INVESTIGATING THE ARCHITECTURAL BASIS OF RPA QUATERNARY REMODELING UPON BINDING ssDNA

By

Chris Arlen Brosey

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

DOCTOR OF PHILOSOPHY

in

Biochemistry
December, 2011
Nashville, Tennessee

Approved by:

Professor Walter J. Chazin, Ph.D.

Professor David Cortez, Ph.D.

Professor Brandt Eichman, Ph.D.

Professor Ellen Fanning, Ph.D.

Professor Jens Meiler, Ph.D.

Professor Charles Sanders, Ph.D.

To my loving parents,

William and Jamie Brosey

ACKNOWLEDGEMENTS

I gratefully acknowledge my advisor, Dr. Walter Chazin, for allowing me the opportunity to pursue my graduate career in his laboratory. I have truly grown into a scientist under his watch and have enjoyed observing his approach to scientific inquiry and the excitement and dedication with which he tackles interesting questions. Much has happened over the past five years, personally and scientifically, and I have been very fortunate to have him as my mentor.

I would also like to thank the members of my committee for their dedication in overseeing and encouraging me through this project. Drs. David Cortez, Brandt Eichman, Ellen Fanning, Jens Meiler and Charles Sanders have been generous with their time and advice both during and outside of our formal committee meetings, and I have greatly appreciated the different perspectives they have brought to this work and my scientific development.

I have had the great pleasure to work with a number of collaborators throughout the course of my graduate career. Dr. Kavita Dorai of the Indian Institute for Science and Education Research (IISER) in Mohali, India was instrumental in piloting and optimizing the relaxation experiments used on RPA during her visit to Vanderbilt in the fall of 2008. Dr. Susan Tsutakawa of the Tainer Laboratory at Lawrence Berkeley National Laboratory has been a superlative collaborator to many members of our laboratory and has been more than generous in aiding in SAXS data collection and in sharing her expertise and insights into the scattering of DNA processing proteins. Dr. William Heller of the Center for Structural and Molecular Biology of Oak Ridge National Laboratory has

graciously worked with us to navigate the complexities of collecting SANS data, from proposal submission to data collection to data reduction and analysis. We would not have been able to pioneer this new technology in our laboratory without his help.

I would also like to thank Dr. Markus Voehler, director for the Biomolecular NMR Center, for his help in the original piloting of the relaxation experiments with Dr. Dorai, his advice and expertise on all things NMR, and most especially, his unending patience in answering five years worth of anxious NMR questions (even one hour prior to his own dissertation defense!). I would also like to thank Susan Meyn, our former laboratory manager, who always had a moment to spare to give a pro's answer to a rookie's question. I hope to emulate her enthusiastic, unflappable approach to laboratory science when I am supervising a lab of my own someday.

It has been a tremendous pleasure (and great fun!) to work with the Chazin laboratory during the past five years. A special thank you goes to Yoana Dimitrova, who has been a generous friend and wonderful colleague over the past five years. She was a great inspiration to me as she pushed through her dissertation research while coping with the loss of her best friend and mother to cancer. Dr. Benjamin and Marie-Eve Chagot graciously taught me all I know about NMR and protein biochemistry, respectively. They, too, were tremendously patient in answering my unending supply of questions. Dr. Sarah Warnick has been more than generous in sharing her expertise on the analysis of relaxation data, as well as countless discussions on data interpretation (and the occasional life-saving chocolate). I have truly enjoyed working alongside members of this laboratory and wish you all the best. You will be missed.

I owe a tremendous debt of gratitude to my friends and family, who have provided a listening ear throughout my time in graduate school. Danielle and Jake Duke have come to know RPA almost as well as I do, and their friendship has really meant a lot to me over the past decade. My extended family, the Sparks and the Broseys, have unfailingly asked after and encouraged my progress these past six years. My brother, Marc, and his wife Alice, and my sister Erin have all provided support and encouragement, as have my parents, Bill and Jamie Brosey. I am especially indebted to my mother, who spent the last two weeks prior to my defense helping me recover from a severe wreck so that I could finish my dissertation on time. The completion of this work owes as much to her dedication as mine.

The seasons of these past six years have been in God's hand, and it is with final gratitude to Him that I have been able to see them come to fruition.

'Though the fig tree does not bud and there are no grapes on the vines,
though the olive crop fails and the fields produce no food,
though there are no sheep in the pen and no cattle in the stalls,
yet I will rejoice in the LORD, I will be joyful in God my Savior.'

'The Sovereign LORD is my strength; he makes my feet like the feet of a deer,
he enables me to go on the heights.'

Habakkuk 3:17-19 (NIV)

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATRIP	ATR-interacting protein
BD	Brownian dynamics
βME	β-mercaptoethanol
BRCA1	Breast cancer type I susceptibility protein
CRINEPT	Cross-correlated relaxation-enhanced polarization transfer
CRIPT	Cross relaxtion-induced polarization transfer
DNA	Deoxyribonucleic acid
EOM	Ensemble optimization method
ER-/PR-/HER2	Estrogen receptor negative/progesterone receptor negative/ human epidermal growth receptor 2
ESI-MS	Electrospray ionization mass spectometry
HSQC	Heteronuclear single quantum coherence
IEC	Ion exchange chromatography
MALS	Multi-angle light scattering
MBP	Maltose binding protein
MD	Molecular dyanmics
MF	ModelFree
Ni-NTA	Nickel nitriloacetic acid
NMR	Nuclear magnetic resonance spectroscopy
NOE	Nuclear Overhauser effect
OB-fold	Oligonucleotide/oligosaccharide binding fold
PCNA	Proliferating cell nuclear antigen

