is for an archaeal (*S. solfataricus*) primase that does not form a pol-prim complex (78, 99). Multiple sequence alignments reveal homology between the p48 subunits and the N-terminal half of the p58 subunit. In the crystal structure of the *S. solfataricus* primase core, the p48 subunit assembles with the N-terminal half of p58 (78). However, in human p58, both the N- and C-terminal regions have contacts with p48 (79). Interestingly, the C-terminal half of p58 also contains a region with homology to a DNA polymerase beta domain; this region was determined to be important for primer synthesis (80), but how it functions is not known.

DNA primase serves as a key target for regulation of DNA replication initiation, telomere maintenance, and response to DNA damage or fork stalling, in part through its physical interactions with other proteins involved in DNA replication and in checkpoint signaling (100). Primase interacts physically with the viral helicase SV40 large T antigen (Tag), eukaryotic replication protein A (RPA) (86, 88), and GINS, a recently identified component that plays a central role in establishment and progression of eukaryotic and archaeal replication forks (101-103). Primase activity is essential for optimal checkpoint signaling at stalled replication forks (104-107) and possibly in rescuing stalled replication fork progression (108), but its interaction partners are not known.

To better understand the role of human primase in these pathways, it will be vital to elucidate its structure and interactions with partner proteins. This strategy has been useful in determining the roles of the Tag-RPA interaction in the context of SV40 DNA

replication (53, 55). In order to facilitate similar experiments with human DNA primase, we sought to characterize the domain architecture of DNA primase. The p58 and p48 subunits can be expressed and purified independently of the other two subunits and retain primase activity *in vitro* at levels similar to those observed for the intact heterotetramer (79, 98). Working from bacterially expressed primase protein, a structured domain in the C-terminus of the p58 subunit (p58C) was identified. Biophysical analysis of this construct showed that the domain is folded and has the characteristics of a [4Fe-4S] high-potential iron protein (HiPIP). The conservation of four cysteines across several species suggests a critical role for the cluster, and this was confirmed by *in vitro* experiments that demonstrate that the [4Fe-4S] cluster is required for primase activity.

Results

Primase Domain Architecture

Proteins involved in DNA replication are often modular, containing several independent domains tethered together (82). In many cases, it is possible to map the location of these structural domains by subjecting the intact protein to very limited proteolytic digestion. The protease will preferentially cleave solvent exposed linkers, leaving structured domains intact. The digestion is then analyzed by SDS-PAGE, the bands are then excised and identified using mass spectrometry and N-terminal sequencing. To test whether DNA primase contained any structured domains, the p48/p58 dimer was initially subjected to limited proteolysis with trypsin, chymotrypsin, and proteinase K (Figure 2.1). Digestion products were monitored as a function of

reaction time by using denaturing gel electrophoresis. As the three proteases have different cleavage site specificities, they were able to cleave primase to varying degrees. Proteinase K produced the best fragmentation profile, and was chosen for further experiments.
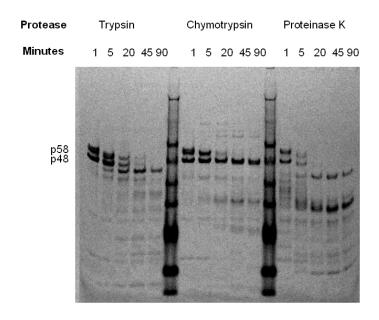
Figure 2.1. Limited proteolysis of p48/p58. The purified primase dimer was subjected to limited proteolytic digestion with trypsin, chymotrypsin, and proteinase K.

As shown in Figures 2.1 and 2.2.A, two stable fragments were produced corresponding to molecular weights of ~42 kDa (band a) and ~28 kDa (band b). Bands were excised from the gel and characterized by MALDI-TOF MS, LC-MS, and N-terminal sequencing in order to determine the identity of the two fragments. The 42 kDa band was found to be the result of a C-terminal truncation of roughly 60 residues from p48. Proteinase K digestion of isolated p48 produced the same fragment (Figure 2.2.B), which confirmed that the 42 kDa fragment is a stable domain of p48. The 28 kDa fragment corresponds to the C-terminal half of p58, residues Gly266-Ser509.
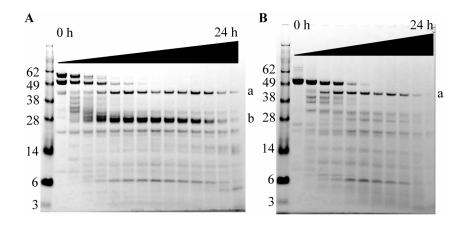
Figure 2.2. Identification of a stable p58C domain. Purified primase dimer (A) and p48 (B) were subjected to limited proteolytic digestion using proteinase K. Aliquots were taken at several time-points up to 24 hours and analyzed by SDS-PAGE. Mass spectrometry and N-terminal sequencing were used to establish that band a results from cleavage near the C-terminus of p48, and band b is the C-terminal half of p58.

Production of Primase Domain Expression Constructs

To evaluate the potential primase domains further, they were sub-cloned into bacterial expression vectors. Two p48 constructs (Met1-Ser361 and Met1-Glu366) were sub-cloned into bacterial expression vectors but did not produce soluble protein, suggesting the C-terminus is an integral part of the protein. This was surprising given that secondary structure prediction and a sequence alignment of human p48 with *P. furiosus* primase also suggest a C-terminal extension (Figure 2.3). Thus, p48 may actually be globular and contain a solvent exposed loop in the vicinity of the proteinase K cleavage site.

```
Human METFDPTELPELLKLYYRRLFPYSQYYRWLNYGGVIKNYFQHREFSFTLK-DDIYIRYQSFNNQSDLEKEMQ-KMNPYKIDIGAVYSHRPNQHNTVKLGA   98
      ---------HHHHHHHHH----HHHHHHH------HHH--EEEEE-- ---EHEHH----HHHHHHHHH HH--EEEEEEEE-----H--------
Pfu   ------MLMREVTKE--ERSEFYSKEWSAKKIPKFIVDTLESREFGFDHNGEGPSDRKNQYSDIRDLEDYIR-ATSPYAVYSSVAFYENPR-----EMEG   85
Pho   ------MLLREVTRE--ERKNFYTNEWKVKDIPDFIVKTLELREFGFDHSGEGPSDRKNQYTDIRDLEDYIR-ATAPYAVYSSVALYEKPQ-----EMEG   85
             ------HH HHHHHHHH--------HHHHHHH--EEEE-------EEE-----HHHHHHHHH HH--EEEEEEE-EEEEE-      -EEE
Sso   ---MGTFTLHQGQTN--LIKSFFRNYY--LNAELELPKDMELREFALQPFGSDTYVRHLSFSSSEELRDYLVNRNLPLHLFYSSARYQLPSARN-MEEKA   92
              H   HHHHHHHHH HHH------HHHHH-EEEE-------------HHHHHHHHHHH---EEEEEEE---------H HHHHH
              : :          : : :        : .:: ***.:    .. *   .:. :*.. :       *   : .... *      : .

Human FQAQEKELVFDIDMTDYDDVRRCCSSADICPKCWTLMTMAIRIIDRALKEDFGFKHRLWVYSGRRGVHCWVCDESVRKLSSAVRSGIVEYLSLVKGGQDV  198
      --HHHEEEEEE-------HHH--------HHHHHHHHHHHHHHHHHHH----EEEEEEE---EEEEE--HHHH---HHHHHHHHHHHHHH------
Pfu   WRG--AELVFDIDAKDLPLKRCNHEPGTVCPICLEDAKELAKDTLIILREELGFENIHVVYSGR-GYHIRILDEWALQLDSKSRERILAFISASEIENVE  183
Pho   WLG--TELVFDIDAKDLPLRRCEHEPGTVCPICLNDAKEIVRDTVIILREELGFNDIHIIYSGR-GYHIRVLDEWALKLDSKSRERILSFVSASEIEDVE  183
      EE-  -EEEEEEE---------------HHHHHHHHHHHHHHHHHHHHH-------EEEEEEE- EEEEEE-------HHHHHHHHHHHHHHH-----HHH
Sso   WMG--SDLLFDIDADHLCKLR-SIRFCPVCGNAVVSEKCERDNVETLEYVEMTSECIKRGLEQTRNLVEILEDDFGLKPKVYFSGNRGFHVQVDCYGNCA  189
      ---  --EEEE-HHHHHHH-H HHHH----------EE----EE-HHHHHHHHHHHHHHHHHHHHH-----EEEE-----EEEEE---HHH
      : .   .:*:****  .   *        :* .    .            :: :       .  .  .  : *: :.        .:.       :

Human KKKVHLSEKIHPFIRKSINIIKKYFEEYALVNQDILENKESWDKILALVPETIHDELQQSFQKSHNSLQRWEHLKKVASRYQNNIKNDKYGPWLEWEIML  298
      -----------HHHHHHHHHHHHHHHHHHHH-----HHHHHHHH-HHHHHHHHHHHH-----HHHHHHHHHHHHH---------HHHHHHHHH
Pfu   EFRRFLLEKRGWFVLK-----HGYPRVFRLRLGYFILR----VNVPHLLSIGIRRNIAKKILDHKEEIYEGFVRKAILASFPEGVGIESMAKLFALSTRF  274
Pho   EFRKLLLNKRGWFVLN-----HGYPRAFRLRFGYFILR----IKLPHLINAGIRKSIAKSILKSKEEIYEEFVRKAILAAFPQGVGIESLAKLFALSTRF  274
      HHHHHHH-HHHHH---    --HHHHHHHHHHHHHHHH    -HHHHHHHHH----HHHHHHHHHHHHHH-------------HHHHHHHHHH
Sso   LLDSDERKEIAEYVMGIG--VPGYPGGSENAPGWVGRKNR—GINGVTIDEQVTIDVKRLIRIPNSLHGKSGLIVKRVPNLDDFEFNETLSPFTGYTIFL  285
      H---HHHHHHHHHH---  ------------HHHHHHH  -------HHHHHH---EEE-------------EE--------HHHHH----EEEEE
      ::  ::        *            .  .         :    :   .: : :  :.  .        .   :     :.  .        :

Human QYCFPRLDINVSKGINHLLK-SPFSVHPKTGRISVPIDLQKVDQFDPFTVPTISFICRELDAI TNEE KEENEAESDVKHRTRDYKKTSLAPYVKVFEH  397
      HH----------H---HE-- ---EE-----EEEEE-----HH--------HHHHHHHHH----H---------H-----------HHHHHHHHHH
Pfu   SKAY--FDGRVTVDIKRILR-LPSTLHSKVGLIATYVGTK----------------EREVMKFN----------PFRHAVPKFRKKEVR---EAYKL  339
Pho   SKSY--FDGRVTVDLKRILR-LPSTLHSKVGLIAKYVGTN----------------ERDVMRFN----------PFKHAVPKFRKEEVK---VEYKK  339
      HHHH  H-HHHHHHH--EEE EE--EEE--EEEEEEE--H            HHHH---H         HHHH-----HHHHH   HHHHH
Sso   PYIT--IETEVLGSIIKLNRGIPIKIKSSIGIYLHLRNLG-------------------------------------EVKAYVR-----------  330
      E---  EEEE----EEEE-----EEEE-HHHHHHHHH---                                  EEEEE--
      :: .*  .: :: :  *  .::.. *      .                                          .  : :

Human FLENLDKSRKGELLKKSDLQKDF  420
      HHHHHHH-H-HHHHHHHHHH--
Pfu   WRESLEYE---------------  347
Pho   FLESLGT---------------  346
      HHHHH--
Sso   ----------------------  330
```

Figure 2.3. p48 sequence analysis. Amino acid sequences for human, *Pfu*, *Pho*, and *Sso* primase catalytic subunits were aligned using ClustalX. Identity is noted by a "*", a conservative substitution is noted by a ":", and a semi-conservative substitution is noted by a ".". Secondary structure is also listed below each sequence (α-helix as "H" and β-strand as "E"). The secondary structure content of the human primase was predicted using the SSpro program. The C-terminus of the two p48 constructs, S361 and E366 are highlighted in green.

Although a stable fragment was not detected in the experiments for the N-terminal half of p58, sequence analysis suggested it should form a structured domain: sequence alignment reveals homology between the N-terminal half of p58 and the region of the large subunit from *S. solfataricus* that has previously been crystallized (Figure 2.4, (78)). In an effort to produce and characterize a p58N construct, a wide range of sub-cloning experiments was performed by Chris Brosey in the Chazin laboratory, but none yielded soluble protein. A similar observation has been made for the full-length p58 subunit.

27

```
Human  MEFSGRKWRKLRLAGDQRNASYPHCLQFYLQPPSENISLIEFENLAIDRVKLLKSVENLGVSYVKGTEQYQSKLESELRKLKFSYRENLEDEYEPRRRDH  100
          -------HEEEEEE------------EEE--------EHHHHHHHHHHHHHHHHHHHH----EE-HHHHHHHHHHHHHHH--------HHHHHHHHHH
Sso    -----------MALDVKKYPFIKSLDDELKKYG---GGITLTDLLLNSTTLIDQAKDRIQKTKSGDELPHYVSYNEPVLVFYTTLLSLAILNDVK----   81
          -------HHH-HHHHHHHHH--  ---HHHHHHHH-HHHHHHHHHHHH-----HHHHH-HHHHHHHHHHHHHHHH--HH
            :* * :: .: :.*: *: . . * : :* :: ..*:...:: . .* * : .* : ::. .*    : :

Human  ISHFILRLAYCQSEELRRWFIQQEMDLLRFRFSILPKDKIQDFLKDSQLQFEAISDEEKTLREQEIVASSPSLSGLKLGFESIYKIPFADALDLFRGRKV  200
          HHHHHHHHH---HHHHHHHHHHHHHHH----HHHHHHHHH---------HHHHHHHH--EE-------EEE----EEE-HHHHHHHHH--EE
Sso    ---LIRRYAYAEAKQFRSLLHTENEE------NLLEISKLLDLKINRCDPIKFYLEKKRRIIQKEFCVHFIDYLKYTKDLKEDWKLS-----------GQ  161
          HHHHHHHH--HHH     HHHHHHHH---EEEEEEEE-----EEEEEEE-HHHHHHH---HHHHH--          --
            :* * **.:::::* : ::: :   .:* .*: *: : :: :::: : ::*: . . . .::. :*:.

Human  YLEDGFAYVPLKDIVAIILNEFRAKLSKALALTARSLPAVQSDERLQPLLNHLSHSYTGQDYSTQGNVGKISLDQIDLLSTKSFPPCMRQLHKALRENHH  300
          EE---EEEE-HHHHHHHHHHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHHHH--------------E-HHHHHHHH----HHHHHHHHHHH--
Sso    ILHKGYVYLDKNQLIGLIAESIKSKIVEMIRP----LNLKEIPEKLKSLIERR-------------------------GIIPPCIENILAKEKLNEE     229
          -EE--EEE-HHHHHHHHHHHHHHHHHHHHH    HHH---HHHHHHHHH
            *..*:.*:  ::::.:* :.::::*: : :    *   :  *:*:.*::: :***:.:: : *..

Human  LRHGGRMQYGLFLKGIGLTLEQALQFWKQEFIKGKMDPDKFDKGYSYNIRHSFGKEGKRTDYTPFSCLKIILSNPPSQGDYHGCPFRHSDPELLKQKLQS  400
          ----HHHHHHHHH-----HHHHHHHHHHHH-----HHHHHH---EEEEE--------------HHHEE-------------------HHHHHHHHHH
Sso    EIR----TLITFYIDIGKGLSGIVSIMKKYNVSNVEDLYRKYRGD-----------KGTRYIVYSCAKMKQLG-----------LCVSSCNVKNPLQL     301
            :       * .** *. :.: *: :.. * : :*      * * * :** *: . .. :*: **

Human  YKISPGGISQILDLVKGTHYQVACQKYFEMIHNVDDCGFSLNHPNQFFCESQRILNGGKDIKKEPIQPETPQPKPSVQKTKDASSALASLNSSLEMDMEG  500
          ----HHHHHHHHHH----EEEHEEEEEE-------------HHHHHHHHH-------------------------HHHHH-------HH-HH-
Sso    YFLSNE--------------------------------------------------------------------------------------------     307
            * :*

Human  LEDYFSEDS   509
          HHHH-----
Sso    ---------   307
```
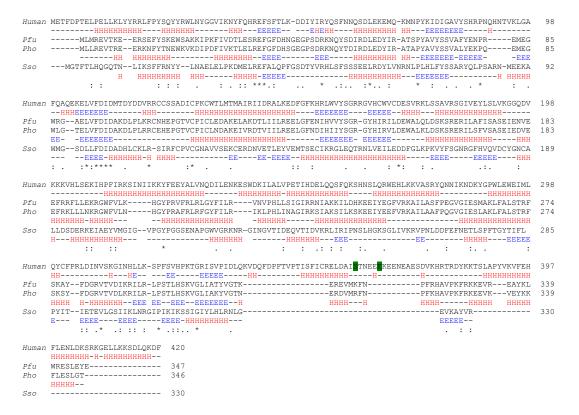
Figure 2.4. p58 sequence analysis. Amino acid sequences for human and *Sso* primase secondary subunits were aligned using ClustalW. Identity is noted by a "*", a conservative substitution is noted by a ":", and a semi-conservative substitution is noted by a ".". Secondary structure is also listed below each sequence (α-helix as "H" and β-strand as "E"). The secondary structure content of the human p58 was predicted using the SSpro program. The construct co-crystallized with the *Sso* primase catalytic subunit is highlighted in green. The p58C construct is highlighted in yellow.

In contrast to p58N, p58C has no significant homology to proteins whose structure is known. However, an extensive amount of secondary structure was predicted for this stable fragment (Figure 2.4). In fact, the p58C domain expressed well and was very soluble, even to very high concentrations (~30 mg/ml).

A combination of circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy was used to characterize the structural integrity of p58C. Minima observed at 208 and 222 nm in the far-UV CD spectrum (Figure 2.5.A) indicate a significant amount of alpha-helical content in p58C. Analysis of the CD spectrum using the K2d program provided an estimate of 31% helix, 14% beta-sheet, and 55% coil. The 600 MHz ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) NMR spectrum of

p58C is shown in Figure 2.5.B. The spectrum contains relatively narrow linewidths and

very good dispersion, which are indicative of a well-folded structural domain.
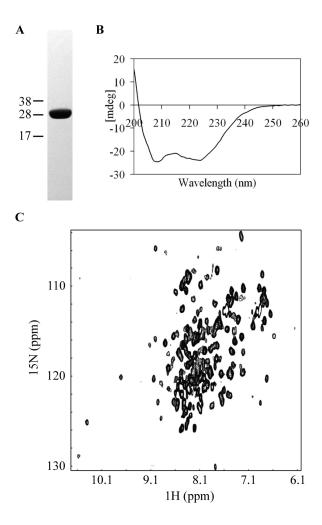
Figure 2.5. Structural characterization of p58C. (A) Purified p58C was visualized by SDS-PAGE and staining with Coomassie. (B) CD spectrum of p58C. Minima at 208 and 222 nm indicate the presence of a significant quantity of helical secondary structure. (C) The 600 MHz ^{15}N-^{1}H HSQC NMR spectrum of p58C obtained at 25 $^{\circ}$C. The relatively narrow lines and wide dispersion of the signals indicate a well-folded tertiary structure.

An Iron-sulfur Cluster in p58C

Interestingly, when either the primase dimer or p58C was purified, the protein

solution had a golden-brown color. The color intensified as the protein was concentrated,

becoming very dark at high concentrations. A UV-visible spectrum of p58C contained a broad peak at 400 nm, similar to spectra from proteins containing iron-sulfur clusters (Figure 2.6.A) [e.g. (109)]. The primase dimer also had this property, while isolated p48 did not (data not shown).
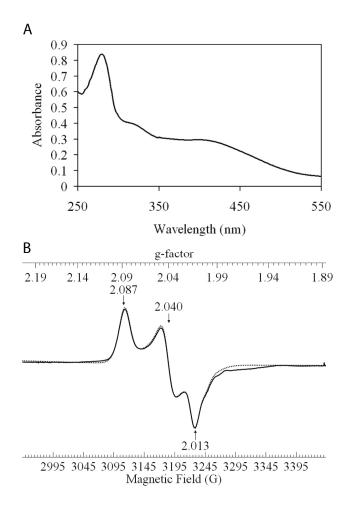
Figure 2.6. An iron-sulfur cluster in p58C. (*A*) The UV-visible spectrum of p58C. The broad peak at ~400 nm is typical of proteins containing iron-sulfur clusters. (*B*) Experimental (solid line) and simulated (dashed line) EPR spectra of p58C obtained at 15 K.

To further investigate the presence of an iron-sulfur cluster in p58C, the protein was analyzed by inductively coupled plasma mass spectrometry (ICP-MS), which enables the content of metal ions to be quantified. This analysis provided an estimate of 3.09 moles iron per mole of protein, which is consistent with the presence of 4 iron atoms in the cluster. A stoichiometry considerably lower than 4 is typical for proteins containing [4Fe-4S] clusters (110, 111), and is attributable to the loss of one iron in the cluster during protein purification and analysis.

EPR spectroscopy is now a well accepted means to confirm the presence of iron-sulfur clusters in proteins and assign them to specific cluster types and oxidation states. The 9.05 GHz X-band EPR spectra of p58C acquired directly on the purified protein lacked any appreciable signal. Reduction of p58C with sodium dithionite produced only a weak, broad signal. However, oxidation of the protein resulted in a strong signal characteristic of a low-spin S = ½ iron-sulfur cluster (Fig 2.6.B). The signal is highly rhombic with $g_1 = 2.087$, $g_2 = 2.040$, and $g_3 = 2.013$. Simulation of the EPR spectrum with SIMPOW6 software (112) reveals two slightly different forms are present in solution, one with $g_1 = 2.0870$, $g_2 = 2.0405$, and $g_3 = 2.0126$ and a second weaker form with $g_1 = 2.0872$, $g_2 = 2.0311$, and $g_3 = 2.0094$ and broader linewidths. Similar results reported for oxidized HiPIPs have been interpreted as resulting from the presence of multiple isomeric states [e.g. (113)]. The average EPR g-factor of 2.046 observed for p58C is much more typical of a $[4Fe-4S]^{3+}$ cluster than a $[3Fe-4S]^{+}$ cluster, as the latter normally exhibit an average g-factor of 2.015 and relatively smaller g anisotropy (111, 114, 115).

The EPR spectrum has many properties that suggest the [4Fe-4S] cluster is HiPIP-like. Like typical HiPIP clusters, the iron sulfur cluster in p58C is only visible by EPR when oxidized, has an average g-value greater than two, is observed best below 30K, and the signal does not readily saturate. While the observed g-values deviate somewhat from the prototypical HiPIP cluster, such differences may be attributable to changes in the environment around the [4Fe-4S] cluster or distortion of the cluster by the protein environment (116, 117). In summary, the biophysical data on p58C are consistent with assignment to the class of high potential iron proteins, thus adding to the growing list of DNA processing proteins with this unique co-factor.

