NMR ANALYSIS OF THE RPA DIMER CORE RPA32D/14

by

Ileana Alers Rivera

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

Master of Science in

Chemical and Physical Biology

August, 2011

Nashville, Tennessee

Approved:

Walter J. Chazin

Albert H. Beth

Lacy D. Borden

ACKNOWLEDGMENTS

I will like to thank my advisor Dr. Walter Chazin. First, for letting me be part of his lab. Without his guidance and mentorship in would have been impossible. He is a great scientist and can transmit his passion for science to us the students.

I thank all the members from the Chazin lab for their help and insightful comments. I want to especially thank Chris Brosey who helped me with processing and analyzing the NMR data. I thank Dr. Siva Vaithiyalingam who helped me on setting up the experiments for the primase project. Also past and present members of the Chazin group who made the lab a great environment to work and some have become great friends.

I thank my committee members Dr. Albert H. Beth and Lacy D. Borden for their input.

I also want to thank my family and friends for their support.

Finally, I want to thank the NIH for providing the funding for my research.

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER I

CHAPTER II

APPENDIX

LIST OF TABLES

CHAPTER I

CHAPTER II

LIST OF ABBREVIATIONS

2D	two dimensional
ATR	ataxia telangiectasia-related protein
ATRIP	ATR-interacting protein
BME	beta mercaptoethanol
BRCA2	breast cancer, gene 2
cdc13	cell division cycle protein 13
CRIPT	cross-correlated relaxation-induced polarization transfer
CRINEPT	cross-correlated relaxation-enhanced polarization transfer
CSA	chemical shift anisotropy
CTD	c-terminal domain
dsDNA	double stranded deoxyribonucleic acid
DTT	dithiothreitol
ERCC1	excision repair cross complementing 1
ES-MS	electrospray mass spectrometry
Estp1	ever shorter telomere, gene 1
G-residues	guanine residues
hPot1	human protection of telomeres
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl thio-beta-D-galactopyranoside
ITC	isothermal calorimetry
LB	lysogeny broth

M9	minimal media
MALS	multi-angle light scattering
MRE11	meiotic recombination, homolog II
NBS1	Nijmegen breakage syndrome 1
NER	nucleotide excision repair
Ni-NTA	nickel nitrilotriacetic acid
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser effect spectroscopy
NP-40	nonidet P-40
OB-Fold	oligosacharide/oligonucleotide binding fold
OD	optical density
PCNA	proliferating cell nuclear antigen
PMSF	phenylmethylsulfonyl fluoride
pol-prim	polymerase α-primase
Pot1	protection of telomeres
RFC	replicator factor C
RMSD	root mean square deviation
RPA	replication protein A
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
ssDNA	single stranded deoxyribonucleic acid
TROSY	transverse relaxation optimized spectroscopy
UNG2	nuclear uracil-DNA glycosylase

WHD	winged helix domain
XPA	xeroderma pigmentosum, complementation group A
XPC	xeroderma pigmentosum, complementation group C
XPF	xeroderma pigmentosum, complementation group F
XPG	xeroderma pigmentosum, complementation group G

Chapter I

Introduction

DNA replication and repair

DNA replication is a fundamental process for passing the genetic information from one generation to the next. DNA replication occurs once per cell cycle in a highly regulated manner. Eukaryotic DNA replication requires the proper assembly of multiple proteins into what is known as the replisome [2]. The initial step for DNA replication involves the recognition of origin sites of replication to form the pre-replicative complex. Transitioning into the synthesis phase of replication requires unwinding of duplex DNA by helicases [3, 4]. Exposed ssDNA is coated by the eukaryotic single-stranded DNA binding protein Replication Protein A (RPA) [5] RPA protects and coordinates the assembly and disassembly of downstream proteins in the replication pathways through protein-protein interactions. The first step in DNA synthesis is to generate a RNA-DNA primer by DNA polymerase α-primase [6], then PCNA, RFC and polymerases δ/ε are loaded to synthesize the DNA in the leading and lagging strand [7].

The process of replication does not go smoothly all the time. For example, sometimes, DNA lesions are encountered by the DNA replication machinery. This can cause stalling of the replication machinery or interfere with proper placement of nucleotides, resulting in mutations. Mutations during replication can lead to genome instability and diseases with fatal consequences such as cancer [8]. Some mutations can interfere with the normal function of a cell, or they can be lethal during early

development of an embryo. The lesions encountered by the DNA replication machinery must be overcome to maintain the fidelity of the process. The cell has developed multiple pathways to remove lesions from DNA, three of these are: nucleotide excision repair, base excision repair and double strand break repair recombination [9]. All three recognize specific kinds of lesions. RPA has been shown to interact with one or more proteins involved in these three different pathways, which places RPA as a key player in the process that couples DNA repair and replication.

Another challenge encountered by DNA replication is telomeres. Eukaryotic cells have developed telomeres at the end of linear chromosomes to prevent their joining or degradation. Telomere ends are of great importance because of their role in aging and cancer [10]. Telomeric ends are composed of repeating sequences of DNA containing alternating short tracks of guanine residues and A/T-rich regions. One of the strands contains a G-rich overhang that is able to form a secondary structure known as G-quadraplex (Fig. 1.4) [11]. Specific proteins are known to interact with G-quadraplexes to protect and regulate access to the DNA by forming packed structures at the end of the chromosomes [12]. These structures represent a topological problem that must be overcome to be able to completely replicate the DNA. The conventional machinery for DNA replication is not able to copy the DNA at these ends and therefore, a whole new machinery has been developed by cells [13]. One of the reasons is because DNA polymerase is not able to replicate 5'-ends of the DNA. Secondly, proteins interacting with G-quadruplexes must first dissociate to give access to the DNA replication machinery, and thirdly the secondary structures formed by the G-rich overhang must be resolved and converted into ssDNA in order to be replicated. Several lines of

investigation place RPA as an important protein in the regulation of telomeric ends, but how RPA regulates these processes is still poorly understood. RPA has been shown to aid in the recruitment of telomerase to telomeric DNA [14]. RPA has also been shown to bind and unfold G-quadruplexes [15] and to have structural homology to telomere specific interacting proteins [16, 17].

Replication Protein A

Replication Protein A (RPA) is the primary eukaryotic ssDNA binding protein [18]. RPA is a heterotrimer composed of subunits RPA70, RPA32 and RPA14, named after their estimated molecular weight. Binding of RPA to ssDNA protects it from nucleases and prevents formation of ssDNA and other secondary structures. RPA is involved in numerous functions involving the metabolism of DNA such as replication, repair, damage response, recombination and telomere processing. RPA is a modular protein with multiple domains connected by flexible linkers [19].

Figure 1.1. The oligosaccharide/oligonucleotide binding fold of RPA70A [1].

Six of the eight RPA domains are oligonucleotide-oligosaccharide binding folds (OB-fold domains). OB-fold domains are known for their interaction with DNA, also RNA and bacterial polysaccharides [20]. They are composed of 5-stranded antiparallel β-sheets, with an α helix between the third and fourth strands [21] (Fig. 1.1). This motif folds into a closed β-barrel, containing a basic patch positioned in between loops 1-2 and 4-5, which is the preferred site of interaction for ssDNA. The properties of each OB-fold differ according to its function and the identity of the amino acids in the cleft defines its affinity for DNA. OB-folds with high affinity for ssDNA exhibit three main features: aromatic residues able to stack with DNA bases, ability to form several hydrogen bonds between side chains and DNA nucleotides, and electrostatic interactions between basic amino acids and the DNA phosphate backbone [21]. In other cases, OB-fold domains serve as protein-protein interaction domains; their specificity can also be dictated by the identity of the amino acids in the OB-fold cleft, although there is much more variability than for interactions with DNA.

RPA70 is composed of domains RPA70N, RPA70A, RPA70B and RPA70C [18]. The N-terminus domain of RPA70 is separated from domain RPA70A by a long linker of 70 amino acids. This allows RPA70N to move independently of the other three domains. Although an OB-fold is found in this domain, evidence from SAXS studies demonstrated that RPA70N plays no role in DNA binding [22]. On the other hand, several proteins involved in DNA repair, checkpoint control, and damage response are known to interact with the basic cleft of RPA70N [23, 24]. Domains RPA70A and RPA70B are separated by a 10 amino acid linker (Fig. 1.2), and provide the initial high affinity interactions with

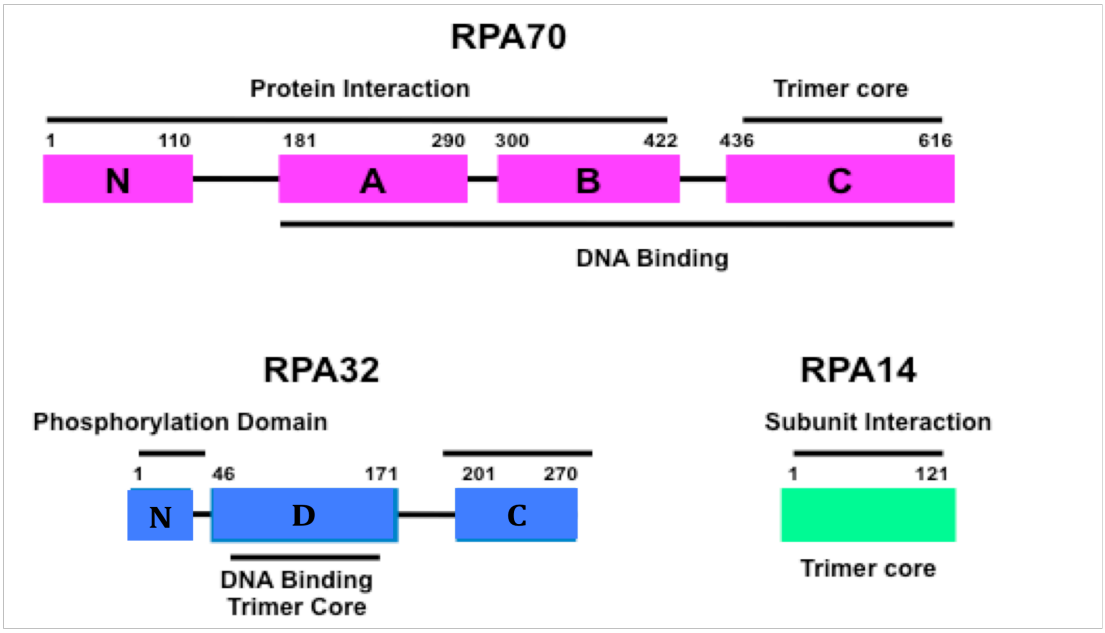


Figure 1.2. Cartoon depiction of RPA subunits and their domains.

ssDNA [25]. There is a 15 amino acid linker separating RPA70C, which also binds ssDNA. Additionally, RPA70C contains a zinc-binding motif within its structure but its functional role is poorly understood. A role in DNA binding has been attributed to the zinc-binding motif based upon comparisons made with other proteins containing a similar motif within their structure [26]. RPA70A and RPA70B also interact with proteins involved in DNA replication [27] and recombination [28].

RPA32 is comprised of an unstructured N-terminus followed by two structured domains, RPA32D and RPA32C, separated by a long ~30 residue linker [29] (Fig 1.2). The N-terminal disordered domain contains eight phosphorylation sites that are phosphorylated in a cell cycle dependent manner [30]. The OB-fold domain RPA32D contributes to the ssDNA binding activity of RPA and also froms the trimer core interfases with RPA14 and RPA70C. RPA32C is a winged helix domain (WHD) involved in the interaction with replication and repair proteins [29, 31]. RPA14 forms an

OB-fold domain required for the formation and solubility of the trimer core [32]. There is no evidence for DNA binding activity or protein interaction of RPA14; its role is still not well understood. RPA32/RPA14 and RPA32D/14 form stable soluble complexes [33]. This RPA 'dimer' is thought to aid folding of RPA70C to produce the active trimer [18]. In fact, RPA32/14 has been found to form *in vitro*, in cells undergoing apoptosis and at the telomeric end of chromosomes [32].

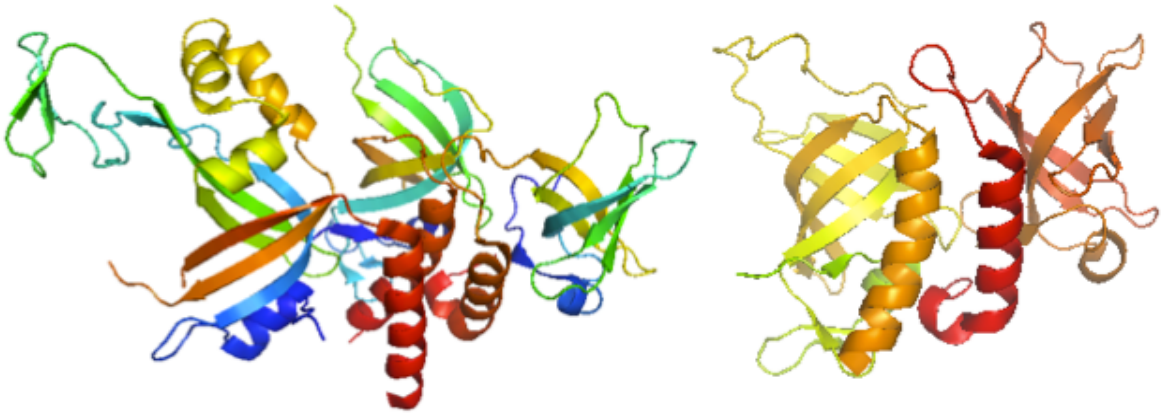


Figure 1.3. Ribbon diagrams of the RPA trimer core and heterodimer. The structures of RPA70C/32/14 (left) and RPA32D/14 (right) are shown (PDB id: 1L1O and 2PQA).

The crystal structures of RPA32D/14 constructs and the structure of the trimer core (RPA70C/32D/14) have been determined by X-ray crystallography [26] (Fig. 1.3). Comparison of the structures of RPA32D/14 alone and in the context of the trimer core showed only small deviations in terms of the contacts between domains and they are discussed in detail by Bochkareva *et. al*. [26]. The interfaces between subunits are mediated by C-terminal α-helices extending from the OB-fold domains. The three α-

helices are arranged in parallel forming a three helical bundle that is stabilized through hydrophobic interactions.

RPA Interaction with ssDNA

RPA binds to ssDNA with a 5' to 3' polarity and independent of the sequence. RPA has been implicated in proper positioning of proteins with respect to the DNA template [34]. Although individually each of the OB-folds binds DNA weakly, full-length RPA binds tightly with a K_d of ~10^{-10} M [35]. Interaction with DNA occurs in three different binding modes. The initial 8-10 nucleotide binding mode involves the tandem RPA70A and RPA70B domains. As more DNA becomes available, RPA70C engages, occluding 14-20 nucleotides. A final 30 nucleotide binding mode utilizes RPA70A, RPA70B, RPA70C and RPA32D [36]. OB-fold domains organized in tandem and separated by short linkers allow RPA to bind tighter to DNA. Exactly how ssDNA threads onto the RPA domains, as well as the spatial arrangement of domains is still not well understood. This is in part due to the difficulties encountered when trying to determine the architecture or quaternary structure of a dynamically tethered multi-domain protein. No structure of full-length RPA is available due largely to the flexibility of the protein. At present, all the structural information available from NMR or crystallography has been gathered from RPA domain constructs.

The fact that RPA binds so tightly to DNA brings up the question of how it is able to release from DNA so that other proteins can gain access and process the DNA. One hypothesis is the remodeling of RPA domains. Various proteins are known to have more than one contact point with RPA domains. This allows proteins involved in different

DNA pathways to rearrange or remodel RPA architecture to decrease its affinity for DNA and provides potential means to compete RPA off [37]. RPA and many other DNA processing proteins are modular and are believed to function via a hand-off mechanism [35,47,48].

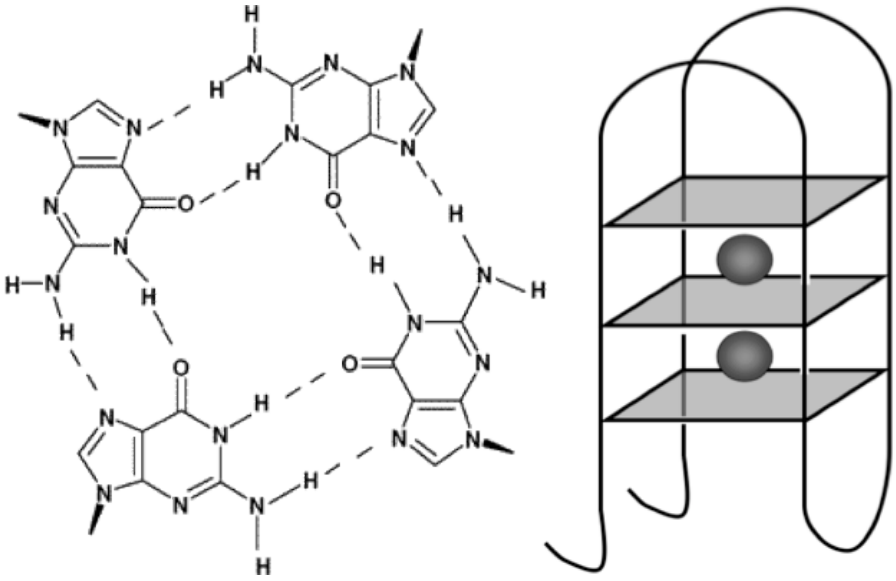


Figure 1.4. DNA G-quadruplex structure. Formation of the structure involves the planar tetrad arrangement of four guanines through Hoogsteen hydrogen bonding. The tetrad can stack on top of each other to form the G-quadruplex which is stabilized by the presence of cations.

Recently, RPA has been shown to bind non-canonical ssDNA sequences capable of forming non B-DNA structures [15, 38, 39]. Clusters of guanines residues are found throughout the genome, and they have a tendency to form G-quadruplex structures (Fig. 1.4). These structures form when four G-residues associate through Hoogsteen base-pairing to form a planar tetrad, which stack on top of each other. Intramolecular hairpin loops lead to parallel and antiparallel G-quadruplexes [11]. The stacking of G tetrads is

stabilized through binding of one or more cations, usually potassium. G-tetrads have been shown to form *in vitro* [11] and *in vivo* in some cases [40]. At telomere ends, overhangs rich in guanine residues are prone to form these secondary structures. G-quadruplexes prevent telomere replication and also recombination of the DNA. These structures have come to light because there are other sequences in human cells that are rich in guanine residues, thus having a high propensity to form quadruplexes. For example, promoters of the oncogenes c-myc, HIF-1α, bcl-2 and c-kit [41] have been shown to be rich in guanine residues. It is intriguing to explore the role G-quadruplex structures may play in the regulation of their transcription.