Pol-prim	Polymerase α/primase
PRE	Paramagnetic relaxation enhancement
RDC	Residual dipolar coupling
RFC	Replication factor C
R_g	Radius of gyration
RNA	Ribonucleic acid
RPA	Replication Protein A
RPA-DBC	RPA DNA binding core
SANS	Small-angle neutron scattering
SAR by NMR	Structure-activity relationship by NMR
SAS	Small-angle scattering
SAXS	Small-angle x-ray scattering
SDF	Spectral density function
SDS-PAGE	Sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SSB	Single-stranded DNA-binding protein
ssDNA	Single-stranded DNA
STEM	Scanning tunneling electron microscopy
SV40	Simian virus 40
Tag, Tag-OBD	Large T-antigen, origin binding domain of large T-antigen
TROSY	Transverse relaxation optimized spectroscopy
XPF/ERCC1	Xeroderma pigmentosum group F/ Excision repair cross-complementing group 1
XPG	Xeroderma pigmentosum group G

CHAPTER I

INTRODUCTION

Overview

Propagation and maintenance of the cellular genome are among the most fundamental biochemical processes executed across all kingdoms of life. Such processes range from the meticulous duplication of the cellular genome, to the high-precision exchange of genetic information between chromosomes, to detection and repair of damaged genetic material. These intricate tasks are carefully guided from start to finish by a highly synchronized army of DNA processing factors that dynamically assemble, exchange, and reorganize at the DNA to advance each biochemical step. The dynamic composition and organization of these multi-protein machines enables efficient and reliable completion of each DNA processing pathway thousands of times each day.

While genomic maintenance is crucial for life, DNA processing itself, paradoxically, is a source of increased genomic vulnerability. Access to individual DNA strands necessitates displacement of the DNA from its protected, stable duplex form, leaving the ensuing single-stranded (ss)DNA exposed to chemical and enzymatic assault and prone to self-entanglement. This precarious situation is remedied by the existence of single-stranded DNA binding proteins (SSBs), which coat and protect ssDNA from these hazards for as long as the DNA remains unwound. In eukaryotes, Replication Protein A (RPA) serves as the primary SSB and coordinates the recruitment and progression of DNA processing complexes in addition to its role in protecting ssDNA. As a key and

early-identified DNA processing protein, RPA has been subject to extensive biochemical characterization, as well as structural investigation for all globular fragments of the protein. In spite of this, the structural and mechanistic basis for how full-length RPA carries out the fundamental biological function of binding ssDNA remains unknown and forms the central line of inquiry for my Ph.D. dissertation research.

Propagating and Maintaining the Genome: DNA Processing and the Role of SSBs

Ever since the debut of the 'double helix' in 1953 (*1*), how DNA is propagated and preserved has remained among the most widely pursued questions in biology. In the decades since publication of this pioneering structure, the landscape for DNA processing has moved beyond the essential mechanics of DNA replication to include a complicated regulatory landscape governing DNA damage recognition and response (*2*), multiple DNA repair pathways (*3-9*), the intracellular exchange of genetic information via homologous recombination (*10*), and numerous points of crossover among all of them. Collectively, this broad assortment of DNA transactions employs a wide variety of DNA biochemistry carried out by a host of structurally diverse multi-protein assemblies. Despite the biochemical and structural diversity populating the DNA processing milieu, all DNA transactions face the common challenge of manipulating single strands of DNA. Not surprisingly, SSBs play a ubiquitous role in nearly every form of DNA processing present within the cell. Numerous reviews on SSBs are available (*11-16*); a brief summary of the structural organization of SSBs and the specific benefits they provide to DNA processing follows.

Structural organization and hierarchy of SSBs

The fundamental structural module of the SSB is the oligonucleotide / oligosaccharide binding domain (OB-fold) (*17-20*). This particular structural motif consists of a coiled, five-strand antiparallel β-barrel capped with an α-helix between strands 3 and 4 (Figure 1.1A) (*17*). Variation among OB-folds is achieved by varying loop length between β-strands or the direct insertion of additional structural motifs between canonical elements of the fold (i.e. zinc ribbons, zinc fingers, helical towers (*17, 18, 20*). The concave surface presented by β-strands 2 and 3 is the primary site for ligand binding with potential contributions from residues in the flanking β-strands 1 and 4, as well as from loops joining β1 and β2 (L_{12}), β3 and α ($L_{3\alpha}$), α and β4 ($L_{\alpha 4}$), and β4 and β5 (L_{45}) (Figure 1.1B) (*18*). Polarity of nucleic acid binding is highly conserved, with the majority of nucleic acids associating with the 5' end toward β-strands 3 and 4 and the 3' end toward β-strands 1 and 2 (Figure 1.1B) (*18*). Notable exceptions to this are the *E. coli* SSB and *O. nova* telomere end-binding protein, for which binding polarity is reversed (*18*). Most importantly, polarity at the OB-fold binding interface is propagated with tandem repetition of the OB-fold, allowing polarity to be maintained across an entire protein. OB-folds are not constrained to exclusive binding of nucleic acid or oligosaccharides and often serve as important protein interaction motifs in SSBs (*18-20*).

The degree of tandem domain architecture and oligomeric assembly of OB-folds forms the basis for structural classification of SSBs (Figure 1.2) (*15*). Simple SSBs, which are found in bacteria and crenarcheons, consist of one or two OB-folds encoded on a single polypeptide (*15*). Functionally, simple SSBs organize into homotetrameric and homodimeric assemblies to bind ssDNA; however, oligomerization is not required to

maintain the stability of their native monomeric folds. Prototypical examples from this group include the SSBs of *E. coli* (bacterial) and *S. solfataricus* (crenarchaeal) (*15*).