Identification of Iron-sulfur Ligands in p58C

Iron-sulfur clusters are typically bound to proteins via four cysteine residues, and p58C contains six cysteine residues. Following the strategy used for other DNA-processing proteins that contain an iron-sulfur cluster, potential cysteine ligands were identified from a multiple sequence alignment of p58 from 5 different species using CLUSTALX (118). The alignment in Figure 2.8.A reveals that four of the cysteine residues (Cys287, Cys367, Cys384, Cys424) are conserved. These are residues most likely responsible for cluster binding.

In order to test this hypothesis, each of the four cysteine residues in p58C was individually mutated to serine. While the mutant constructs expressed at levels similar to the WT protein (data not shown), they were poorly soluble (Figure 2.7). Overnight dialysis resulted in total loss of each mutant protein, likely due to precipitation.
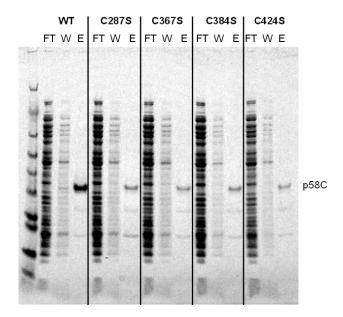
Figure 2.7. Ni-NTA purification of p58C mutants. The four Cys-to-Ser mutants were expressed as described for the WT p58C domain. The soluble cellular lysate was passed over a gravity flow Ni-NTA column. The lanes represent the column flow-through (FT), a three column volume wash (W) and a 2 CV elution with 250 mM imidazole (E).

Consequently, an alternate strategy was used involving mutation in the context of the p48/p58 dimer. Data is shown here for the Cys367Ser mutant primase dimer, which expressed at levels comparable to the wildtype primase, remained soluble, and co-purified with the p48 subunit (Figure 2.8.B). Analysis of this mutant primase dimer by UV-visible spectroscopy provided a spectrum in which the broad peak at 400 nm was clearly absent (Figure 2.8.C), confirming this mutant primase does not contain an iron-sulfur cluster.

A

```
                     ::  *        .  . *::   : :: ***::: .* * .*******:: :::: :  : * :  ::  :** : * *.: : :* :
Human  266 GNVGKISLD--QIDLLSTK------SFPP●MRQLHKALRENHHLRHGGRMQYGLFLKGIGLTLEQALQFWKQEFIKG-KMDPDKFDKGYSYNIRHSFGKE
Mouse  266 KNTGKISLD--QIDSLSTK------SFPP●MRQLHKALRENHHLRHGGRMQYGLFLKGIGLTLEQALQFWKQEFIRG-KMDPDKFDKGYSYNIRHSFGKE
Dros   278 RDA-AVPIE--SLDQLSKT------SMPL●MRMCHEHIRANHHIKHSGRMQYGLFLKGIGVALEDSLRFWREEFTK--KMDADKFTRSYEYNIYHNYGKK
Cele   286 QPDGSIIVTRHMIDKLSRR------SFAP●MKQMHNHLRVNHHLKYGARRQYGLFLKGIGLSLDEAIAMMRDEFTK--KITSDKFDKEYGYNIRYMYGKE
Sacch  307 KANQFSENVDDEINAQSVWSEEISSNYPL●IKNLMEGLKKNHHLRYYGRQQLSLFLKGIGLSADEALKFWSEAFTRNGNMTMEKFNKEYRYSFRHNYGLE

              *       ..  *  .*:     ..  * ****::.  . *.  :: *    *  :    . .*  .* : *::  **              : *** ::
Human  357 GKRTDYTPFS●LKIILSNPPSQGDYHG●PFRHSDPELLKQKLQSYKISP-GGISQILDLVKGTHYQVA●QKYFEMIHNVDD------CGFSLNHPNQFFC
Mouse  357 GKRTDYTPFS●MKIILTNPPGQGDYHG●PFRHSDAELLKQKMQSYKIPA-SGISQILDLVKGNHYQVA●QKYFEMTHNVDD------CGFSLNHPNQFFF
Dros   367 GSMVNYTPYS●AKIIK-DVAGPGDCHG●PYKNMDQGSLKTKLSSYGLSA-SAIDEVMFFVSRGHYQIG●GKYFQLTHNSP-------VEPSINHPNNYFE
Cele   378 GRRVAQTAMS●ATIILRNPPSAVDCHG●PFRHSEKQVLKQKLGSDAILKKEQIERIAELAELNQYDKA●TRYFEFMHNLEEG----GLGQLITHPNQFYE
Sacch  407 GNRINYKPWD●HTILSKPRPGRGDYHG●PFRDWSHERLSAELRSMKLTQ-AQIISVLDSCQKGEYTIA●TKVFEMTHNSASADLEIGEQTHIAHPNLYFE

               .*: :
Human  450 ESQRILNGGKDIKKEPIQPETPQPKPSVQKTKDASSALAS-------LNSSLEMDMEG---LEDYFSEDS-------
Mouse  450 ESQRILTGGKDIKKEISQPETPQHKPSTQKTRDAASALAS-------LDSSLEMDLEG---LEEYF----------
Dros   458 ESQ-IAMGNRQKRANGSAPPKARIRPDIKGHGDRSMLMGDDDDELWRIAETQERIMQSQKDISEAFNDDLDLTQIDY
Cele   474 ESQKIVAERIAAKQESQEAPAPKIEKMDTE----------------------------------------------
Sacch  506 RSRQLQKKQQKLEKEKLFNNGNH------------------------------------------------------
```
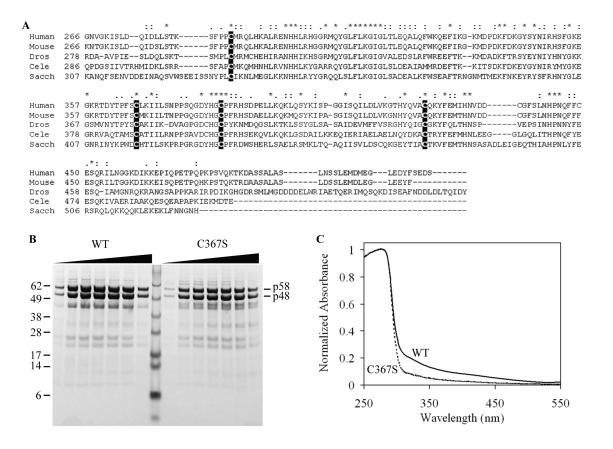
B WT   C367S

C

Figure 2.8. The p58C iron-sulfur cluster is coordinated by four conserved cysteines. (A) Sequences are shown for: *H. sapiens* (Human), *M. musculus* (Mouse), *D. melanogaster* (Dros), *C. elegans* (Cele), and *S. cerevisiae* (Sacch). The four highlighted cysteine residues are conserved, suggesting they serve as ligands for the iron-sulfur cluster. (B) Ni-purified p48/p58 WT (left) and p48/Cys367Ser-p58 mutant (right) were visualized by SDS-PAGE and Coomassie staining. Lanes represent fractions collected during Ni-NTA purification as the concentration of imidazole is increased using a linear gradient. (C) UV-visible spectrum of WT p48/p58 (solid line) and p48/Cys367Ser-p58 mutant (dashed line).

A Role for the Iron-sulfur Cluster in Primase Function

To initially assess the functional relevance of the iron-sulfur cluster in p58, the primase activity of wildtype and p48/Cys367Ser-p58 primase dimers on a natural ssDNA template was assayed as a function of protein concentration. Radiolabeled CTP was incorporated into RNA primers of 8-10 nucleotides by the wildtype primase (Figure 2.9.A, lanes 1-3), and small amounts of larger products were detectable, as observed

34

previously (79, 81). No RNA primers were observed in the absence of enzyme (Figure 2.9.A, lane 7). Products of the mutant primase that lacks the iron-sulfur cluster were barely detectable above background (Figure 2.9.A, lanes 4-6), and the level of reaction product was not proportional to the amount of mutant primase in the reaction. Quantification of the products as a function of primase concentration revealed that the specific activity of the mutant primase was reduced at least 5-fold (Figure 2.9.B). These data suggest that the iron-sulfur cluster in p58 is crucial for the primase activity of p48/p58 in this assay.
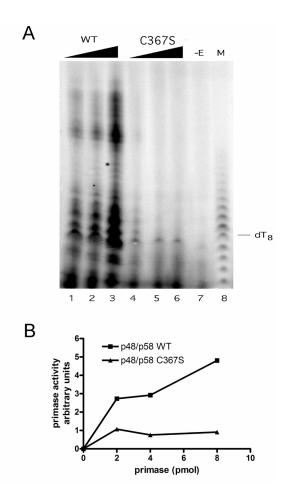
Figure 2.9. Primase activity of p48/p58 requires the iron-sulfur cluster. (A) Radiolabeled primers synthesized by wildtype (WT) p48/p58 (lanes 1-3) or mutant Cys367Ser (lanes 4-6) on M13 ssDNA template were visualized by denaturing gel electrophoresis and phosphorimaging. Products of a control reaction without enzyme (-E) are shown in lane 7. M, radiolabeled dT_{4-22} markers (lane 8). (B) Primase products were quantified using phosphorimager software. Incorporation in the negative control reaction (A, lane 7) were subtracted from those in lanes 1-6 and graphed as a function of primase concentration. Data collected by H. Huang, Fanning laboratory.

Discussion

Our studies show that DNA primase contains a structured domain in the C-terminus of the p58 subunit and that this domain contains an iron-sulfur cluster. This is the first report of an iron-sulfur cluster in a DNA replication protein. Analysis of p58C by ICP-MS, UV-Vis, EPR, and phylogenetic amino acid sequence comparisons are

consistent with the presence of a HiPIP [4Fe-4S] cluster coordinated by four conserved cysteine residues.

Iron-sulfur clusters have been reported in DNA repair proteins, including DNA glycosylases MutY (119), endonuclease III (120), and family 4 uracil-DNA glycosylases (110). Recently, several members of a family of DNA helicases involved in DNA repair, including the nucleotide excision repair helicase XPD and the Fanconi J crosslink repair helicase, have also been shown to contain an iron-sulfur cluster that is essential for helicase activity, but not for ssDNA-dependent ATPase activity (109). Remarkably, each of these proteins contains a [4Fe-4S] cluster coordinated by four conserved cysteines like DNA primase.

Iron sulfur clusters in proteins are traditionally associated with electron transport and redox chemistry (121). The presence of clusters in proteins involved in several aspects of DNA metabolism points to a different and possibly common function. However, a specific role for iron-sulfur clusters in DNA repair proteins has not yet been determined. Evidence has accumulated showing that the clusters are required for enzymatic activity, e.g. for MutY (122) and XPD (109). In addition, x-ray crystal structures of MutY (123) and endonuclease III (124) have been determined. These structures show that the iron-sulfur clusters are too far from the active site to participate directly in catalysis. Thus, the influence of the cluster on enzymatic activity appears to arise from an allosteric effect.

In support of this proposal, studies conducted on MutY revealed that the cluster was critical for orienting key residues that contact the distorted DNA (122, 123, 125). Also, modeling studies on the uracil-DNA glycosylases point to a role in substrate

37

recognition rather than catalysis (110). The lack of structural data on XPD makes determining the role of the iron-sulfur cluster considerably more difficult. Hence, the bulk of the data available to date point to the iron-sulfur clusters having an influence on structural features of these proteins as opposed to participating directly in aspects of protein chemistry (109). Since iron-sulfur clusters are invariably integrated into protein structure, it makes sense that they modulate the structure and stability. However, it should be noted that it is not readily possible to manipulate the redox state of HiPIP proteins *in vivo* or in cells. Moreover, it is difficult to design experiments that directly address functional questions that do not result in complete loss of the cluster. Thus, while the importance of iron-sulfur clusters in the structure DNA processing proteins is evident, additional redox-mediated functional roles cannot yet be ruled out.

If the role of the iron-sulfur cluster is purely structural, why was such a complex co-factor chosen to serve for this purpose? Iron sulfur clusters are inserted into proteins via a multi-step process involving several proteins (126), which seems like an excessive utilization of cellular resources for a purpose that could be attained at a lower energetic cost. Recently, studies on MutY have led to another theory of iron-sulfur cluster function in DNA glycosylases. Initially, the cluster was not thought to be involved in redox chemistry as it is not redox active *in vitro* (127). However, a recent report that binding of DNA shifts the redox potential of the cluster (128). This observation led to a complex model of glycosylase function in which electron transfer between iron-sulfur clusters on separate glycosylases occurs through the DNA; the cluster in the remote glycosylase is reduced and the enzyme dissociates from the DNA. Aberrant DNA would not have this property, and the glycosylase would remain oxidized and bound to DNA, promoting

DNA repair. Although this model has a number of appealing features, it remains controversial.

While we have shown that the iron-sulfur cluster in DNA primase is required for enzymatic activity, the specific function of the cluster remains unclear. The decreased solubility and stability of the cysteine-to-serine mutants in p58C suggest that the cluster does provide some level of structural stability to the protein. However, full-length p58 is still sufficiently structured in the absence of the iron-sulfur cluster to bind p48. Interestingly, a budding yeast mutant *pri2-2* that encodes Tyr instead of one of the conserved Cys ligands (Cys434Tyr, Figure 2.6.A) in p58 displays a temperature-sensitive, slow-growth phenotype (129, 130). This substitution would be expected to result in loss of the iron-sulfur cluster from p58. The observation that the mutation is not lethal suggests that in the context of the pol-prim complex in vivo, the cluster may serve primarily a regulatory as opposed to a purely structural function in p58.

The possibility that the DNA primase iron-sulfur cluster becomes redox active when bound to DNA cannot be ruled out. The idea that this might account for some of the unique properties of p58, such as its ability to regulate the length of unit length primers, is intriguing. The above noted 5'-3' iron-sulfur repair helicases as well as both subunits of primase (81) bind ssDNA. Moreover, a 5' overhanging, 3' recessed primer-template junction is known to be especially important for processing DNA by these proteins. The helicases have ssDNA-dependent ATPase motor domains that translocate the protein 5' to 3' along ssDNA until it encounters the 3' end of the complementary strand in duplex DNA. It is known that without the iron-sulfur cluster, the helicase cannot unwind the DNA, perhaps because its interaction with the junction is weak or

defective. Similarly, pol-prim bound to ssDNA template polymerizes NTPs into an oligoribonucleotide-template that remains bound to p58 (81). These authors proposed that the completed RNA primer-template likely remained bound to p58C, based on the sequence homology to the 8-kDa domain of DNA pol-beta that is known to enhance pol-beta processivity.

Negative regulation of primer polymerization beyond unit length requires a stable primer-template (71, 81), implying that the unit-length primer-template forms a stable complex with p58C prior to internal transfer of the primer to the polymerase active site in intact pol-prim. Thus, there is a pause in the reaction until the transfer occurs, relieving the negative regulation. In the absence of the polymerase subunit, primase activity resumes only after the primase dissociates from the primer-template and rebinds to ssDNA. Taken together, these studies strongly suggest that p58C interacts with primer-template during and after primer synthesis. If the iron-sulfur cluster in p58C is important for primer-template binding, the Cys367Ser primase might dissociate too frequently to enable efficient polymerization to create the full-length primer. This interpretation could also partially explain the temperature-sensitive phenotype of the *pri2-2* p58 yeast mutant, which would be expected to dissociate from the DNA more frequently at higher temperature and lose efficiency in generating the primer. In terms of a redox function for the iron-sulfur cluster in p58C, one could speculate that control of the redox state may enable primase to be retained on the growing primer-template either until the unit-length primer is completed and transferred to p180 for elongation, or primase dissociates from the primer-template.

The discovery of an iron-sulfur cluster in DNA primase offers new horizons in which to investigate the function of [4Fe-4S] clusters in DNA replication and repair machinery. Our findings imply that in addition to modulating the structural stability, the cluster in eukaryotic DNA primases may well function in some form of regulatory role, perhaps in controlling the length of the primer strand. As is evidenced by the study of MutY and endonuclease III, high-resolution structural analysis would aide significantly in investigating the role of the iron-sulfur cluster in DNA primase function. To this end, further integrated structural and functional analyses are currently in progress in our laboratories.

Note: After this work had been completed and submitted for publication, Klinge and colleagues reported the presence of an iron-sulfur cluster in DNA primase from *S. solfataricus* and *S. cerevisiae* (131). Our findings are in agreement with their data and extend it by identifying the cluster in human DNA primase and by showing that the cluster is contained in a distinct and well-folded structural domain within the p58 subunit.

Experimental Procedures

Primase Construct Design

     The recombinant human p48/p58 primase expression plasmid used in this study has been previously described (79). Full-length p48 cDNA was amplified using the dimer plasmid as the template. Using BamHI and EcoRI restriction enzymes, it was then inserted into the in-house pBG100 vector (L. Mizoue, Center for Structural Biology, Vanderbilt University), which contains an N-terminal His_6 tag. A p58 (266-509) construct (p58C) was subcloned from the dimer plasmid into pET15b (Novagen) using NdeI and XhoI. This construct also contains an N-terminal His_6 tag. Mutations were generated in the p58C expression plasmid by site-directed mutagenesis (Quikchange, Stratagene), and verified by DNA sequencing. An SphI-BglII fragment of p58C containing the mutation was then used to replace the corresponding wild-type fragment in the dimer p48/p58 expression plasmid. The oligonucleotides used to generate these constructs are listed in Appendix A.

Protein Expression and Purification

     Each construct was expressed in BL21 (DE3) cells. Cells were grown at 37 °C in LB to an OD_{600} of approximately 0.6. The temperature was then lowered to 22 °C and the cells were allowed to equilibrate for 30 minutes. Expression was induced using 1 mM isopropyl thio-beta-D-galactopyranoside (IPTG). Cells were harvested by centrifugation four hours post induction. Pelleted cells were resuspended in lysis buffer containing 50 mM Tris-HCl (pH 8), 300 mM NaCl, 20 mM imidazole, 3 mM 2-

mercaptoethanol (BME), 1% Nonidet P-40 (NP-40), 0.5 mg/ml lysozyme, ~10 mg DNase I, and one Complete Mini EDTA-Free protease inhibitor cocktail tablet (Roche). Cells were lysed by sonication at 4 °C. Insoluble material was removed by centrifugation.

The primase polypeptides were purified using Ni-NTA affinity chromatography. The bound proteins were eluted using a linear imidazole gradient ranging from 20 mM to 250 mM. Fractions containing the primase polypeptides were pooled and dialyzed overnight at 4 °C into buffer containing 30 mM MES (pH 6.5), 50 mM NaCl, and 3 mM BME. The sample was then further purified using a MonoS column (Amersham Biosciences) equilibrated in the same buffer and eluted with a linear gradient to 1 M NaCl.

Limited Proteolysis

p48/p58 and p48 primase preparations were exchanged into buffer containing 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 3 mM BME. Digestions using proteinase K were performed at room temperature using a 1:1000 ratio of protease:protein. At several time points over the 24-hour course of the digestion, an aliquot was removed, mixed with SDS sample loading buffer and boiled for 10 minutes. The aliquots were then stored on ice until the end of the experiment, when they were analyzed by SDS-PAGE. Bands of interest were excised and submitted for in-gel trypsin digestion and mass spectrometry analysis, or they were transferred to polyvinylidene fluoride (PVDF) membrane and submitted for Edman degradation.

Circular Dichroism (CD)

p58C was buffer exchanged into 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, and 3 mM BME. The protein concentration was ~10 μM. The far-UV CD spectrum was scanned at room temperature from 190-260 nm using a Jasco J-180 spectrophotometer (Easton, MD). Secondary structure content was estimated using the K2d web server (132).

Nuclear Magnetic Resonance (NMR)

To obtain ^{15}N-enriched p58C, the protein was expressed in an M9 minimal media with [15]NH$_4$Cl as the nitrogen source. The protein was buffer exchanged into 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 3 mM BME, and 5% D$_2$O. [15]N-^{1}H heteronuclear single quantum coherence (HSQC) spectra were acquired at 25 °C on a Bruker Avance 600 MHz spectrometer equipped with a cryoprobe. A total of 64 scans were recorded in the direct (^{1}H) dimension for each of the 128 points sampled in the indirect (^{15}N) dimension. The data were processed using NMRPipe (133) and analyzed and displayed using NMRView (134).

UV-Visible Spectrophotometry

Spectra were recorded using a Varian Cary 100 Bio spectrophotometer. Samples were scanned from 550 nm to 250 nm at room temperature in buffer containing 20 mM Tris (pH 7.5), 50 mM NaCl, and 3 mM BME.

Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

Concentrated p58C was diluted to 10 mL in 99% chelex H_2O and 1% nitric acid (optima). This gave a final p58C concentration of 2.16 μM. A blank sample containing buffer but no protein was prepared in the same manner. The sample was submitted to the Environmental Engineering facility at Vanderbilt University for ICP-MS analysis. The Fe57 content was measured and the [Fe]/[p58C] ratio was calculated.

Electron Paramagnetic Resonance (EPR)

Samples were prepared at 0.5 mM p58C in 20 mM MES (pH 6.5) and 50 mM NaCl, then freshly oxidized with 2.5 mM potassium ferricyanide. Experiments were collected by M. Nilges at the Illinois EPR Research Center. Spectra were recorded at X-band (~9.05 GHz) on a Varian E-122 spectrometer. The data were acquired from frozen glasses at ~15 K using an Air Products Helitran cryostat with liquid helium. The magnetic fields were calibrated with a Varian NMR Gauss meter and the microwave frequency was measured with an EIP frequency meter.

Multiple Sequence Alignment

Amino acid sequences of p58 from *H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*, and *S. cerevisiae* were aligned using ClustalX (118).

Primase Assay

The activity of wildtype and mutant primases was tested on M13mp18 ssDNA (USB Corp., Cleveland, OH). Reactions (20 μL) contained 0 to 8 pmol of primase, 100

ng of M13 DNA, in reaction buffer (30 mM HEPES-KOH (pH 7.9), 1 mM DTT, 7 mM Mg-acetate, 4 mM ATP, 0.2 mM UTP, 0.2 mM GTP, 0.01 mM CTP) and 20 μCi of [alpha-^{32}P] CTP (3000 Ci/mmol; Dupont NEN, Boston, MA). Reactions were assembled on ice and incubated at 37 °C for 90 min. Reaction products were precipitated with 2% NaClO$_4$ in acetone, washed with acetone, and dried. Reaction products were dissolved in formamide loading buffer (45% vol/vol formamide, 5 mM EDTA) at 65 °C for 10 min and resolved by denaturing 20% polyacrylamide gel electrophoresis for 4-5 h at 500V, monitored using 0.08% wt/vol xylene cyanol and 0.08% bromophenol blue as markers. The reaction products were visualized and quantified by phosphorimaging.