RPA has been found to interact with G-quadruplexes and to resolve them into ssDNA [15]. RPA shows similarities in terms of structure and sequence to telomere specific interacting proteins [12]. Pot1 (protection of telomeres) has been identified as one of these proteins in humans and yeast [42]. The human POT1 (hPOT1) structure has been determined in complex with telomeric DNA [17]. The structure shows two OB-fold domains in tandem separated by a linker (Fig. 1.5). Comparison between hPOT1 and RPA domains RPA70A and RPA70B bound to DNA shows a difference in the orientation of the two domains. In the case of RPA, the two OB-folds are oriented in the same direction [1] while hPOT1 has the OB-fold binding cleft connected to form a continuous channel. Yeast proteins stn1/Ten1 form a RPA32D/RPA14-like complex that has been shown to interact specifically with telomeric DNA through its OB-fold domains. High structural homology has been found between RPA32D and stn1, and RPA14 and Ten1 with C_α RMSD alignments of 2.4 and 2.8 Å, respectively [16]. Moreover, recent studies indicate that RPA is able to specifically bind and unfold G-quadruplex structures

through RPA32D [43]. All this evidence suggests a role for RPA in the regulation of telomeres at the end of the synthesis phase of DNA replication, and opening G-quadruplex structures to form ssDNA so that proteins involved in telomere maintenance can gain access to the DNA [44].

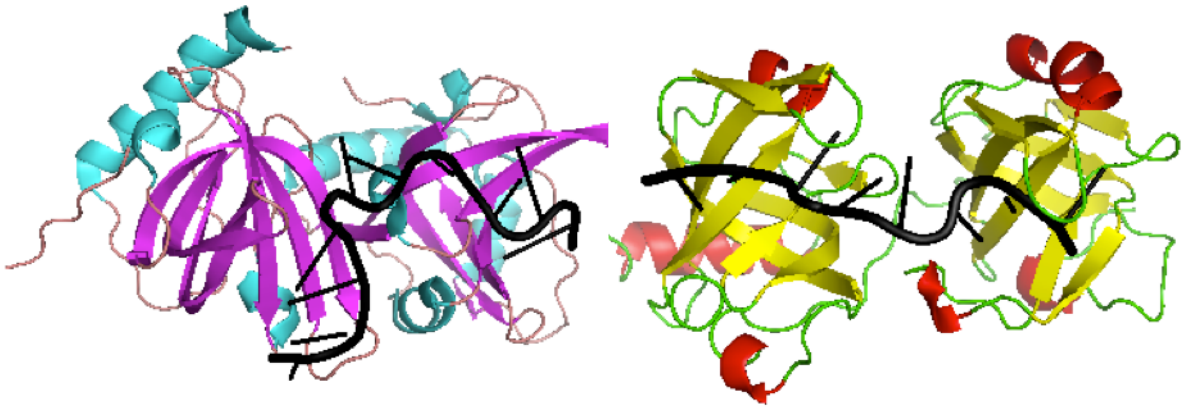


Figure 1.5. Ribbon diagram of hPOT1 (left) and RPA70AB (right) bound to ssDNA.

RPA-protein Interactions

RPA-protein interactions are an important aspect of modulating protein function and are thought to be one of the main ways for RPA to remodel and dissociate from ssDNA [45, 46]. Domains RPA70N, RPA70A, RPA70B and RPA32C are the mediators of these interactions and do so with high specificity.

During DNA replication, dynamic assembly of proteins at the replication fork is required for progression and regulation of replication. Protein-protein interactions mediated by RPA recruit proteins to the replication fork, but they must first dissociate from RPA so that the next protein in line can then associate. For example, polymerase α–primase [31], RFC, and Pol δ [47, 48] interact with RPA during replication at different

steps of the process. A competition-based protein switch mechanism has been proposed to explain how RPA may serve as an exchange point for multiple proteins involved in the same process [29].

Different proteins involved in different DNA repair pathways are known to interact with RPA [29]. This suggests a role for RPA in coordinating the assembly of proteins at sites of damage. One example is the interaction of RPA with three different proteins involved in the nucleotide excision repair pathway (NER). XPA and XPG are involved in the recognition of damaged DNA. Two contact points between XPA and RPA have been characterized, one with RPA32C [29] and the second with RPA70AB [49]. An interaction between RPA and XPG has been reported in the literature [50]. The nuclease ERCC-1/XPF has been shown to interact with RPA later on in the NER pathway. This interaction is proposed to be important in positioning ERCC-1/XPF with respect to the DNA strand to be cleaved [46]. RPA also interacts with RAD52 a protein essential in dsDNA break repair and UNG2 a protein involved in the base excision repair pathway. All of these are mediated through RPA32C WHD [29].

RPA is able to interact with cell cycle checkpoint control proteins including ATRIP, Mre11 and Rad9 [24]. Interaction of ATRIP with RPA, at least in part, through its 70N domain activates ATR mediated checkpoint pathways in response to DNA damage. The MRE11/Rad50/NBS1 complex as well as BRCA2 proteins, Rad51 and Rad52 are all involved in early stages of double strand break repair [28]. All these protein interact with RPA through its 70N domain, suggesting a common mechanism for function.

Recently, RPA was found to play a role in telomere maintenance. Interactions between RPA and telomere maintenance proteins have not been characterized

biophysically. Studies in human and yeast have demonstrated that RPA plays an important role at telomere ends [14, 43, 51]. Schramke et al. have shown that RPA is able to promote telomerase activity through Estp1 and cdc13, telomere specific binding proteins found at the G-rich overhang [14]. Their studies suggest an interaction between RPA and Estp1. In the proposed a mechanism of action, Estp1 is loaded in a RPA dependent manner, then Estp1 recruits telomerase to telomeric ends.

RPA studies by NMR

The structures of all the domains of RPA have been determined either by crystallography (RPA70AB, RPA70C/32/14) [1, 26] or NMR (RPA70N and RPA32C) [29, 52]. Although RPA is a large protein, it does not tumble as a single globular protein. Flexible linkers allow domains to tumble independently making feasible its study by NMR as demonstrated by Brosey *et. al.* [19]. However, to conduct such studies by NMR, resonance specific assignments must be made. These assignments will be explained in the next few sections. One strategy to ease the assignment process is to use fragments of RPA. Assignments are available for many RPA domains including RPA70N [29], RPA70AB [53] and RPA32C [29], but assignments for the trimer core are yet to be determined. This portion of the molecule is of special importance due to its role in binding ssDNA and DNA secondary structures (*i.e.* G-quadruplexes).

This thesis focuses on the assignment of a portion of the trimer core, RPA32D/14, known at the RPA dimer core. Beyond their intrinsic value, the assignments obtained for the dimer can be transferred to the trimer core for eventual study of this portion of the molecule. Thus, assignment of RPA32D/14 can be used for the study of the DNA

binding properties of the heterodimer and the trimer core. For example, NMR can be used to study how RPA's trimer core tumbling properties are changed upon DNA binding. Moreover, studies can be performed to determine conformational changes in RPA32/14 due to protein-protein interactions and gain more insight into the molecular hand-off mechanism.

NMR Spectroscopy - HSQC

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for the biophysical study of the structure, dynamics and chemical kinetics of biological macromolecules. These subjects have been amply covered in various reviews and books [54, 55]. NMR studies have been intensely developed over the past 30 years. The major step forward came from the ability to apply multi-nuclear, multi-dimensional NMR experiments. This enabled a quantum leap forward for the determination of structures, analysis of internal and global motions, and the analysis of inter-molecular interactions.

The 2D heteronuclear single quantum coherence (HSQC) experiment is the most basic heteronuclear NMR spectrum, and the ^{15}N-^{1}H HSQC spectrum is the most commonly used. In this experiment the ^{15}N resonance is exited in the first evolution indirect evolution period, then the ^{1}H resonances are detected by the receiver. By convention I will show the X axis for the hydrogen dimension and the Y axis for the nitrogen dimension. Essentially, the HSQC correlates the amide hydrogen resonance frequency with the nitrogen atom resonance frequency for each amide bond (Fig. 1.6). In the HSQC experiment one peak per backbone amide is is obtained except for proline. Additional signals are typically observed for the side chains of glutamine, asparagine and

tryptophan. At pH's below 7, other side chain NH resonances can also be observed. The power of the NMR spectrum is that the resonance frequencies (chemical shifts) are very dependant on the electronic environment of each nucleus so that this experiment serves as a fingerprint for the protein structure. The simplicity of this experiment, having one signal per amino acid, allows one to rapidly assess the structural integrity of the protein and binding events.

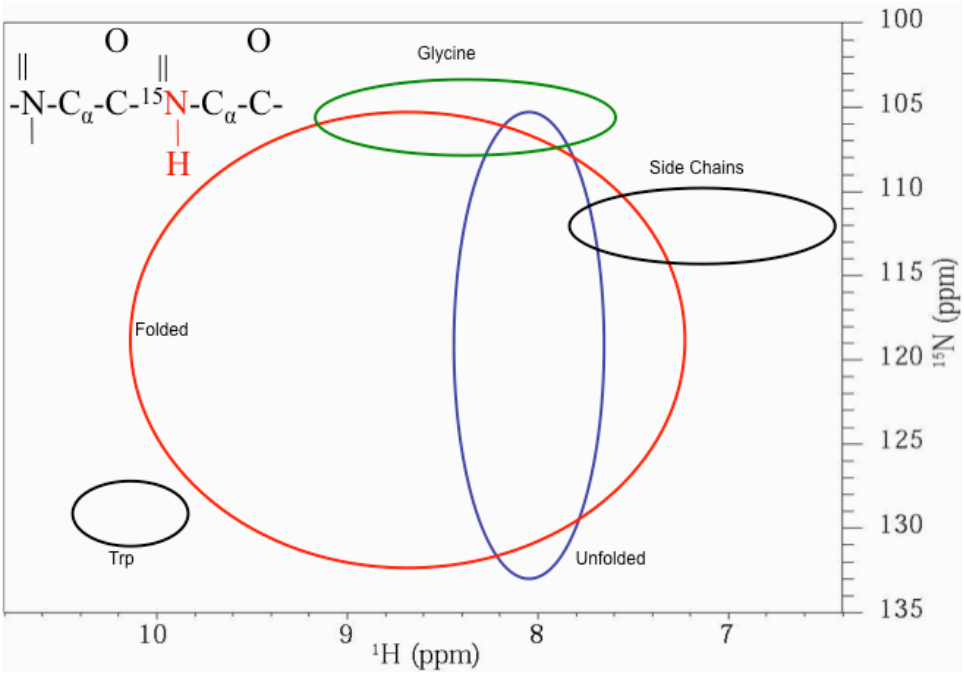


Figure 1.6. Characteristic regions of ^{15}N-^{1}H HSQC spectra. Glycines are found at the top of the spectrum and depicted by the green oval. Side chain signals for glutamine and asparagine as well as tryptophans are depicted with black ovals. Folded proteins have signals spread as depicted by the red oval, while unfolded proteins will have all the signals in one region of the spectrum and shown by the blue oval.

A HSQC spectrum is often used to assess the applicability of NMR analysis. As shown in Figure 1.6, HSQC spectra give information on whether the protein is folded or unfolded. The hallmark for an unfolded protein is that all signals will be localized in a certain region of the spectrum (blue oval). Thus the HSQC can be used to search for conditions where the protein is well folded.

The NMR TROSY technique

The sensitivity and resolution of heteronuclear experiments can be increased through the use of the transverse relaxation optimized spectroscopy (TROSY) technique [56]. TROSY takes advantage of the interference between dipolar and chemical shift anisotropy (CSA) relaxation pathways. These two pathways of relaxation can add or cancel each other depending on the spin state of the coupled spins. CSA is dependant on the magnetic field; at high magnetic field (~1 Ghz), the oscillation for CSA becomes equal to the field from dipolar coupling, causing a more effective cancellation, leaving only the most narrow resonance line. The effect of TROSY will be discussed next using the HSQC experiment as an example. The TROSY strategy can be incorporated into most NMR experiments and is especially valuable for triple resonance experiments to make resonance assignments for large molecules.

The HSQC spectrum of large proteins can be very crowded and have very broad resonances. In the non-decoupled HSQC spectra there are four signals shown for each ^{1}H-^{15}N pair (Fig. 1.7A). Each signal differs in intensity due to non, partial or complete cancelation of CSA and dipolar relaxation. When the decoupled HSQC is recorded the signal intensity of the four components is averaged which results in partial cancellation

but only one signal per ^{15}N-^{1}H pair (Fig. 1.7B). When a TROSY pulse sequence is applied, the signal from the three broader components is filtered out and only the sharpest resonance is left (Fig. 1.7C). This results in higher spectral resolution [56].

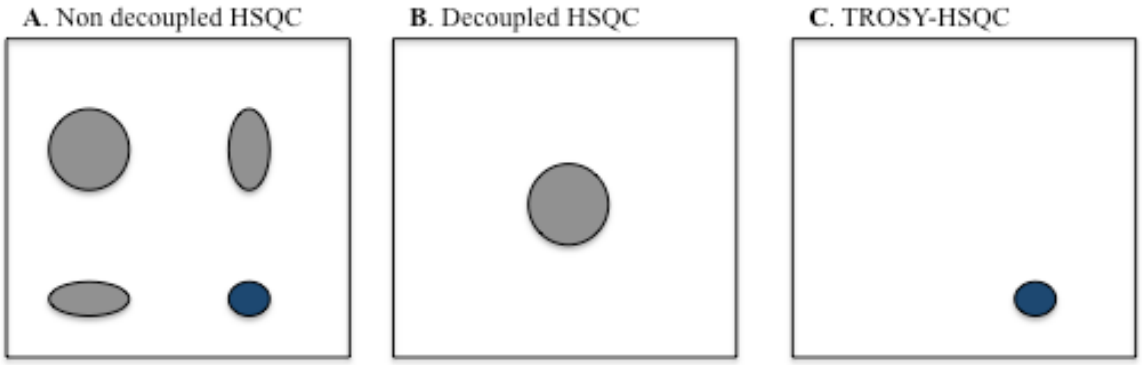


Figure 1.7. The TROSY effect. Four signals are observed for the non-decoupled HSQC spectra (A), decoupling averages all four signals resulting in a single broad resonance (B). TROSY provides "spectral editing", selecting the sharpest of the four signals (C).

NMR resonance assignments: Principles and experiments

The very first step for any NMR experiment is the assignment of resonances in the spectrum. NMR assignments correlate each signal to a specific nucleus in the molecule. NMR provides essentially two types of information useful in the resonance assignment process: through-bond interactions (via scalar coupling), and through-space interactions (via dipolar coupling). The process of assignment also utilizes the protein sequence. During the assignment process resonance signals belonging to the same amino acid are identified. This collection of specific residue signals is referred to as a spin system. With this information, the next step is to identify which amino acid in the protein sequence a given spin belongs to. The identity of the residue is determined based on the number of

resonances and their chemical shifts. The combination of ^{1}H, ^{13}C and ^{15}N chemical shifts can be used to determine the type of a spin system.

This section will focus on the discussion of triple resonance NMR experiments for backbone assignment of perdeuterated proteins. The ^{1}H-^{13}C and ^{1}H-^{15}N dipolar interactions provide a very dominant relaxation pathway for ^{13}C and ^{15}N. Major relaxation of ^{1}H spins also comes from the dipolar interaction, typically with aliphatic ^{1}H spins in close proximity. Perdeuteration reduces these relaxation pathways improving the resolution and sensitivity of the triple resonance experiments. These allow NMR assignments for proteins larger than 20-30 kDa.

Triple resonance backbone experiments correlate $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, ^{13}CO, $^{1}H^{\beta}$, and $^{13}C^{\beta}$ resonances using one-bond and two-bond scalar coupling interactions. Scalar coupling between heteronuclear spins is larger than those between homonuclear spins providing efficient transfer of magnetization through the backbone (Fig. 1.8). Another feature exploited by these experiments is the fact that $^{13}C_{\alpha}$ and CO can be treated as independent nuclei due to their very different chemical shift range, allowing magnetization to be transferred from ^{15}N-^{13}CO or ^{15}N-$^{13}C_{\alpha}$. The nuclear spin labeled during indirect evolution periods or the acquisition period are listed as HN, N, CA, CO and CB in the experiment name to represent $^{1}H^{N}$, ^{15}N, $^{13}C_{\alpha}$, ^{13}CO, and $^{13}C_{\beta}$ spins, respectively. Spins shown in parenthesis represent those through which coherence is transferred but are not frequency labeled (no signal is observed in the spectrum).

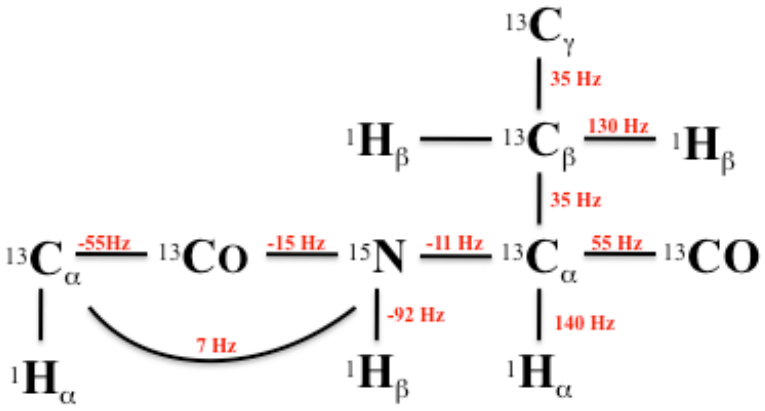
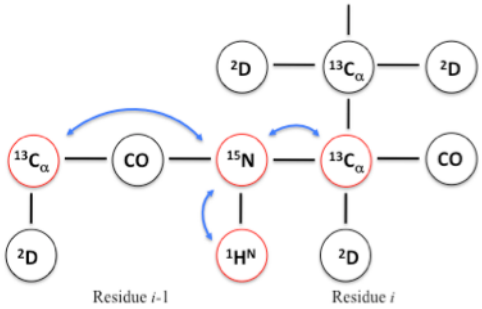
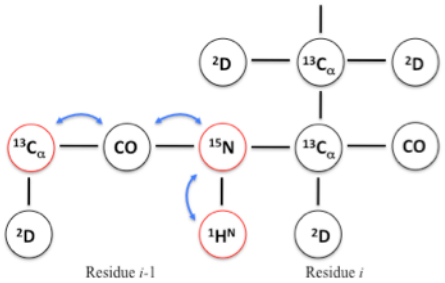
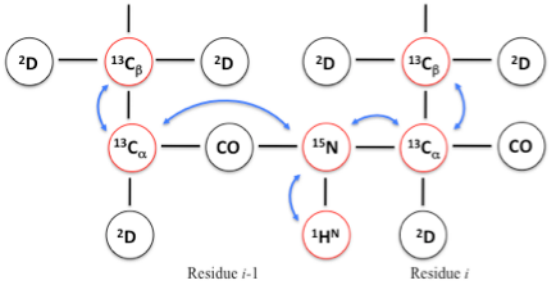
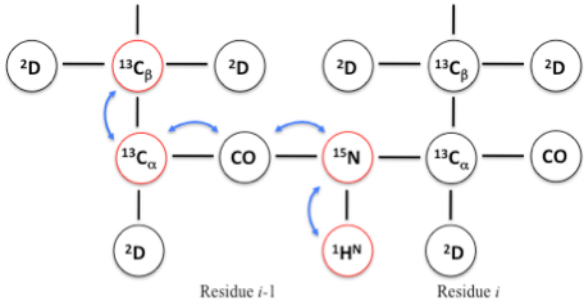


Figure 1.8. Scalar coupling values in polypeptides.