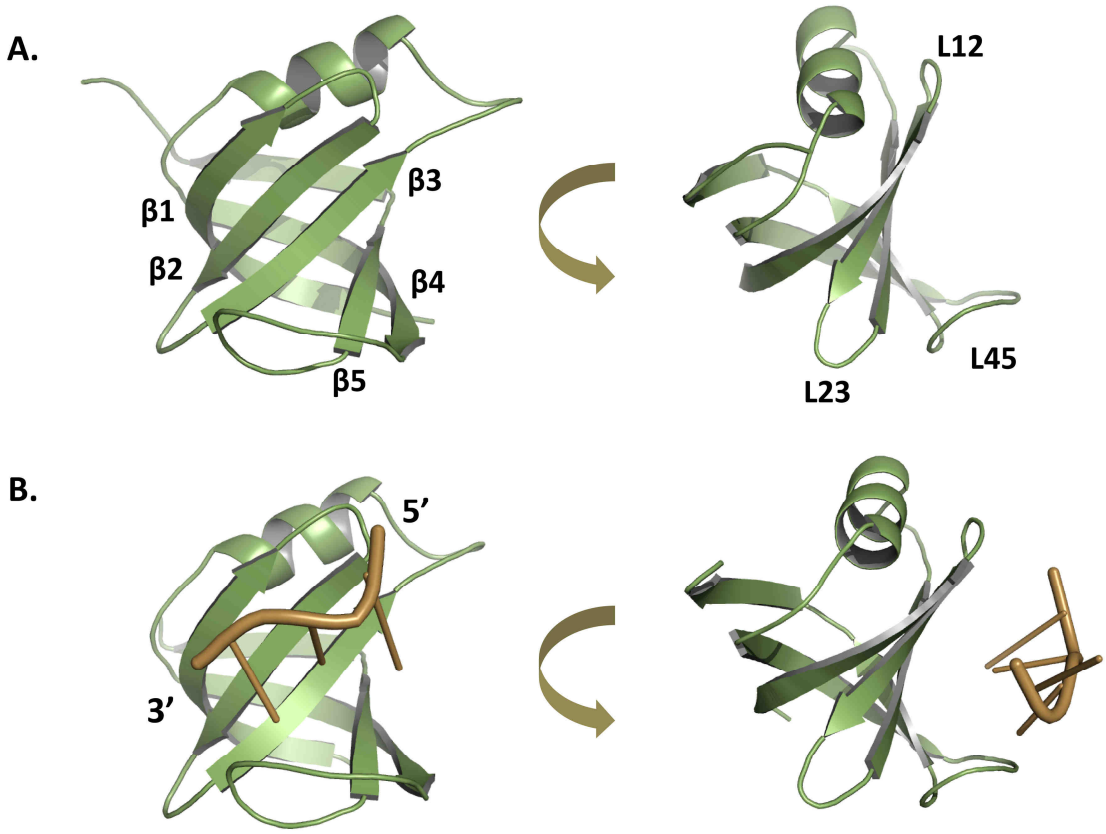


Figure 1.1. OB-fold structure and function. A) Archetypal OB-fold topology. OB-fold domain from aspartate t-RNA synthetase, *S. cerevisiae*. PDB entry 1ASY. B) Nucleic acid interaction surface and binding polarity.

Complex SSBs, also referred to as the RPA sub-group, function exclusively in eukaryotes and euryarcheons (*15*). They comprise multiple, tandem OB-folds distributed across more than one polypeptide and require oligomeric assembly to generate a native protein structure (*15*). Because of the complexity of their inter-domain architectures, higher-order oligomerization of individual RPA molecules is not required for functional

ssDNA binding, as it is with simple SSBs. Examples from this group include human RPA and the *A. fulgidus* SSB from euryarchaea (*15*). Limited cross-over from simple SSBs to higher-order organisms is present, as demonstrated by human mitochondrial SSB (mtSSB) (*21*) and the recently discovered human SSBs (hSSB1, hSSB2) (*22*).

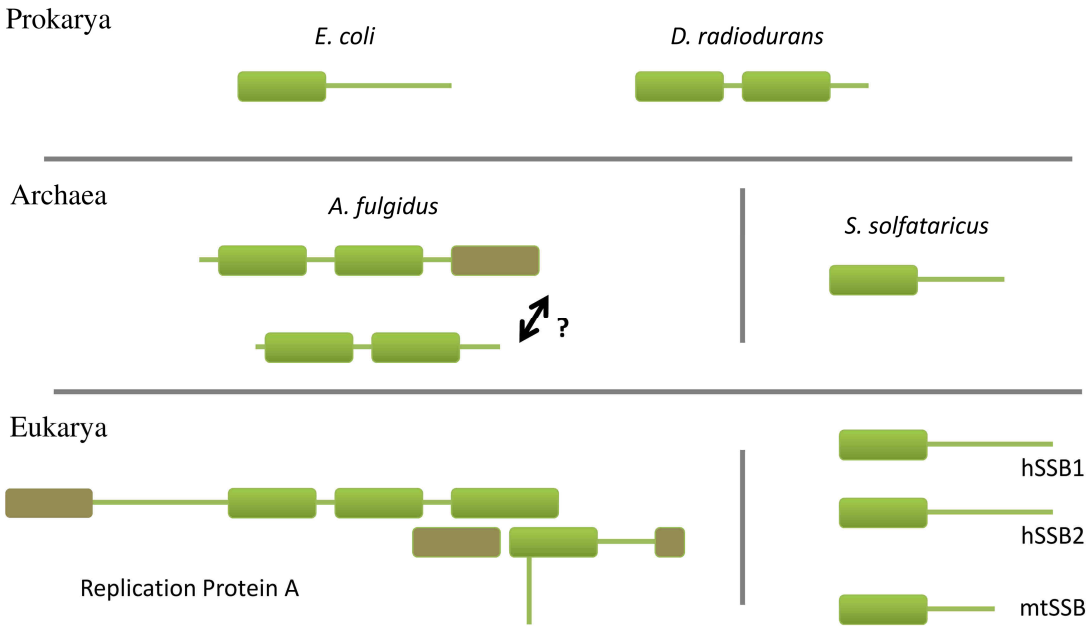


Figure 1.2. Structural classification of SSBs. Structured regions are represented as boxes and disordered regions as lines. DNA-binding OB-fold domains are green; non-DNA binding or unidentified domains are brown. Higher order oligomerization for *A. fulgidus* and other euryarcheaons has not been characterized. Adapted from Richards, et al (*15*).