CHAPTER III

AN INTERACTION BETWEEN THE p68 SUBUNIT OF DNA POLYMERASE
ALPHA AND SV40 T ANTIGEN

Introduction

DNA polymerase alpha (pol-prim) is an essential component of the eukaryotic replication machinery, serving to lay down chimeric RNA-DNA primers. The function of the p68 subunit of pol-prim is not yet established. Previous studies have established that while p68 is not required for polymerase or primase enzymatic activities, it is essential for SV40 DNA replication (63). These authors showed that a trimeric pol-prim construct lacking the p68 subunit was defective in both initiation and primer elongation.

p68 has no known catalytic activity but has been shown to be involved in protein-protein interactions in both eukaryotic DNA replication and the SV40 model system (59, 64, 65). Protein interactions have been mapped to different regions of p68, and among these studies, the N-terminal 240 residues have been shown to interact with SV40 Tag (59). The functional relevance of this region is underscored by studies showing that residues 141-160 contain sites that are phosphorylated by cyclin dependent kinases (67), which presumably regulates p68 function in a cell-cycle dependent manner. More detailed investigations of these interactions are required in order to understand the function of p68.

Our laboratory and others have observed that the majority of proteins that comprise DNA processing machinery exhibit multiple contact sites between structural domains. These domain-domain contact points can be studied to build a view of the

47

structural basis for function. For example, the domain structures of RPA and Tag,
combined with studies of interactions using NMR spectroscopy and site-directed
mutagenesis, have allowed for the mapping of specific binding interfaces (53, 55). The
structure-based design of mutants deficient in binding provided a highly efficient
approach to understand the function of specific protein-protein interactions in the context
of the SV40 replisome (53, 55). Our laboratory is currently using similar approaches to
study the function of pol-prim.

Results

p68(1-107) Interacts with Tag HD

In an effort to more finely map the portion of the N-terminal region involved in
interacting with SV40 Tag, a series of yeast two-hybrid experiments were conducted. Of
the N-terminal 240 residues of p68 previously shown to interact with Tag (59), the N-
terminal 107 residues were predicted to be structured using secondary structure
prediction. We first sought to determine if p68(1-107) could interact with Tag and to
identify which Tag domain was responsible for the interaction. A yeast two-hybrid
screen revealed that p68(1-107) interacted with Tag, while p68(108-598) did not (Figure
3.1). Additionally, Tag(357-627), which corresponds to D2 and D3 of the helicase
domain (HD), was sufficient for the interaction. GST-pulldown experiments were then
performed with GST-Tag(357-627) and His$_6$-p68(1-107) to confirm that p68(1-107)
physically interacts with Tag(357-627) (data not shown). The finding that the Tag HD is

the domain responsible for the interaction is not surprising given that an antibody that

recognizes this domain prevents binding between Tag and pol-prim (135).
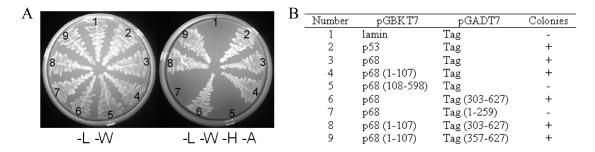
Figure 3.1. p68(1-107) interacts with Tag(357-627). (A) Yeast two-hybrid experiment with p68 and control constructs fused to the Gal4 DNA-binding domain. The Tag constructs were fused to the Gal4 activation domain. Colony growth was observed on selective media (Leu (L), Trp (W), His (H), and Ade (A)). (B) Table summarizing data from A. Data collected by H. Huang, Fanning laboratory.

p68(1-107) is Essential for SV40 DNA Replication

In order to further evaluate the role of the p68(1-107) in SV40 DNA replication, a truncated pol-prim lacking the N-terminal 107 residues of p68 (p68ΔN pol-prim) was purified and tested in a monopolymerase SV40 replication assay (Figure 3.2). Similar to the observation of the trimeric pol-prim, p68ΔN pol-prim was defective in SV40 DNA replication. Thus, the N-terminal 107 residues of p68 have an important functional role in SV40 DNA replication.
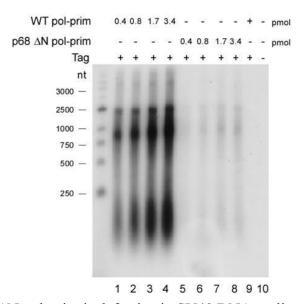
Figure 3.2. p68ΔN pol-prim is defective in SV40 DNA replication. WT (lanes 1-4) and p68ΔN (lanes 5-8) pol-prim were tested in the monopolymerase assay for SV40 initiation coupled with elongation. Reaction products were visualized by alkaline agarose electrophoresis and autoradiography. Data collected by H. Zhang, Fanning laboratory.

p68(1-107) is Structured

Given the evidence that p68(1-107) has a direct functional role in SV40 DNA replication, I set out to more fully characterize this construct. p68(1-107) was initially expressed using the pET32a vector. This vector contains three N-terminal tags: a thioredoxin tag, a His$_6$ tag, and an S-tag. The construct was expressed in BL21 (DE3) cells and purified using Ni-NTA chromatography. The tags were removed via enterokinase digestion. p68(1-107) was then purified away from the tag using a SourceQ column, giving a sample of sufficient purity for further studies (Figure 3.3.A).
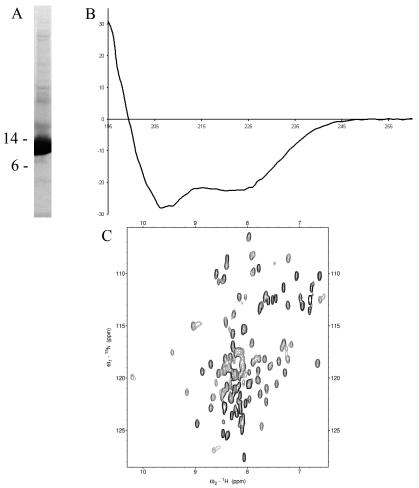
Figure 3.3. Structural analysis of p68(1-107). (A) SDS-PAGE of purified p68(1-107) stained with Coomassie blue. (B) Far-UV CD spectrum of p68(1-107) shows characteristic alpha-helical minima. (C) ^{15}N-^{1}H HSQC spectrum of p68(1-107) contains dispersed peaks indicative of a folded protein.

Circular dichroism was used to assess secondary structure content in p68(1-107). Similar to what was observed for p58C in Chapter II, minima were present at 208 nm and 222 nm in the far-UV CD spectrum (Figure 3.3.B), which indicates the presence of alpha helical secondary structure and suggests that there is a reasonable probability that p68(1-107) is folded.

Given the promising CD data, p68(1-107) was transferred to another expression vector to simplify the purification process. The construct was inserted into the pBG100

vector (L. Mizoue, Center for Structural Biology, Vanderbilt University), which contains an N-terminal His_6 tag with an H3C protease cleavage site. The pBG100-p68(1-107) fusion protein was expressed in BL21 (DE3) cells and purified to homogeneity using a standard Ni-NTA chromatography protocol. To further analyze the protein, p68(1-107) was expressed in ^{15}N enriched minimal media, and an ^{15}N-^{1}H HSQC was recorded. The spectrum contained a significant number of dispersed peaks with narrow linewidths which indicated the presence of a folded, globular structure (Figure 3.3.C). Together, the CD and NMR data suggest that p68(1-107) contains a globular domain. However, it was also clear from the presence of a significant number of strong peaks between 8.1 and 8.5 ppm in the ^{1}H dimension, that a considerable portion of the construct was not stably folded. A calculation of the CD mean residue ellipticity indicated only ~33 of the 107 residues are involved in helical structures. The estimate from the NMR analysis was that about 1/3 of the construct was unfolded. These observations prompted efforts to determine if the minimal Tag binding region was smaller than p68(1-107). Based on previous observations of interactions between many DNA processing proteins involving stable folded domains, our next step involved more precisely delineating the putative folded domain in the N-terminal region of p68, which we term p68N.

p68(1-78) Contains the Structured Region of p68(1-107)

The initial approach used to try to identify p68N was to assign the backbone NMR resonances of p68(1-107). Unfortunately, p68(1-107) was not sufficiently stable to undertake these experiments: the protein degraded after several days, even in the presence of protease inhibitors. This observation of sample degradation suggested that

p68(1-107) contained several unstructured residues at either the N- or C-terminus. This prompted a detailed analysis of the p68(1-107) sequence. Our initial analysis of the p68(1-107) sequence using several secondary structure prediction algorithms predicted that helical elements extended through the entire sequence (Figure 3.4). However, these analyses were performed by restricting the sequence analysis to the residues of interest. Since secondary structure prediction algorithms consider a sliding window of residues to obtain the most accurate results, these initial analyses could be improved upon. In fact, repeating the analysis using the entire p68 amino acid sequence produced a different result: no element of secondary structure was predicted beyond Ser75; thus the C-terminal tail of p68(1-107) is likely to be unstructured (Figure 3.4). Consequently, two new constructs, p68(1-78) and p68(1-87) were designed to eliminate the unstructured C-terminal tail, which would presumably result in a more stable construct. p68(1-87) was designed based on mass spectrometry data collected on a fragment produced by p68(1-107) degradation, which had a mass consistent with residues 1-87.

```
MSASAQQLAEELQIFGLDCEEALIEKLVELCVQYGQNEEQ       40
----HHHHHHHHH------HHHHHHHHHHHHH----HHH       FL
----HHHHHHHHH------HHHHHHHHHHHHH-----HHH    1-107


MVGELIAFCTSTHKVGLTSEILNSFEHEFLSKRLSKARHS       80
HHHHHHHHHHH-------HHHHHHHHHHHHHHHHH-----       FL
HHHHHHHHH---------HHHHHHHHHHHHHH-------    1-107


TCKDSGHAGARDIVSIQELIEVEEEEEILLNSYTTPSKGS      120
-----------EE------------------------       FL
-----------HHHHHHHHHHH----                 1-107
```

Figure 3.4.   Secondary structure prediction of the N-terminus of p68. The secondary structure content of p68 was predicted using the Jpred server (136). The entire p68 sequence was analyzed (middle row, labeled "FL"), however only the N-terminal 120 residues are shown. Analysis of only residues 1-107 (bottom row, labeled "1-107") predicted a fifth helix. Secondary structure is listed below the sequence (α-helix as "H" and β-strand as "E"). The p68N construct is represented in blue, and the p68(1-107) construct contains the blue and green residues.

The p68(1-78) construct was selected for analysis first because it was the minimal size for p68N based on the structure prediction. The construct was sub-cloned, expressed and purified using the same protocol as for p68(1-107). Curiously, two protein bands were observed. Both forms co-purified in the primary Ni-NTA chromatography step (Figure 3.5.A). The lower MW band ran at a size consistent with the expected p68N monomer, whereas the higher MW band ran at a size consistent with a p68N dimer. The gels were run under denaturing conditions, and since the p68(1-107) construct contains three Cys residues, it was assumed the dimer arose from a stable inter-molecular disulfide. MALDI mass spectrometry (MS) confirmed that the molecular weights correspond to the monomer and dimer states. In fact, the two species could be separated using an S75 gel filtration column (Figure 3.5.B), and once separated, they did not re-equilibrate.
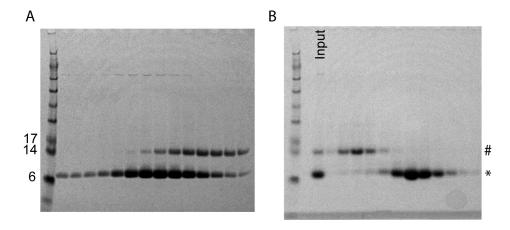
Figure 3.5. p68N purification. (A) Fractions eluted off a Ni-NTA column as the concentration of imidazole is increased using a linear gradient (from left to right). Two bands co-purify. Band * is the putative p68N monomer, and band # is the putative dimer. (B) Fractions from the Ni-NTA column were pooled and loaded onto an S75 gel filtration column. Two separate peaks eluted at retention volumes consistent with a monomer and dimer. The gel confirms the species were separated, and additional gels demonstrated that they do not re-equilibrate.

Since the overall yield of the protein was very high and the dimer was much less abundant than the monomer, all further analyses were performed with the monomer. Comparison of p68(1-78) with p68(1-107) were made by NMR. Figure 3.6.A shows an overlay of the ^{15}N-^{1}H HSQC spectra of these two constructs. The overlay demonstrates that all of the disperse peaks in p68(1-107) are present in the spectrum of p68(1-78); all p68(1-78) peaks that are missing from the spectrum of p68(1-107) are located in the central region, which contains peaks primarily from unstructured residues. This implies that p68(1-78) contains the structured p68N domain.

I then set out to further characterize the two bands observed in the gel. Since MS data suggested the upper band was twice the size of the lower band, an ^{15}N-^{1}H HSQC spectrum was acquired to determine if it was comprised of the same protein. The overlay of these two spectra (Figure 3.6.B) shows that the two species are in fact different forms

of the same protein as many of the dispersed chemical shifts are identical. This observation provides strong support for the proposal that the upper band is a p68N dimer.
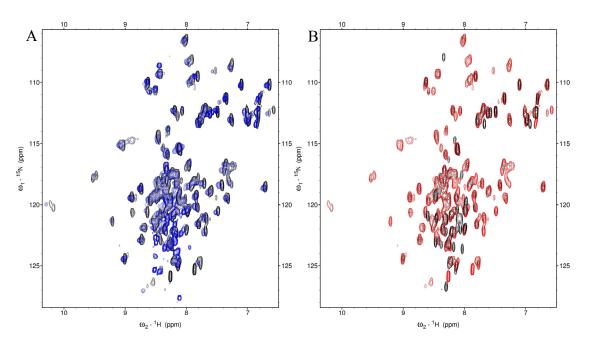
Figure 3.6. NMR characterization of p68(1-78). (A) The ^{15}N-^{1}H HSQC spectrum of p68(1-107) (blue) overlaid on top of p68(1-78) (black). All dispersed peaks from p68(1-107) are present in the p68N spectrum which indicates that the folded portion is present in both constructs. (B) The ^{15}N-^{1}H HSQC spectrum of the p68N dimer (black) overlaid on top of the monomer spectrum (red). The spectra overlay extremely well, confirming that the high MW band is a p68N dimer.

More detailed analysis of the HSQC spectrum of p68N revealed several residues that exhibited more than one peak, indicating sample heterogeneity (Figure 3.7.A). In the initial spectra, there was clear evidence for a major and minor species. Over time, the equilibrium distribution changed, shifting from a near equal distribution to almost only one form (Figure 3.7.A). This shift was not accompanied by degradation. Thus, even though the dimer p68N had been removed from the sample, the p68N monomer was not

homogeneous. The most likely explanation for this observation is a slow shift to a

reduced state of p68N, resulting in the loss of an intra-molecular disulfide.
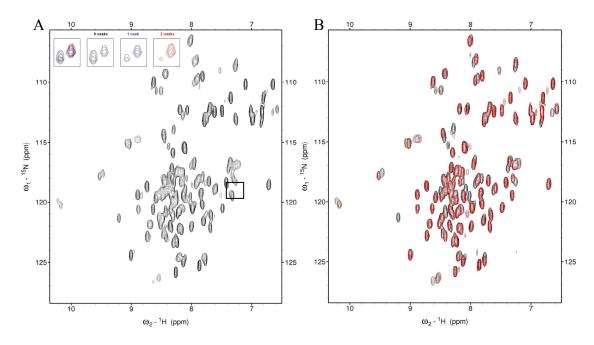
Figure 3.7. Redox-mediated heterogeneity of p68N. (A) ^{15}N-^{1}H HSQC spectrum of p68N. In the inset is a close-up view of the boxed peak. p68N spectra were recorded at 0 weeks, 1 week, and 2 weeks after purification. The spectra are overlaid (left column) and shown individually (right three insets) to highlight how the intensity of one peak decreases as the other increases. This indicates a shift in p68N sample population that occurs on a slow timescale. (B) Spectra of p68N in the presence (black) and absence (red) of BME. Purification and analysis of p68N in the absence of reducing agent resulted in a homogenous sample.

To further probe the cause of p68N heterogeneity, a sample was purified in the

absence of reducing agent. The ^{15}N-^{1}H HSQC spectrum of this p68N preparation was

homogeneous (Figure 3.7.C). Addition of 20 mM BME to the solution reproduced the

heterogeneous peaks, presumably due to incomplete reduction of the disulfide (Figure

3.7.C). The p68N peaks in the sample lacking reducing agents matched the initial

population (i.e. the peaks that slowly disappeared) observed previously in Figure 3.7.A.

These observations support the proposal that p68N forms an intramolecular disulfide

bond that it is slowly reduced over the two week timecourse when BME was present in the sample. Based on this finding, we omitted reducing agent from further p68N preparations and structural studies. Importantly, unlike p68(1-107), p68N was stable for 2-3 weeks at room temperature. We also noted that p68N was still somewhat susceptible to degradation after cleavage of the His_6 tag, so the tag was left intact for experiments requiring long lifetimes, such as the NMR structure determination.

Biophysical Characterization of the p68N-Tag HD Interaction

Since initial studies of the p68-Tag interaction identified p68(1-107) as the region responsible for binding Tag (Figure 3.2), the question remained whether p68N would be sufficient for the interaction. It was possible that Tag could bind to the unstructured residues in p68(1-107). However, further yeast two-hybrid experiments confirmed that p68N was sufficient for the interaction with Tag (data not shown). Thus the interaction is mediated through two structured domains, similar to what has been observed for the other well studied protein-protein interactions in SV40 DNA replication.

In order to obtain a more detailed understanding of the interaction, isothermal titration calorimetry (ITC) was used to quantify binding parameters. Initial attempts to work with Tag(357-627) were unsuccessful, as the protein aggregated upon cleavage of the GST tag. This problem was alleviated by switching to the longer Tag(303-627) construct. This construct disrupts the Zn subdomain (D1) of the HD, and is primarily monomeric in solution (32). Tag(303-627) requires both a high ionic strength (150-250 mM NaCl) and the presence of a reducing agent (1 mM DTT) to remain soluble. Due to the difficulty of producing Tag(303-627) and its limited solubility, 50 μM Tag(303-627)

was placed in the ITC sample cell at 25 ºC, and 0.735 mM p68N was titrated into the cell.
Both proteins were in an identical buffer, which contained 25 mM Tris (pH 8.0), 250 mM
NaCl, and 1 mM DTT. The data were fit to a single site binding model to calculate a K_d
of 22 ± 1 µM, a ΔH of 1713 ± 20 kcal/mol, and a ΔS of 27.1 e.u. for the interaction
(Figure 3.8).
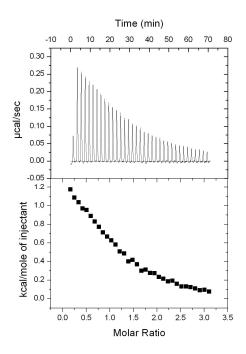
Figure 3.8. p68N interacts with Tag(303-627) with an affinity of 22 µM. p68N was
titrated into Tag(303-627) and was monitored using ITC. The raw data is shown on top,
and the converted data is shown at bottom. The data were fit to a single site binding
model to calculate the dissociation constant.

Mapping the Binding Surface

The solution structure of p68N was determined using NMR methods as will be
described in detail in Chapter IV. This structure was used in combination with NMR
titration data to map the p68N surface involved in the Tag interaction. NMR is a

powerful tool for studying protein-protein interactions (137). A common approach used to study protein-protein interactions is the chemical shift perturbation assay (CSPA). In this experiment, one protein is ^{15}N-labeled and an ^{15}N-^{1}H HSQC spectrum is recorded. The unlabeled binding partner (which produces no signals in the HSQC spectrum) is then added to the sample. The chemical shift is a very sensitive parameter, so peaks corresponding to residues at the binding interface will be selectively perturbed. Notably, structural changes induced by the binding event will also cause chemical shift changes, so the assay requires careful consideration of both factors. If a high resolution structure and resonance assignments are available, the binding interface can be mapped onto the structure by identifying the residues that are perturbed when the binding partner is added to the solution (e.g. (55)).

There are many instances where the standard CSPA is not practical however, including cases of intermediate exchange where many signals are broadened beyond detection or in the formation of large complexes (137). The p68N-Tag HD interaction is an example of the latter case, as the complex is ~48 kDa, which is large by NMR standards. Thus, as unlabeled Tag (303-627) is titrated into ^{15}N-labeled p68N, a large majority of p68N peaks are broadened beyond detection at a molar ratio of 1:1 (Figure 3.9). Consequently, although there is unambiguous evidence for binding, no specific binding interface can be defined based on CSPA data alone.
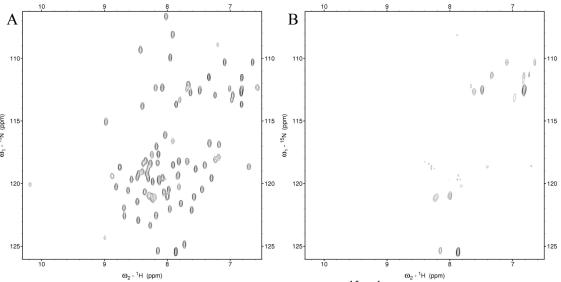
Figure 3.9. Tag-binding broadens p68N peaks. (A) ^{15}N-^{1}H HSQC of p68N alone. (B) ^{15}N-^{1}H HSQC of p68N with an equimolar amount of unlabeled Tag (303-627).

Differential line broadening (DLB) is a less frequently used and more complicated approach that can be applied to systems such as p68N-Tag(303-627). This technique involves the addition of sub-stoichiometric ratios of ligand to observed protein (138). When a large, unlabeled protein is added to the smaller, labeled protein, the relaxation rate of the small protein is dramatically increased, causing line broadening. This, in effect, renders the bound state invisible to NMR. At substoichiometric amounts however, line broadening can be used to indirectly assess resonance shifts caused by binding. When there is a difference in chemical shift between the free and bound state, the peak is more severely broadened due to exchange effects. Since only peaks at the binding interface have a chemical shift difference between the two states, severe line broadening can be used as an indicator for involvement at the protein-protein interface (reviewed in (137)).