The following experiments were used in this thesis work for sequence specific backbone assignments: HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, and HNCO. The first two, then third and fourth experiments are run in pairs to obtain correlations for intra-residue spins (*i*) and inter-residue spins (*i*-1). Analyzing a 3D experiment can be quite complicated. To ease the analysis of the spectrum, 2D slices are chosen during the process. Usually a slice from the ^{15}N dimension is chosen and ^{13}C signals are observed as a function of ^{1}H chemical shift. A summary of the experiments and their correlations are presented in Table 1.1. One additional experiment to be discussed is NOESY.

Table 1.1. Summary of triple resonance experiments used for sequential backbone assignment

Experiment	Magnetization transfer	^{13}C signals observed
HNCA		$C_{\alpha(i)}$, $C_{\alpha(i-1)}$
HN(CO)CA		$C_{\alpha(i-1)}$
HNCACB		$C_{\alpha(i)}$,$C_{\alpha(i-1)}$,$C_{\beta(i)}$,$C_{\beta(i-1)}$
HN(CO)CACB		$C_{\alpha(i-1)}$,$C_{\beta(i-1)}$

HNCA and HN(CO)CA

In the HNCA experiment the $^{1}H^{N}$ and ^{15}N chemical shifts are correlated with the $C_{\alpha(i)}$ and $C_{\alpha(i-1)}$. This experiment is an out-and-back triple resonance experiment, where the magnetization is transferred for the first two steps without detection as follows:

$$^{1}H^{N} \rightarrow {}^{15}N \rightarrow {}^{13}\underline{C}_{\alpha} \rightarrow {}^{15}\underline{N} \rightarrow {}^{1}\underline{H}^{N}$$

The chemical shifts for ^{15}N and $^{1}H^{N}$ can be displayed in the same fashion as the 2D-HSQC (Fig. 1.9A), separated by the C_{α} chemical shift in the third dimension. For each residue, except prolines, two signals will be observed for $C_{\alpha(i)}$ and $C_{\alpha(i-1)}$ (Fig. 1.9B). Identification of the intra and inter C_{α} can be based on the intensity of the resonances, however ambiguities arise from similarities in the intensities. The addition of the HN(CO)CA completely resolves these ambiguities.

In the HN(CO)CA experiment, the transfer in magnetization is as follows:

$$^{1}H^{N} \rightarrow {}^{15}N \rightarrow {}^{13}CO \rightarrow {}^{13}\underline{C}_{\alpha(i-1)} \rightarrow {}^{15}\underline{N} \rightarrow {}^{1}\underline{H}^{N}$$

The experiment will show only one C_{α} signal, that for the previous residue, which when compared to the HNCA data allows the intra and inter residue C_{α} to be distinguished (Fig. 1.9C). In the case of prolines, C_{α} signals will be observed when they are in the *i*-1 position.

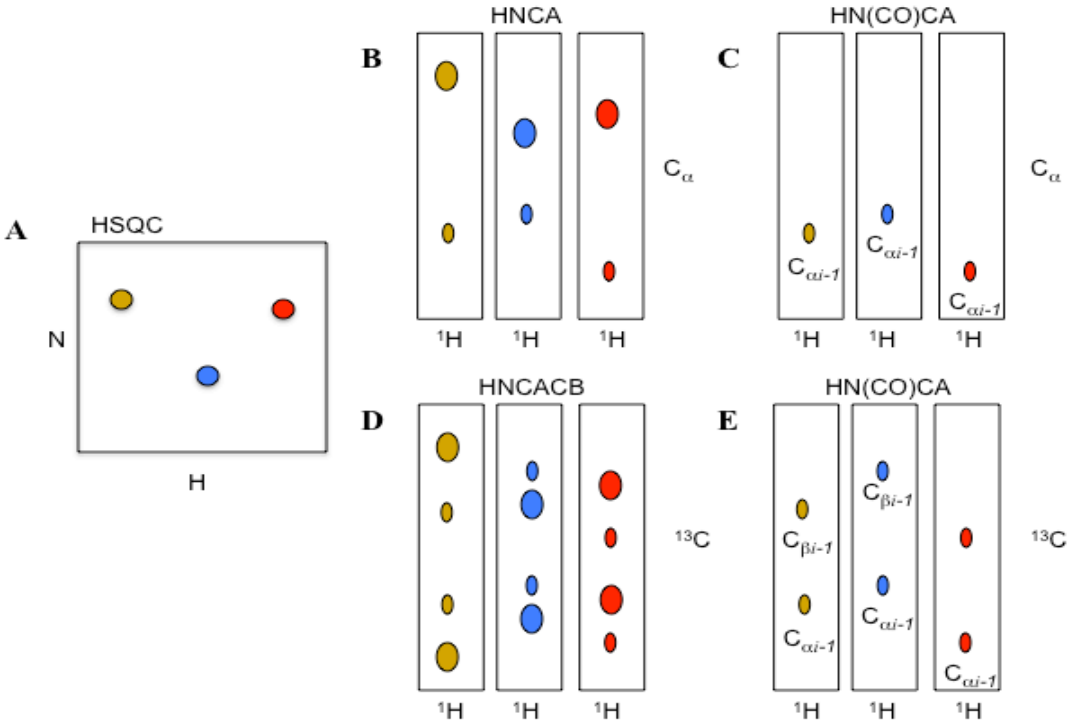


Figure 1.9. NMR triple resonance experiments. Complementarity of each experiment is shown to demonstrate how identification of resonance for amino acids *i* and *i*-1 is performed**.**

Although the HNCA and the HN(CO)CA can be use in conjunction to perform sequential assignment identification of the amino acids, using only the C_α chemical shift is not possible because overlap in the C_α chemical shift often occurs. In order to be able to identify all amino acids in most proteins it is necessary to add information from additional nuclei.

HNCACB and HN(CO)CACB

The HNCACB is very similar to the HNCA but provides additional resonances for the C_β side chain. Transfer of magnetization for the HNCACB follows:

$$^1H^N \rightarrow {}^{15}N \rightarrow C_\alpha \rightarrow \underline{C}_\beta \rightarrow \underline{C}_\alpha \rightarrow {}^{15}\underline{N} \rightarrow {}^1\underline{H}^N$$

Four signals are observed in the spectrum per residue. The fact that the ^{15}N from the amide bond correlates stronger to the intra C_α results in a stronger resonance signal for the C_α and C_β and weaker for the inter-residue resonances (Fig. 1.9). This experiment is quite informative for amino acid identification because the combination of C_α and C_β is reasonably specific. Glycines lack a C_β carbon and the chemical shift for the C_α is unique, which make them good candidates for starting the strip plot search to connect residue *i* with residue *i*-1 and be able to make a backbone walk.

The HN(CO)CACB is the complementary experiment. The magnetization path is as follows:

$$^1H^N \rightarrow {}^{15}N \rightarrow CO \rightarrow C_{\alpha(i\text{-}1)} \rightarrow \underline{C}_{\beta(i\text{-}1)} \rightarrow \underline{C}_{\alpha(i\text{-}1)} \rightarrow {}^{15}\underline{N} \rightarrow {}^1\underline{H}^N$$

Analogous to the HN(CO)CA, only information for the *i*-1 residue is obtained. This along with the HNCACB allows the identification of pairs of sequential residues. If this is a unique pair in the amino acid sequence, direct assignment can be made.

NOESY

The experiments discussed so far make use of through-bond (scalar) couplings between spins. In the case of the NOESY, through-space (dipolar) coupling are used.

The NOESY experiment is very dependent on the distance between the interacting spins. A crosspeak is observed between spins at a distance of ~5 Å or less (Table 1.1). The NOESY experiment is very important for obtaining distance restraints between ^{1}H when determining protein structures. In the cases, of perdeuterated proteins only $^{1}H^{N}$-$^{1}H^{N}$ information can be obtained. In such case, the information can be used to confirm sequential assignment. Additional cross peaks may be observed by amino acids brought close in space as a result of the protein fold.

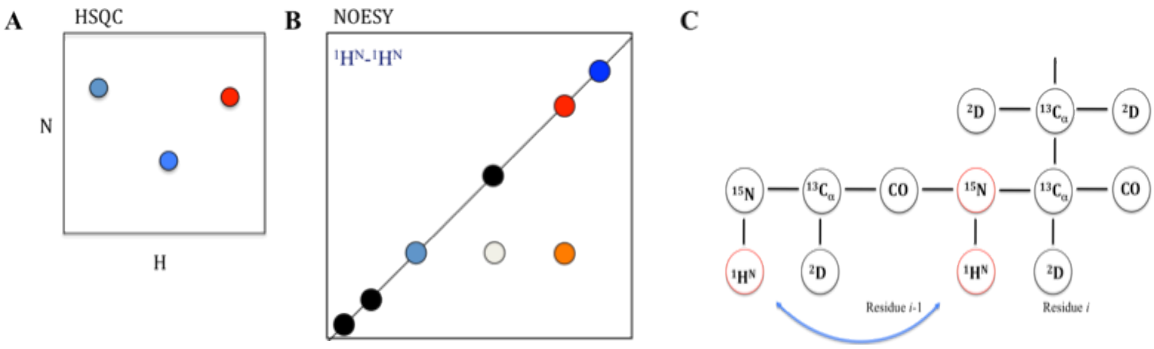


Figure 1.10. NOESY experiment. Signals in he ^{15}N-^{1}H HSQC (A) are observed in the NOESY spectra (B) as peaks in the diagonal line. Only the region of the spectrum for the $^{1}H^{N}$-$^{1}H^{N}$ signals is depicted. Amino acids less than 5Å apart in space will give rise to crosspeaks

The final step in the assignment process is placing the pieces of the puzzle together. This starts with identifying the spin system and pairs of sequentially adjacent spins systems. This involves searching for a strip plot in different planes to identify residues i and *i*-1. The reverse search can also be done. In the example shown in Figure 1.11, the HNCACB was used for a strip plot search. Notice that both $C_{\alpha(i)}$ and $C_{\beta(i)}$ of residue F align with the $C_{\alpha(i-1)}$ and $C_{\beta(i-1)}$ of the strip plot for residue D. This way we keep

connecting the puzzle pieces until we are able to "walk" a stretch of the amino acid sequence. In practice, "walking" of the entire amino acid sequence is not possible. In such cases, a new plane is picked and the search is started in the same manner. In practice the process of obtaining assignments is iterative, with multiple starting points, and a slow build up of unique assignments to the sequence from stretches of consecutive spin systems that are related to each other.

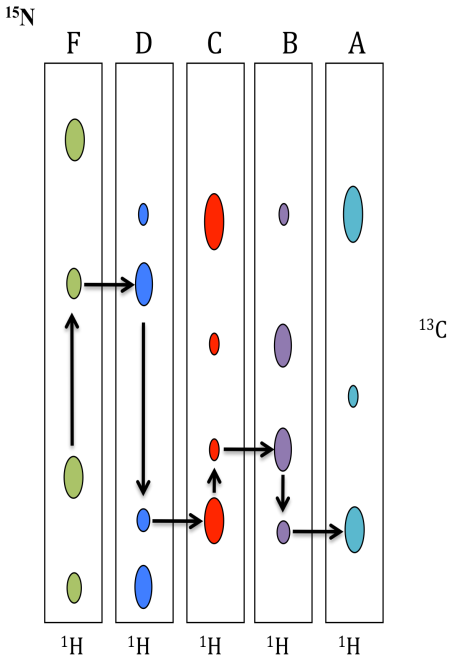


Figure 1.11. Backbone walk using the HNCACB experiment. One of the ^{13}C spins from the HNCACB experiment was used to demonstrate how the amino acid strip plots in the sequence ABCDF are connected.

Chapter II

Sequence-Specific Resonance Assignment of RPA32D/14

Introduction

DNA replication is a fundamental process for passing the genome information from one generation to the next. DNA replication needs to occur in a controlled manner where proteins are loaded dynamically and at the right time to ensure the process goes to completion [2]. The process of replication does not go smoothly all the time. Failure of the process causes genome instability and the onset and development of related diseases such as cancer [8].

Replication protein A (RPA) is the eukaryotic primary ssDNA binding protein [18]. RPA is present during DNA metabolic processes such as replication, repair, recombination and telomere maintenance. RPA is a modular heterotrimeric protein composed of subunits RPA70, RPA32 and RPA14 named after their molecular weight (Fig. 1.2). RPA binds to ssDNA through four of its six OB-fold domains: RPA70A, RPA70B, RPA70C, and RPA32D [18]. The three subunits of RPA come together to form what is known as the trimer core, involving the OB-fold domains RPA70C, RPA32D and RPA14. A soluble and stable heterodimer has also been found to form *in vivo* between RPA32 and RPA14 (RPA32/14) [33]. Recently, RPA32/14 has been shown to play a role in telomere maintenance [15].

Telomeres are formed at the end of linear chromosomes to protect the DNA from joining and degradation. Telomeres are of special importance due to their role in cancer and aging [10]. Telomere ends are composed of alternating AT and G nucleotides with

an overhang that is rich in guanine residues. Guanine rich stretches are capable of forming secondary structures known as G-quadruplexes. Four G-residues associate through Hoogsteen basepairing and these planar structures stack on top of each other. The hairpin loops linking them lead to formation of parallel and antiparallel G-quadruplexes, which are stabilized by binding of one or more cations [11]. These secondary structures have been shown to form *in vitro* [11] and *in vivo* in some cases [40]. G-quadruplexes represent a topological problem for DNA replication and recombination. They need to be resolved in order for the telomere maintenance machinery to gain access to the DNA . These structures have come to light because there are other sequences in human cells that are rich in guanine residues, thus having a high propensity to form quadruplexes. For example, promoters of the oncogenes c-myc, HIF-1α, bcl-2 and c-kit [41] have been shown to be rich in guanine residues. These findings raise the question whether formation of secondary structures play a regulatory role in their transcription.

A number of proteins have been found to interact with G-quadruplexes. Their function ranges from capping and protecting G-quadraplexes to resolving them into ssDNA [12]. All these telomere interacting proteins have in common OB-fold domains in their structure. Yeast proteins stn1/Ten1 interact specifically with telomeric DNA through OB-fold domains. High structural homology has been found between RPA32D and stn1, and RPA14 and Ten1, with C_α RMSDs of 2.4 and 2.8 Å, respectively [16]. Pot1 (protection of telomeres) is another OB-fold containing protein identified in human and yeast [42] as a telomere specific interacting protein. Studies have demonstrated that RPA also plays a direct role in maintaining telomere length. Recently, it was shown by

Pakrash *et al* that RPA is able to specifically bind and unfold G-quadraplex structures through its RPA32D/14 domains [43]. All this evidence suggests a role for RPA in the regulation of telomeres during the synthesis phase of the DNA replication cycle, by opening G-quadraplex structures to form ssDNA so that proteins involved in telomere maintenance can gain access to the DNA template [44].

NMR provides a powerful tool for the dynamic study of molecules. The NMR signals are specific to each nucleus in the molecule and exquisitely sensitive to the chemical environment. RPA constructs have been studied by NMR to understand their interaction with proteins [29, 31] and DNA [57]. These studies have provided insight into interactions at the atomic level. The purpose of my research was to obtain NMR specific assignments for RPA32D/14. Three crystal structures for the RPA heterodimer have been determined using the following constructs: RPA32/14, RPA32D/14, and RPA70C/32D/14. These static pictures of RPA32D/14 have led to speculation about the structural basis for binding ssDNA. This chapter will describe the NMR triple resonance experiments HNCA, HN(CO)CA, HNCACB and HN(CO)CACB recorded for sequential backbone assignment of RPA32D/14. With the information obtained we were able to assign 60% of the resonances. Additionally, a discussion is included about additional experiments that can be performed to further the assignments and how the NMR assignments can be used to perform more thorough NMR studies on RPA32D/14.

Methods

Expression and Purification of RPA32D/14

The RPA32D14 plasmid was transformed into the BL21 (DE3) expression cell line. 10 mL LB (rich media) was then inoculated and grown overnight at 37 ^{o}C. The following day, the culture was added to 1L M9 (minimal media) and grown at 37 ^{o}C. The D_2O was used in place of H_2O to prepare the minimal media and obtain a perdeuterated sample. Media was supplemented with ^{15}N-NH_4Cl and ^{13}C-glucose for ^{15}N and ^{13}C incorporation as required. Protein expression was induced with 1 mM IPTG when the culture reached an OD_{600} of ~0.6. Cells were then harvested ~18 hrs post-induction. Pelleted cells were resuspended in lysis buffer containing 10 mM Hepes at pH 7.5, 500 mM NaCl, 10 mM ZnCl, 10 μM BME, 10 % glycerol, 10 mM imidazole, 100 mM L-arginine. Cells were lysed by sonication at 4 °C. Insoluble material was removed by centrifugation. RPA32D/14 was purified using Ni-NTA affinity chromatography. Bound protein was eluted using a buffer containing 500 mM imidazole. Fractions containing RPA32D/14 were pooled and dialyzed overnight at 4 °C into buffer containing 20 mM Hepes at pH7.5, 100 mM NaCl, 10 mM MgCl, 5 mM BME and 200 mM arginine. The sample was then concentrated to 5 mL and loaded onto an S75 gel filtration column. Fractions containing RPA32D/14 were pooled and dialyzed into NMR buffer containing 10 mM HEPES at pH 7.5, 100 mM NaCl, 10 mM MgCl, 5 mM BME and 200 mM arginine. The sample was then concentrated to 506 μM. Marie-Eve Chagot produced and prepared the sample used for the NMR studies.