Specific roles for SSBs during DNA processing

From a biochemical perspective, SSBs assist DNA processing by providing two key functions at the site of action: stabilization of ssDNA substrates and coordinated recruitment of a diverse array of DNA processing factors. The binding of ssDNA by SSBs occurs with high affinity (K_d ~10^{-9} M) and low sequence specificity (*12*, *16*).

Conversely, interactions with other DNA processing proteins possess much lower affinities in comparison (K_d ~10^{-6} M), consistent with the need for dynamic protein interchange during DNA processing (*23*, *24*).

SSB interaction with ssDNA provides multiple benefits to the forward progression of DNA processing assemblies. As mentioned previously, high affinity, nonspecific binding of ssDNA by SSBs protects against nucleolytic damage and chemical modification to DNA bases, as well as inappropriate binding by other DNA processing factors. Coating ssDNA with SSBs also prevents formation of DNA secondary structure and duplex reannealment, which can hinder or stall DNA processing machinery (*23*). Importantly, SSB binding of ssDNA is critical for ensuring the efficiency of DNA processing by suppressing dynamic fluctuations of non-duplex DNA and thus providing a stable template for enzymatic action. SSB stimulation of helicase or polymerase activity is one important consequence of this ability to organize ssDNA (*25*-*27*).

In addition to providing chemical and structural stabilization to ssDNA templates, SSBs also interact with a wide variety of DNA processing factors, serving as a platform for the recruitment and interchange of these proteins (*11*, *14*). The role of SSB protein interaction in recruitment and exchange of DNA processing factors is seen most vividly with polymerase switching at the eukaryotic replication fork, where polymerase δ assumes control of the ssDNA template from polymerase α through a concerted series of contacts through RPA, RFC, and PCNA (*23*). Protein interaction with SSBs, which maintain a pre-defined DNA binding polarity, is also critical for ensuring the proper physical orientation of biochemical action on ssDNA templates. Nucleotide excision repair relies upon RPA DNA-binding polarity to recruit and orient excision enzymes

XPF/ERCC1 and XPG to 5' and 3' sites of the target lesion, respectively (*28*). Perhaps the most direct outcome of contact between SSBs and other DNA processing factors is the initial loading of SSBs onto newly unwound ssDNA and their subsequent displacement to permit access to ssDNA templates. This aspect of SSB protein interaction has been characterized in detail for the SV40 system of replication, where interaction between the SV40 large T-antigen helicase and the human SSB RPA results in efficient loading of RPA onto emerging ssDNA. Subsequent contact between these two proteins enables displacement of RPA from ssDNA substrates to permit loading of polymerase α/primase to initiate primer synthesis (*29*, *30*).

The interplay between SSB DNA binding and interaction with other proteins provides an ideal platform for the assembly, coordination, and regulation of DNA transactions. To serve as a common hub for very different types of DNA metabolism requires great structural versatility from SSBs. How SSBs adapt and evolve architecturally with the forward progression of a particular DNA transaction remains a central question for DNA processing. The remainder of this chapter focuses upon the current state of these questions for RPA prior to undertaking this dissertation project and the experimental methodology employed to probe the plasticity of RPA architecture and its DNA-binding trajectory.

Replication Protein A

Replication Protein A (RPA) was originally discovered over twenty years ago as an essential biochemical factor necessary for *in vitro* reconstitution of SV40 replication (*31*, *32*). Isolation of this eukaryotic counterpart to the bacterial SSBs was somewhat late

in coming, as almost all other eukaryotic components complementary to the prokaryotic replication system had been identified some years previously. The delayed entrance of RPA led some researchers to speculate that eukaryotes may not even require SSBs (*33*). In the years since then, RPA has not only been shown to be indispensible to replication of DNA, but has also been established as a key signaling factor in DNA damage recognition and response (*4*); an essential component for nearly every DNA repair pathway (*4*); and an important participant in homologous recombination (*10*, *34*), telomere maintenance (*35*), and transcriptional regulation (*12*, *36*). Intensive biochemical and structural investigation of this protein has provided a fundamental understanding of SSB function (*12-14*, *37*).

Structural and functional organization of RPA

A heterotrimer, RPA consists of three polypeptide subunits, denoted by molecular weight: RPA70, RPA32, and RPA14 (Figure 1.3A) (*12*, *13*). Proteolytic and structural studies on RPA fragments and the intact protein have revealed a modular inter-domain organization with a total of seven globular domains connected by flexible linkers and a single disordered domain (RPA32N) (*38-49*). Three of these domains (70C, 32D, 14) interface through a hydrophobic three-helix bundle to form the trimeric core of RPA (Figure 1.3) (*43*), from which emanates the N-terminal domains of RPA70 (70B, 70A, and 70N, all flexibly linked in tandem) and the C-terminal domain of RPA32 (RPA32C). Individual expression for each RPA subunit is insoluble; however, RPA32 and RPA14 are able to form a soluble subcomplex in the absence of RPA70 (*50*, *51*). As with the majority of SSBs, all domains are OB-folds with the exception of the C-terminal winged

helix domain of RPA32, and have been characterized individually or in tandem at high structural resolution (Figure 1.3C) (*39*, *41-45*, *48*).