DLB was used to study the p68N-Tag interaction. First, an HSQC spectrum of 100 μM p68N alone was recorded. Tag (303-627) was then added to the sample at 0.1:1 and 0.3:1 Tag:p68N molar ratios, and additional HSQC spectra were recorded. The expected general decrease in p68N peak intensities was observed as Tag was added to the solution (Figure 3.10.A). However, certain peaks appeared to be selectively broadened more drastically relative to the average. Since Tag(303-627) has only limited solubility, buffer conditions in these experiments were not the same as those used for resonance assignments. The buffer for the DLB experiment contained 25 mM Tris (pH 8.0), 150 mM NaCl, and 3 mM BME, while the structural studies were conducted in 20 mM sodium phosphate (pH 6.5) and 50 mM NaCl. Consequently, some of the p68N peaks could not be unambiguously assigned: Met1-Gln13, Cys19, Leu27, Val28, Cys31, Gln36, His53-Val55, Leu57, Ser59, Glu60, Leu62, Ser64, Glu66, His67, Leu70 and Lys76. While at least sixteen of these residues have peaks present in the HSQC spectrum which could not be assigned, the remaining residues are likely broadened beyond detection due to solvent exchange at the higher pH.

In order to make a detailed assessment of the effects of the titration, peak intensities were measured for all assigned resonances in all three spectra. Peak intensities were measured as opposed to peak volumes because the accuracy of volume measurements is lower than intensities and intensities are much easier to measure. Figure 3.10.B and 3.10.C shows a plot of the intensities versus the sequence, normalized against the control spectrum without Tag(303-627). Ten peaks were found to be selectively broadened when Tag(303-627) was added to the sample based on reduction in relative intensities that exceeds one standard deviation below the mean.

The residues identified in the DLB experiment as significantly perturbed include: Ile14, Phe15, Leu17, Glu20, Leu45, Ile46, Ala47, Cys49, Gly56, and Ser71. Given that the effect of dynamic line broadening was subtle, we repeated this experiment using Tag(251-627) Val350Glu/Pro417Glu obtained from Xiaojiang Chen's laboratory. This double point mutant is also a monomer in solution but retains the Zn domain (X. Chen, unpublished data). The two experiments produced very similar results, confirming our interpretation of which residues were significantly perturbed.
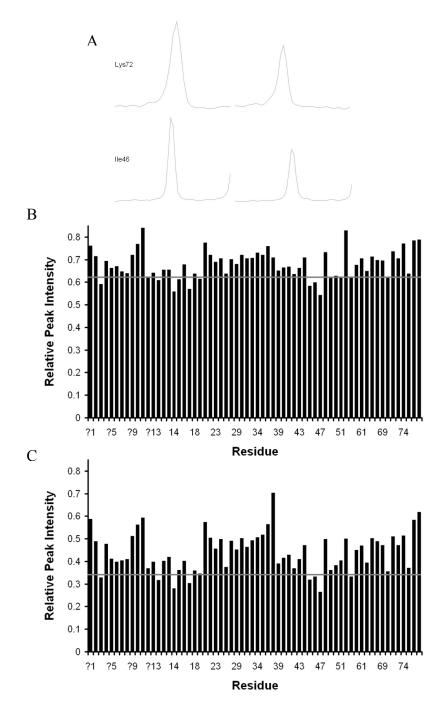
Figure 3.10. Tag(303-627) differentially broadens p68N signals. (A) Selected 1D slices of ^{15}N-^{1}H spectra without Tag (left) and at a ratio of 0.1:1 Tag(303-627):p68N (right) Although subtle, the I46 peak experiences a more significant intensity decrease. Peak intensity at Tag(303-627):p68N ratios of 0.1:1 (B) and 0.3:1 (C) are plotted relative to the intensity without Tag present. Peaks that could not be assigned (left) are underlined with a dashed line. Residues with peak intensities below the gray line (mean minus one standard deviation) were considered significant.

Figure 3.11 shows the perturbed residues mapped onto the p68N. A cluster of effected residues around the C-terminal ends of helices I and III suggest a possible binding surface. Ser71, and to a lesser extent Glu20, were remote from the potential interface and these are assumed to arise from structural perturbations allosterically induced by the binding of Tag(303-627).
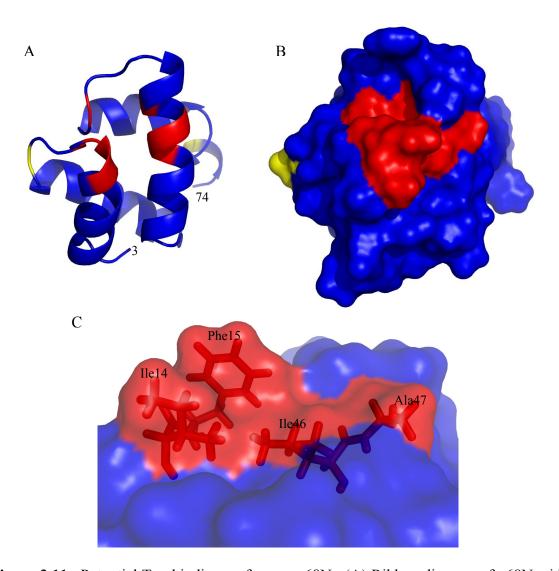
Figure 3.11. Potential Tag-binding surface on p68N. (A) Ribbon diagram of p68N with significant residues identified from the DLB experiment colored in red. Glu20 and Ser71 are colored in yellow as they are remote from the other perturbed residues. (B) Same as in A, but with a surface representation. (C) Close-up view of the surface, which is primarily hydrophobic and is composed of Ile14, Phe15, Ile46, and Ala47.

The potential binding surface is flat, primarily hydrophobic, and flanked by helix I and helix III (Figure 3.11.C). Ile46 is primarily buried, but the CG2 methyl group is surface exposed. Ile14, Phe15, and Ala47 are solvent exposed, and thus are good candidates for site directed mutagenesis. These residues, as well as additional residues near this surface are being mutated to confirm our analysis of the Tag binding site, and for subsequent use in functional assays to test the importance of this contribution to the interaction of p68 with SV40 Tag.

p68N mutants will be initially assayed for Tag-binding using a yeast two-hybrid assay. Those p68N mutants defective in Tag-binding will be assessed for structural integrity by collecting a ^{15}N-^{1}H HSQC spectrum. If the mutation does not perturb p68N structure, there should be a few chemical shift perturbations corresponding to residues near the mutation, but overall, the spectrum should match that of the WT. The binding affinity for the interaction between the p68N and Tag HD will be determined using ITC in order to quantify the reduction in affinity. These mutants will then be incorporated into the full-length pol-prim tetramer for functional assays, such as the monopolymerase assay (Figure 3.2) and for the ability of Tag to relieve RPA inhibition of pol-prim on a ssDNA template.

Experimental Procedures

Yeast Two-Hybrid Assay

Tag fragments were fused to the Gal4 activation domain in the pGADT7 vector (Trp selection). The control proteins and p68 fragments were fused to the Gal4 DNA

binding domain in the pGBKT7 vector (Leu selection). These plasmids are part of the Matchmaker Two-Hybrid System 3, designed by Clontech (Mountain View, CA) (139). The plasmids were co-transformed into the yeast AH109 strain which contains three reporter genes: His3, Ade2, and LacZ. The cells were grown for three days on -Leu -Trp plates. Positive colonies were streaked onto -Leu -Trp plates and -Leu -Trp -His -Ade plates. Cells were grown for three days and analyzed for the presence of colonies.

Monopolymerase Assay

The monopolymerase assay was carried out as previously described (63). Briefly, increasing amounts of WT and Δp68N pol-prim were incubated with SV40 origin dsDNA, Topo I, 1000 ng RPA, 1200 ng Tag, and 0.10 mM dNTPs (including 3 μCi of [α-^{32}P]dATP. Reaction products were precipitated with 2% NaClO$_4$ in acetone, washed with acetone, and dried. Products were then resuspended in loading buffer and analyzed by agarose gel electrophoresis and autoradiography.

p68N Construct Design

p68(1-107), p68(1-87), and p68(1-78) insert DNA was amplified from the Trx-tagged p68(1-107) construct obtained from the Fanning laboratory. Oligonucleotides used to amplify the insert DNA are listed in Appendix A. The inserts were then cloned into the pBG100 vector (L. Mizoue, Center for Structural Biology, Vanderbilt University) using NdeI and BamHI restriction enzymes. This vector introduces an N-terminal H3C protease-cleavable His$_6$ tag.

p68N Expression and Purification

The p68N plasmid was transformed into the BL21 (DE3) expression cell line. 10 mL LB was then inoculated and grown overnight at 37°C. The following day, the culture was added to 1 L LB and grown at 37°C. Protein expression was induced with 1 mM IPTG when the culture reached an OD_{600} of ~0.6. Cells were then harvested four hours post-induction. Isotopically labeled p68N was grown in a similar manner except that 1 L M9 minimal media was substituted for the LB. The minimal media was supplemented with ^{15}N-NH$_4$Cl and/or ^{13}C-glucose for ^{15}N/^{13}C incorporation as required.

Pelleted cells were resuspended in lysis buffer containing 50 mM Tris-HCl (pH 8), 300 mM NaCl, 20 mM imidazole, 1% Nonidet P-40 (NP-40), 0.5 mg/ml lysozyme, and one Complete Mini EDTA-Free protease inhibitor cocktail tablet (Roche). As mentioned in the results section, initial purifications also contained 3 mM BME, however this component was found to introduce sample heterogeneity and was subsequently removed. Cells were lysed by sonication at 4 °C. Insoluble material was removed by centrifugation.

p68N was purified using Ni-NTA affinity chromatography. The bound proteins were eluted using a linear imidazole gradient ranging from 20 mM to 250 mM. Fractions containing p68N were pooled and dialyzed overnight at 4°C into NMR buffer, which contains 20 mM sodium phosphate (pH 6.5) and 50 mM NaCl. The sample was then concentrated to 3 mL and loaded onto an S75 gel filtration column. Fractions containing p68 were pooled and concentrated to 0.5 mL, to give a p68N NMR sample of ~0.75-1 mM.

CD

10 μM p68(1-107) was analyzed by CD as described for p58C in Chapter II. p68(1-107) was buffer exchanged into 20 mM sodium phosphate (pH 7.5), 50 mM NaCl, and 3 mM BME. The far-UV CD spectrum was scanned at room temperature from 190-260 nm using a Jasco J-180 spectrophotometer (Easton, MD). Secondary structure content was estimated using the K2d web server (132).

ITC

p68N and Tag (303-627) were buffer exchanged into 25 mM Tris (pH 8.0), 250 mM NaCl, and 3 mM BME. Tag (303-627) was concentrated to 50 μM, and p68N was concentrated to 730 μM. Data were collected at 25 ºC using a MicroCal VP-ITC and analyzed using the accompanying software. Tag (303-627) was placed in the sample cell, and p68N was injected during the run. The data were fitted to a single site binding model to generate calculate several thermodynamic values, including K_d.

NMR

^{15}N-^{1}H HSQC spectra were recorded at 25 ºC on a Bruker Avance 500 MHz spectrometer equipped with a cryoprobe. ^{15}N-labeled p68N and unlabeled Tag (303-627) were buffer exchanged into 25 mM Tris (pH 8.0), 150 mM NaCl, 3 mM BME, and 5% D_2O. The initial p68N concentration was 100 μM. ^{15}N-^{1}H HSQC spectra were recorded as described in Chapter II at the following Tag:p68N molar ratios: 0:1, 0.1:1, 0.3:1, 1:1. Spectra were processed using NMRPipe (133) and analyzed using SPARKY (140). The

data were processed with a sine-squared bell window function followed by zero filling

prior to the Fourier transformation in both dimensions.

CHAPTER IV

THE SOLUTION STRUCTURE OF THE N-TERMINAL DOMAIN OF THE p68
SUBUNIT OF DNA POLYMERASE ALPHA

Introduction

High resolution structures of SV40 DNA replication proteins have provided insight into many aspects of initiation, ranging from enzymatic activity to protein-protein interactions. For example, x-ray crystal structures of the Tag helicase domain in various nucleotide binding modes have provided a model for Tag helicase activity (38). Additionally, the structures of the RPA70AB domain and the Tag OBD combined with biophysical data, have allowed for the elucidation of the role of this domain-domain interaction in loading RPA onto ssDNA (53).

Chapter III describes studies of the interaction between p68N and Tag HD. NMR data was used to map the potential binding interface onto the p68N structure. Currently there is no published structure available for any pol-prim domain. Thus we sought to determine a high resolution structure of a pol-prim domain with the goal of further elucidating the structural basis for p68N-Tag HD interaction and ultimately for pol-prim activity in the SV40 replisome. This chapter describes the methods used to determine the p68N structure as well as a detailed analysis of the structure itself.

Results

Resonance Assignments

The critical first step in any detailed NMR analysis is the assignment of the resonances. The standard multi-dimensional heteronuclear NMR approach on isotopically enriched (^{13}C, ^{15}N) protein was used for p68N. The backbone and side chain resonances were assigned using standard triple resonance NMR experiments on ^{13}C, ^{15}N-enriched protein. The following five spectra were recorded to assign the backbone resonances: HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, and HNCO. From these data, 78/78 H_N, 78/78 N, 78/78 C_α, and 75/78 C' resonances were assigned. Of the three missing C' resonances, one is the C-terminal residue, and two are ambiguous due to overlapped H_N and N resonances of the following residues. The HNCO experiment used to assign the C' resonances transfers magnetization to the C' nucleus of the previous residue. An HCACO experiment could conceivably be used to obtain these last three assignments, but it was not performed because these were not required for the structure determination. Thus in cases of overlapped or missing H_N and N resonances, the C' resonance of the previous residue cannot be assigned. To demonstrate the quality of the 3D triple resonance data, Figure 4.1 shows paired strips from the HNCACB and CBCA(CO)NH spectra for Gly43-Ile46. A ^{15}N-^{1}H HSQC spectrum of p68N is shown in Figure 4.2 with all peaks assigned.
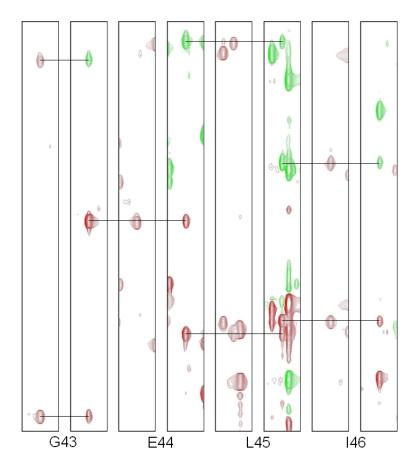
Figure 4.1. Strip plots from the 600 MHz 3D CBCA(CO)NH and HNCACB spectra. Spectra aquired for ~750 µM p68N at 25 ºC in a buffer containing 20 mM sodium phosphate (pH 6.5) and 50 mM BME. Strip plots from CBCA(CO)NH (left) and HNCACB (right) spectra are shown for residues Gly43-Ile46. Positive contours are colored red, and negative contours (which correspond to C_β resonances in the HNCACB) are colored green.
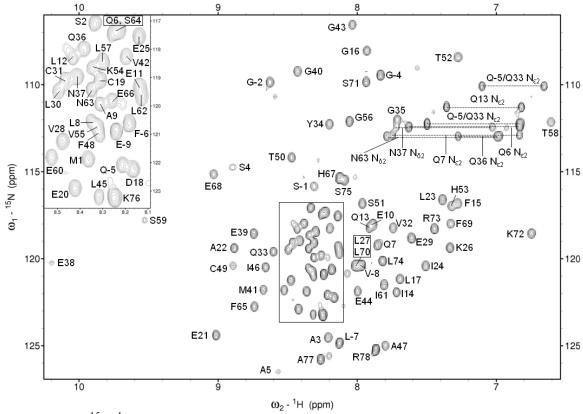
Figure 4.2. ^{15}N-^{1}H HSQC spectrum of p68N with assignments. Spectrum acquired for ~750 μM p68N at 25 ºC in a buffer containing 20 mM sodium phosphate (pH 6.5) and 50 mM BME. Residues with negative numbers are from the uncleaved His$_6$ tag. The central region is expanded in the upper left.

Additional spectra were recorded to complete the assignments of the side chain resonances. An HBHA(CO)NH spectrum was useful for providing many of the ^{1}H assignments for the C_α and C_β protons. The majority of the remaining side chain ^{13}C and ^{1}H resonances were assigned from (H)CC(CO)NH-TOCSY and H(CC)(CO)NH-TOCSY spectra. Aromatic side chain resonances were assigned using a combination of homonuclear NOESY and COSY experiments. In all, 96% of the side chain ^{13}C and ^{1}H resonances were assigned. The missing assignments are: Gln6 C_γ, Phe15 H_ε, Phe15 H_ζ, Lys26 H_ε, Gln33 H_γ, Gln33 H_ε, Gln36 H_γ, Gln36 C_γ, Phe48 H_ζ, His53 $H_{\varepsilon 1}$, Lys54 H_δ,

Lys54 H$_\varepsilon$, Leu57 H$_\gamma$, Thr58 H$_\alpha$, Ser59 C$_\beta$, Ile61 H$_{\gamma 2}$, Ile61 C$_{\gamma 1}$, Leu62 H$_\gamma$, Ser64 C$_\beta$, His67 H$_{\varepsilon 1}$, Arg78 H$_\delta$, Arg78 C$_\gamma$, and Arg78 C$_\delta$.

Structural Restraints

<u>Backbone Dihedral Angle Restraints</u>

The chemical shifts for N, C', C$_\alpha$, C$_\beta$, and H$_\alpha$ resonances were input into TALOS (141) to generate constraints on φ and ψ angles along the polypeptide backbone. TALOS compares chemical shifts to a database of known structures and chemical shifts to predict the backbone dihedral angles. In the case of p68N, TALOS predicted φ and ψ values for 61 out of the 78 residues (Table 4.5). These correspond well to the four helical elements ultimately found in the structure (Figure 4.3).

```
MSASAQQLAEELQIFGLDCEEALIEKLVELCVQYGQNEEQ    40
----HHHHHHHHHHH-----HHHHHHHHHHHHHH---HHH

MVGELIAFCTSTHKVGLTSEILNSFEHEFLSKRLSKAR     78
HHHHHHHHHHH----HHHHHHHHHHH-HHH------
```
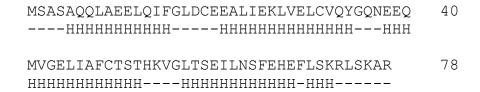
Figure 4.3. TALOS secondary structure prediction for p68N. "H" represents an α-helix and "-" represents no prediction.

<u>NOE Distance Restraints</u>

The following NOESY experiments were collected to generate NOE distance restraints: 2D ^{1}H NOESY, 3D ^{15}N-edited NOESY, 3D ^{13}C-edited NOESY, and 4D ^{13}C, ^{13}C-edited NOESY. Peaks were picked in the 2D and 3D experiments to give 3489 potential proton-proton connectivities. Rather than performing the very time consuming process of manually assigning each peak, we turned to CYANA (142) for automated

assignment. The CYANA approach involves an iterative process to assign the NOESY peaks, which is directly coupled with structure calculations. This corresponds to an automated version of "distance filtering" (e.g. (143)).

Of the 3489 peaks picked from the NOESY spectra, 2648 were eventually unambiguously assigned according to CYANA, and 987 distance restraints were generated. 533 of the restraints are short range (between protons on the same or sequential residue), 233 are medium range (between protons 2 to 4 residues apart), and 221 are long range (between protons greater than 4 residues apart). The distribution of NOEs is shown in Figure 4.4.A. This figure shows that the regions with the fewest long range NOEs (black bars) correspond to loops in p68N. This is expected, as regions with few NOEs typically map to flexible loops and termini in NMR structures. There is a good distribution of NOEs along the polypeptide chain, which is important to obtain an accurate tertiary structure. Some residues, such as Leu45 and Phe65, have a large number of NOEs because they are centrally packed in the hydrophobic core. Medium range NOEs are useful for defining secondary structure (Figure 4.4.B), while long range NOEs drive the overall fold of the protein. Helical elements of secondary structure are defined by i, $i+3$ ($d_{\alpha\beta}$, $d_{\alpha N}$) NOEs and specific assignment of an α-helix was based on i, $i+4$ ($d_{\alpha N}$) NOEs. The assignments of long range NOEs were verified manually, using the 4D NOESY when possible to ensure the accuracy of these critical distance restraints.
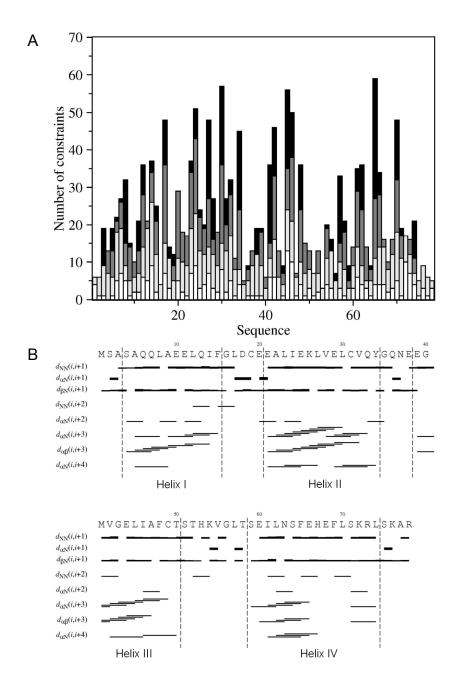
Figure 4.4. Distribution of NOEs observed for p68N. (A) Plot of number of NOEs for each residue. White bars represent intramolecular NOEs, light gray bars are sequential NOEs, dark grey bars are medium range NOEs, and black bars are long range NOEs. (B) Plot of sequential and medium range NOEs. Vertical dotted lines are drawn to delineate the four different helices that can be assigned from the analysis of this data.

77

TALOS predicted four helices corresponding to Ala5-Phe15, Glu21-Tyr34, Glu38-Thr52, and Ser59-Lys72. The medium range NOEs also define four helices, corresponding to Ser4-Phe15, Glu21-Tyr34, Glu39-Thr50, and Ser59-Leu74. Both methods for determining secondary structure are in good agreement as the boundaries for the four helices are extremely similar. Interestingly, both TALOS and the NOESY data suggest a break in helix IV at Phe69.

In addition to distance restraints, CYANA was also valuable for generating 18 sets of stereospecific assignments. The software uses the GLOMSA algoritm (144), taking into account NOE distance restraints, dihedral angle restraints (not directly measured for p68N), and the features of the structure. Stereospecific H_β assignments were obtained for 15 residues: Leu8, Phe15, Leu17, Glu20, Leu23, Leu27, Leu30, Tyr34, Glu38, Leu45, Phe48, Phe65, Phe69, and Leu70. Three additional stereospecific assignments were obtained for Glu25 H_γ, Val32 H_γ, and Gly43 H_α. Only the Val32 methyls could be stereospecifically assigned at the current stage of analysis.