HSQC

The spectra were acquired at 295 K using a Bruker DRX 800 MHz spectrometer equipped with a cryoprobe. $^{2}H,^{15}N,^{13}C$-RPA32D/14 was prepared at a concentration of 506 μM in buffer containing 10 mM Hepes at pH 7.5, 100 mM NaCl, 10 mM MgCl, 5 mM BME and 200 mM arginine. Two-dimensional TROSY ^{15}N-^{1}H-heteronuclear single quantum coherence (HSQC) spectra were acquired with 32 scans over 2k points in the direct dimension and 128 points in the indirect dimension. The proton sweep width was set at 14 ppm. The nitrogen sweep width was set at 32 ppm. The data were processed using NMRpipe [58], a sine-squared bell window function followed by zero filling prior to the Fourier transformation in both dimensions. Sparky (University of California) was used to analyze the data. The NMR experiments were set up by Chris Brosey and Markus Voehler.

Triple Resonance Experiments

The following experiments were collected using an 800 MHz spectrometer equipped with cryoprobe to complete backbone resonance assignments: HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, and NOESY. All of these experiments were recorded using TROSY enhancement to get sharper peaks for the 28 kDa protein. The experimental parameters are listed in Table 2.1. 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 (^{1}H) dimension was processed first, followed by F2 (^{15}N) and then F1 (^{13}C). NMR pipe was used for processing and Sparky to analyze the data. Chris Brosey and Markus Voehler set up the NMR

experiments.

Table 2.1. Experimental parameters for backbone resonance assignment spectra

Experiment	Data Points			Sweep Width			Time
	F3	F2	F1	F3	F2	F1	
HNCACB	2k	48	128	14	32	64	3d 23h
HN(CO)CACB	2k	48	128	14	32	68	3d 23h
HNCA	2k	48	96	14	29	3	1d 9h
HN(CO)CA	2k	48	96	14	32	68	1d 9h
NOESY	2k	64	236	14	32	12.2	4d 15h

Results

HSQC

The HSQC is one of the basic NMR experiments. A ^{15}N-^{1}H HSQC spectrum was recorded for ^{2}H,^{15}N,^{13}C-RPA32D/14. The spectrum (Fig. 2.1) contained 267 dispersed peaks including side chains. The protein contains 264 amino acids excluding prolines. Dispersed narrow linewidths are observed, which is indicative of a well-folded protein. Peak overlap was observed in the spectrum mainly due to the size of the protein; RPA32D/14 is 28 kDa. ^{15}N-^{1}H HSQC spectra were recorded through the course of the triple resonance experiments to ensure signals were not shifting or disappearing as a consequence of degradation. We also noticed that some peaks appear as doublets in the HSQC spectrum, and these peaks have weak intensity in the triple resonance experiments. This is due to the presence of prolines close in sequence to these residues. Prolines can isomerize into trans and cis conformations causing the splitting of the signals of adjacent residues. The presence of an amino acid closer in sequence to the

proline caused their signal to disappear.

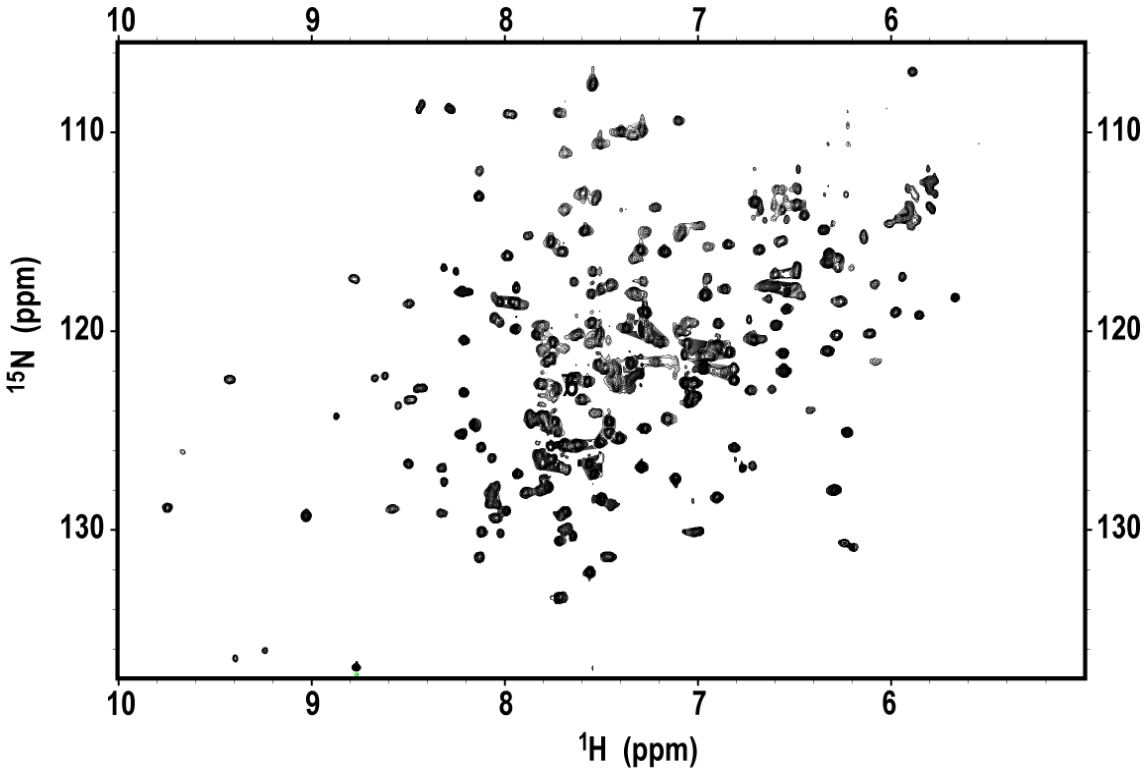


Figure 2.1. ^{15}N-^{1}H TROSY-HSQC spectrum for ^{2}H, ^{15}N, ^{13}C-RPA32D/14

Triple resonance experiments for the sequential assignment of RPA32D/14

Pick peaking for the various experiments was performed using the HSQC as a reference. The HNCA showed 576 peaks. Since there can be a C_α for the intra and inter residue, twice the number of peaks from the HSQC were expected. The number obtained exceeded that number which could be from the selection of "peaks" that were noise. In the case of the higher sensitivity HN(CO)CA spectrum, which contains signals only for the inter C_α, 317 peaks were found which is close to the total expected number.

Peak picking using the HSQC as a reference resulted in 1053 peaks in both the HNCACB and the HN(CO)CACB spectra. Backbone resonances were manually assigned. 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 (e.g. Thr and Ser C_β resonances are far downfield, for example) allowed the sequence-specific assignments. Figure 2.2 shows an example of a stretch of four residues in sequence and how they were connected using these experiments.

Finally, a NOESY spectrum was recorded on the triple labeled sample. The experiment provided NOEs for protons 5 Å or less in distance from each other. Short mixing times were used for this experiment. In the case of non-deuterated proteins this is done to avoid spin diffusion, but the opposite is followed for deuterated samples because only the amide proton can provide a signal. The use of short mixing times caused the signals in this spectrum to be weak. Figure 2.3 shows an example of connectivities between A14, G15, M16 and L17 of RPA14, showing cross peaks at their respective nitrogen planes. With this experiment we were able to confirm the sequential assignment of 32 amino acids for RPA14 and 25 for RPA32D (Summarized in Figure 2.4).

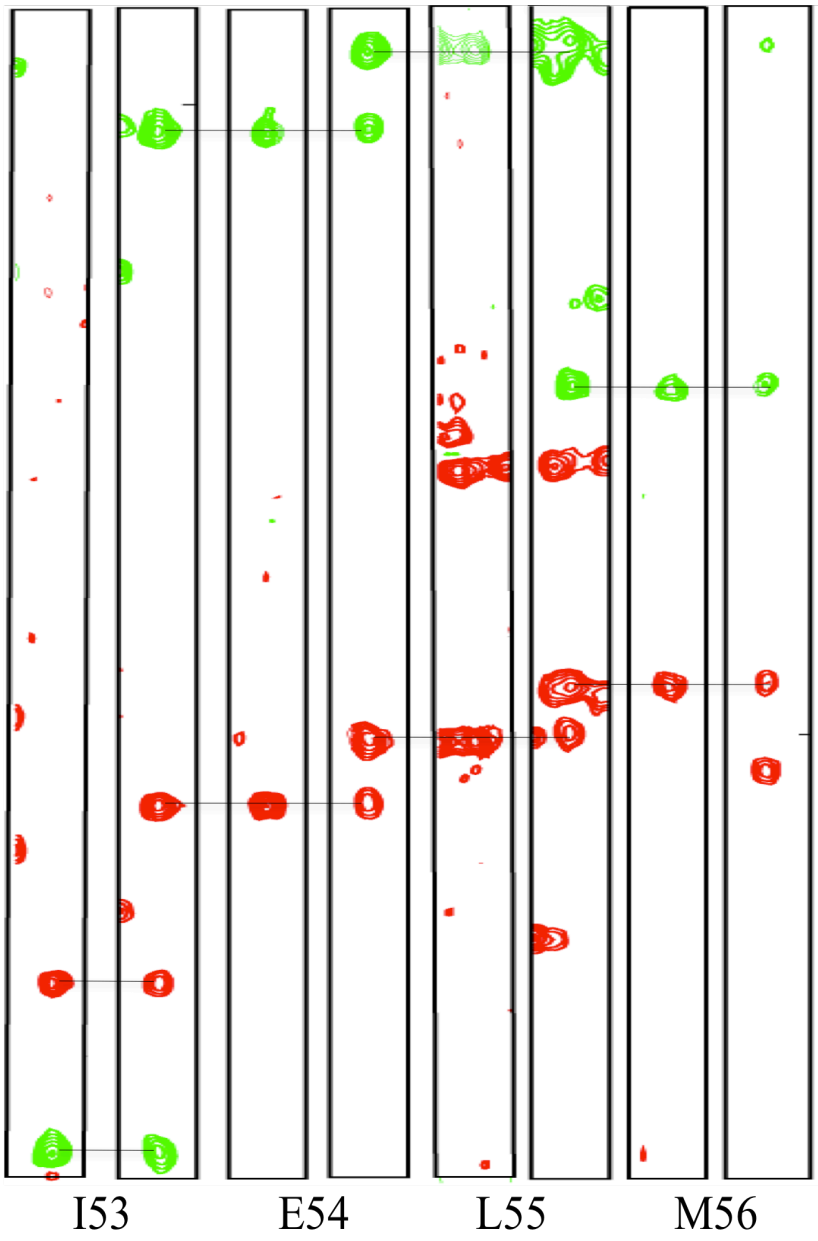


Figure 2.2. Strip plots for the sequential assignment of RPA32D/14. HN(CO)CACB (left) and HNCACB (right) strip plots for the sequential assignment of amino acids I53-M56 of RPA14.

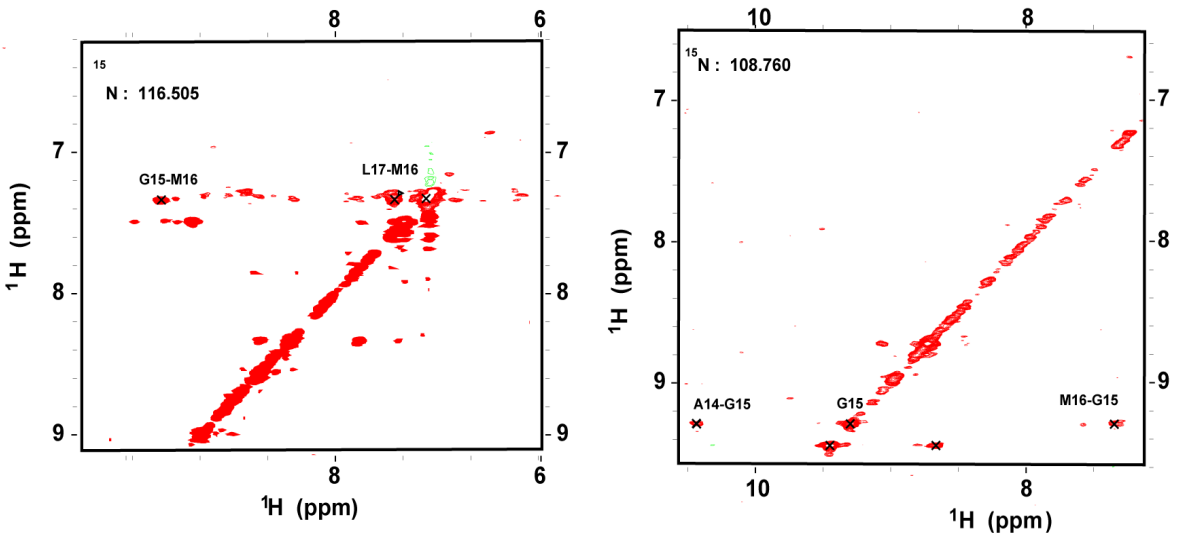


Figure 2.3. NOESY planes for residues G15 and M16 of RPA14. Crosspeaks for the G15-M16 NOE are found at the NH plane for G15 (left) and a crosspeak is also found at the M16 NH plane.

Summary of assignments

A total of 136 residues were assigned using the three-dimensional experiments. A summary of assigned residues is shown in Table 2.1 and Table 2.2 with the values for C_α, C_β, N and H^N chemical shifts for RPA14 and RPA32D, respectively. Assigned residues were mapped into the structure of RPA32D/14 (Fig. 2.4). All the assignments refer back to the ^{15}N-^{1}H HSQC (Fig. 2.5).

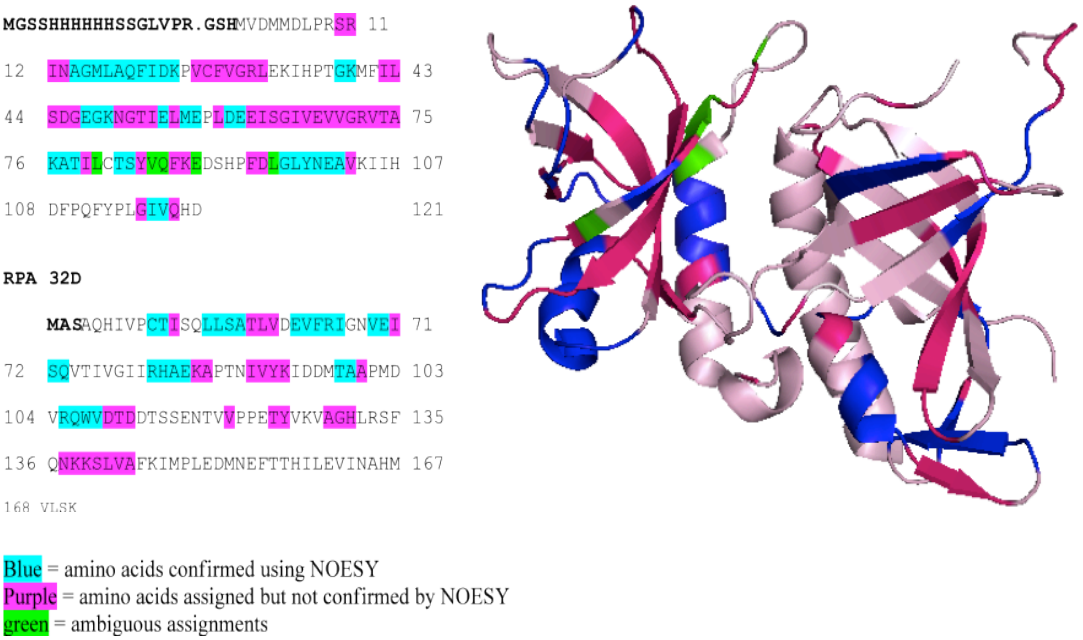


Figure 2.4. Assigned residues mapped onto the structure of RPA32D/14. Residues confirmed by NOESY are shown in blue, non-confirmed in purple and ambiguous in green.