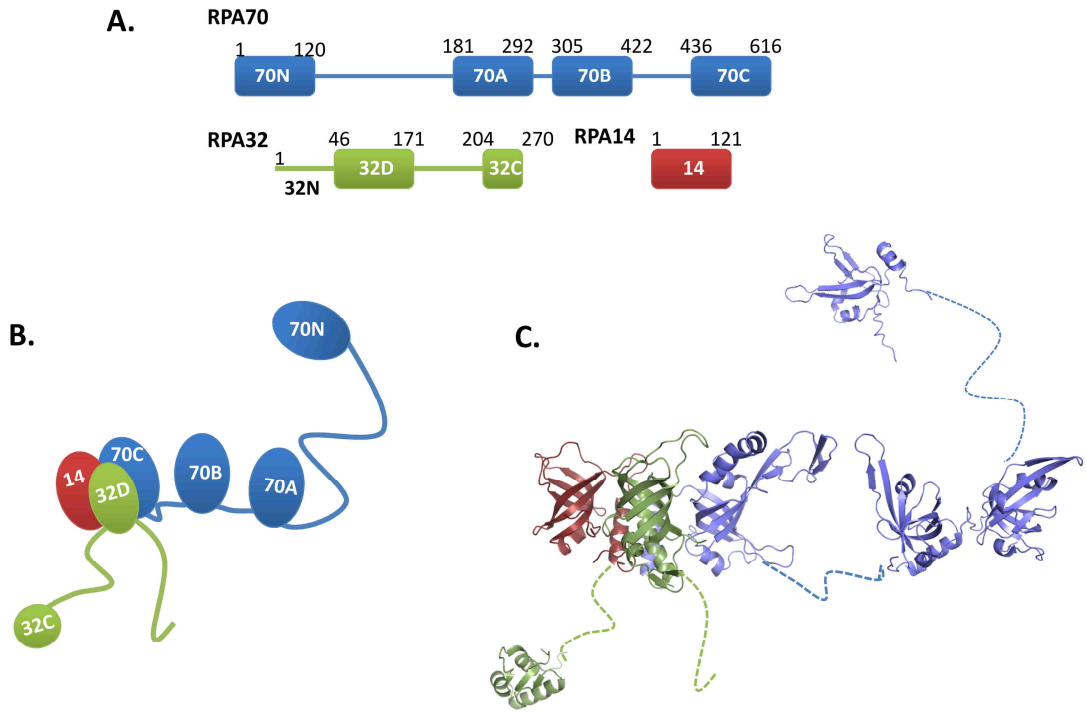


Figure 1.3. Structural organization of RPA. A) RPA domain map. B) Cartoon representation of RPA inter-domain organization. C) Representative high resolution structures of RPA domains. PDB entries 1EWI (70N), 1FGU (70AB), 1L1O (70C/32D/14), 1DPU (32C).

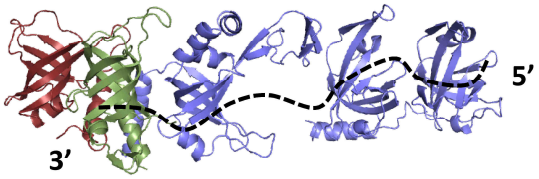
DNA binding is mediated by the four central OB-fold domains – 70A, 70B, 70C, and 32D – with 5' to 3' polarity from 70A to 32D (Figure 1.4) (*28*, *52*). Contact with ssDNA is maintained through base-stacking by two conserved aromatic residues in each domain, as well as hydrogen bonding and electrostatic interactions from basic residues lining each OB-fold cleft (*39*, *43*). *In vitro* cross-linking studies, mutational analysis, and crystal structures from the tandem 70AB domains have indicated that RPA proceeds through three different interaction modes upon binding ssDNA: an initial 8-10 nucleotide

binding mode that includes domains 70A and 70B, a 12-23 nucleotide mode that proceeds to engage 70C, and a final 28-30 nucleotide mode that encompasses all four DNA-binding domains (Figure 1.4) (*20*, *53*, *54*). This final binding mode complements RPA's occluded site size of 30 nucleotides (*55*, *56*). Collectively, these discrete interaction modes are thought to form a 'DNA-binding trajectory' that RPA traverses each time it interacts with ssDNA (*14*).

DNA-binding occurs with high affinity (K_d ~10^{-9} – 10^{-11} M) and sequence non-specificity, but there exists a fifty-fold preference for poly-pyrimidine over poly-purine substrates (*55*, *56*, *59*). The binding affinities of individual DNA-binding domains are quite weak in comparison and are believed to decrease in strength from 5' to 3' along RPA's DNA binding channel (Figure 1.4B) (*46*, *57*, *60*, *61*). High affinity for ssDNA is achieved through the close tethering of these domains (10 residues for 70A to 70B; 15 residues for 70B to 70C), increasing the local concentration of the otherwise weak binding sites (*62*). Diminishing domain affinity, as well as the modularity of the DNA binding channel, is believed to be critical for enabling other DNA processing proteins to displace RPA from ssDNA, as this presents a series of manageable, weaker interactions to be shifted.

As with all SSBs, RPA makes multiple contacts with a host of DNA processing factors. An excellent summary of proteins known to interact with RPA and their target interaction sites, if known, can be found in the review work of Fanning and colleagues (*14*). RPA domains 70N and 32C serve as primary protein interaction sites; however, protein binding is also targeted to the principal DNA-binding domains 70A and 70B to

A.



B.

Domain	K_d	Method[1]
70A	1.7 μM	NMR[2]
70B	16.8 μM	NMR[2]
70C	N.D.[3]	-------
32D	N.D.[4]	--------

C.

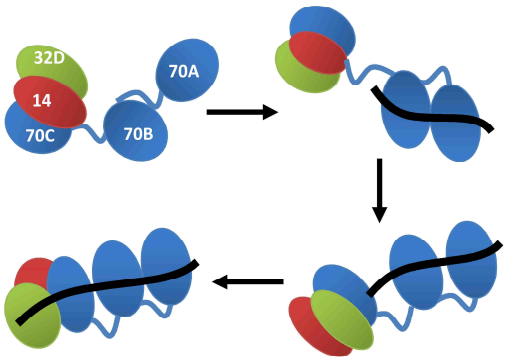
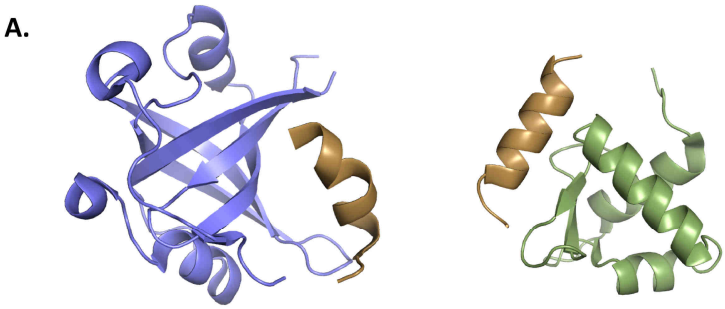