In initial p68N structures, helix II was bent, and consistent dihedral angle violations in structures arose for Leu27 and Val28. This was surprising given that both TALOS and medium range NOEs suggested a continuous helix in this region. This kink in helix II was retained through several rounds of CYANA calculations and even after restrained molecular dynamics refinement. Analysis of the distance violations did not reveal "bad" NOEs; indeed, there were many restraints forcing the helix to bend. Our initial concern was that resonances had been misassigned, especially given that Leu27 is overlapped with Leu70 in the ^{15}N-^{1}H HSQC (Figure 4.2), which would complicate the assignment of the Lys26 side chain resonances. However, careful re-examination of the

assignment data did not turn up any misassigned resonances. Returning to the NOEs, a detailed manual analysis of the long range NOEs was performed. This revealed that a particular resonance that had several long range NOEs was being inconsistently assigned in CYANA calculations. Further inspection identified this resonance as Met41 H_ε, which had not been previously assigned. Assigning this resonance and rerunning CYANA resulted in structures with no kinks in helix II as expected. This example highlights one of the severe drawbacks of the network anchoring approach used by CYANA. While it allows for the structures to more quickly converge, this approach also generates the possibility of amplifying the negative (or adverse) effects of incorrect or incomplete assignments, which in turn results in an inaccurate structure. Our assignment completeness was quite high at 96%, which one might anticipate is sufficient to avoid this problem. The example of p68N makes clear that this is not the case. While the value of automating the NOE assignment process is clear, it is absolutely essential to recognize the potential for this process to generate erroneous structures. Manual examination of NOEs after they are assigned, in particular the critical long range NOEs, is recommended as one means to check for errors in the NOE assignments made by automated approaches.

Restrained Molecular Dynamics (rMD)

After CYANA analysis was completed, the 50 structures with the lowest CYANA target energy were used as starting structures for rMD simulations using AMBER (145). In order to optimize the protocol and weighting factors for the refinement of p68N data, six rMD simulations were run with different distance restraint sets based on the CYANA output. These tests involved altering the upper limit for each restraint, multiplying the

CYANA output by 0.9, 0.95, 1, 1.05, 1.1, and 1.2. The objective of this test was to find a proper balance between restraining conformation based on experimental data and retaining the chemical accuracy of the local conformational features (bond lengths, bond angles, etc.) that is optimized by the force field alone.

In order to evaluate the six rMD simulations, the effect of the distance restraints on total energy, violation energy, and root mean square deviation (RMSD) was plotted versus weighting factor (Figure 4.5). The expected trend is that as the structures become over-restrained, the restraints will cause chemically unreasonable representations of the structure, driving the total AMBER energy up (Figure 4.5.A). On the other hand, if the restraints are too loose, then the protein will be poorly restrained and the RMSD of the ensemble rises (Figure 4.5.B). The curves in Figure 4.5 show that a 5% increase in the distance restraints is optimal, resulting in low total AMBER and violation energies without significantly inflating the RMSD. Thus the restraints generated by CYANA were loosened by 5% in the final rMD analysis.
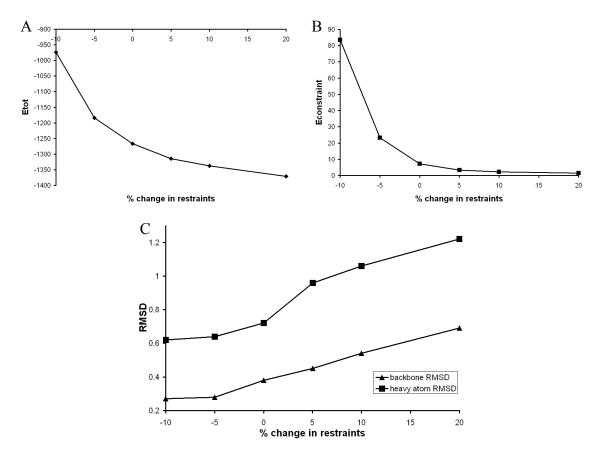
Figure 4.5. Effects of altering the upper bounds of NOE distance restraints. Six identical simulations were performed changing only the upper bounds on the distance restraints. The upper bounds generated by CYANA were factored as indicated along the x-axis of each plot. (A) Total energy, (B) violation energy, and (C) RMSD (for residues 5-68) are plotted versus the percent change to the restraints.

Analysis of the Structure

After completion of the rMD refinement, the 50 structures were ordered by AMBER restraint violation energy and the first 20 structures were selected for further analysis. The extreme termini of p68N (Met1-Ser2, Ser75-Arg78) did not contain any medium or long range NOEs (Figure 4.4.A) and as a result occupied a wide range of conformations after rMD refinement. These six residues were removed for clarity in all subsequent analyses. Figure 4.6 shows the representative ensemble of 20 p68N conformers along with a ribbon diagram of the representative single conformer (that

which is closest to the mean). p68N is seen to adopt a compact structure composed of

four helices, in good agreement with the secondary structure prediction (Figure 3.4).
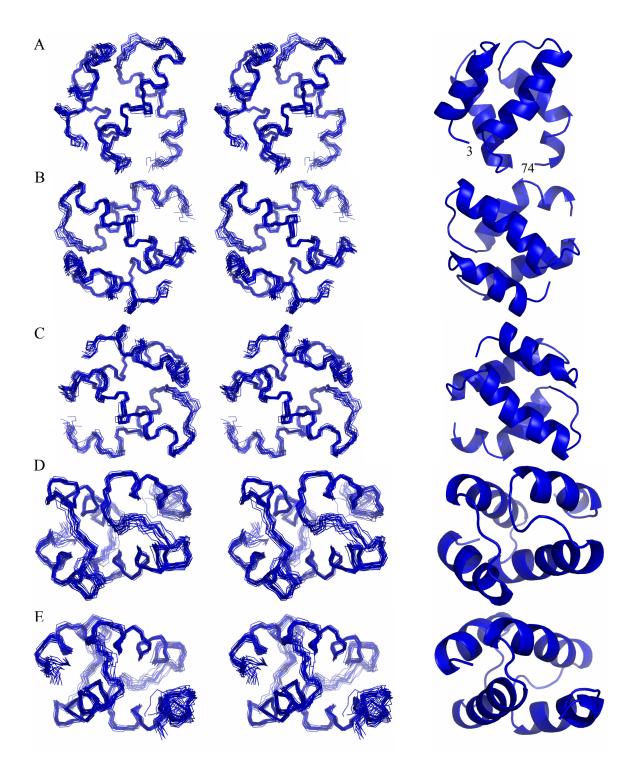
Figure 4.6. NMR solution structure of p68N determined at pH 6.5 and 25 ºC. (A) Ensemble of the conformers representing the structure of p68N (left, in stereo). Ribbon diagram of the structure closest to the mean structure, with the helical elements of secondary structure labeled (right). (B) Rotation of 90º along the Z-axis. (C) Rotation of -90º along the Z-axis. (D) Rotation of 90º along the X-axis. (E) Rotation of -90º along the X-axis.

The structural statistics for the ensemble (Table 4.1) shows that the structure of p68N is reasonably well defined, consistent with the average of 14.2 restraints/residue. The average total energy of -1315 ± 23 kcal/mol is largely negative indicating a stable globular structure, and the average constraint energy of only 3.4 ± 0.4 kcal/mol indicates excellent agreement with the experimental data. There were no distance constraints violated by more than 0.2 Å nor angle constraints violated by more than 5°. PROCHECK-NMR (146) analysis revealed 98% of the backbone psi-phi angles in the most favored and additionally allowed region, and only 2% in the generously allowed region of the Ramachandran plot. These measures indicate a high quality structure has been determined.

Table 4.1. Structural statistics for p68N.

p68N Structural Statistics	
Total restraints	1109
NOE restraints	987
Short range	533
Medium range	233
Long range	221
Dihedral angle restraints	122
Restraint violations (mean ± s.d.)	
Distance restraints ≥0.2 Å	0
Dihedral angle restraint violations ≥5°	0
Maximum distance restraint violation	0.13 ± 0.02
Maximum dihedral restraint violation	3.6 ± 0.7
AMBER energies (kcal/mol)	
Restraint energy	3.4 ± 0.4
Total energy	-1314 ± 23
Ramachandran statistics (%)	
Most favored	88
Additionally allowed	10
Generously allowed	2
Disallowed	0
RMSD from mean structure (Å)[1]	
Backbone atoms	0.53
Heavy atoms	1.02
Backbone atoms in helices	0.41
Heavy atoms in helices	0.90

[1]Residues used for RMSD calculations include Ala3-Leu74.

The RMSD from the mean structure for residues Ala3-Leu74 is 0.53 Å for the backbone atoms and 1.02 Å for all heavy atoms. The corresponding values for the residues in the four helices (Ala5-Phe15, Glu21-Tyr34, Glu38-Thr52, and Ser59-Glu68) are 0.41 Å and 0.91 Å, respectively. These RMSDs indicate that the structure is reasonably well converged. Figure 4.7 contains a plot of the RMSD from the mean structure per residue which shows the expected trend of lower RMSD in the helical regions, which reflects the distribution of restraints (Figure 4.4.A). Additionally, residues

that form the hydrophobic core of the protein are less variable (4.8.A), while surface

exposed residues have a greater spatial variability (4.8.B).
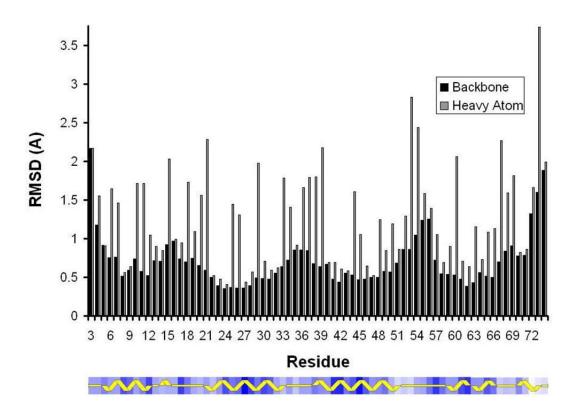
Figure 4.7. RMSD per residue. The backbone and heavy atom RMSD is plotted for each residue in p68N. Below the chart, the average secondary structure content is shown in yellow, with solvent accessibility shaded in blue (white is surface exposed, and the darkest shades of blue represent the most buried side chains, generated using PROCHECK-NMR (146)). As expected, the buried helical residues are regions of the protein with a low RMSD.
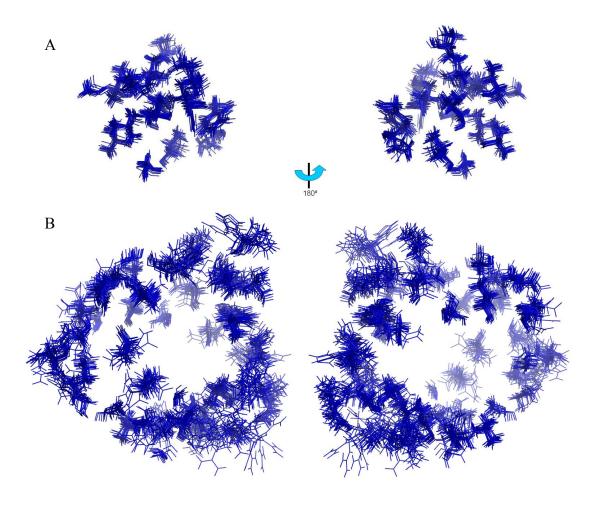
Figure 4.8. Resolution of side chains in the structure of p68N. (A) Side chains contributing to the hydrophobic core are well ordered with a low RMSD (0.93 Å). The view on the left displays p68N in the same orientation as in Figure 4.6. (B) Surface exposed side chains are more variable, with an RMSD of 1.36 Å.

    p68N is a small, and highly compact/near spherical globular domain with specific hydrophobic patches (Figure 4.9.A). The surface properties of p68N were examined to discern if there were any remarkable features. One feature is the potential Tag-binding site described in Chapter III, which is primarily hydrophobic surrounded by negative patches (Figure 3.11). p68N contains 13 Glu and Asp residues resulting in a pI of 4.6. These negatively charged residues are distributed evenly across the surface in the protein (Figure 4.9.B). This suggests that p68N would be unlikely to interact with negatively

charged molecules, such as DNA. On the other hand, the high negative charge may enable p68N to interact with DNA binding domains, in the manner of the p53-RPA70N interaction and others, which have been termed DNA mimic domains (50).
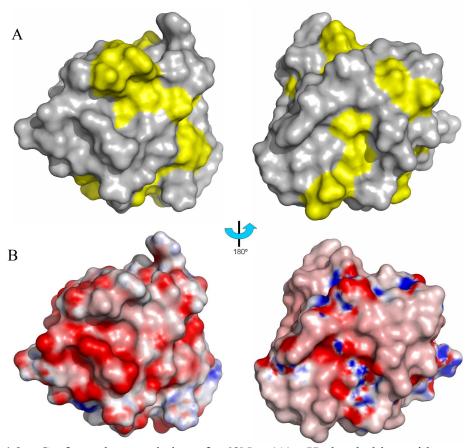
Figure 4.9. Surface characteristics of p68N. (A) Hydrophobic residues colored in yellow to highlight the hydrophobic surface. The orientation on the left is the same as shown in Figure 4.6.A. (B) Electrostatic potential with negative charge colored in red and positive charge colored in blue.

p68N lacks sequence similarity with other proteins. However, now that the structure has been determined, structure-based alignments can be used to identify other proteins that adopt a similar fold. Given the relatively simple fold of p68N, it is not surprising that a DALI server (147) query using the p68N structure revealed many similar

proteins. Interestingly, many of these proteins, including the closest match, RuvB, are ATP-binding proteins. In the case of RuvB, the p68N structure aligns with domain II, which contains the AAA+ sensor 2 motif (148). In addition to missing the key arginine in the sensor 2 motif, the p68N surface is negatively charged, and would likely repel ATP molecules. While p68 is not predicted to be an ATPase, it does interact with the AAA+ protein, Tag, and it is intriguing that it adopts a fold that is found in some of these proteins.

In conclusion, p68N is a compact domain that adopts a typical four-helix bundle architecture. There are distinguishing features on the surface of the protein, such as a diffuse negative charge, which suggest a function in binding but not catalysis. All evidence points to p68N being a protein-protein interaction domain of pol-prim.

Experimental Procedures

p68N Expression and Purification

p68N was expressed and purified as described in Chapter III. Briefly, p68N was overexpressed in BL21 DE3 cells, and then purified using Ni-NTA chromatography followed by gel filtration. Cells were grown in minimal media enriched with $^{15}NH_4Cl$ and/or ^{13}C-glucose as required. Growth in 1L of media typically produced a 500 μL NMR sample of ~750 μM.

NMR Spectroscopy

NMR spectra were recorded on Bruker Avance 500 MHz, 600 MHz, and 800 MHz spectrometers equipped with cryoprobes. Spectra were recorded at 25 °C. p68N NMR buffer contained 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 5% D_2O, and was supplemented with 0.02% NaN_3 for experiments lasting more than one week. Data were processed using NMRPipe (133) and analyzed using Sparky (140).

HSQC

^{15}N-^1H HSQC was obtained using 1024 points in the direct dimension and 128 points in the indirect dimension. Typically eight scans were collected, which results in an experiment time of ~23 minutes. The proton sweep width was set at 12 ppm. The nitrogen sweep width was set at 23 ppm using a carrier frequency of 116 ppm. The data were processed with a sine-squared bell window function followed by zero filling prior to the Fourier transformation in both dimensions.

Backbone Resonance Assignment Experiments

The following experiments were collected at 600 MHz to complete backbone resonance assignments: HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, and HNCO. The experimental parameters are listed in Table 4.2. The data were processed similar to the HSQC experiment, with a sine-squared bell window function followed by zero filling prior to the Fourier transformation in all three dimensions. The F3 (^1H) dimension was processed first, followed by F2 (^{15}N) and then F1 (^{13}C).

Table 4.2. Backbone resonance assignment experimental parameters.

Experiment	Data Points			Sweep Width			Time
	F3	F2	F1	F3	F2	F1	
HNCACB	1k	48	128	11.97 (H)	23 (N)	62 (C)	4d 1h
CBCA(CO)NH	1k	48	128	11.97 (H)	23 (N)	62 (C)	1d 9h
HNCA	1k	48	128	11.97 (H)	23 (N)	30 (C)	1d 8h
HN(CO)CA	1k	48	128	11.97 (H)	23 (N)	30 (C)	1d 9h
HNCO	1k	48	128	11.97 (H)	23 (N)	11.2 (C)	1d 8h

Side Chain Resonance Assignment Experiments

The following experiments were collected at 600 MHz to complete side chain resonance assignments: H(CC)(CO)NH-TOCSY, (H)CC(CO)NH-TOCSY, and HBHA(CO)NH. The experimental parameters are listed in Table 4.3. The data were processed using the same procedure as the backbone resonance assignment experiments.

Table 4.3. Side chain resonance assignment experimental parameters.

Experiment	Data Points			Sweep Width			Time
	F3	F2	F1	F3	F2	F1	
H(CC)(CO)NH-TOCSY	1k	48	160	11.97 (H)	28 (N)	10 (H)	3d 12h
(H)CC(CO)NH-TOCSY	1k	48	160	11.97 (H)	28 (N)	75 (C)	2d 19h
HBHA(CO)NH	1k	48	160	11.97 (H)	23 (N)	8 (H)	1d 18h

NOESY Spectra

The following NOESY experiments were collected: 2D ^{1}H NOESY (800 MHz), 3D ^{15}N-edited NOESY (600 MHz), 3D ^{13}C-edited NOESY (600 MHz), and a 4D ^{13}C, ^{13}C-edited NOESY (500 MHz) with the experimental parameters listed in Table 4.4. The 2D and 3D NOESY experiments were processed as described for the resonance assignment experiments. The 4D NOESY experiment additionally included linear prediction in both ^{13}C dimensions.

Table 4.4. NOESY experimental parameters.

Experiment	Data Points				Sweep Width				Time
	F4	F3	F2	F1	F4	F3	F2	F1	
1H NOESY			2k	1k			12.01 (H)	12.01 (H)	20h
^{15}N HSQC-NOESY		1k	48	192		11.97 (H)	28 (N)	11.97 (H)	3d 2h
^{13}C HSQC-NOESY		1k	48	192		11.97 (H)	75 (C)	11.97 (H)	3d 20h
^{13}C, ^{13}C-edited NOESY	1k	40	128	40	12.01 (H)	70 (C)	12.01 (H)	70 (C)	~6 d

Resonance Assignments

Backbone resonances were manually assigned using the previously mentioned triple resonance experiments. Briefly, for the HNCACB and CBCA(CO)NH spectra, strip plots were created in SPARKY (140), as shown in Figure 4.1. This was possible due to the extremely high quality of the 3D triple resonance data for this small protein. Since the HNCACB spectrum contains C_α and C_β peaks for both the intra-residue and sequential residues for each NH strip, connectivity chains can be constructed to connect the resonances of residues that are adjacent in the sequence. These chains, combined with the characteristic chemical shifts of some residues (Thr and Ser C_β resonances are far downfield, for example) allow for the sequence-specific assignments.

Side chain resonances were primarily assigned using the 3D TOCSY experiments. These experiments utilize a spin-lock to excite all ^{13}C nuclei within a side chain, then relaying this information to the NH of the sequential residue. Thus the side chain resonances can be quickly assigned to corresponding using the backbone resonances in the corresponding N and H_N strips. For example, the (H)CC(CO)NH-TOCSY spectrum contains all peaks seen in the CBCA(CO)NH spectrum plus additional peaks corresponding to additional side chain carbon resonances (Figure 4.10)
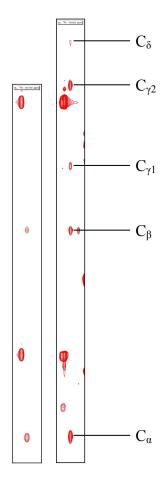
Figure 4.10. Strip plots of Ala47 from the 600 MHz 3D CBCA(CO)NH and (H)CC(CO)NH-TOCSY spectra. The CBCA(CO)NH spectrum is on the left, and the (H)CC(CO)NH-TOCSY spectrum is on the right. C_α and C_β peaks for Ile46 are in both spectra, but the (H)CC(CO)NH-TOCSY experiment also contains Ile46 $C_{\gamma1}$, $C_{\gamma2}$, and C_δ peaks.

Aromatic side chain resonances were assigned using COSY and homonuclear ^{1}H NOESY experiments, because the TOCSY experiments do not excite magnetization in the far downfield aromatic region. Aromatic protons are shifted downfield from the aliphatic protons, so they can be identified in 2D spectra on the basis of their unique chemical shifts and coupling patterns. NOEs from aromatic protons to the H_β resonances enable the connectivity to the other resonances with the given amino acid.

Determination of the Structure

<u>Dihedral Angle Restraints</u>

The goal of restraining dihedral angles is to assist in limiting the search of conformational space in the structure calculations. TALOS (141) is a program that predicts backbone dihedral angle values by comparing NMR chemical shift data to a database of published chemical shifts and structures. Backbone and C_β chemical shift values were input into the program TALOS to generate backbone φ and ψ dihedral angle restraints (Table 4.5). The restraints generated by TALOS were doubled in size prior to structure calculation to be conservative in restraining these angles and therefore not unduly biasing the structures.

Table 4.5. Output from TALOS calculations providing target values and standard deviations for the backbone φ and ψ dihedral angles.