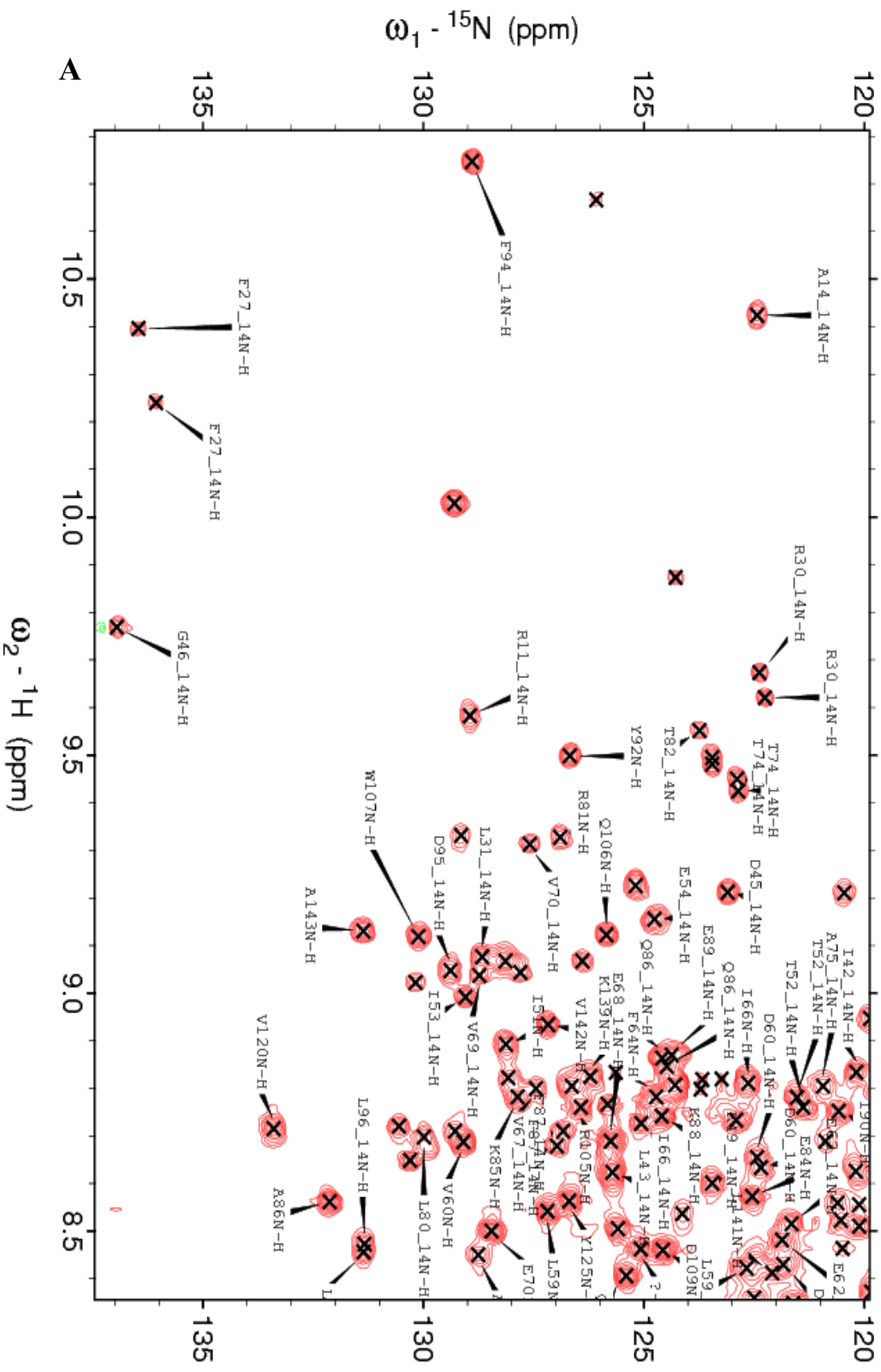
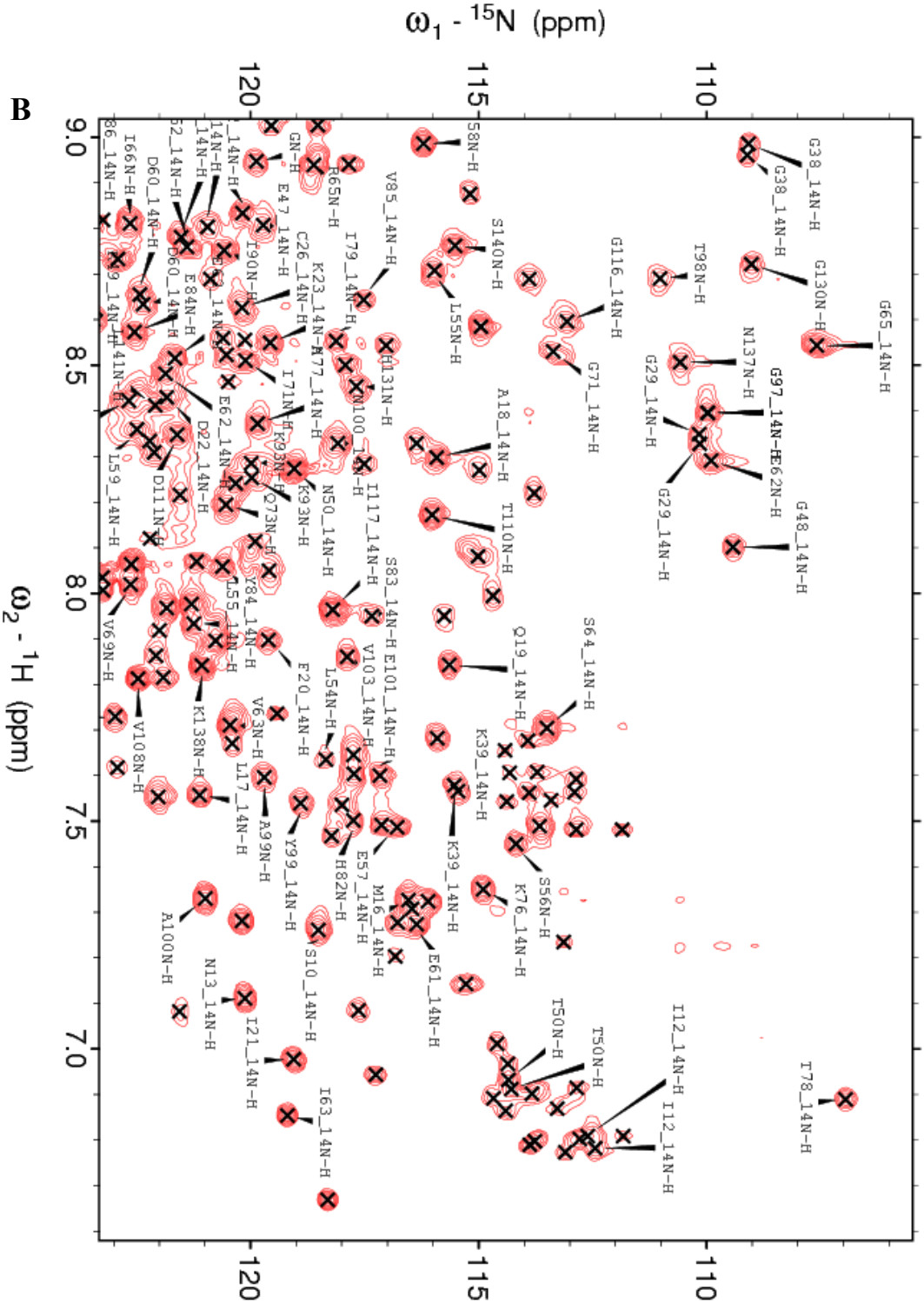
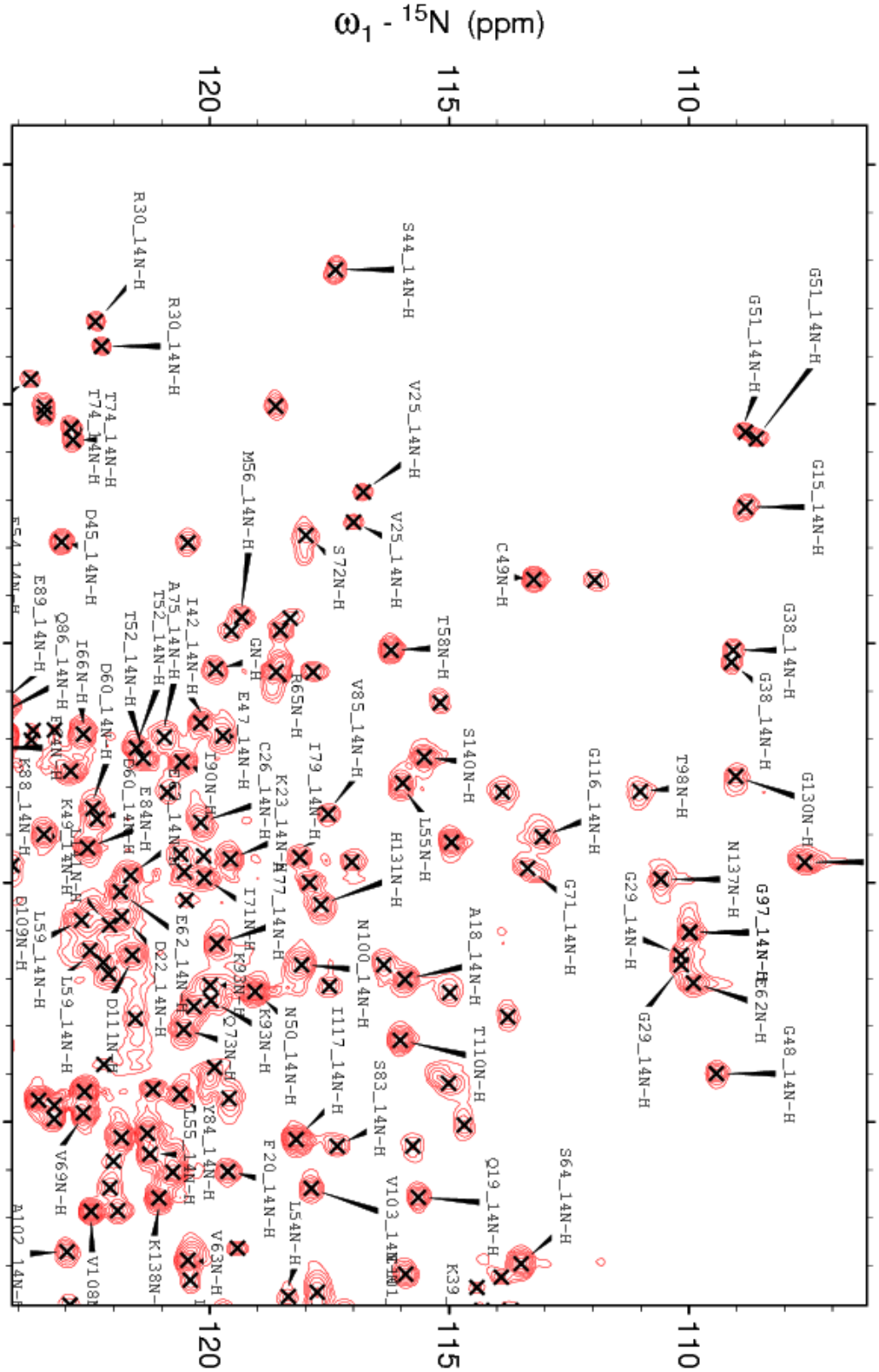


Figure 2.5. Assigned RPA32D/14 ^{15}N-^{1}H HSQC. Figures A-D show sections of the RPA32D/14 ^{1}H-^{15}N-HSQC with the assignments for simplification.

B


ω1 - 15N (ppm)
ω2 - 1H (ppm)
110
115
120
9.0
8.5
8.0
7.5
7.0
G65_14N-H
G38_14N-H
G130N-H
G38_14N-H
N137N-H
G97_14N-HE62N-H
T98N-H
G29_14N-H
G48_14N-H
G29_14N-H
T78_14N-H
G116_14N-H
G71_14N-H
I12_14N-H
I12_14N-H
S140N-H
L55N-H
A18_14N-H
T110N-H
T50N-H
T50N-H
S64_14N-H
Q19_14N-H
K39_14N-H
S56N-H
K76_14N-H
K39_14N-H
E61_14N-H
M16_14N-H
E57_14N-H
H82N-H
S10_14N-H
I63_14N-H
I21_14N-H
N13_14N-H
A100N-H
R99N-H
Y99_14N-H
L17_14N-H
K138N-H
V63N-H
F20_14N-H
L54N-H
S83_14N-H
E101_14N-H
V103_14N-H
I117_14N-H
N50_14N-H
K93N-H
Q73N-H
Y84_14N-H
L55_14N-H
V69N-H
V108N-H
D111N-H
D22_14N-H
E62_14N-H
E84N-H
D60_14N-H
K23_14N-H
C26_14N-H
I77_14N-H
I71N-H
K93N-H
R65N-H
I79_14N-H
Q100_14N-H
Q131N-H
V85_14N-H
58N-H
E47_14N-H
T90N-H
I66N-H
L59_14N-H

C


ω1 - 15N (ppm)
ω2 - 1H (ppm)
110
115
120
10.0
9.5
9.0
8.5
8.0
G51_14N-H
G15_14N-H
G38_14N-H
G130N-H
G97_14N-HE2N-H
G48_14N-H
N137N-H
G29_14N-H
T98N-H
G116_14N-H
G71_14N-H
S64_14N-H
C49N-H
A18_14N-H
T110N-H
Q19_14N-H
S44_14N-H
V25_14N-H
T58N-H
S140N-H
L55N-H
H131N-H
V85_14N-H
I79_14N-H
S72N-H
R65N-H
M56_14N-H
R30_14N-H
T74_14N-H
D45_14N-H
E89_14N-H
Q86_14N-H
I66N-H
T52_14N-H
A75_14N-H
I42_14N-H
GN-H
E47_14N-H
I90N-H
C26_14N-H
K23_14N-H
I77_14N-H
I71N-H
K93N-H
N100_14N-H
I117_14N-H
N50_14N-H
S83_14N-H
V103_14N-H
L54N-H
F20_14N-H
L55_14N-H
Y84_14N-H
V63N-H
K138N-
V108
A102_14N-
K39_
V69N-H
L59_14N-H
D109N-H
D111N-H
D22_14N-H
E62_14N-H
E84N-H
D60_14N-H
K88_14N-H
K49_14N-H

D

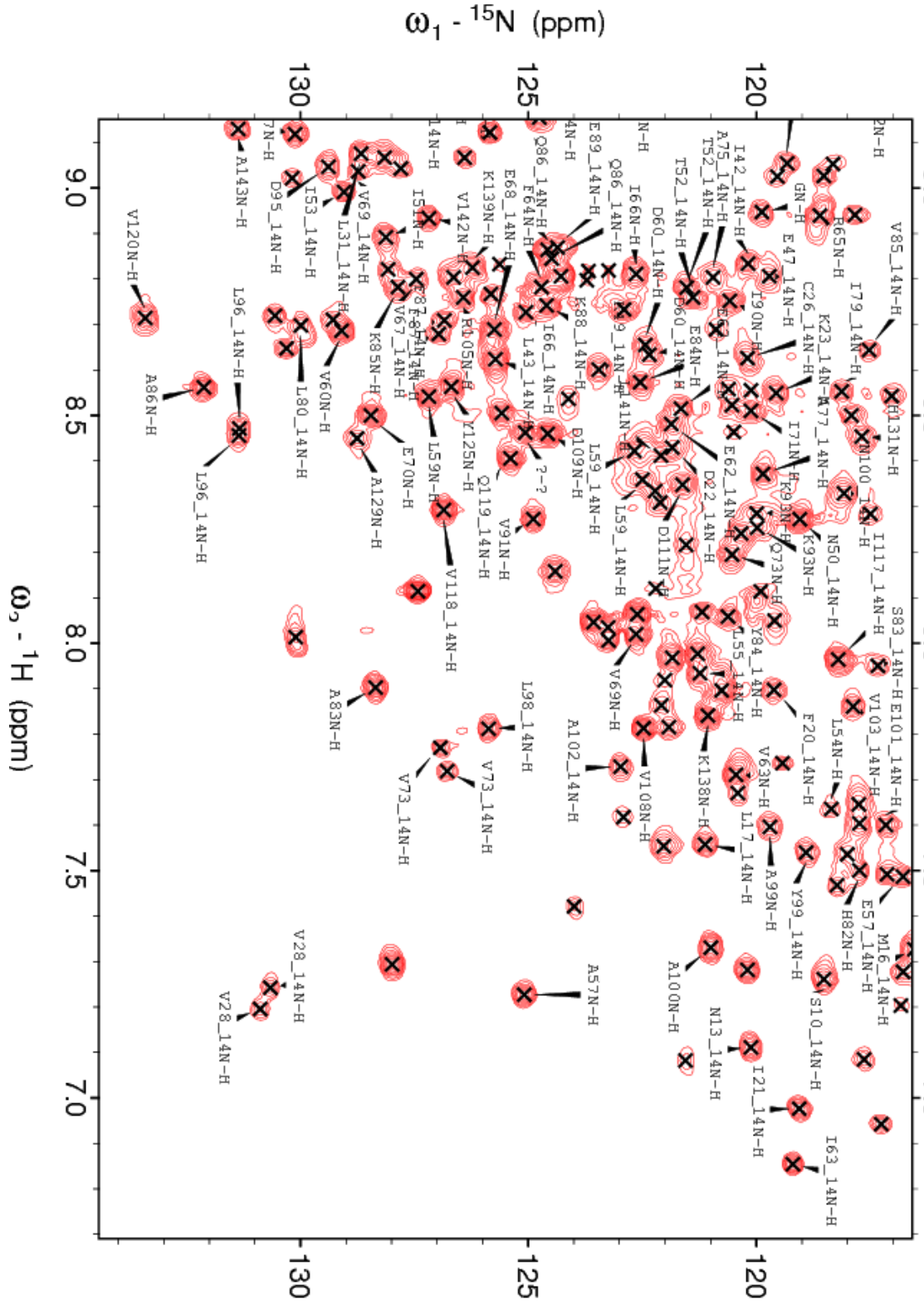

ω1 - 15N (ppm)
ω2 - 1H (ppm)
9.0
8.5
8.0
7.5
7.0
120
125
130
V85_14N-H
H131N-H
S83_14N-H
E101_14N-H
M16_14N-H
E57_14N-H
H82N-H
S10_14N-H
I63_14N-H
I21_14N-H
N13_14N-H
R100N-H
Y99_14N-H
R99N-H
F20_14N-H
L54N-H
V103_14N-H
I117_14N-H
N50_14N-H
K93N-H
Q73N-H
Y84_14N-H
L55_14N-H
V63N-H
L17_14N-H
K138N-H
V108N-H
V69N-H
L59_14N-H
D111N-H
D22_14N-H
E62_14N-H
D109N-H
K88_14N-H
L43_14N-H
I66_14N-H
F64N-H
Q86_14N-H
E68_14N-H
K139N-H
V142N-H
R105N-H
Y125N-H
Q119_14N-H
V91N-H
L59N-H
V118_14N-H
A102_14N-H
L98_14N-H
V73_14N-H
A57N-H
A83N-H
E70N-H
K85N-H
A129N-H
V67_14N-H
L31_14N-H
V69_14N-H
I53_14N-H
V60N-H
L80_14N-H
D95_14N-H
L96_14N-H
A143N-H
A86N-H
V120N-H
V28_14N-H
V28_14N-H
I79_14N-H
E47_14N-H
C26_14N-H
K23_14N-H
A77_14N-H
I90N-H
I71N-H
E84N-H
D60_14N-H
T52_14N-H
T52_14N-H
A75_14N-H
I42_14N-H
I66N-H
Q86_14N-H
E89_14N-H
L59_14N-H
I51N-H

Table 2.2. RPA14 backbone chemical shift

Residue	CA	CB	N	H
S10	59.834	63.746	118.478	7.256
R11	56.49	428.384	128.850	9.583
I12	58.829	41.785	112.637	6.799
N13	51.032	38.203	120.111	7.111
A14	55.216	17.166	122.371	10.420
G15	45.820		108.887	9.289
M16	55.815	33.363	116.592	7.334
L17	60.168	41.190	121.063	7.558
A18	54.052	17.069	115.899	8.300
Q19	56.117	28.319	115.622	7.847
F20	57.986	39.245	119.614	7.898
I21	64.129	36.209	119.028	6.976
D22	56.309	40.119	121.803	8.438
K23	53.375	31.874	119.552	8.553
V25	58.811	37.004	117.010	9.257
C26	56.450	27.927	120.145	8.630
F27	56.403	42.861	135.959	10.245
V28	60.234	32.288	130.517	7.245
G29	45.511		110.180	8.328
R30	54.307	30.549	122.390	9.677
L31	55.802	39.668	128.492	9.075
G38	45.262		109.187	8.964
K39	56.212	31.449	115.392	7.564
I42	59.224	39.814	120.190	8.837
L43	51.707	45.478	125.650	8.629
S44	55.374	64.783	117.280	9.778
D45	52.007	41.620	123.083	9.216
G46	46.455		136.878	9.774
E47	54.564	29.052	119.701	8.806
G48	45.552		109.469	8.102
K49	54.706	32.665	123.399	8.603
N50	52.420	39.870	119.049	8.275
G51	43.376		108.813	9.442
T52	62.963	68.664	121.328	8.762
I53	57.038	34.466	128.982	8.995
E54	54.920	31.901	124.762	9.156
L55	53.511	43.409	121.136	7.932
M56	56.390	32.075	119.275	9.056
E57	52.546	30.326	116.921	7.491
L59	55.577	42.673	122.453	8.360
D60	53.516	40.933	122.482	8.657
E61	53.941	31.346	116.396	7.275
E62	56.019	28.724	121.804	8.483
I63	59.633	39.962	119.174	6.855
S64	57.334	64.862	113.525	7.703
G65	45.572		107.626	8.545
I66	57.458	34.906	124.945	8.730

Table 2.2. Continued.