Figure 1.4. RPA and ssDNA binding. A) RPA DNA-binding domains (70A, 70B, 70C, 32D) form a modular binding channel. RPA70ABC (blue), RPA32D (green), RPA14 (red). Tentative path for ssDNA is indicated by black dotted line. PDB entries 1JMC (70AB) and 1L1O (70C/32D/14). B) Dissociation constants for individual DNA-binding domains. N.D. (not determined). [1]Method used to determine affinity. [2]Arunkumar, et al (*46*). [3]A quantitative measurement of 70C affinity is unavailable in the literature. Qualitative comparisons of binding curves suggest that its affinity is comparable to that of 70B (*57*). [4]Reported attempts to quantify 32D binding have been unsuccessful thus far (*58*). C) RPA binds ssDNA in three discrete binding modes: (1) an initial 8-10 nucleotide mode that engages 70A and 70B, (2) an intermediate 12-23 nucleotide mode that includes 70A-70C, and (3) a final 28-30 nucleotide mode that includes all four DNA-binding domains (70A-70C, 32D).

promote or compete with DNA binding (*12-14*). Structural characterization of complexes formed between proteins and peptide fragments targeting 70N or 32C have revealed common binding surfaces on each domain – either the basic cleft of the 70N OB-fold or the surface of the three-strand β-sheet of 32C (Figure 1.5A) (*45*, *48*, *63*, *64*). Common sequence motifs targeting each domain have also been characterized for a select number

of these interactions (Figure 1.5B) (*48*, *64*); however, universality of this consensus has yet to be established. As mentioned above, protein interactions with RPA serve to recruit other DNA processing proteins to sites of action and to ensure their proper orientation with respect to the DNA. Similarly, protein interaction also facilitates loading or displacement of RPA from ssDNA (*29*, *30*).



B.

RPA70N	ATRIP(54-68):	DFTADDLEELDTLAS
	MRE11(539-553):	AFSADDLMSIDLAEQ
	RAD9(297-311):	DFANDDIDSYMIAME
	P53(44-58):	MLSPDDIEQWFTEDP
RPA32C	UNG2(73-88):	RIQRNKAAALLRLAAR
	XPA(29-46):	ERKRQRALMLRQARLAAR
	RAD52(257-274):	RKLRQKQLQQQFRERMEK

Figure 1.5. RPA and protein interaction. A) Protein interaction surfaces of 70N (blue) and 32C (green) making contact with peptide fragments of p53 (*45*) and UNG2 (*48*) in gold, respectively. PDB references 2B3G (70N) and 1DPU (32C). B) Peptide binding motifs for 70N and 32C. Identical or similar residues are highlighted in red (adapted from (*64*) and (*48*), respectively).

Emerging therapeutic implications for RPA research

Due to its centrality in DNA processing, genomic loss of any single RPA subunit is lethal (*65*, *66*). However, even subtle mutations can have dramatic effect as witnessed by the RPA70 missense mutant L221P in the DNA-binding domain RPA70A. This particular mutant, initially characterized in yeast, abrogates DNA binding for the entire

protein (*67-69*). Homozygosity for this mutation is embryonic lethal in mice; heterozygous carriers are subject to increased incidence of lymphomas, where tumors are characterized by gross chromosomal breaks and rearrangements (*65*). Loss of the RPA70 chromosomal locus (17p13.3) has been noted in a number of cancers; exactly how this loss impacts cancer progression remains to be determined, but the role of RPA in maintaining genomic stability is likely to be relevant (*70*, *71*).

Recent efforts have begun to explore the potential for exploiting RPA's role in genome maintenance for the development of chemotherapeutic compounds. As an essential factor for the repair of DNA damage inflicted by traditional chemotherapeutic agents, RPA presents a particularly attractive secondary target to increase the efficacy of these compounds (*72*). Moreover, RPA expression is upregulated in certain types of cancer. As an example, triple-negative (ER-/PR-/HER2-negative), BRCA1-deficient breast tumors exhibit elevated levels of RPA, and knock-down of RPA in BRCA1-deficient HeLa cells has been shown to reduce cell survival in culture (*73*).

Small molecule screening by Turchi and colleagues has identified inhibitors capable of disrupting DNA-binding by RPA domains 70A and 70B. When target compounds are combined with cisplatin or etoposide, a synergistic decrease in cellular viability is observed in cultured cells (*72*, *74*, *75*). Small molecule screening for the protein interaction domain 70N is currently on-going in our lab, following the well-established SAR by NMR methodology (*76*). RPA70N screening has also been reported in the generation of compounds capable of disrupting interaction between 70N and damage sensor Rad9, as well as tumor suppressor p53, though an *in vivo* effect has yet to be tested (*77*). In addition to their therapeutic potential, small molecule inhibitors of

RPA function are valuable research tools for dissecting relationships between ligand targeting of a particular domain(s) and specific functional outcomes in DNA processing.

Mechanistic Integration of RPA Structure and Function: How Architectural Remodeling Drives DNA Processing

As mentioned previously, how SSBs adapt to the shifting substrate landscapes of DNA metabolism remains a key question in DNA processing biology. The modular domain organization of human RPA and other SSBs would indicate that inter-domain rearrangement is a vital component of this structural adaptability. Global characterization of RPA's modular architecture and its response to interaction with ssDNA or protein, however, has remained limited. The prevailing understanding of RPA architectural remodeling at the time this project was initiated is discussed below.