Res.	Phi (°)	Δphi (±°)	Psi (°)	Δpsi (±°)
Ala3	-71.52	9.93	-30.29	21.79
Ala5	-60.93	5.15	-36.99	5.25
Gln6	-64.62	5.38	-40.94	5.77
Gln7	-63.68	5.41	-41.27	5.71
Leu8	-62.05	4.44	-42.07	5.32
Ala9	-63.37	5.8	-39.17	6.16
Glu10	-63.25	7.13	-44.92	7.6
Glu11	-65.51	6.6	-41.29	6.97
Leu12	-61.49	6.92	-41.86	6.58
Gln13	-66.29	6.28	-40.52	6.45
Ile14	-70.92	12.19	-38.86	8.9
Phe15	-92.63	15.79	4.27	14.89
Gly16	82.73	13.04	17.83	12.56
Leu17	-96.71	19.31	138.56	30.18
Glu20	-88.24	27.98	149.46	21.87
Glu21	-58.34	4.4	-39.35	12.45
Ala22	-60.92	3.93	-37.87	5.93
Leu23	-66.31	5.59	-42.61	5.59
Ile24	-63.52	7.92	-41.63	7.08
Glu25	-60.39	3.29	-41.97	7.03
Lys26	-64.91	3.39	-41.4	6.48
Leu27	-65.78	6.13	-41.2	5.07
Val28	-64.33	4.7	-42.77	5.06
Glu29	-61.86	4.9	-39.34	6.77
Leu30	-62.6	6.06	-39.04	4.86
Cys31	-63.4	8.04	-40.57	6.26
Val32	-66.6	3.99	-41.33	2.29
Gln33	-66.04	10.84	-37.48	6.37
Tyr34	-88.63	16.34	0.51	9.21
Gln36	-118.08	23.52	149.63	10.31
Asn37	-83.39	16.26	157.74	19.58
Glu38	-58.95	6.39	-36.17	6.18
Glu39	-62.44	3.53	-37.03	12.52
Gly40	-67.26	8.55	-38.2	6.71
Met41	-66.87	5.46	-40.28	6.67
Val42	-63.57	4.39	-44.39	4.44
Gly43	-64.83	4.17	-39.63	6.34
Glu44	-63.88	3.47	-41.01	8.47
Leu45	-59.51	6.16	-46.77	5.06
Ile46	-63.04	3.01	-41.62	3.92
Ala47	-62.29	6.28	-41.66	4.14
Phe48	-70.25	6.87	-40.3	8.41
Cyc49	-60.86	4.89	-44.29	7.58
Thr50	-64.13	5.25	-41.79	5.9
Ser51	-67.93	12.68	-34.46	9.75
Thr52	-103.18	13.26	3.3	8.53
Leu57	-103.4	27.9	144.04	15.48
Thr58	-102.94	24.1	162.68	11.63
Ser59	-62.55	8.1	-32.57	8.77
Glu60	-65.7	4.98	-39.43	6.06
Ile61	-64.32	3.05	-41.37	5.31
Leu62	-64.4	3.3	-41.04	8.02
Asn63	-65.95	6.82	-37.82	3.09
Ser64	-63.64	2.86	-39.62	9.09
Phe65	-60.53	9.15	-44.77	6.73
Glu66	-58.64	7.72	-40.98	13.48
His67	-67.85	8.55	-31.2	11.26
Glu68	-82.45	11.78	-29.77	17.3
Leu70	-59.31	5.21	-40.57	6.62
Ser71	-61.45	5.81	-30.72	12.24
Lys72	-78.7	14.55	-24.5	13.18

CYANA Calculations

      Distance restraints were generated by CYANA (142) using peak lists from the 2D ^{1}H NOESY, 3D ^{15}N-edited NOESY, and 3D ^{13}C-edited NOESY experiments. The 4D ^{13}C, ^{13}C-edited NOESY was used to validate CYANA assignments and to check for any overlapped peaks in the 3D ^{13}C-edited NOESY. Each CYANA run was set to generate 100 structures, and the 50 best were selected based on CYANA target energy. The chemical shift tolerances were set 0.02 ppm for ^{1}H resonances and 0.2 ppm for ^{15}N and ^{13}C resonances.

Restrained Molecular Dynamics Refinement

      The 50 structures generated by CYANA were used as starting structures for restrained molecular dynamics (rMD) simulations in AMBER (145). The structures first underwent a short energy minimization and then were subjected to simulated annealing. The AMBER files used for both these processes are listed in Figures 4.11 and 4.12. The energy minimization protocol is 5000 steps and incorporates the experimental restraints. The simulated annealing protocol heats the system to 1000K, loosely incorporating the experimental restraints, and then gradually increasing the restraints until they are fully enforced at 5 ps. The system is then slowly cooled to the starting temperature for another 15 ps, allowing the protein to sample to sample energetically favorable conformations. The structure present at the final step of the calculation is then contributed to the ensemble and the process repeated with a new starting structure.

```
energy minimization for starting structures

 &cntrl

maxcyc=5000, ncyc=50, pencut=0.1, nmropt=1, imin=1,
temp0=0.5, tautp=0.02, ntpr=50, ntt=1,
nsnb=50, ntb=0, cut=10.0, scee=2.0,

 /
 &ewald
        eedmeth=5,
 /

 &wt type='END'   /

LISTOUT=POUT
DISANG=RST
```

Figure 4.11. AMBER file for performing energy minimization. This protocol was used to regularize p68N starting structures in the AMBER force-field prior to simulated annealing.

```
20 ps vacuum simulated annealing for initial NMR structure refinement

 &cntrl
   imin=0, nstlim=20000
   irest=0, ntx=1,
   ntb=0, cut=10.0,
   ntt=1, tautp=0.2,
   igb=0, saltcon=0.0, gbsa=0,
   ntr=0, nmropt=1,
   ig=23491765, vlimit=12, scee=2.0,
   ntwe=100, ntwx=100, ntpr=100,
 /
 &ewald
   eedmeth=5,
 /
#
#Simple simulated annealing algorithm:
#
#from steps 0 to 1000: heat the system to 1000K with a slower
#  temperature coupling paramter:
#from steps 1001 to 3100: equilibration with increasing constraints
#from steps 3101 to 5000: equilibration with full constraints
#from steps 5001 to 19000: re-cool to low temperatures
#from steps 19001 to 20000: final cooling with short tautp
#
 &wt type='TEMP0',istep1=0,istep2=1000,value1=0.0,
    value2=1000.0,  /
 &wt type='TEMP0',istep1=1001,istep2=5000,value1=1000.0,
    value2=1000.0,  /
  &wt type='TEMP0',istep1=5001,istep2=17000,value1=1000.0,
    value2=200.0,  /
  &wt type='TEMP0',istep1=17001,istep2=20000,value1=0.0,
    value2=0.0,  /
  &wt type='TAUTP',istep1=0,istep2=1000,value1=0.4,
    value2=0.4,  /
  &wt type='TAUTP',istep1=1001,istep2=5000,value1=0.2,
    value2=0.2,  /
  &wt type='TAUTP',istep1=5001,istep2=13100,value1=2.0,
    value2=2.0,  /
  &wt type='TAUTP',istep1=13101,istep2=16000,value1=1.0,
    value2=1.0,  /
  &wt type='TAUTP',istep1=16001,istep2=19000,value1=0.5,
    value2=0.5,  /
  &wt type='TAUTP',istep1=19001,istep2=20000,value1=0.05,
    value2=0.05,  /

  &wt type='REST',istep1=0,istep2=3100,value1=0.1,
    value2=1.0,  /
  &wt type='REST',istep1=3101,istep2=20000,value1=1.0,
    value2=1.0,  /

 &wt type='END'  /
LISTOUT=POUT
DISANG=RST
```

Figure 4.12. AMBER file for performing simulated annealing refinement. This protocol was used to refine the energy minimized p68N starting structures.

<u>Structure Analysis</u>

PROCHECK-NMR (146) was used to generate structural statistics for the ensemble, including secondary structure assignments and phi, psi, and chi dihedral angle distributions. p68N PDB coordinates were submitted to the DALI server (147) to identify proteins with similar structures. Final coordinates, restraints, and NMR chemical shifts will be deposited in the PDB and BMRB databases, respectively.

CHAPTER V

DISCUSSION AND FUTURE DIRECTIONS

An Iron-sulfur Cluster in DNA Primase

Human DNA primase was found to contain an essential iron-sulfur cluster in the C-terminal domain of the p58 subunit. Limited proteolysis, CD, and NMR experiments identified p58C as a structured domain. Purified p58C was golden-brown in color, which prompted the collection of a UV-Vis spectrum. The spectrum was similar to those reported for proteins containing an iron-sulfur cluster. ICP-MS confirmed the presence of iron, and EPR spectra were consistent with the presence of a HiPIP [4Fe-4S] cluster. Four conserved cysteine residues were identified which likely bind to the iron-sulfur cluster. Mutating one of these, Cys367, to serine in the context of the primase dimer, prevents cluster binding. Additionally, p48/Cys367Ser-p58 had dramatically reduced primase activity when compared to the WT, suggesting that the cluster is required for the p58 stimulation of p48 activity.

While primase is the first DNA replication protein found to contain an iron-sulfur cluster, there are examples of iron-sulfur proteins involved in other aspects of DNA processing. DNA glycosylases, including Endo III (119) and MutY (120), are the best characterized examples. More recently, iron-sulfur clusters have also been identified in the XPD (Rad3) and FancJ helicases (109). In addition to being involved in DNA metabolism, these proteins share two important properties with DNA primase. First, they

contain HiPIP [4Fe-4S] clusters.   Second, the clusters are essential for full enzymatic activity.

Iron-sulfur clusters are traditionally associated with electron transport and redox chemistry; however they are becoming recognized as much more versatile cofactors. Perhaps the most interesting example of this diversity is the activity of the iron regulatory protein (IRP) (Reviewed in (149)). When iron content is high, IRP contains a [2Fe-2S] cluster. However, when the cellular iron content drops, IRP loses its cluster. This results in a conformational change and the protein interacts with iron-responsive elements (IREs) on specific mRNAs. IRP binding to the transferrin receptor mRNA protects it from degradation, and binding to ferritin mRNA blocks translation.   This serves to promote import of iron into the cell and prevent ferritin from sequestering free iron, thus raising the intracellular iron concentration.   The iron-sulfur clusters in the DNA processing enzymes are not predicted to have this type of property, but this example illustrates that functions beyond electron transfer must be considered, especially in the case of HiPIP clusters.

Proteins containing [4Fe-4S] clusters can be classified into two types: ferredoxin-like, and HiPIP-like. [4Fe-4S] clusters in ferredoxin proteins exist in either the +2 or +1 oxidation state with typical redox potentials ranging from -250 to -650 mV (150). Ferredoxin-like proteins are typically involved in electron transport. HiPIPs on the other hand, exist in either the +2 or +3 oxidation state *in vitro* and have a much higher redox potential, +50 to +450 mV (150). HiPIPs were first identified in photosynthetic bacteria, but the functions of these proteins, including whether they are redox active *in vivo*, are not clear. [4Fe-4S] clusters in ferredoxins and HiPIPs are structured similarly, however

the ferredoxin clusters tend to be more surface exposed and in a less hydrophobic environment. The [4Fe-4S] clusters in the DNA processing enzymes appear to be HiPIP-like primarily based on defining characteristics in the EPR spectra described in Chapter II. Additionally, only the oxidized [4Fe-4S] is EPR-active, suggesting that the cluster is in the +3 oxidation state.

Despite the findings that the [4Fe-4S] cluster is required for activity in these DNA processing enzymes, relatively little is known about their direct role in binding and/or catalysis. The clusters in MutY and Endo III are remote from the active site, but they are important for orienting key residues involved in substrate recognition (122, 123, 125). XPD and FancJ require the iron-sulfur cluster for helicase activity but not for ssDNA-dependent ATPase activity (109). Interestingly, recent studies of the archaeal XPD protein also point to a substrate recognition role, as the presence of the iron-sulfur cluster targeted the helicase to ssDNA-dsDNA junctions (151), and these findings were supported by recent XPD crystal structures (152, 153). Mutating any of the four cysteine ligands in p58C resulted in substantially reduced solubility of the protein (Figure 2.7). This also suggests a structural role for the [4Fe-4S] in DNA primase. However, since HiPIP clusters are typically buried, it is not surprising that they would significantly impact the overall fold of the protein. This was directly observed in the case of the archaeal XPD protein using CD measurements (151).

While the bulk of the available data point to a structural role for the [4Fe-4S] clusters in the DNA processing enzymes, the use of this specific cofactor is intriguing. Iron-sulfur clusters are actively assembled via several proteins in a multi-step process (126). This seems like an extensive utilization of cellular resources for a purpose that

could be achieved in a simpler manner. Another role for the iron-sulfur clusters in the DNA glycosylases has been proposed. HiPIP clusters have redox potentials that are outside the range of what is typically considered accessible for cellular redox activity. However, there is evidence that DNA-binding may shift the redox potential of the iron-sulfur cluster, such that redox chemistry becomes possible (128), which be relevant for p58C, as it has been shown to interact with DNA (S. Vaithiyalingam & W. J. Chazin, unpublished results). This property has led to a complex and controversial model of how DNA glycosylases locate sites of DNA damage. In this model, one glycosylase can donate an electron to a nearby glycosylase via the dsDNA helix. The cluster is reduced in the nearby glycosylase prompting the enzyme to dissociate and scan elsewhere. A DNA lesion would not transmit the electron, leaving the other glycosylase associated with the dsDNA to promote repair. Given the difficulty in studying the redox properties of HiPIP proteins, the possibility that the iron-sulfur cluster of DNA primase can be redox active cannot be ruled out.

We have shown that human DNA primase cannot initiate primer formation on M13 ssDNA without the presence of the iron-sulfur cluster (Figure 2.9). Klinge, *et al*. reported similar findings for the yeast primase using a synthetic template (131). Interestingly, they also found that the cluster was not required for extension of a pre-formed primer-template substrate. This suggests that the cluster is involved in the initial dinucleotide formation, however this remains to be conclusively demonstrated. p58C interacts with both p48 and DNA ((79, 81), S. Vaithiyalingam & W. J. Chazin, unpublished results). Thus p58C may facilitate nucleotide binding or properly orient the

p48 active site on the template ssDNA. This is in agreement with the ability of p58 to stimulate p48 activity (79).

A previous study identified a region of p58 with homology to a pol β domain (80). Deletion of the homologous region, residues 288-313, results in reduced enzymatic activity coupled with the inability to create unit-length primers (i.e. counting). Interestingly, this deletion is directly adjacent to one of the conserved cysteine ligands, Cys287. Thus it is possible that the authors are actually observing the consequences of the loss of the iron-sulfur cluster. The potential role of the cluster in counting is intriguing given the previously discussed role of substrate recognition in the HiPIP glycosylases and helicases. In the context of the pol-prim heterotetramer, the p58 subunit has also been proposed to mediate the transfer of the RNA primer to the p180 subunit (81). p58C may only be able to interact with the 5' end of the primer when it is 7-10 nucleotides long (Figure 5.1). This interaction could cause a conformational change to inhibit further primer extension and promote primer transfer. Primer extension occurs on the 3' end, which is the end that must be transferred to the p180 subunit.
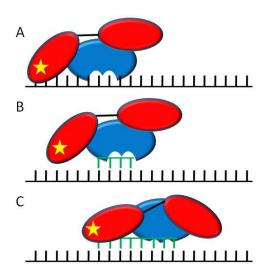
Figure 5.1. Proposed role of the iron-sulfur cluster in primase activity. (A) DNA primase bound to a ssDNA template. p48 is represented in blue, p58 is represented in red, and the [4Fe-4S] cluster is represented as a yellow star. (B) The p48/p58 dimer initiates priming. (C) When the primer is unit-length, p58C contacts the primer-template substrate and undergoes a conformational change. This conformational change propagates to the p48 subunit, inhibiting elongation which stalls pol-prim to promote primer handoff.

If p58C is involved in dinucleotide formation as well as counting, it must "sense" both ends of the RNA primer. One possibility is that the primer-template is wrapped around the p58C subunit, allowing for direct contacts with both the 5' and 3' ends. Another possibility is that it directly contacts one end, and has effects on the other end via an allosteric mechanism. For example, once the RNA primer is unit length, it may interact with the p58C domain. The 5' end of the primer-template junction would be analogous to the ssDNA-dsDNA junction that is involved in binding the iron-sulfur domain in XPD. This interaction could prompt a conformational change, negatively regulating the p48 subunit, causing the enzyme to stall (Figure 5.1). This promotes transfer rather than extension, and the primer is transferred, perhaps via the p58N domain to the p180 subunit. This proposed model is highly speculative, and several additional

experiments would be needed to validate the model, some of which will be discussed in a following section.

High Resolution Structure of p58C

Currently there are no published high resolution structures of a eukaryotic primase. Moreover, of the three archaeal primase structures, none contain the domain homologous to the eukaryotic p58C domain. Given that p58C is required for primase activity, determination of the structure would be highly beneficial. For example, the structure may help identify the role of the iron-sulfur cluster, in much the same way the structures of MutY (123) and Endo III (124) have led to the proposal of a substrate recognition role for those proteins. The strategy of either x-ray crystallography or NMR can be used to determine the structure. Once available, one could then begin a series of biochemical and structural experiments to map out the DNA and RPA binding regions of the domain and, in turn, develop hypotheses for the mechanism of action of p58C.

NMR studies of iron-sulfur proteins are difficult. The cluster is paramagnetic and broadens NMR signals from nearby nuclei in the protein. Missing signals complicate the resonance assignment process, and also result in few NOEs to residues near the cluster. Nonetheless, solution structures have been determined for small iron-sulfur proteins (154). Additionally, recent advances in ^{13}C detection based experiments have facilitated resonance assignments of paramagnetic proteins (155). The large size of p58C (~28 kDa) further complicates the possibility of NMR structure determination as signal overlap becomes more of a concern. Given these drawbacks, x-ray crystallography would be the preferred method of determining the structure.

Attempts were made throughout the thesis period to crystallize p58C. This construct was tested in several crystal screens, including an automated screen of more than 1500 conditions at the Hauptman-Woodward Institute. For Hampton and Wizard screens, various protein concentrations were tried, with the maximum being ~ 30 mg/ml. Despite extensive efforts, none of the conditions produced protein crystals. Careful analysis of the secondary structure prediction, as well as CD and NMR data suggest that the C-terminal region of the p58C construct may be unstructured. Consequently, two C-terminal truncations have been created, and one of these, p58(266-424) has been promising in initial crystal screens (S. Vaithiyalingam, W. J. Chazin, and B. F. Eichman, unpublished results). Future work in the Chazin laboratory will focus on optimizing conditions to produce diffraction-quality crystals, determining the structure, and characterizing the structural basis for activity.

p58C Interactions

Given that the iron-sulfur clusters in the DNA glycosylases and repair helicases have been implicated in DNA substrate recognition, it is possible that the cluster has a similar role in p58C. WT primase, p48/C367S-p58 primase, and p58C could be tested for their ability to interact with various DNA substrates including ssDNA, dsDNA, forked DNA, 5'-overhang DNA, and 3'-overhang DNA. The electromobility shift assay (EMSA) could be used for its simplicity. For a more quantitative analysis, fluorescence anisotropy measurements should be recorded. These data would help refine the model of the iron-sulfur cluster's role in primase activity described above in Figure 5.1.

DNA primase has been shown to interact with both Tag (86) and RPA (88-90), however the regions of primase involved in these interactions have not been determined. Since primase domain architecture is now better characterized, identifying the Tag and RPA binding domains is warranted. Initial work in our laboratory using affinity chromatography has verified the p48/p58-RPA70NAB and p48/p58-RPA32C interactions. Additionally, p58C has been shown to physically interact with RPA32C using ITC and NMR (S. Vaithiyalingam & W. J. Chazin, unpublished results). Initial assays to identify other domain-domain contacts could include yeast two-hybrid, protease protection, and pulldown experiments. Once domain-domain interactions are identified, they would be studied biophysically as described for the p68N-Tag HD interaction in Chapter III.

The Iron-sulfur Complex in the Context of Pol-prim

We have shown that the iron-sulfur cluster in p58C is required for human primase activity. This has been confirmed in a complementary study in yeast (131). These authors also showed that the cluster is dispensable for primer extension. The authors used a synthetic poly(dT) template, while our studies utilized M13 ssDNA. When an $(A)_{15}$ primer was annealed to the template, primase activity was restored. This experiment should be repeated using additional templates and primers to further confirm this finding. The primer length should also be varied to include a preformed dinucleotide as well as a unit-length primer. The initial dinucleotide formation is the rate limiting step in primase activity (71). Perhaps in the absence of the iron-sulfur cluster, the template-

primase interaction is not sufficiently stable to allow for dinucleotide formation, but since polymerization is rapid, it is not affected.

These studies were conducted on mutants in the context of the primase dimer. To obtain a more complete understanding of the role of the cluster, the effects of removing the cluster in the pol-prim heterotetramer can be studied. Since p58 has been implicated in the transfer of the RNA primer to the p180 active site, it would be interesting to study the role of the iron-sulfur cluster in this context (81).

An Interaction between p68N and Tag HD

The N-terminal region of the p68 subunit of pol-prim was shown by a yeast two-hybrid experiment in the Fanning laboratory to interact with the helicase domain of SV40 Tag. Biochemical studies showed that the first 78 residues of the 107 residue p68N construct was responsible for this activity. The solution structure of this domain was determined using NMR techniques. p68N is a compact, globular domain containing four helices, and NMR experiments identified a potential binding site for Tag on p68N that is primarily hydrophobic.

The solution structure of p68N represents the first high-resolution structure of a pol-prim domain. Secondary structure prediction suggests that the p68N domain may be flexibly linked to the rest of the p68 subunit, although this has not been confirmed experimentally. Given the ability of p68N to interact with Tag, it may function in a manner similar to the RPA70N and RPA32C domains. These two domains are protein interaction domains that are flexibly linked to the ssDNA-binding core of the protein. The linkers allow the protein-protein interaction domains the required freedom to explore

for potential binding partners while other domains of the protein are interacting with DNA. The C-terminus of p68 is responsible for its interaction with p180, and p68 does not directly associate with either p58 or p48. Thus p68N may also be remote from the polymerase active sites, available to interact with Tag or other undetermined proteins.

As discussed in Chapter I, protein-protein interactions allow the four discrete proteins of SV40 DNA replication initiation to assemble into a molecular machine. Protein-protein interactions in DNA processing machinery are often composed of multiple domain-domain contacts (82). These domain-domain interactions have specific roles in the context of the SV40 replisome. The role of the p68N-Tag interaction remains to be determined, and experiments are currently underway to study the interaction functionally. However, given the current data, it is possible to speculate on the role of the interaction of p68N with the helicase domain (HD) of Tag in the context of SV40 DNA replication.

Based on fluorescence measurements, the stoichiometry of pol-prim binding to Tag has been calculated to be one molecule of pol-prim per six molecules of Tag (156). Thus in the context of the Tag double hexamer assembling at the origin of replication, two pol-prim molecules would be bound, presumably one for each fork. This is interesting given our observation that p68N can interact with the monomeric mutant of Tag HD. It is possible that in the context of the assembled Tag hexamer, only one molecule of pol-prim may bind due to steric occlusion (Figure 5.2). H. Huang in the Fanning laboratory has also collected preliminary evidence suggesting that the region of Tag responsible for binding p68N is on the back surface of the helicase domain, i.e. the region away from the center of the double hexamer. Topo I has also recently been shown

to interact with the back surface of the helicase domain (84), and it also binds with a stoichiometry of one molecule per Tag hexamer (157). Whether both proteins can bind simultaneously has not been studied, and it would be interesting to determine if this is the case or if a hand-off mechanism is utilized.

Figure 5.2. Schematic view of the p68N-Tag interaction. Although the p68N:Tag HD stoichiometry is 1:1, in the context of hexameric Tag (red), only one pol-prim tetramer (blue) may bind due to the large size of pol-prim.