Residue	CA	CB	N	H
E68	54.174	32.160	125.555	8.695
V69	60.603	34.327	128.631	9.041
V70	59.672	32.560	127.487	9.317
G71	45.985		113.359	8.531
R72	53.830	32.480	120.774	7.893
V73	62.741	29.774	126.708	7.718
T74	60.853	71.622	122.860	9.455
A75	54.033	17.834	120.839	8.813
K76	55.153	31.237	114.894	7.353
A77	53.028	80.913	119.827	8.374
T78	58.937	70.424	107.010	6.893
I79	60.221	38.992	118.179	8.554
L80	53.460	40.073	129.863	8.704
T82	63.496	69.284	123.659	9.544
S83	57.300	64.776	117.292	7.948
Y84	55.946	41.221	120.564	8.059
V85	60.255	35.379	117.471	8.647
Q86	54.805	30.299	124.452	8.846
F87	53.795	36.352	126.844	8.714
K88	56.867	32.070	124.535	8.745
E89	55.552	30.250	124.375	8.873
F94	58.572	39.225	128.798	10.752
D95	52.420	39.253	129.364	9.048
L96	56.043	42.161	131.331	8.479
G97	47.014		110.029	8.398
L98	56.648	39.997	125.848	7.813
Y99	61.956	37.257	118.963	7.542
N100	57.448	38.438	118.033	8.332
E101	58.717	27.790	117.189	7.603
A102	54.808	82.076	122.958	7.732
V103	66.087	30.078	117.910	7.862
G116	44.559		113.104	8.596
I117	60.490	37.726	118.177	7.966
V118	61.560	32.219	126.796	8.295
Q119	54.837	28.944	125.316	8.406

Table 2.3. RPA32D backbone chemical shift.

Residue	CA	CB	N	H
C49	57.976	25.694	113.239	9.136
T50	62.025	69.403	114.290	6.912
I51	59.580	39.794	128.063	8.894
L54	57.370	39.046	118.377	7.640
L55	56.854	40.646	115.997	8.705
S56	58.908	63.699	114.183	7.451
A57	52.495	18.577	125.037	7.229
T58	60.243	70.801	116.198	8.989
L59	54.247	41.648	127.126	8.544
V60	61.485	32.932	129.008	8.689
E62	57.744	27.636	109.852	8.290
V63	60.404	34.074	120.424	7.712
F64	57.996	40.468	124.655	8.785
R65	54.224	33.483	118.550	8.942
I66	59.054	38.499	122.689	8.820
V69	62.355	32.261	122.553	8.020
E70	55.976	29.497	128.400	8.502
I71	58.654	40.275	120.140	8.515
S72	59.529	63.488	117.897	9.237
Q73	55.188	29.301	120.501	8.195
R81	54.180	30.692	126.845	9.334
H82	56.238	33.813	117.616	7.503
A83	51.094	22.697	128.267	7.902
E84	54.634	31.887	122.516	8.575
A86	49.936	18.455	132.031	8.564
I90	59.754	39.697	120.550	8.753
V91	60.371	32.579	124.863	8.275
Y92	56.081	41.837	126.640	9.501
K93	54.159		119.931	8.286
T98	64.947	69.587	111.113	8.687
A99	50.248	20.414	119.676	7.593
A100	51.203	16.274	120.955	7.327
R105	53.751	32.603	126.394	8.762
Q106	53.678	31.456	125.766	9.125
W107	57.947	28.990	130.032	9.122
V108	60.685	32.981	122.402	7.815
D109	53.369	40.881	124.686	8.463
T110	62.367	68.887	116.008	8.172
D111	54.342	40.591	121.633	8.351
V120	60.239	31.217	133.287	8.714
T124	56.514	64.395	114.932	8.586
Y125	56.955	40.302	126.609	8.566
A129	49.272	21.155	128.621	8.447
G130	46.337		109.100	8.724
H131	54.871	32.829	117.669	8.447
N137	54.417	37.370	110.596	8.515
K138	54.833	33.188	121.038	7.842
K139	55.869	33.449	126.224	8.829
S140	57.219	65.660	115.521	8.764
L141	52.928	44.222	122.012	8.421
V142	61.669	30.831	127.104	8.937
A143	51.804	19.932	131.232	9.135

Chapter III

Discussion and Future Directions

RPA is involved in virtually all DNA transactions in the cell. RPA binds to ssDNA in three different modes with very high overall affinity. The first mode involves RPA70AB, the second adds RPA70C and the third engages RPA32D. Recent studies have also revealed that RPA32/14 may specifically bind and resolve certain DNA secondary structures, in particular G-quadruplexes. This potential biological role for RPA32/14 makes it a relevant target for study. Our laboratory uses NMR to study the structural dynamics and binding properties of RPA32/14. Any thorough study of proteins by NMR requires specific resonance assignments, so this has been the focus of this thesis research.

RPA32/14 is comprised of the dimer core (RPA32D/14) and the RPA32 N- and C-terminal domains (RPA32N, RPA32C). RPA32N is a disordered functional domain of 40 residues that contains a large number of serine and threonine residues that are targets for phosphorylation by kinases. RPA32C is a winged-helix protein recruitment module connected to the dimer core by a long ~30 residue linker. NMR assignments are already available for both RPA32N and RPA32C, so the focus of this thesis research was the dimer core.

Due to its molecular mass of 28 kDa, the assignment of RPA32D/14 represented a challenge. Several approaches were taken to obtain well-resolved spectra with the goal of obtaining as many assignments as possible. Protein perdeuteration, as well as the use of TROSY, allowed us to assign about 60% of the amino acid sequence. Signal overlap in many of the triple resonance experiments made it difficult to obtain more assignments.

To further the analysis, a sample of RPA32D/14 was sent to SECNMR at the University of Georgia in Athens to record data on a 900 MHz spectrometer. These data were processed and partially analyzed. The advantages of the spectra obtained with the 900 MHz instrument is improvement of the sensitivity, increase in the signal-to-noise ratio, and higher spectral resolution. Unfortunately, these data are insufficient for obtaining assignments because the experiment was run to with optimized excitation of C_β signals but the C_α signals are needed for this large protein. This can be partially overcome when the HNCA experiment is added, but alignment in all three dimensions proved challenging. Although these data can be helpful in making a few more assignments as well as confirming some of the already assigned peaks, additional experiments will be required for the complete assignment of the NMR resonances of RPA32D/14.

Several other approaches are available for the facilitating assignment of large proteins [59], such as specific amino acid labeling. For example, methyl groups of amino acids such as valine, isoleucine and leucine can be specifically labeled and with appropriate experiments these resonances can be correlated back to the protein backbone to help make assignments. The power of selective labeling is the reduced complexity of the spectrum.

Once resonance assignments are complete, there are many ways they can be used to perform more in depth studies. One of these could be structure determination by NMR. The crystal structure of RPA32D/14 has been determined using three different constructs (see introduction). A solution structure could help in resolving some discrepancies observed among the structures of different crystal forms obtained for the same protein.

In general, the fold adopted by the domains in the crystal structure is the same. Differences arise in the contacts found between the alpha helices found at the interface. One of these structures supports the idea of helix bundle formation to mediate the subunits coming together. While in other crystal forms, the inter helical angles differ thus changing the contacts between helices. These observations raised concerns about the importance of crystal packing and questions about the biologically relevant conformation. Solution structure and dynamics studies by NMR can potentially resolve these concerns. A solution structure can be determined to see if it is in agreement with one of conformations, while relaxation studies can shed light on the dynamics of the alpha helices of RPA32D/14 .

DNA binding activity is another feature we can study using NMR. DNA titrations can be made to determine specific amino acids involved in DNA binding activity of RPA32D/14. With the assigned HSQC spectra we can refer back to specific amino acids in the RPA14 and RPA32D subunits. Chemical shift perturbation analysis of a titration of RPA32D/14 with ssDNA should give information on the specific amino acids involved in ssDNA binding and if RPA14 plays a role in DNA binding activity.

It has been demonstrated that RPA domains tumble independently, which creates the possibility to transfer the assignments from smaller constructs into larger constructs. The same approach could be used for transferring RPA32D/14 assignments from the heterodimer to the trimer core. This leaves only one domain from the trimer core to be assigned, RPA70C, which should make the overall process of assignments for the trimer core considerably easier. One objective would be to learn more about the effect of ssDNA binding upon the trimer core. For example, titrations with ssDNA should reveal

if the presence of the third domain does or does not affect the ssDNA properties of the heterodimer.

Resonance assignment transfer can be useful for additional studies. For example, RPA32D/14 assignments can be transferred into a spectrum containing full-length heterodimer (RPA32/14). One of RPA's important functions is protein interaction; proteins involved in different aspects of DNA repair and replication have been identified to interact with RPA. Protein interactions can remodel RPA architecture and cause its release from DNA. Performing protein titrations into RPA32/14 in the presence and absence of ssDNA will enhance our understanding of RPA function and shed light on how it uses the hand-off mechanism. The approach involves following chemical shift perturbations at the DNA binding interface as protein binding partner are added to the solution. Chemical shift perturbations can be used to determine if ssDNA is released upon protein titration or if a different architectural rearrangement is observed. The same type of experiment can be performed in the context of the trimer core to learn more about domain rearrangement from protein and DNA binding. These kind of studies will require the use of NMR pulse sequences that alleviate the effect of line broadening from slow tumbling biomolecular complexes such as CRIPT, CRINEPT or methyl TROSY depending on the size of the complex [60].

Recently, it was shown that RPA32D/14 is able to bind and resolve DNA secondary structure. Guanine stretches are able to form G-quadruplexes at the end of chromosomes and possibly at some oncogene promoter regions. NMR studies can be used to determine if indeed only this portion of RPA is able to bind or if we can get binding from other RPA domains known to bind DNA (i.e. RPA70AB). Analysis of the crystal structure of

hPOT1 bound to telomeric DNA shows a different rearrangement than what is observed from the structure of RPA70AB bound to ssDNA (see Fig. 1.5). In the first case the two OB-fold domains rearrange in such as way that they form a continuous channel. In the case of RPA70AB, the two OB-fold domains are oriented in the same direction. It will be interesting to verify if the same rearrangement holds in the case of RPA32D/14 bound to telomeric DNA.

Appendix

Characterization of the p48 subunit of DNA primase and its interaction with Replication Protein A

Introduction

Eukaryotic DNA replication is a highly regulated process that requires the proper assembly of multiple proteins into a molecular machine termed the replisome [2, 61]. The initial step for DNA replication involves the recognition of origin sites of replication to form the pre-replicative complex. Transitioning into replication requires unwinding of DNA by helicases [59, 62]. Exposed ssDNA is then coated with the eukaryotic single-stranded DNA binding protein Replication Protein A (RPA) [5, 63]. The first step for the polymerization reaction to initiate is loading of DNA polymerase α-primase (pol-prim) to synthesize a ~30 nucleotide RNA-DNA chimera [64]. Loading of PCNA, RFC and processive polymerase ε/δ is followed to extend the RNA-DNA primer and to synthesize the majority of the DNA during replication [65].

DNA polymerase α-primase (pol-prim) is the very first protein loaded during DNA replication initiation. DNA polymerases lack the ability to synthesize DNA *de novo,* directly from a ssDNA template, a task that can only be performed by DNA primases. RPA recruits pol-prim to the replication fork [66]. Pol-prim exists as a complex of four subunits, named after their estimated molecular weights: p180, p68, p58, and p48. Polymerase α is comprised of subunits p180 and p68, subunits p58 and p48 form the DNA primase portion of the heterotetramer [64]. DNA primase synthesizes small stretches of 8-10 nt RNA primers in both the leading and lagging strands. Then polymerase α extends the primer with DNA to a total length of ~30 nucleotides. The

catalytic core for DNA primases is found in the small p48 subunit of the human DNA primase. p48 activity is significantly enhanced by p58. Other functions of subunit p58 are: stabilize the p48 subunit, determining the length of the primer [64], and transfer of the RNA primed template to p180 for further elongation of the primer [67] by acting as a bridge between the two polymerases to couple their function. Pol-prim also plays a role in telomere maintenance and intra S-phase checkpoint activity [68, 69].

The priming cycle consists of basically three steps: initiation, elongation and termination. Eukaryotic primase preferentially uses ribonucleotides, two of these bind to DNA primase, which catalyzes dinucleotide formation to initiate the process. [70]. During the elongation phase the growing ribonucleotide binds to the initiation nucleotide-binding site of primase, thereby creating space for the next ribonucleotide to bind to the elongation binding site and subsequent bond formation. The nucleotide is incorporated at the 3'-hydroxy end of the elongating strand. The strand is extended until a length of 8-10 RNA nucleotides is reached. The mechanism of how DNA primase determines the length of the primer, also known as the counting mechanism, remains elusive. Once the primer has been extended to a defined length, primase dissociates from the growing chain and transfers RNA-primed template to polymerase α for further elongation [6, 71]. How the internal switching from primase to polymerase α occurs is still not well understood in eukaryotes, mainly due to the lack of a high resolution structures for pol-prim in higher organisms.

A considerable amount of structural and biochemical knowledge has accumulated for DNA primases from bacteria and archaea. However, most knowledge from such systems is not applicable to human or higher eukaryotes because there is low sequence and

structure conservation [72, 73]. In fact, the structural organization and function of bacterial primase is entirely different from eukaryotes. To date, the only structure available for higher eukaryotic primases is that of the human p58 C-terminal domain from our laboratory [74]. Comparison to the structure of PriL-CTD, the yeast counterpart for p58C, was published and showed important differences at the DNA binding region. These differences in DNA function highlight the importance of having a high-resolution structure of human DNA primase to enhance our understanding of the process of priming. Moreover, knowledge of the primosome is key to understanding the structural basis for DNA replication initiation.

Replication protein A (RPA) is the primary ssDNA binding protein in eukaryotes and is involved in virtually all DNA transactions in the cell [5]. Binding of RPA to ssDNA protects it from nucleases and prevents formation of ssDNA secondary structures [35]. In addition, RPA is known to physically interact with a large number of genome maintenance proteins, serving as a scaffold protein [29, 75, 76]. RPA is a modular protein composed of three polypeptides: RPA70, RPA32, and RPA14. Each of these subunits contains oligonucleotide binding-fold (OB-fold) domains: RPA70N, RPA70AB, RPA70C, RPA 32D, and RPA14 [36]. OB-folds A through D are mainly involved in DNA binding. RPA14 has a role in stabilizing the trimer core but no DNA binding activity has been associated with it. RPA32C consists of a winged helix domain [29]; this domain and RPA70N make protein-protein contacts necessary for DNA processing. Other protein-protein contacts were mapped to RPA70AB [29, 76, 77].

One critical binding partner of RPA is pol-prim. Interactions between RPA and intact pol-prim or just primase were mapped to RPA70 subunit [78]. However, detailed

knowledge of these interactions is limited. Recently, the interaction between RPA32C and p58C was determined by pull down assays and studied in detail using NMR [74]. In the SV40 system, interactions between RPA, and large Tag are required for efficient initiation of DNA replication [78-80]. Physical interaction between RPA32C with T-antigen allows pol-prim gain access to the 3' end of DNA for primer synthesis initiation. The higher affinity of p58C for RPA32C relative to Tag suggests that p58C can out-compete Tag-OBD for binding to RPA-coated ssDNA. To gain additional insight into this model, it is important to determine all interactions between primase and RPA subunits and their implications for the assembly of the DNA replication initiation machinery.

Our laboratory has purified a stable construct of the p48 subunit of DNA primase. We have characterized this construct and performed proteolysis protection assays to identify interacting domains between RPA and p48. A stronger interaction was found between RPA70NAB and p48. To further characterize this interaction, isothermal calorimetry (ITC) experiments were carried out. Unfortunately, we did not obtain a good titration curve from the ITC experiments. Since NMR is a more sensitive method for the study of weak interactions, a preliminary titration of RPA70AB into p48 was performed. The modest chemical shift changes of discrete peaks indicate an interaction between these two domains. Finally, crystallization screens were set up in an effort to determine the structure of p48.

Methods

Expression and Purification of p48

Recombinant full-length human p48 has been inserted into an in house vector pBG100 (L. Mizoue, Center for Structural Biology, Vanderbilt University). This vector contains an H_3C cleavable His_6 tag. p48 was expressed in BL21 (DE3) cells. Overnight cultures were prepared, then ~10 mL was used to inoculate 1L of fresh media. Cells were grown at 37 °C in TB or M9 media to an OD_{600} of approximately 0.5. The temperature was then lowered to 18 °C and the cells were allowed to equilibrate for 30 minutes. Expression was induced using 0.5 mM isopropyl thio-beta-D-galactopyranoside (IPTG). Cells were harvested by centrifugation 18 hours post induction and stored at -20 ^{o}C. Pelleted cells were resuspended in lysis buffer containing 50 mM Hepes (pH 7.5), 500 mM NaCl, 0.1% Nonidet P-40 (NP-40), and 1μL of phenylmethylsulfonyl fluoride (PMSF) per mL of buffer was added. Cells were lysed by sonication at 4 °C. Insoluble material was removed by centrifugation. The p48 subunit was purified using Ni-NTA affinity chromatography. The bound protein was eluted using a buffer containing 500 mM imidazole. Fractions containing the primase polypeptides were pooled and dialyzed overnight at 4 °C into buffer containing 20 mM Hepes (pH 7.0), 500 mM NaCl, 10% glycerol and 5 mM BME. H_3C protease was added to the dialysis bag to cleave the His_6 tag off. The sample was then repassed through the Ni-NTA column to remove the tag and other impurities. Further purification of the sample was done using an S200 column (Amersham Biosciences) equilibrated in buffer containing 20 mM Hepes (pH 7.5), 150 mM NaCl and 2 mM DTT. Samples were pooled and store at -80^{o}C.

A similar approach was used to produce a sample enriched in ^{2}H and ^{15}N. In this case M9 media was used to grow cells until an OD_{600} of ~0.5 was reached. Cells were then pelleted and transferred to M9 media supplemented with $^{15}NH_4Cl$ as the sole source of nitrogen and in D_2O for perdeuteration. Cells were allowed to equilibrate for an hour at 37 ^{o}C then transferred to 18 ^{o}C. After 30 min equilibration, cells were induced with 0.5 mM IPTG. The same procedure used for the unlabeled protein was followed for harvesting and purifying $^{2}H,^{15}N$-p48.