RPA quaternary architecture

Coincident with the publication of proteolytic mapping experiments highlighting the modular domain organization of RPA (*38*), scanning transmission electron microscopy (STEM) studies on full-length RPA reported visual changes to the protein's conformation when engaged in the 8-10 and 28-30 nucleotide binding modes (*78*). Specifically, the authors noted a transition from a 'globular' to 'elongated' profile upon binding ssDNA substrates of increasing length (*78*). A number of researchers have embraced and propagated this 'compact' view of RPA architecture in the absence of ssDNA (*20*, *43*, *53*, *79*), suggesting that modular RPA domains must engage in inter-domain interactions that are dispersed upon DNA binding (Figure 1.6A).

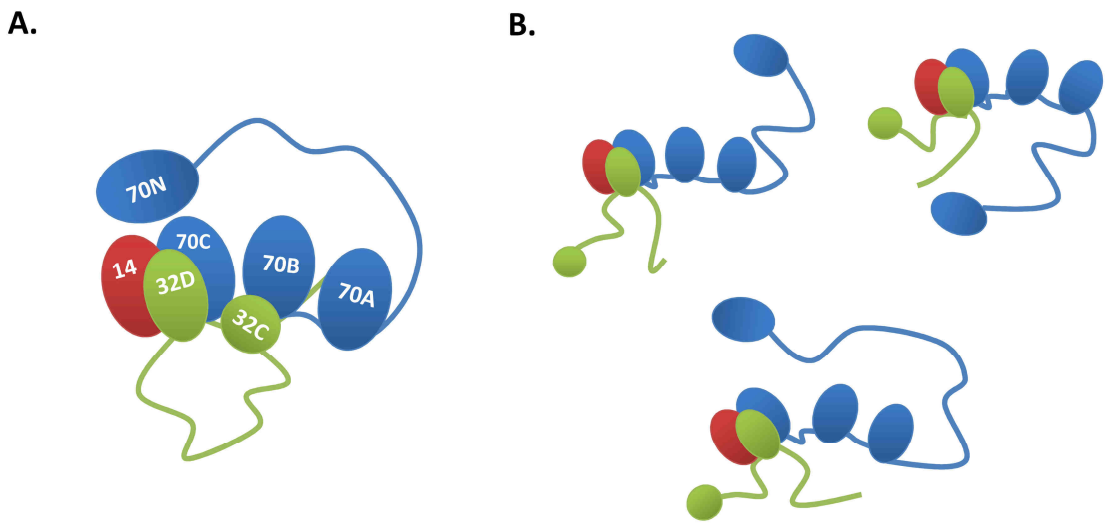


Figure 1.6. Alternate views of RPA quaternary architecture. A) 'Compact' inter-domain organization. B) 'Independent' inter-domain organization.

The harsh fixation techniques of electron microscopy are not well suited to preserving the native structure of flexible, modular proteins, which would indicate that the question of native RPA architecture is better addressed under solution conditions. Subsequent NMR studies on tandem domain constructs, RPA70AB, RPA70NA[1], and RPA32/14 revealed a high degree of inter-domain independence in the absence and presence of ssDNA (*46-48*). This argued for a more flexible and dynamic model of RPA architecture in solution (Figure 1.6B). The crystal structure of RPA70AB in the absence of ssDNA also lent support to an independent arrangement of domains, as 70A and 70B assumed dissimilar inter-domain orientations in two separate crystal forms (*42*).

Remodeling RPA architecture upon binding ssDNA

While advocating conflicting views on RPA's DNA-free architecture, neither model (compact versus flexibly independent) has been validated for full-length RPA

[1] The RPA70NA construct used in the cited study includes approximately half of the 70B sequence as well.

under native solution conditions. A similar lack of structural clarity also persists in descriptions of architectural changes to RPA upon passing through its DNA binding trajectory. The existing biochemical evidence supports the presence of progressive architectural change as RPA binds ssDNA, but provides limited structural information as to the nature of these changes. Protease protection assays of RPA with and without ssDNA reveal increased proteolytic resistance of RPA's modular structure when bound to ssDNA (*38*). UV cross-linking studies of RPA and various multi-domain fragments demonstrate a shift in RPA subunits that contact a primer-template junction (from RPA70 to RPA32) as a ssDNA overhang is increased from 4 to 31 nucleotides (*53*, *80*). These studies, along with earlier UV cross-linking experiments (*81*), provide biochemical delineation of RPA's three modes of DNA binding (8-10 nt, 13-14 nt, 30 nt). Subsequent experiments on RPA mutants, with targeted disruption of DNA-binding activity in individual domains, allowed correlation of RPA binding modes with the activity of specific domains and also provided alternative boundaries for each mode (8-10 nt for 70AB, 12-23 nt for 70ABC, 23-27 nt for 70ABC/32D) (*54*). More recently, fluorescence quenching studies have suggested an upper limit for the intermediate mode of 16 nucleotides and a lower limit for the final mode of 26 nucleotides (*82*).

Existing structural information on these binding modes is limited to the original crystal structures of RPA70AB bound to an octanucleotide substrate (*39*, *40*) and the STEM study outlined above (*78*). The RPA70AB/dC_8 crystal structure provides valuable insight into the how domains 70A and 70B mediate RPA's first interaction mode with ssDNA; the question as to whether this is their native architecture in the context of full-length RPA remains. This is particularly salient as the biochemical studies detailed

above have proposed that the interaction modes of RPA correlate with the globular (8-10 nt), contracted elongated (12-23 nt), and extended elongated conformations (28-30 nt) observed in the STEM studies, suggesting that RPA70AB makes additional inter-domain contacts in the initial binding mode (*53*, *79*). However, if inter-domain flexibility persists upon DNA-binding, as indicated for the DNA-free protein, this would again render RPA susceptible to STEM preparation artifacts. The broad and occasionally conflicting nucleotide ranges reported for the intermediate (12-23 nt) and final binding modes (23-27 nt or 28-30 nt) relative to that of the initial binding interaction (8-10 nt) are particularly suggestive of additional dynamic plasticity in the DNA-bound forms of RPA.