In addition to interacting with the p68 subunit, Tag also interacts with p180 (85) and the p58/p48 dimer (86). The p180 interaction is thought to be important for the ability of Tag to stimulate pol-prim's activity. This is supported by the observation that Tag can stimulate pol-prim activity of a recombinantly expressed trimer pol-prim lacking the p68 subunit (63). Additionally, a peptide designed to inhibit the Tag-p180 interaction eliminates the stimulatory effect of Tag on pol-prim (158). The interaction with p58/p48 dimer has not been well characterized. In addition to demonstrating that p68N interacts with Tag, the Fanning laboratory has shown that deleting this domain in the context of the SV40 replisome does not allow for the initiation of DNA replication. Although confirmatory experiments are currently underway, this defect is almost surely due to abolishment of the p68-Tag interaction. We can therefore expand upon the model for the

initiation of SV40 DNA replication as diagramed in Figure 1.6. The pol-prim loading

step from Figure 1.6.E to 1.6.F is modified to incorporate this proposal in Figure 5.3.
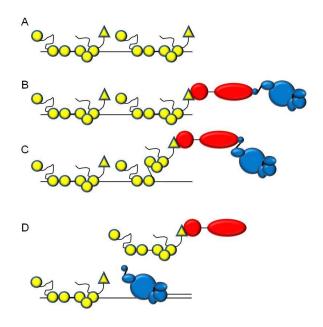
Figure 5.3. Tag-mediated loading of pol-prim onto ssDNA. (A) RPA (yellow) coated ssDNA. (B) Tag (red) interacts with RPA32C via the OBD and with p68N via the HD. pol-prim is shown in blue. (C) The RPA32C-Tag OBD interaction facilitates RPA displacement. (D) Pol-prim can then be loaded onto the newly freed ssDNA to initiate primer synthesis. Figure adapted from (55).

The p68-Tag interaction may be required for simply tethering pol-prim to the

replication fork. By localizing pol-prim to the replication fork, primer formation is

indirectly promoted. Another possibility is that Tag may actively load pol-prim onto the

template DNA, similar to its ability to load RPA onto emerging ssDNA via the Tag

ODB-RPA70AB interaction (53). It is well established that RPA inhibits pol-prim

activity on a ssDNA template and that Tag is able to relieve this inhibition (55).

Additionally, this property is dependent on the Tag OBD-RPA32C interaction as well as

the presence of the p68 subunit of pol-prim (55, 63). It is possible that Tag promotes

RPA displacement from ssDNA via the origin binding domain while simultaneously loading pol-prim via the helicase domain.

Role of the p68N-Tag HD Interaction

The identification of the Tag-binding interface in Chapter III sets the stage for future functional studies. The binding interface will be confirmed via site-directed mutagenesis. Mutation of key residues at the binding interface should eliminate or reduce Tag binding. The p68N mutants will initially be screened using a yeast two-hybrid assay. Those defective in binding Tag will be assayed for structural integrity by NMR, and then ITC will be used to confirm the binding defect. These mutations will then be incorporated into the pol-prim heterotetramer. The mutant pol-prim(s) can then be assessed for functional defects in the monopolymerase assay and for Tag-mediated stimulation on an RPA-coated ssDNA template. We hypothesize that the pol-prim point mutants will show the same phenotype as the p68ΔN pol-prim in these assays, which would confirm our hypothesis that the interaction is essential.

Role of p68N in Eukaryotic DNA Replication

p68N is required for SV40 DNA replication initiation (59), but its role in eukaryotic DNA replication has not been studied. p68 has been previously shown to interact with ORC and Cdc45, so it would be interesting to determine if either of these interactions are mediated through p68N. Other potential binding partners include Mcm10, GINS, or other proteins in the initiation complex. Yeast two-hybrid assays could be used as an initial screen with positive hits confirmed by pulldown experiments.

Similar studies as are currently underway for the Tag HD could then be utilized to map the binding interface and create binding-deficient mutants. These mutant pol-prim constructs could be studied *in vivo*, perhaps using a yeast system. This would allow for a direct observation of the role of p68N interactions in cellular DNA replication.

TABLE OF POL-PRIM CONSTRUCTS

Table A.1. Pol-prim construct design. For the p58C point mutations, reverse complements to the oligonucleotides listed were also used.

Construct	Vector	Enzymes	Oligonucleotides
p48	pBG100	BamHI EcoRI	5'-TATTATGGATCCATGGAGACGTTTGACCCCACC-3' 5'-ATAATAGAATTCCTATCAGAAATCTTTTTGTAA-3'
1-361	pBG100	BamHI EcoRI	5'-TATTATGGATCCATGGAGACGTTTGACCCCACC-3' 5'-ATAATAGAATTCCTATCAGGAAATGGCATCCAATTC-3'
1-366	pBG100	BamHI EcoRI	5'-TATTATGGATCCATGGAGACGTTTGACCCCACC-3' 5'-ATAATAGAATTCCTATCATTCCTCTTCATTAGTGGA-3'
p58	pET15b	EcoRI XhoI	5'-TATTATGAATTCATGGAGTTTTCTGGAAGAAAG-3' 5'-TATTATCTCGAGTCACTAAGAATCTTCACTAAA-3'
266-509	pET15b	NdeI XhoI	5'-TATCATATGGGAAATGTTGGGAAGATTTCT-3' 5'-ATACTCGAGCTAAGAATCTTCACTAAAGTA-3'
C287A	pET15b		5'-AAATCCTTCCCACCTGCAATGCGTCAGTTACAT-3'
C287S	pET15b		5'-AAATCCTTCCCACCTTCAATGCGTCAGTTACAT-3'
C367A	pET15b		5'-TATACACCTTTCAGTGCACTGAAGATTATTCTG-3'
C367S	pET15b		5'-TATACACCTTTCAGTTCACTGAAGATTATTCTG-3'
C384A	pET15b		5'-GGGGATTATCATGGGGCACCATTCCGTCACAGT-3'
C384S	pET15b		5'-GGGGATTATCATGGGTCACCATTCCGTCACAGT-3'
C424A	pET15b		5'-CATTACCAGGTAGCCGCACAAAAATACTTTGAG-3'
C424S	pET15b		5'-CATTACCAGGTAGCCTCACAAAAATACTTTGAG-3'
p68			
1-107	pBG100	BamHI XhoI	5'-TATTATGGATCCATGTCCGCATCCGCCCAG-3' 5'-GAATTCCTACTATTCCTCTTCTTCTTC-5'
1-87	pBG100	BamHI XhoI	5'- TATTATGGATCCATGTCCGCATCCGCCCAG -3' 5'-TATTATGAATTCCTATCAATGGCCACTGTCCTTGCA-3'
1-78	pBG100	BamHI XhoI	5'- TATTATGGATCCATGTCCGCATCCGCCCAG -3' 5'-TATTATGAATTCCTATCACCTGGCTTTCGATAATCT-3'

TABLE OF p68N CHEMICAL SHIFT ASSIGNMENTS

Table B.1. p68N backbone chemical shifts.

Residue	CA	CB	H	N
M1	55.68	33.3	8.364	122
S2	58.74	64.19	8.338	117.1
A3	53.32	19.69	8.219	124.7
S4	58.68	64.27	8.886	114.8
A5	55.31	18.43	8.565	126.5
Q6	59.29	28.71	8.255	117.5
Q7	58.79	28.76	7.855	119.3
L8	59.1	42.44	8.346	120.6
A9	55.68	18.08	8.324	120.1
E10	59.37	29.44	7.9	118.2
E11	58.79	29.2	8.139	119.4
L12	58.39	41.37	8.438	118.4
Q13	58.88	28.36	7.904	118.2
I14	64.65	37.4	7.722	121.8
F15	58.91	39.74	7.284	116.9
G16	46.17		7.932	108.1
L17	53.82	43.51	7.706	121.3
D18	53.72	41.95	8.17	122.3
C19	57.9	30.12	8.314	119.2
E20	56.59	30.08	8.427	123
E21	60.43	30.04	9.017	124.5
A22	55.18	18.45	8.89	119.5
L23	56.86	41.6	7.389	116.7
I24	64.67	36.4	7.509	120.5
E25	59.73	29.13	8.144	117.6
K26	58.02	30.84	7.336	119.5
L27	58.33	42.75	8.018	120.5
V28	67.56	31.57	8.478	121.3
E29	59.67	29.41	7.613	118.9
L30	58.16	42.09	8.497	119.5
C31	65.34	26.98	8.469	119.2
V32	66.1	32.18	7.741	118.3
Q33	59.03	29.02	8.605	119.7
Y34	59.14	37.65	8.203	112.3
G35	47.74		7.714	112.2
Q36	53.63	33.33	8.385	118.1
N37	51.55	39.13	8.419	119
E38	62.19	28.58	10.2	120.4

Table B.1 continued.

Residue	CA	CB	H	N
E39	59.68	29.52	8.743	118.6
G40	46.54		8.433	109.3
M41	57.29	32.08	8.677	121.9
V42	67.46	30.89	8.197	118.3
G43	47.49		8.035	106.6
E44	59.01	29.02	8.001	122
L45	57.8	41.72	8.319	123.3
I46	63.72	36.21	8.659	120.5
A47	55.59	17.64	7.801	125.1
F48	61.13	39.28	8.319	121
C49	64.6	26.54	8.897	120.5
T50	65.87	68.88	8.476	114.3
S51	61.27	63.33	7.969	116.9
T52	61.42	68.93	7.28	108.5
H53	56.4	26.34	7.324	117
K54	54.68	33.58	8.349	118.8
V55	61.7	33	8.333	120.8
G56	44.63		8.059	112.2
L57	55.58	43.78	8.305	118.6
T58	59	73.55	6.607	112.2
S59	62.39		9.533	117.8
E60	60.18	29.33	8.529	121.9
I61	65.06	38.75	7.811	121.6
L62	58.14	41.51	8.136	119.7
N63	56.89	38.55	8.353	119.5
S64	62.79		8.239	117.4
F65	59.21	39.1	8.747	122.9
E66	59.57	29.02	8.272	119.9
H67	59.41	30.05	8.133	115.3
E68	56.72	30.3	9.039	115.2
F69	60.22	41.7	7.334	118
L70	58.42	40.45	7.995	120.4
S71	60.55	62.91	7.933	109.9
K72	56.65	32.78	6.748	118.7
R73	56.95	30.95	7.449	118.3
L74	54.71	42.8	7.819	120.2
S75	58.68	63.83	8.099	115.6
K76	56.52	33.07	8.254	123.2
A77	52.56	19.24	8.268	125.8
R78	57.37	31.52	7.869	125.3

Table B.2. p68N side chain ^{1}H chemical shifts.

Residue	HA	HA2	HA3	HB	HB2	HB3	HD	HD1	HD2	HD21	HE	HE2	HE21	HE22	HG	HG1	HG2	HG3	HZ
M1	4.522			2.115							2.084				2.571				
S2	4.415			3.965															
A3	4.398			1.418															
S4	4.533				3.889	4.025													
A5	3.813			1.451															
Q6	4.014				2.032	2.137							7.629		2.426				
Q7	4.088			2.190									6.833	7.729	2.416				
L8	3.865				1.356	1.964	0.855								1.741				
A9	4.000			1.531															
E10	4.015			2.146											2.485				
E11	4.200			2.210											2.599				
L12	4.055			1.937				0.733	0.692						1.775				
Q13	4.084			2.228									6.819	7.357			2.451	2.573	
I14	3.651			1.880				1.087								1.607	0.373		
F15	4.457				2.693	3.422	7.323												
G16		3.840	4.037																
L17	4.530				1.328	1.583	0.777								1.556				
D18	4.604				2.554	2.651													
C19	4.668				2.657	2.808													
E20	4.350				2.111	2.248									2.398				
E21	3.880			2.106											2.340				
A22	4.179			1.449															
L23	4.149				1.376	1.940	0.844								1.692				
I24	3.567			2.162				0.760								1.542	0.974		
E25	3.965			2.090													2.266	2.433	
K26	4.297			2.021			1.626								1.415				
L27	4.104				1.456	2.155		0.841	0.798						1.787				
V28	3.402			2.356												0.986	0.969		
E29	4.029			2.279											2.491				
L30	4.292				1.369	2.271	0.944								2.097				

Table B.2 continued

Residue	HA	HA2	HA3	HB	HB2	HB3	HD	HD1	HD2	HD21	HE	HE2	HE21	HE22	HG	HG1	HG2	HG3	HZ
C31	3.934				3.207	3.364													
V32	3.695			2.197												1.088	0.944		
Q33	3.869				1.792	2.144													
Y34	4.437				2.643	3.260	7.256				6.592								
G35	3.878																		
Q36	4.740			2.282															
N37	4.769				2.966	3.214				7.029									
E38	3.856				1.642	2.198											2.354	2.749	
E39	3.873			2.016											2.273				
G40	3.920																		
M41	4.502			2.183							1.957						2.293	2.769	
V42	3.380			2.117												1.002	0.815		
G43		3.585	3.940																
E44	4.288			2.439											2.260				
L45	3.239				0.819	1.728		0.039	1.088						0.496				
I46	3.496			1.663				0.363								1.048	0.626		
A47	4.098			1.575															
F48	3.976				2.677	3.028	7.088				7.331								
C49	3.595			2.992															
T50	4.015			4.224													1.240		
S51	4.149			3.962															
T52	4.279			4.118													0.862		
H53	4.242				3.284	3.405			7.155										
K54	4.543				1.712	1.570									1.258				
V55	4.199			2.044											0.881				
G56		3.777	4.004																
L57	4.413			1.519			0.799								1.531				
T58				4.764													1.277		
S59	3.949			3.961															
E60	4.179			2.035											2.378				

Residue	HA	HA2	HA3	HB	HB2	HB3	HD	HD1	HD2	HD21	HE	HE2	HE21	HE22	HG	HG1	HG2	HG3	HZ
I61	3.947			1.821				0.968								1.000			
L62	4.108				1.595	1.929	0.857												
N63	4.646			2.987					7.788										
S64	4.229			4.234															
F65	3.030				2.719	3.041	6.543				7.032								7.229
E66	3.365				1.834	2.045											1.826	2.036	
H67	4.304			3.246					7.264										
E68	4.338			2.205											2.451				
F69	4.583				2.657	2.915	7.015				7.418								7.300
L70	3.385				0.354	1.074	0.373								0.329				
S71	4.176			3.962															
K72	4.227				1.577	1.812	1.613				2.972						1.256	1.372	
R73	4.145				1.494	1.775	2.743										1.399	1.348	
L74	4.445			1.661			0.884								1.590				
S75	4.453			3.870															
K76	4.241			1.798			1.686				2.965				1.342				
A77	4.287			1.367															
R78	4.289			1.687											1.390				

Table B.3. p68N side chain ^{13}C and ^{15}N chemical shifts.

Residue	CD	CD1	CD2	CE	CG	CG1	CG2	ND2	NE2
M1					32.46				
S2									
A3									
S4									
A5									
Q6									112.4
Q7					34.21				112.9
L8		25.34			28.36				
A9									
E10					36.18				
E11					34.76				
L12		24.09	26.32		26.11				
Q13					34.47				111.3
I14		12.78				28.83	16.69		
F15									
G16									
L17		24.85			26.85				
D18									
C19									
E20					36.57				
E21					36.17				
A22									
L23		25.62	22.79		27.64				
I24		12.04				28.86	18.56		
E25					35.98				
K26	27.90			45.79	24.70				
L27		24.87			27.49				
V28						23.04	22.01		
E29					36.27				
L30		23.02			25.37				
C31									
V32						22.33	21.39		
Q33					33.18				
Y34									
G35									
Q36									
N37								112.5	
E38					37.81				
E39					36.00				
G40									
M41				16.88	31.45				
V42						24.27	21.97		
G43									
E44					36.02				
L45		22.04			26.17				
I46		11.32				27.88	17.04		
A47									

Table B.3 continued.

Residue	CD	CD1	CD2	CE	CG	CG1	CG2	ND2	NE2
F48									
C49									
T50							21.49		
S51									
T52							21.45		
H53									
K54	28.44			42.16	24.05				
V55						20.69			
G56									
L57		24.91			27.58				
T58							21.78		
S59									
E60					36.63				
I61		13.79					17.59		
L62		24.82			26.50				
N63								113.0	
S64									
F65									
E66					34.79				
H67									
E68					36.46				
F69									
L70		22.97			25.97				
S71									
K72	28.75			42.22	24.95				
R73	42.99				27.01				
L74		25.05	23.13		27.43				
S75									
K76	28.91			42.00	24.66				
A77									

REFERENCES

1.   Cvetic, C., and Walter, J. C. (2005) Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* 16, 343-353

2.   Aladjem, M. I., and Fanning, E. (2004) The replicon revisited: an old model learns new tricks in metazoan chromosomes. *EMBO Rep* 5, 686-691

3.   Bell, S. P., and Dutta, A. (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71, 333-374

4.   Mendez, J., and Stillman, B. (2003) Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins. *Bioessays* 25, 1158-1167

5.   Thomae, A. W., Pich, D., Brocher, J., Spindler, M. P., Berens, C., Hock, R., Hammerschmidt, W., and Schepers, A. (2008) Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. *Proc Natl Acad Sci U S A* 105, 1692-1697

6.   Tanaka, T., Knapp, D., and Nasmyth, K. (1997) Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. *Cell* 90, 649-660

7.   Nishitani, H., Lygerou, Z., Nishimoto, T., and Nurse, P. (2000) The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* 404, 625-628

8.   Blow, J. J., and Dutta, A. (2005) Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol* 6, 476-486

9.   Randell, J. C., Bowers, J. L., Rodriguez, H. K., and Bell, S. P. (2006) Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell* 21, 29-39

10.   DePamphilis, M. L., Blow, J. J., Ghosh, S., Saha, T., Noguchi, K., and Vassilev, A. (2006) Regulating the licensing of DNA replication origins in metazoa. *Curr Opin Cell Biol* 18, 231-239

11.   Wohlschlegel, J. A., Dhar, S. K., Prokhorova, T. A., Dutta, A., and Walter, J. C. (2002) Xenopus Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45. *Mol Cell* 9, 233-240

12.   Walter, J., and Newport, J. (2000) Initiation of eukaryotic DNA replication: origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase alpha. *Mol Cell* 5, 617-627

13.   Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev* 17, 1153-1165

14.   Moyer, S. E., Lewis, P. W., and Botchan, M. R. (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A* 103, 10236-10241

15.   Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R. D., and Labib, K. (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* 8, 358-366

16.   Zegerman, P., and Diffley, J. F. (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445, 281-285

17.   Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007) CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* 445, 328-332

18.   Botchan, M. (2007) Cell biology: a switch for S phase. *Nature* 445, 272-274

19.   Wold, M. S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* 66, 61-92

20.   Feng, W., Rodriguez-Menocal, L., Tolun, G., and D'Urso, G. (2003) Schizosacchromyces pombe Dpb2 binds to origin DNA early in S phase and is required for chromosomal DNA replication. *Mol Biol Cell* 14, 3427-3436

21.   Hubscher, U., Maga, G., and Spadari, S. (2002) Eukaryotic DNA polymerases. *Annu Rev Biochem* 71, 133-163

22.   Garg, P., and Burgers, P. M. (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit Rev Biochem Mol Biol* 40, 115-128

23.   Ahuja, D., Saenz-Robles, M. T., and Pipas, J. M. (2005) SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene* 24, 7729-7745

24.   Engels, E. A., Chen, J., Hartge, P., Cerhan, J. R., Davis, S., Severson, R. K., Cozen, W., and Viscidi, R. P. (2005) Antibody responses to simian virus 40 T antigen: a case-control study of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 14, 521-524

25.   Engels, E. A., Katki, H. A., Nielsen, N. M., Winther, J. F., Hjalgrim, H., Gjerris, F., Rosenberg, P. S., and Frisch, M. (2003) Cancer incidence in Denmark

following exposure to poliovirus vaccine contaminated with simian virus 40. *J Natl Cancer Inst* 95, 532-539

26.   Daniels, R., Sadowicz, D., and Hebert, D. N. (2007) A very late viral protein triggers the lytic release of SV40. *PLoS Pathog* 3, e98

27.   Khalili, K., White, M. K., Sawa, H., Nagashima, K., and Safak, M. (2005) The agnoprotein of polyomaviruses: a multifunctional auxiliary protein. *J Cell Physiol* 204, 1-7

28.   Li, J. J., and Kelly, T. J. (1984) Simian virus 40 DNA replication in vitro. *Proc Natl Acad Sci U S A* 81, 6973-6977

29.   Matsumoto, T., Eki, T., and Hurwitz, J. (1990) Studies on the initiation and elongation reactions in the simian virus 40 DNA replication system. *Proc Natl Acad Sci U S A* 87, 9712-9716

30.   Kim, H. Y., Ahn, B. Y., and Cho, Y. (2001) Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *Embo J* 20, 295-304

31.   Luo, X., Sanford, D. G., Bullock, P. A., and Bachovchin, W. W. (1996) Solution structure of the origin DNA-binding domain of SV40 T-antigen. *Nat Struct Biol* 3, 1034-1039

32.   Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J. A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X. S. (2003) Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* 423, 512-518

33.   Weisshart, K., Friedl, S., Taneja, P., Nasheuer, H. P., Schlott, B., Grosse, F., and Fanning, E. (2004) Partial proteolysis of simian virus 40 T antigen reveals intramolecular contacts between domains and conformation changes upon hexamer assembly. *J Biol Chem* 279, 38943-38951

34.   Gai, D., Li, D., Finkielstein, C. V., Ott, R. D., Taneja, P., Fanning, E., and Chen, X. S. (2004) Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen. *J Biol Chem* 279, 38952-38959

35.   Sullivan, C. S., Cantalupo, P., and Pipas, J. M. (2000) 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

36.   Spence, S. L., and Pipas, J. M. (1994) Simian virus 40 large T antigen host range domain functions in virion assembly. *J Virol* 68, 4227-4240

37.   Tevethia, M. J., Pipas, J. M., Kierstead, T., and Cole, C. (1988) 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

38. Gai, D., Zhao, R., Li, D., Finkielstein, C. V., and Chen, X. S. (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* 119, 47-60

39. Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* 9, 27-43

40. Gomez-Lorenzo, M. G., Valle, M., Frank, J., Gruss, C., Sorzano, C. O., Chen, X. S., Donate, L. E., and Carazo, J. M. (2003) Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. *Embo J* 22, 6205-6213

41. Weisshart, K., Taneja, P., Jenne, A., Herbig, U., Simmons, D. T., and Fanning, E. (1999) 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

42. Valle, M., Gruss, C., Halmer, L., Carazo, J. M., and Donate, L. E. (2000) Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. *Mol Cell Biol* 20, 34-41

43. Leppard, J. B., and Champoux, J. J. (2005) Human DNA topoisomerase I: relaxation, roles, and damage control. *Chromosoma* 114, 75-85

44. Stewart, L., Ireton, G. C., and Champoux, J. J. (1997) Reconstitution of human topoisomerase I by fragment complementation. *J Mol Biol* 269, 355-372

45. Mo, Y. Y., Wang, C., and Beck, W. T. (2000) A novel nuclear localization signal in human DNA topoisomerase I. *J Biol Chem* 275, 41107-41113

46. Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G., and Champoux, J. J. (1998) A model for the mechanism of human topoisomerase I. *Science* 279, 1534-1541

47. Fanning, E., Klimovich, V., and Nager, A. R. (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* 34, 4126-4137

48. Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct* 32, 115-133

49. Kolpashchikov, D. M., Khodyreva, S. N., Khlimankov, D. Y., Wold, M. S., Favre, A., and Lavrik, O. I. (2001) Polarity of human replication protein A binding to DNA. *Nucleic Acids Res* 29, 373-379

50. Bochkareva, E., Kaustov, L., Ayed, A., Yi, G. S., Lu, Y., Pineda-Lucena, A., Liao, J. C., Okorokov, A. L., Milner, J., Arrowsmith, C. H., and Bochkarev, A.