Expression and purification of RPA70 constructs

Recombinant human RPA70AB ($RPA70_{181–422}$) was expressed from a pSV281 plasmid containing a TEV cleavable His_6 tag at the N-terminus and RPA70NAB ($RPA70_{1–422}$) from a pBG100 plasmid containing an H_3C His_6 tag also at the N-terminus. BL21 (DE3) cell were transformed and grown in TB medium containing kanamycin at 37 °C, induced with 1mM IPTG at OD_{600} of 0.6, and harvested after 3 hours by centrifugation or left overnight at 18 ^{o}C. Pellets were stored at −20 °C. Pelleted cells were resuspended in lysis buffer containing 50 mM Hepes (pH 7.5), 500 mM NaCl, 0.1% Nonidet P-40 (NP-40), and 1μL of phenylmethylsulfonyl fluoride (PMSF) for every 1mL of buffer was added. Cells were lysed by sonication at 4 °C. Insoluble material was removed by centrifugation. Then the RPA construct was purified using Ni-NTA affinity chromatography. The bound protein was eluted using a buffer containing 500 mM imidazole. Cleavage of the His_6 tag with TEV protease or H_3C was performed through overnight dialysis in a buffer containing 20 mM Hepes (pH 7.5), 200 mM NaCl, 5 mM BME, and 10% glycerol. A second Ni-NTA purification step was used to remove the

His_6 tag. Size exclusion chromatography (SEC) using a Superdex S75 column or Superdex S200, for RPA70AB and RPA70NAB respectively, was equilibrated with 20 mM Hepes (pH 7.5), 150 mM NaCl and 2 mM DTT. Proteins were stored at −80°C.

ES-MS

p48 was concentrated to 10 mg/mL in a buffer containing 20 mM Hepes at pH 7.5, 150 mM NaCl and 2 mM DTT. Research assistant Victoria Flatt ran the experiment at the Vanderbilt Mass Spectrometry Research Center.

Size Exclusion Chromatography via Multi-Angle Light Scattering

The monodispersity of the p48 sample was verified by multi-angle light scattering connected in line with SEC (SEC-MALS). All experiments were performed using a Wyatt Technology instrument, and data were analyzed using ASTRA version 16.25. Samples were analyzed using a 2.4 mL Superdex75 column. These experiments were performed by Dr. Sivaraja Vaithiyalingam.

Limited Proteolysis

p48 primase, RPA70N, RPA70AB and RPA70NAB preparations were exchanged into buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5% glycerol, and 2 mM DTT. Digestions using proteinase K, Trypsin and Chymotrypsin were performed at room temperature using a 1:1 ratio of primase:RPA and 1:1000 protein:protease. At several time points over the 100 minute course of the digestion, an aliquot was removed, mixed

with SDS sample loading buffer and boiled for 10 minutes or using 2 mM PMSF. Finally, samples were analyzed by SDS-PAGE. The experiments were performed by Amalchi Castillo, a summer intern.

ITC experiments

Samples were dialyzed into a buffer containing 20 mM MES, 75 mM NaCl and 5 mM BME at 15 ^{o}C. Data were acquired using a MicroCal VP-isothermal titration calorimeter by first injecting 2 μL of 500 μM RPA70NAB into 40 μM p48 contained in the sample cell followed by additional 10 μL injections. The data were analyzed using the Origin software provided by MicroCal. The binding constant (K_d) and thermodynamic parameters were calculated by fitting the data using a nonlinear least-square fitting algorithm.

NMR experiments

Spectra were acquired using a Bruker DRX 800 MHz spectrometer equipped with a cryoprobe. ^{15}N-enriched p48 was prepared to 170 μM in a solution containing 20 mM Hepes (pH 7.5), 100 mM NaCl, 100mM ammonium sulfate, and 2 mM DTT. Two-dimensional ^{15}N-^{1}H TROSY-HSQC spectra were acquired with 1,024 and 128 complex points in the ^{1}H and ^{15}N dimension, respectively. A one point titration experiment was obtained by addition of unlabeled RPA70NAB into labeled p48. Data were processed using Topspin (Bruker) and analyzed with Sparky (University of California).

Results

Characterization of the p48 construct

DNA primase subunit p48 was purified as described in the methods section. We are able to express and purify p48 in high quantities with yields of 10 mg/L from rich and minimal media. Electrospray mass spectrometry analysis was used to characterize the p48 protein. The chromatogram shows only one species with a molecular weight of ~50 kDa (Fig. A.1), which is consistent with the calculated molecular weight of p48. We have also performed size exclusion chromatography in conjunction with multi-angle light scattering (SEC-MALS) to determine monodispersity (data not shown). We observed one peak, which is indicative of the presence of only one species of p48.

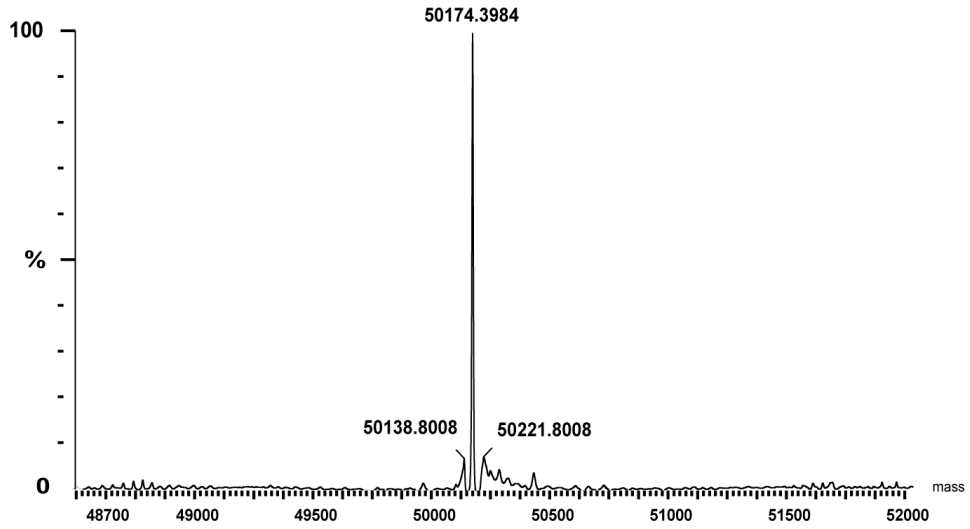


Figure A.1. ES-MS analysis of the p48 subunit of DNA primase. Calculated Mw is 50,174 Da.

Identification of RPA and p48 interacting domains using the protection from proteolysis assay

The first indication that there was an interaction between RPA and DNA primase was in 1992 [81]. The interaction was mapped to RPA70, but studies to define the specific interacting domains were not performed. The first specific domain interaction between DNA primase and RPA was recently studied in detailed for the C-terminal domain of p58 and RPA32C [31]. We investigated the interaction of the p48 subunit of DNA primase with RPA. We chose to perform proteolysis protection assays to determine interacting partners. RPA constructs containing RPA70N, RPA70AB and RPA70NAB were used in the presence and absence of p48 and incubated with trypsin or proteinase K.

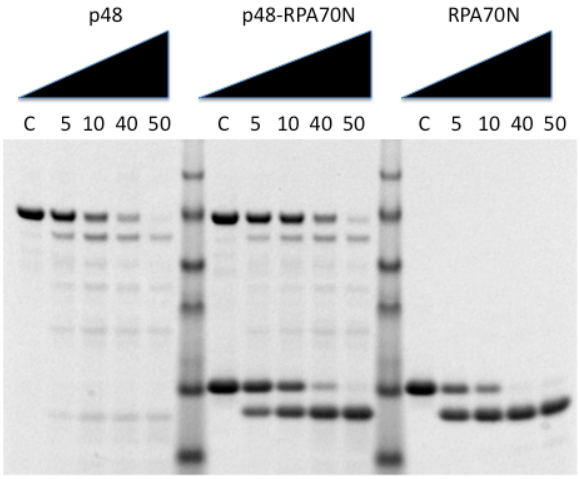


Figure A.2. Proteolysis protection assay for p48 and RPA70N using Trypsin. Purified p48 was subjected to limited proteolytic digestion for p48, RPA70N and both using Trypsin. Aliquots were taken at several time points up to 50 minutes and analyzed by SDS-PAGE.

Cleavage patterns for the incubation p48 and RPA70N (Fig. A.2) in the presence of trypsin are shown. They show disappearance of the p48 and RPA70N at 10 min and 40 min, respectively. The rate of cleavage for p48 and RPA70N are reduced when incubated together in the presence of trypsin, this is most evident for p48 where protein is still

present after 50 min of incubation. Similar results were obtained with proteinase K the only difference being the rate of cleavage was slower for all three reactions (data not shown). These results suggest there is an interactions between these two domains.

The next step was to look at the interaction between RPA70AB and p48. The same approach as with RPA70N was used. The cleavage pattern for RPA70AB incubated with trypsin showed a significant reduction of the band after 40 min (Fig. A.3). The cleavage rate was reduced for p48 in the presence of RPA70AB, we are able to observed a more intense band after 40 minutes of exposure with trypsin (Fig. A.3). Similar studies were performed with proteinase K. They showed a dramatic reduction in the intensity of the bands for RPA70AB after 30 min incubation with proteinase K, but this change is not observed when RPA70AB and p48 are mixed. These results are indicative of an interaction between p48 and the tandem RPA70AB domains.

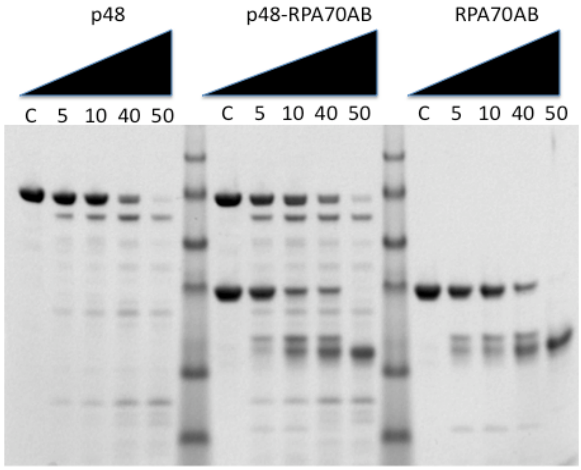


Figure A.3. Proteolysis protection assay for p48 and RPA70AB using Trypsin. Purified p48 was subjected to limited proteolytic digestion for p48, RPA70AB and both using Trypsin. Aliquots were taken at several time points up to 50 minutes and analysed by SDS-PAGE.

After concluding that there is an interaction between domains RPA70N and RPA70AB with p48, we reason that having all domains together will show an even

stronger effect if they are binding to a different p48 surface. The next step was to use RPA70NAB to determine its effect in proteolysis. Figure A.4 shows the results obtained for the proteolysis protection assay of these two domains. Once again p48 is mostly gone after 40 min incubation with trypsin. The band for RPA70NAB disappears after 10 min incubation with trypsin, but in the presence of p48 is still present at 40 min. The two species had very close molecular weights which made it difficult to resolve them in the SDS-PAGE.

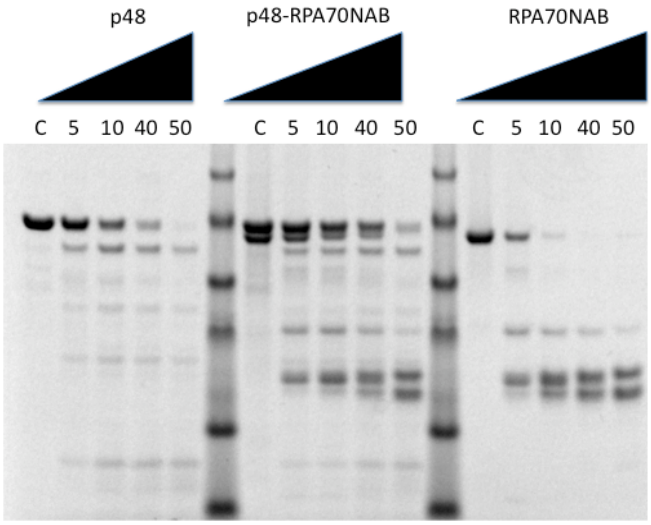


Figure A.4. Proteolysis protection assay for p48 and RPA70NAB using Trypsin. Purified p48 was subjected to limited proteolytic digestion for p48, RPA70NAB and both using Trypsin. Aliquots were taken at several time points up to 50 minutes and analyzed by SDS-PAGE.

Additionally, we performed a control experiment to determine the interactions observed were specific. A proteolysis protection assay was performed between RPA32C and p48 (Fig. A.5). Interestingly, interaction of RPA32C seems to enhance p48 cleavage. On the other hand, no change was observed in the case of RPA32C.

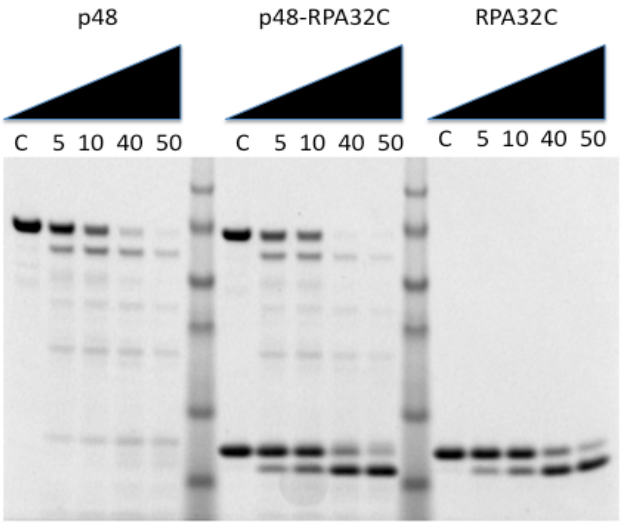


Figure A.5. Proteolysis protection assay for p48 and RPA32C using Trypsin. Purified p48 was subjected to limited proteolytic digestion for p48, RPA32C and both using Trypsin. Aliquots were taken at several time points up to 50 minutes and analyzed by SDS-PAGE.

Study of RPA and p48 interaction using ITC

Finally, we wanted to further characterized these interactions and obtain binding constants. Isothermal calorimetry (ITC) experiments were performed on RPA70NAB and p48. We chose to use this construct since it showed stronger binding to p48 based on the proteolysis protection experiments. RPA was titrated into a cell containing p48. The titration was performed twice. The result for one of the titrations is shown in Figure A.6. Using the conditions described in the methods section, we observed that the changes in heat for the titration were small. Alternate buffer conditions and reduced temperature (data not shown) were tested to determine if improved results could be obtained. However, sufficient sensitivity could not be obtained to be able to draw any definite conclusions. Other experimental techniques, such as NMR which is very sensitive for

low affinity interactions, can be use to study the interaction between RPA domains and p48.

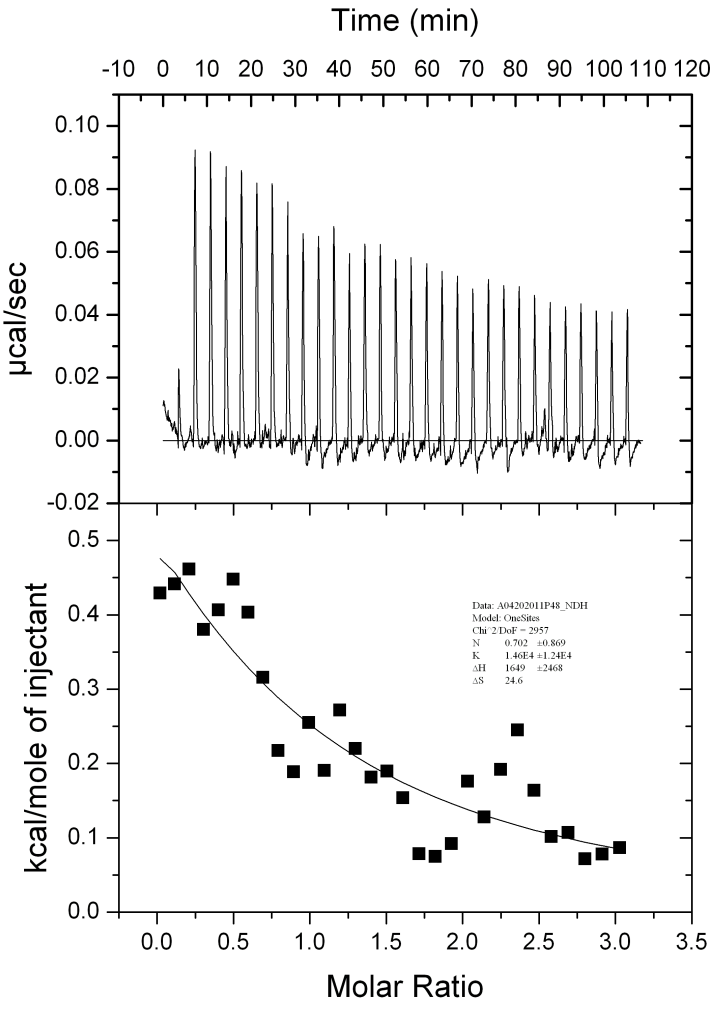


Figure A.6. ITC data for the titration of RPA70NAB into p48

NMR titration of RPA70NAB into p48

A TROSY HSQC for ^{2}H, ^{15}N-p48 was recorded for p48 using the conditions described in the method section. The spectrum showed well-dispersed peaks for the 50 kDa protein indicating it is well-folded (Fig. A.7). A one point titration of RPA70NAB into p48 at a 1:1 ratio was performed to determine an interaction between these two

domains. The spectrum for the titration (Fig. A.7, red spectrum) showed a modest change in the chemical shift of discrete amino acids (Fig. A.7), which is indicative of a weak interaction. A full titration until saturation is reached needs to be performed to be able to make a more thorough study of the interaction.