(2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc Natl Acad Sci U S A* 102, 15412-15417

51.   Ball, H. L., Ehrhardt, M. R., Mordes, D. A., Glick, G. G., Chazin, W. J., and Cortez, D. (2007) Function of a conserved checkpoint recruitment domain in ATRIP proteins. *Mol Cell Biol* 27, 3367-3377

52.   Stauffer, M. E., and Chazin, W. J. (2004) Physical interaction between replication protein A and Rad51 promotes exchange on single-stranded DNA. *J Biol Chem* 279, 25638-25645

53.   Jiang, X., Klimovich, V., Arunkumar, A. I., Hysinger, E. B., Wang, Y., Ott, R. D., Guler, G. D., Weiner, B., Chazin, W. J., and Fanning, E. (2006) Structural mechanism of RPA loading on DNA during activation of a simple pre-replication complex. *Embo J* 25, 5516-5526

54.   Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C. J., Edwards, A. M., and Chazin, W. J. (2000) Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell* 103, 449-456

55.   Arunkumar, A. I., Klimovich, V., Jiang, X., Ott, R. D., Mizoue, L., Fanning, E., and Chazin, W. J. (2005) Insights into hRPA32 C-terminal domain--mediated assembly of the simian virus 40 replisome. *Nat Struct Mol Biol* 12, 332-339

56.   Bochkarev, A., Pfuetzner, R. A., Edwards, A. M., and Frappier, L. (1997) Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385, 176-181

57.   Bochkareva, E., Korolev, S., Lees-Miller, S. P., and Bochkarev, A. (2002) Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *Embo J* 21, 1855-1863

58.   Mer, G., Bochkarev, A., Chazin, W. J., and Edwards, A. M. (2000) Three-dimensional structure and function of replication protein A. *Cold Spring Harb Symp Quant Biol* 65, 193-200

59.   Collins, K. L., Russo, A. A., Tseng, B. Y., and Kelly, T. J. (1993) The role of the 70 kDa subunit of human DNA polymerase alpha in DNA replication. *Embo J* 12, 4555-4566

60.   Mizuno, T., Yamagishi, K., Miyazawa, H., and Hanaoka, F. (1999) Molecular architecture of the mouse DNA polymerase alpha-primase complex. *Mol Cell Biol* 19, 7886-7896

61.   Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H., and Steitz, T. A. (1997) Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell* 89, 1087-1099

62.   Brautigam, C. A., and Steitz, T. A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr Opin Struct Biol* 8, 54-63

63.   Ott, R. D., Rehfuess, C., Podust, V. N., Clark, J. E., and Fanning, E. (2002) Role of the p68 subunit of human DNA polymerase alpha-primase in simian virus 40 DNA replication. *Mol Cell Biol* 22, 5669-5678

64.   Uchiyama, M., and Wang, T. S. (2004) The B-subunit of DNA polymerase alpha-primase associates with the origin recognition complex for initiation of DNA replication. *Mol Cell Biol* 24, 7419-7434

65.   Kukimoto, I., Igaki, H., and Kanda, T. (1999) Human CDC45 protein binds to minichromosome maintenance 7 protein and the p70 subunit of DNA polymerase alpha. *Eur J Biochem* 265, 936-943

66.   Mizuno, T., Ito, N., Yokoi, M., Kobayashi, A., Tamai, K., Miyazawa, H., and Hanaoka, F. (1998) The second-largest subunit of the mouse DNA polymerase alpha-primase complex facilitates both production and nuclear translocation of the catalytic subunit of DNA polymerase alpha. *Mol Cell Biol* 18, 3552-3562

67.   Voitenleitner, C., Rehfuess, C., Hilmes, M., O'Rear, L., Liao, P. C., Gage, D. A., Ott, R., Nasheuer, H. P., and Fanning, E. (1999) Cell cycle-dependent regulation of human DNA polymerase alpha-primase activity by phosphorylation. *Mol Cell Biol* 19, 646-656

68.   Takemura, M., Yoshida, S., Akiyama, T., Kitagawa, M., and Yamada, Y. (2006) Role of the second-largest subunit of DNA polymerase alpha in the interaction between the catalytic subunit and hyperphosphorylated retinoblastoma protein in late S phase. *Biochim Biophys Acta* 1764, 1447-1453

69.   Copeland, W. C., and Tan, X. (1995) Active site mapping of the catalytic mouse primase subunit by alanine scanning mutagenesis. *J Biol Chem* 270, 3905-3913

70.   Kirk, B. W., and Kuchta, R. D. (1999) Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: primase and the family X polymerases share significant sequence homology. *Biochemistry* 38, 7727-7736

71.   Sheaff, R. J., and Kuchta, R. D. (1993) Mechanism of calf thymus DNA primase: slow initiation, rapid polymerization, and intelligent termination. *Biochemistry* 32, 3027-3037

72. Copeland, W. C., and Wang, T. S. (1993) Enzymatic characterization of the individual mammalian primase subunits reveals a biphasic mechanism for initiation of DNA replication. *J Biol Chem* 268, 26179-26189

73. Sheaff, R. J., Kuchta, R. D., and Ilsley, D. (1994) Calf thymus DNA polymerase alpha-primase: "communication" and primer-template movement between the two active sites. *Biochemistry* 33, 2247-2254

74. Keck, J. L., and Berger, J. M. (2001) Primus inter pares (first among equals). *Nat Struct Biol* 8, 2-4

75. Augustin, M. A., Huber, R., and Kaiser, J. T. (2001) Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nat Struct Biol* 8, 57-61

76. Ito, N., Nureki, O., Shirouzu, M., Yokoyama, S., and Hanaoka, F. (2001) Crystallization and preliminary X-ray analysis of a DNA primase from hyperthermophilic archaeon Pyrococcus horikoshii. *J Biochem (Tokyo)* 130, 727-730

77. Ito, N., Nureki, O., Shirouzu, M., Yokoyama, S., and Hanaoka, F. (2003) Crystal structure of the Pyrococcus horikoshii DNA primase-UTP complex: implications for the mechanism of primer synthesis. *Genes Cells* 8, 913-923

78. Lao-Sirieix, S. H., Nookala, R. K., Roversi, P., Bell, S. D., and Pellegrini, L. (2005) Structure of the heterodimeric core primase. *Nat Struct Mol Biol* 12, 1137-1144

79. Copeland, W. C. (1997) Expression, purification, and characterization of the two human primase subunits and truncated complexes from Escherichia coli. *Protein Expr Purif* 9, 1-9

80. Zerbe, L. K., and Kuchta, R. D. (2002) The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. *Biochemistry* 41, 4891-4900

81. Arezi, B., Kirk, B. W., Copeland, W. C., and Kuchta, R. D. (1999) Interactions of DNA with human DNA primase monitored with photoactivatable cross-linking agents: implications for the role of the p58 subunit. *Biochemistry* 38, 12899-12907

82. Stauffer, M. E., and Chazin, W. J. (2004) Structural mechanisms of DNA replication, repair, and recombination. *J Biol Chem* 279, 30915-30918

83. Roy, R., Trowbridge, P., Yang, Z., Champoux, J. J., and Simmons, D. T. (2003) The cap region of topoisomerase I binds to sites near both ends of simian virus 40 T antigen. *J Virol* 77, 9809-9816

84.   Khopde, S., and Simmons, D. T. (2008) Simian virus 40 DNA replication is dependent on an interaction between topoisomerase I and the C-terminal end of T antigen. *J Virol* 82, 1136-1145

85.   Dornreiter, I., Copeland, W. C., and Wang, T. S. (1993) Initiation of simian virus 40 DNA replication requires the interaction of a specific domain of human DNA polymerase alpha with large T antigen. *Mol Cell Biol* 13, 809-820

86.   Weisshart, K., Forster, H., Kremmer, E., Schlott, B., Grosse, F., and Nasheuer, H. P. (2000) Protein-protein interactions of the primase subunits p58 and p48 with simian virus 40 T antigen are required for efficient primer synthesis in a cell-free system. *J Biol Chem* 275, 17328-17337

87.   Weisshart, K., Taneja, P., and Fanning, E. (1998) The replication protein A binding site in simian virus 40 (SV40) T antigen and its role in the initial steps of SV40 DNA replication. *J Virol* 72, 9771-9781

88.   Dornreiter, I., Erdile, L. F., Gilbert, I. U., von Winkler, D., Kelly, T. J., and Fanning, E. (1992) Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen. *Embo J* 11, 769-776

89.   Nasheuer, H. P., von Winkler, D., Schneider, C., Dornreiter, I., Gilbert, I., and Fanning, E. (1992) Purification and functional characterization of bovine RP-A in an in vitro SV40 DNA replication system. *Chromosoma* 102, S52-59

90.   Braun, K. A., Lao, Y., He, Z., Ingles, C. J., and Wold, M. S. (1997) Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase alpha by multiple mechanisms. *Biochemistry* 36, 8443-8454

91.   Simmons, D. T., Gai, D., Parsons, R., Debes, A., and Roy, R. (2004) Assembly of the replication initiation complex on SV40 origin DNA. *Nucleic Acids Res* 32, 1103-1112

92.   Foiani, M., Marini, F., Gamba, D., Lucchini, G., and Plevani, P. (1994) The B subunit of the DNA polymerase alpha-primase complex in Saccharomyces cerevisiae executes an essential function at the initial stage of DNA replication. *Mol Cell Biol* 14, 923-933

93.   Foiani, M., Liberi, G., Lucchini, G., and Plevani, P. (1995) Cell cycle-dependent phosphorylation and dephosphorylation of the yeast DNA polymerase alpha-primase B subunit. *Mol Cell Biol* 15, 883-891

94.   Schub, O., Rohaly, G., Smith, R. W., Schneider, A., Dehde, S., Dornreiter, I., and Nasheuer, H. P. (2001) Multiple phosphorylation sites of DNA polymerase alpha-primase cooperate to regulate the initiation of DNA replication in vitro. *J Biol Chem* 276, 38076-38083

95.   Voitenleitner, C., Fanning, E., and Nasheuer, H. P. (1997) Phosphorylation of DNA polymerase alpha-primase by cyclin A-dependent kinases regulates initiation of DNA replication in vitro. *Oncogene* 14, 1611-1615

96.   Arezi, B., and Kuchta, R. D. (2000) Eukaryotic DNA primase. *Trends Biochem Sci* 25, 572-576

97.   Frick, D. N., and Richardson, C. C. (2001) DNA primases. *Annu Rev Biochem* 70, 39-80

98.   Schneider, A., Smith, R. W., Kautz, A. R., Weisshart, K., Grosse, F., and Nasheuer, H. P. (1998) Primase activity of human DNA polymerase alpha-primase. Divalent cations stabilize the enzyme activity of the p48 subunit. *J Biol Chem* 273, 21608-21615

99.   Lao-Sirieix, S. H., Pellegrini, L., and Bell, S. D. (2005) The promiscuous primase. *Trends Genet* 21, 568-572

100.   Foiani, M., Lucchini, G., and Plevani, P. (1997) The DNA polymerase alpha-primase complex couples DNA replication, cell-cycle progression and DNA-damage response. *Trends Biochem Sci* 22, 424-427

101.   De Falco, M., Ferrari, E., De Felice, M., Rossi, M., Hubscher, U., and Pisani, F. M. (2007) The human GINS complex binds to and specifically stimulates human DNA polymerase alpha-primase. *EMBO Rep* 8, 99-103

102.   Kamada, K., Kubota, Y., Arata, T., Shindo, Y., and Hanaoka, F. (2007) Structure of the human GINS complex and its assembly and functional interface in replication initiation. *Nat Struct Mol Biol* 14, 388-396

103.   Labib, K., and Gambus, A. (2007) A key role for the GINS complex at DNA replication forks. *Trends Cell Biol* 17, 271-278

104.   Michael, W. M., Ott, R., Fanning, E., and Newport, J. (2000) Activation of the DNA replication checkpoint through RNA synthesis by primase. *Science* 289, 2133-2137

105.   Byun, T. S., Pacek, M., Yee, M. C., Walter, J. C., and Cimprich, K. A. (2005) Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev* 19, 1040-1052

106.   Cortez, D. (2005) Unwind and slow down: checkpoint activation by helicase and polymerase uncoupling. *Genes Dev* 19, 1007-1012

107.   Paulsen, R. D., and Cimprich, K. A. (2007) The ATR pathway: Fine-tuning the fork. *DNA Repair (Amst)* 6, 953-966

108. Lambert, S., Froget, B., and Carr, A. M. (2007) Arrested replication fork processing: Interplay between checkpoints and recombination. *DNA Repair (Amst)* 6, 1042-1061

109. Rudolf, J., Makrantoni, V., Ingledew, W. J., Stark, M. J., and White, M. F. (2006) The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol Cell* 23, 801-808

110. Hinks, J. A., Evans, M. C., De Miguel, Y., Sartori, A. A., Jiricny, J., and Pearl, L. H. (2002) An iron-sulfur cluster in the family 4 uracil-DNA glycosylases. *J Biol Chem* 277, 16936-16940

111. Boal, A. K., Yavin, E., Lukianova, O. A., O'Shea, V. L., David, S. S., and Barton, J. K. (2005) DNA-bound redox activity of DNA repair glycosylases containing [4Fe-4S] clusters. *Biochemistry* 44, 8397-8407

112. Hemminga, M. A., and Berliner, L. J. (2007) *ESR spectroscopy in membrane biophysics*, Springer, New York

113. Dilg, A. W., Mincione, G., Achterhold, K., Iakovleva, O., Mentler, M., Luchinat, C., Bertini, I., and Parak, F. G. (1999) Simultaneous interpretation of Mossbauer, EPR and 57Fe ENDOR spectra of the [Fe4S4] cluster in the high-potential iron protein I from Ectothiorhodospira halophila. *J Biol Inorg Chem* 4, 727-741

114. Laurent Le Pape, B. L., Jean-Marie Mouesca, and Gérard Rius (1997) Paramagnetic States of Four Iron-Four Sulfur Clusters. 1. EPR Single-Crystal Study of 3+ and 1+ Clusters of an Asymmetrical Model Compound and General Model for the Interpretation of the g-Tensors of These Two Redox States. *J. Am. Chem. Soc* 119, 9757 -9770

115. Brown, N. M., Kennedy, M. C., Antholine, W. E., Eisenstein, R. S., and Walden, W. E. (2002) Detection of a [3Fe-4S] cluster intermediate of cytosolic aconitase in yeast expressing iron regulatory protein 1. Insights into the mechanism of Fe-S cluster cycling. *J Biol Chem* 277, 7246-7254

116. Coldren, C. D., Hellinga, H. W., and Caradonna, J. P. (1997) The rational design and construction of a cuboidal iron-sulfur protein. *Proc Natl Acad Sci U S A* 94, 6635-6640

117. Staples, C. R., Ameyibor, E., Fu, W., Gardet-Salvi, L., Stritt-Etter, A. L., Schurmann, P., Knaff, D. B., and Johnson, M. K. (1996) The function and properties of the iron-sulfur center in spinach ferredoxin: thioredoxin reductase: a new biological role for iron-sulfur clusters. *Biochemistry* 35, 11425-11434

118. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25, 4876-4882

119.   Michaels, M. L., Pham, L., Nghiem, Y., Cruz, C., and Miller, J. H. (1990) MutY, an adenine glycosylase active on G-A mispairs, has homology to endonuclease III. *Nucleic Acids Res* 18, 3841-3845

120.   Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Munck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) Endonuclease III is an iron-sulfur protein. *Biochemistry* 28, 4450-4455

121.   Johnson, D. C., Dean, D. R., Smith, A. D., and Johnson, M. K. (2005) Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem* 74, 247-281

122.   Porello, S. L., Cannon, M. J., and David, S. S. (1998) A substrate recognition role for the [4Fe-4S]2+ cluster of the DNA repair glycosylase MutY. *Biochemistry* 37, 6465-6475

123.   Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, S., and Tainer, J. A. (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat Struct Biol* 5, 1058-1064

124.   Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *Embo J* 14, 4108-4120

125.   Chepanoske, C. L., Golinelli, M. P., Williams, S. D., and David, S. S. (2000) Positively charged residues within the iron-sulfur cluster loop of E. coli MutY participate in damage recognition and removal. *Arch Biochem Biophys* 380, 11-19

126.   Lill, R., and Muhlenhoff, U. (2006) Iron-sulfur protein biogenesis in eukaryotes: components and mechanisms. *Annu Rev Cell Dev Biol* 22, 457-486

127.   Lukianova, O. A., and David, S. S. (2005) A role for iron-sulfur clusters in DNA repair. *Curr Opin Chem Biol* 9, 145-151

128.   Boon, E. M., Livingston, A. L., Chmiel, N. H., David, S. S., and Barton, J. K. (2003) DNA-mediated charge transport for DNA repair. *Proc Natl Acad Sci U S A* 100, 12543-12547

129.   Francesconi, S., Longhese, M. P., Piseri, A., Santocanale, C., Lucchini, G., and Plevani, P. (1991) Mutations in conserved yeast DNA primase domains impair DNA replication in vivo. *Proc Natl Acad Sci U S A* 88, 3877-3881

130.   Longhese, M. P., Jovine, L., Plevani, P., and Lucchini, G. (1993) Conditional mutations in the yeast DNA primase genes affect different aspects of DNA metabolism and interactions in the DNA polymerase alpha-primase complex. *Genetics* 133, 183-191

131. Klinge, S., Hirst, J., Maman, J. D., Krude, T., and Pellegrini, L. (2007) An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis. *Nat Struct Mol Biol* 14, 875-877

132. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng* 6, 383-390

133. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-293

134. Johnson, B. A., and Blevins, R. A. (1994) NMR View - A computer-program for the visualization and analysis of NMR data. *J Biomol NMR* 4, 603-614

135. Dornreiter, I., Hoss, A., Arthur, A. K., and Fanning, E. (1990) SV40 T antigen binds directly to the large subunit of purified DNA polymerase alpha. *Embo J* 9, 3329-3336

136. Cole, and Barton (2008)

137. Takeuchi, K., and Wagner, G. (2006) NMR studies of protein interactions. *Curr Opin Struct Biol* 16, 109-117

138. Matsuo, H., Walters, K., Teruya, K., Tanaka, T., Gassner, G., Lippard, S., Kyogoku, Y., and Wagner, G. (1999) Identification by NMR spectroscopy of residues at contact surfaces in large, slowly exchanging macromolecular complexes. *J Am Chem Soc* 121, 9903-9904

139. Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340, 245-246

140. Goddard, T. D., and Kneller, D. G. SPARKY 3. *University of California, San Francisco*

141. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13, 289-302

142. Guntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278, 353-378

143. Sastry, M., Ketchem, R. R., Crescenzi, O., Weber, C., Lubienski, M. J., Hidaka, H., and Chazin, W. J. (1998) The three-dimensional structure of Ca(2+)-bound calcyclin: implications for Ca(2+)-signal transduction by S100 proteins. *Structure* 6, 223-231

144. Guntert, P., Braun, W., and Wuthrich, K. (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol* 217, 517-530

145. Case, D. A. (2006) AMBER 9. *University of California, San Francisco*

146. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8, 477-486

147. Holm, L., and Sander, C. (1996) Mapping the protein universe. *Science* 273, 595-603

148. Putnam, C. D., Clancy, S. B., Tsuruta, H., Gonzalez, S., Wetmur, J. G., and Tainer, J. A. (2001) Structure and mechanism of the RuvB Holliday junction branch migration motor. *J Mol Biol* 311, 297-310

149. Beinert, H., Holm, R. H., and Munck, E. (1997) Iron-sulfur clusters: nature's modular, multipurpose structures. *Science* 277, 653-659

150. Bian, S., and Cowan, J. (1999) Protein-bound iron-sulfur centers. Form, function, and assembly. *Coordin Chem Rev* 192, 1049-1066

151. Pugh, R. A., Honda, M., Leesley, H., Thomas, A., Lin, Y., Nilges, M. J., Cann, I. K., and Spies, M. (2008) The iron-containing domain is essential in Rad3 helicases for coupling of ATP hydrolysis to DNA translocation and for targeting the helicase to the single-stranded DNA-double-stranded DNA junction. *J Biol Chem* 283, 1732-1743

152. Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., Cooper, P. K., and Tainer, J. A. (2008) XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* 133, 789-800

153. Liu, H., Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., McRobbie, A. M., Brown, S. E., Naismith, J. H., and White, M. F. (2008) Structure of the DNA repair helicase XPD. *Cell* 133, 801-812

154. Bertini, I., Couture, M. M., Donaire, A., Eltis, L. D., Felli, I. C., Luchinat, C., Piccioli, M., and Rosato, A. (1996) The solution structure refinement of the paramagnetic reduced high-potential iron-sulfur protein I from Ectothiorhodospira halophila by using stable isotope labeling and nuclear relaxation. *Eur J Biochem* 241, 440-452

155. Bermel, W., Bertini, I., Felli, I. C., Kummerle, R., and Pierattelli, R. (2006) Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *J Magn Reson* 178, 56-64

156. Huang, S. G., Weisshart, K., Gilbert, I., and Fanning, E. (1998) Stoichiometry and mechanism of assembly of SV40 T antigen complexes with the viral origin of DNA replication and DNA polymerase alpha-primase. *Biochemistry* 37, 15345-15352

157. Gai, D., Roy, R., Wu, C., and Simmons, D. T. (2000) Topoisomerase I associates specifically with simian virus 40 large-T-antigen double hexamer-origin complexes. *J Virol* 74, 5224-5232

158. Taneja, P., Nasheuer, H. P., Hartmann, H., Grosse, F., Fanning, E., and Weisshart, K. (2007) Timed interactions between viral and cellular replication factors during the initiation of SV40 in vitro DNA replication. *Biochem J* 407, 313-320