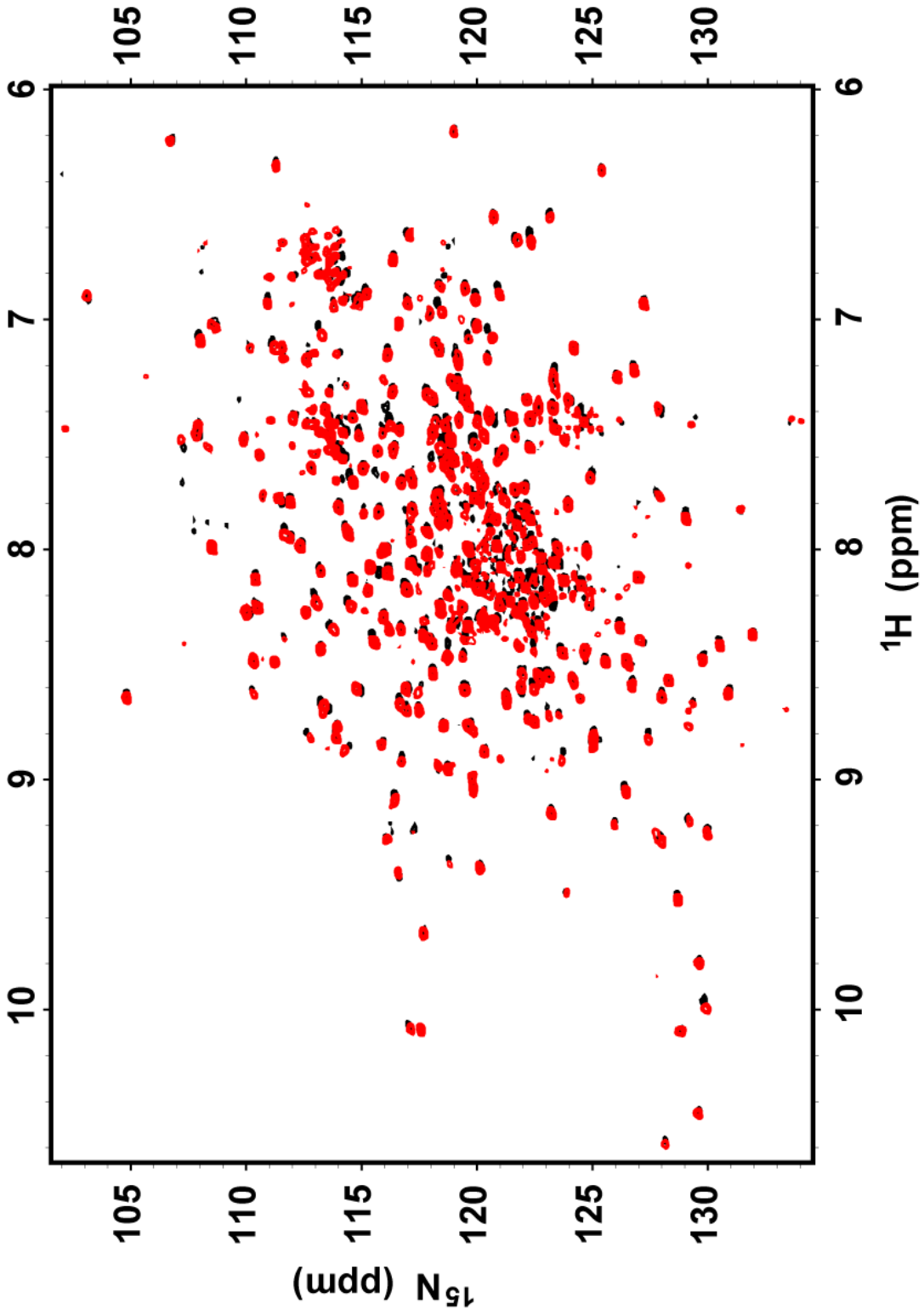


Figure A.7. NMR titration of RPA70NAB into ^{2}H, ^{15}N-p48. The spectra for p48 (black) and p48 in complex with RPA70NAB

p48 crystallization trial

Structural information on eukaryotic DNA primase is only available for the C-terminus of p58 [31]. In an effort to obtain the structure of p48, crystallization trials were performed. The p48 was obtained from cells grown in unlabelled M9 media. Several commercially available conditions for crystal formation were screened. An initial hit was observed in a buffer containing 0.2 M triammonium citrate at pH 7.0 and 20% PEG 3350. The crystal grew large enough to be mounted and a diffraction pattern was collected (Fig. A.8) on an in house X-ray beam source. The crystal diffracted to 9 Å.

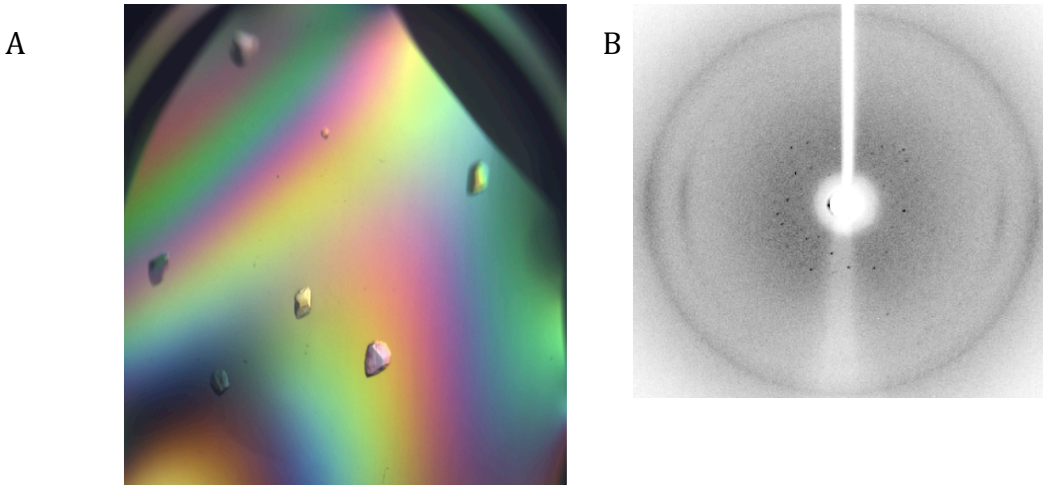


Figure A.8. Crystallization of p48. Crystals of p48 (A) where mounted and shot with an in house X-ray beam source to obtain a diffraction pattern (B)

The dispersion of the diffraction pattern indicated the presence of protein. Atomic resolution is needed to be able to determine a structure of p48. In order to improve the quality of the crystal, optimization trials were performed. The same sample was used to

perform the optimization, but we were not able to improve the crystals, and the sample was about a week old when the optimization trays were set up and some precipitation was observed in the sample. We tried to optimize with a different sample but were not able to reproduce crystal formation. For future directions, set up of crystal trays using the same screen where the hit was observed should be performed as well as testing of additional crystallization conditions.

REFERENCES

1. Bochkarev, A., et al., *Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA.* Nature, 1997. **385**(6612): p. 176-81.
2. Aves, S.J., *DNA replication initiation.* Methods in molecular biology, 2009. **521**: p. 3-17.
3. Bell, S.P. and A. Dutta, *DNA replication in eukaryotic cells.* Annu Rev Biochem, 2002. **71**: p. 333-74.
4. Sclafani, R.A. and T.M. Holzen, *Cell cycle regulation of DNA replication.* Annu Rev Genet, 2007. **41**: p. 237-80.
5. Wold, M.S., *Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism.* Annu Rev Biochem, 1997. **66**: p. 61-92.
6. Arezi, B. and R.D. Kuchta, *Eukaryotic DNA primase.* Trends Biochem Sci, 2000. **25**(11): p. 572-6.
7. Hamdan, S.M. and C.C. Richardson, *Motors, switches, and contacts in the replisome.* Annu Rev Biochem, 2009. **78**: p. 205-43.
8. Hoeijmakers, J.H., *Genome maintenance mechanisms for preventing cancer.* Nature, 2001. **411**(6835): p. 366-74.
9. Tuteja, N. and R. Tuteja, *Unraveling DNA repair in human: molecular mechanisms and consequences of repair defect.* Critical reviews in biochemistry and molecular biology, 2001. **36**(3): p. 261-90.
10. Blackburn, E.H., *Telomere states and cell fates.* Nature, 2000. **408**(6808): p. 53-6.

11. Sundquist, W.I. and A. Klug, *Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops.* Nature, 1989. **342**(6251): p. 825-9.

12. Wei, C. and M. Price, *Protecting the terminus: t-loops and telomere end-binding proteins.* Cellular and molecular life sciences : CMLS, 2003. **60**(11): p. 2283-94.

13. McEachern, M.J., A. Krauskopf, and E.H. Blackburn, *Telomeres and their control.* Annual review of genetics, 2000. **34**: p. 331-358.

14. Schramke, V., et al., *RPA regulates telomerase action by providing Est1p access to chromosome ends.* Nature genetics, 2004. **36**(1): p. 46-54.

15. Salas, T.R., et al., *Human replication protein A unfolds telomeric G-quadruplexes.* Nucleic acids research, 2006. **34**(17): p. 4857-65.

16. Sun, J., et al., *Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres.* Genes & development, 2009. **23**(24): p. 2900-14.

17. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nature structural & molecular biology, 2004. **11**(12): p. 1223-9.

18. Wold, M.S., *Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism.* Annual review of biochemistry, 1997. **66**: p. 61-92.

19. Brosey, C.A., et al., *NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA.* Journal of the American Chemical Society, 2009. **131**(18): p. 6346-7.

20. Murzin, A.G., *OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences.* The EMBO journal, 1993. **12**(3): p. 861-7.

21. Bochkarev, A. and E. Bochkareva, *From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold.* Current opinion in structural biology, 2004. **14**(1): p. 36-42.

22. Pretto, D.I., et al., *Structural dynamics and single-stranded DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering.* Biochemistry, 2010. **49**(13): p. 2880-9.

23. Ball, H.L., et al., *Function of a conserved checkpoint recruitment domain in ATRIP proteins.* Molecular and cellular biology, 2007. **27**(9): p. 3367-77.

24. Xu, X., et al., *The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling.* Molecular and cellular biology, 2008. **28**(24): p. 7345-53.

25. Pfuetzner, R.A., et al., *Replication protein A. Characterization and crystallization of the DNA binding domain.* The Journal of biological chemistry, 1997. **272**(1): p. 430-4.

26. Bochkareva, E., et al., *Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA.* The EMBO journal, 2002. **21**(7): p. 1855-63.

27. Braun, K.A., et al., *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, 1997. **36**(28): p. 8443-54.

28. Stauffer, M.E. and W.J. Chazin, *Physical interaction between replication protein A and Rad51 promotes exchange on single-stranded DNA.* The Journal of biological chemistry, 2004. **279**(24): p. 25638-45.

29. Mer, G., et al., *Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA.* Cell, 2000. **103**(3): p. 449-56.

30. Binz, S.K., A.M. Sheehan, and M.S. Wold, *Replication protein A phosphorylation and the cellular response to DNA damage.* DNA repair, 2004. **3**(8-9): p. 1015-24.

31. Vaithiyalingam, S., et al., *Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(31): p. 13684-9.

32. Deng, X., et al., *Structure of the full-length human RPA14/32 complex gives insights into the mechanism of DNA binding and complex formation.* Journal of molecular biology, 2007. **374**(4): p. 865-76.

33. Henricksen, L.A., C.B. Umbricht, and M.S. Wold, *Recombinant replication protein A: expression, complex formation, and functional characterization.* The Journal of biological chemistry, 1994. **269**(15): p. 11121-32.

34. de Laat, W.L., et al., *DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair.* Genes Dev, 1998. **12**(16): p. 2598-609.

35. Kim, C., R.O. Snyder, and M.S. Wold, *Binding properties of replication protein A from human and yeast cells.* Mol Cell Biol, 1992. **12**(7): p. 3050-9.

36. Fanning, E., V. Klimovich, and A.R. Nager, *A dynamic model for replication protein A (RPA) function in DNA processing pathways.* Nucleic Acids Res, 2006. **34**(15): p. 4126-37.

37. Stauffer, M.E. and W.J. Chazin, *Structural mechanisms of DNA replication, repair, and recombination.* J Biol Chem, 2004. **279**(30): p. 30915-8.

38. Wu, Y., et al., *Human replication protein A melts a DNA triple helix structure in a potent and specific manner.* Biochemistry, 2008. **47**(18): p. 5068-77.

39. Fan, J.H., et al., *Circular dichroism spectra and electrophoretic mobility shift assays show that human replication protein A binds and melts intramolecular G-quadruplex structures.* Biochemistry, 2009. **48**(5): p. 1099-111.

40. Lipps, H.J. and D. Rhodes, *G-quadruplex structures: in vivo evidence and function.* Trends in cell biology, 2009. **19**(8): p. 414-22.

41. Patel, D.J., A.T. Phan, and V. Kuryavyi, *Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics.* Nucleic acids research, 2007. **35**(22): p. 7429-55.

42. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. **292**(5519): p. 1171-5.

43. Prakash, A., Natarajan, A., Marky, L.A., Ouellette, M.M., Borgstahl, G.E.O, *Identification of the DNA-biding domain of human replication protein A that recognize G-quadruplex DNA.* Journal of Nucleic Acids, 2011. **2011**.

44. Loayza, D., et al., *DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding*

to multimeric sites. The Journal of biological chemistry, 2004. **279**(13): p. 13241-8.

45. Stauffer, M.E. and W.J. Chazin, *Structural mechanisms of DNA replication, repair, and recombination.* The Journal of biological chemistry, 2004. **279**(30): p. 30915-8.

46. Fanning, E., V. Klimovich, and A.R. Nager, *A dynamic model for replication protein A (RPA) function in DNA processing pathways.* Nucleic acids research, 2006. **34**(15): p. 4126-37.

47. Yuzhakov, A., Z. Kelman, and M. O'Donnell, *Trading places on DNA--a three-point switch underlies primer handoff from primase to the replicative DNA polymerase.* Cell, 1999. **96**(1): p. 153-63.

48. Yuzhakov, A., et al., *Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme.* The EMBO journal, 1999. **18**(21): p. 6189-99.

49. Daughdrill, G.W., et al., *Chemical shift changes provide evidence for overlapping single-stranded DNA- and XPA-binding sites on the 70 kDa subunit of human replication protein A.* Nucleic acids research, 2003. **31**(14): p. 4176-83.

50. He, Z., et al., *RPA involvement in the damage-recognition and incision steps of nucleotide excision repair.* Nature, 1995. **374**(6522): p. 566-9.

51. Rubtsova, M.P., et al., *Replication protein A modulates the activity of human telomerase in vitro.* Biochemistry. Biokhimiia, 2009. **74**(1): p. 92-6.

52. Jacobs, D.M., et al., *Human replication protein A: global fold of the N-terminal RPA-70 domain reveals a basic cleft and flexible C-terminal linker.* Journal of biomolecular NMR, 1999. **14**(4): p. 321-31.

53. Bhattacharya, S., et al., *1H, 13C and 15N assignments of single-stranded DNA binding domains from the 70 kDa subunit of human replication protein A.* Journal of biomolecular NMR, 2004. **28**(2): p. 195-6.

54. Rule, G.S. and T.K. Hitchens, *Fundamentals of protein NMR spectrosopy*. Focus on structural biology2006, Dordrecht: Springer. xxvi, 530 p.

55. Cavanagh, J., *Protein NMR spectroscopy : principles and practice*. 2nd ed2007, Amsterdam ; Boston: Academic Press. xxv, 885 p.

56. Pervushin, K., et al., *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(23): p. 12366-71.

57. Lee, J.H., et al., *NMR study on the interaction between RPA and DNA decamer containing cis-syn cyclobutane pyrimidine dimer in the presence of XPA: implication for damage verification and strand-specific dual incision in nucleotide excision repair.* Nucleic acids research, 2003. **31**(16): p. 4747-54.

58. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of biomolecular NMR, 1995. **6**(3): p. 277-93.

59. Tzakos, A.G., et al., *NMR techniques for very large proteins and rnas in solution.* Annual review of biophysics and biomolecular structure, 2006. **35**: p. 319-42.

60. Riek, R., et al., *Solution NMR techniques for large molecular and supramolecular structures.* Journal of the American Chemical Society, 2002. **124**(41): p. 12144-53.

61. Huang, H., et al., *Structure of a DNA polymerase alpha-primase domain that docks on the SV40 helicase and activates the viral primosome.* The Journal of biological chemistry, 2010. **285**(22): p. 17112-22.

62. Kuchta, R.D. and G. Stengel, *Mechanism and evolution of DNA primases.* Biochimica et biophysica acta, 2010. **1804**(5): p. 1180-9.

63. Dixon, K. and E. Kopras, *Genetic alterations and DNA repair in human carcinogenesis.* Seminars in cancer biology, 2004. **14**(6): p. 441-8.

64. Sauguet, L., et al., *Shared active site architecture between the large subunit of eukaryotic primase and DNA photolyase.* PloS one, 2010. **5**(4): p. e10083.

65. Mott, M.L. and J.M. Berger, *DNA replication initiation: mechanisms and regulation in bacteria.* Nature reviews. Microbiology, 2007. **5**(5): p. 343-54.

66. Arunkumar, A.I., et al., *Insights into hRPA32 C-terminal domain--mediated assembly of the simian virus 40 replisome.* Nature structural & molecular biology, 2005. **12**(4): p. 332-9.

67. Arezi, B., et al., *Interactions of DNA with human DNA primase monitored with photoactivatable cross-linking agents: implications for the role of the p58 subunit.* Biochemistry, 1999. **38**(39): p. 12899-907.

68. Diede, S.J. and D.E. Gottschling, *Telomerase-mediated telomere addition in vivo requires DNA primase and DNA polymerases alpha and delta.* Cell, 1999. **99**(7): p. 723-33.

69. Michael, W.M., et al., *Activation of the DNA replication checkpoint through RNA synthesis by primase.* Science, 2000. **289**(5487): p. 2133-7.

70. Sheaff, R.J. and R.D. Kuchta, *Mechanism of calf thymus DNA primase: slow initiation, rapid polymerization, and intelligent termination.* Biochemistry, 1993. **32**(12): p. 3027-37.

71. Frick, D.N. and C.C. Richardson, *DNA primases.* Annu Rev Biochem, 2001. **70**: p. 39-80.

72. Augustin, M.A., R. Huber, and J.T. Kaiser, *Crystal structure of a DNA-dependent RNA polymerase (DNA primase).* Nat Struct Biol, 2001. **8**(1): p. 57-61.

73. Desogus, G., et al., *Identification and characterization of a DNA primase from the hyperthermophilic archaeon Methanococcus jannaschii.* Nucleic Acids Res, 1999. **27**(22): p. 4444-50.

74. Vaithiyalingam, S., et al., *Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase.* Proc Natl Acad Sci U S A, 2010. **107**(31): p. 13684-9.

75. Nakaya, R., et al., *Identification of proteins that may directly interact with human RPA.* J Biochem.

76. Xu, X., et al., *The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling.* Mol Cell Biol, 2008. **28**(24): p. 7345-53.

77. Iftode, C., Y. Daniely, and J.A. Borowiec, *Replication protein A (RPA): the eukaryotic SSB.* Crit Rev Biochem Mol Biol, 1999. **34**(3): p. 141-80.

78. Dornreiter, I., et al., *Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen.* EMBO J, 1992. **11**(2): p. 769-76.

79. Weisshart, K., P. Taneja, and E. Fanning, *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, 1998. **72**(12): p. 9771-81.

80. Arunkumar, A.I., et al., *Insights into hRPA32 C-terminal domain--mediated assembly of the simian virus 40 replisome.* Nat Struct Mol Biol, 2005. **12**(4): p. 332-9.

81. Dornreiter, I., et al., *Interaction of DNA polymerase alpha-primase with cellular replication protein A and SV40 T antigen.* The EMBO journal, 1992. **11**(2): p. 769-76.