Interplay between protein-induced architectural remodeling of RPA and ssDNA binding

More recently, attention has focused upon the ability of specific protein interactions to prompt architectural remodeling of RPA, particularly to stimulate or inhibit DNA-binding. A typical example is the interaction between the origin-binding domain of SV40 large T-antigen (Tag-OBD) and RPA70AB (*29*). Contact between Tag-OBD and a contiguous surface formed by 70A and 70B opposite their DNA-binding clefts stimulates RPA's ability to bind ssDNA (*29*). This Tag-induced remodeling of inter-domain orientation purportedly offsets the entropic cost of aligning free 70A and 70B to engage the ssDNA substrate, thus increasing DNA binding (*29*). The interaction between Tag and RPA is maintained for substrates representative of RPA's initial DNA binding mode (dT_8), but is subsequently lost upon progression to intermediate (dT_{15}) and final binding modes (dT_{30}), suggesting that structural transitions accompanying RPA's DNA binding trajectory are ultimately responsible for disrupting this contact (*29*).

Likewise, interaction between Tag-OBD and the protein interaction domain 32C facilitates removal of RPA from ssDNA to allow initiation of primer synthesis by polymerase α/primase (*30*). Here the correlation between protein interaction and interruption of DNA binding is not intuitive, as a flexible 30-residue linker separates 32C and the final DNA-binding domain 32D. However, the authors note that the weak binding affinities of 32D and neighboring 70C would be capable of allowing transient access to the 3' side of substrate ssDNA. They hypothesize that capture of 32C and subsequent tension on the intervening 32DC linker may increase this transient accessibility enough to allow loading of polymerase α/primase. Subsequent initiation of primer synthesis would then be sufficient to drive the equilibrium towards release of RPA from the DNA.

Toward a structural understanding of RPA architecture and its DNA binding trajectory

As mentioned previously, the precise nature of the structural transitions accompanying RPA binding of DNA or the architectural influence of protein interaction on these transitions remains to be determined. It is clear, however, that a structural understanding of RPA's DNA-binding trajectory has broader implications for the progression of DNA processing. Despite the importance of this information, structural characterization of full-length RPA remains challenging. Flexible, modular proteins are typically not ideal candidates for crystallization and study by x-ray diffraction. Notably, published crystal structures of the subcomplex RPA32/14 possess diffuse density for the flexibly linked 32N and 32C domains (*49*). At 116 kDa, RPA is also too small for shape reconstruction analysis by cryo-electron microscopy. Moreover, evidence for inter-

domain flexibility from studies of smaller RPA fragments would indicate that solution approaches are to be preferred (*46-48*). This thesis utilizes the combination of NMR spectroscopy and small-angle scattering, which has proved an increasingly powerful approach for providing both local and global descriptions of the architectures of flexible, modular proteins and dynamic multi-protein assemblies (*83*, *84*). A basic overview of these experimental techniques is provided in the next section, as well as the strategies employed for applying this methodology to RPA.

Experimental methods for structural and dynamics analysis

Integration of detailed atomic-resolution analysis with global macromolecular conformation has provided unprecedented insights into the architecture and dynamics of macromolecular assemblies (*84*, *85*). Traditional structural tools, such as x-ray crystallography and NMR spectroscopy, have afforded detailed insight into individual structural modules of macromolecules, their ability to engage ligands, and their capacity to form higher-order complexes. The extensive library of high-resolution structures for component domains of RPA demonstrates the past fruitfulness of these structural approaches (*40-45*, *48*, *49*).

Of key interest, though, is how this collection of individual domain structures organizes in the solution environment of the cellular milieu. For macromolecules possessing internal flexibility or large regions of disorder, describing a global architecture is not simply a matter of geometric reconstruction, but rather requires an understanding of how the global macromolecular conformation evolves dynamically over time as it diffuses through solution. In other words, rather than existing as single, fixed

architectures, flexible, modular proteins are more accurately described by a conformational ensemble, populated by multiple, interchanging conformations (*84*, *86*, *87*). A complete understanding of such macromolecular architectures requires knowledge of the relative frequency and distribution of macromolecular conformations within the conformational ensemble.

In recent years, small-angle scattering (SAS) has emerged as a powerful technique for experimentally characterizing conformational ensembles of flexibly modular and disordered proteins (*84*, *88*, *89*). The low resolution spatial information encoded in the scattering curve provides important insight into the global conformations of macromolecules. However, its ability to capture this structural information across multiple states within a conformational ensemble provides an invaluable tool for probing global protein dynamics. This perspective on global macromolecular dynamics can then be integrated with more local information on inter-domain motion and distribution using high-resolution techniques such as NMR. The resulting 'structural dynamics' description of a protein conformational ensemble can then be probed for insight into biological function.

Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for extracting high-resolution structural and dynamic information from macromolecules in native solution environments. From three-dimensional determination of structures, to characterization of macromolecular interactions, to probing global and local biomolecular motions, NMR has granted atomic-level access to fundamental biological processes. The

basic physical principles that underlie this versatile biophysical technique, as well as NMR experiments most relevant to structural biology are described below.

Of all the spectroscopic methodologies, magnetic resonance alone relies upon manipulation of the fundamental quantum mechanical properties of atomic nuclei to provide unique information on their electronic environment (*90*). In this case, the quantum mechanical property of interest is nuclear spin angular momentum. Placement of nuclei with magnetically active spins (where the spin quantum number is I = ½) within a static magnetic field of sufficient strength induces the spin angular momentum vector to rotate about an axis parallel to the magnetic field (Figure 1.7) (*91-94*)[2]. The frequency of this rotational motion (the resonance frequency) depends upon the electronic environment surrounding a given nucleus, whether it forms part of a covalent bond, participates in non-covalent interactions, or is simply proximate to other atoms. Subsequent excitation of these precessing spins by an oscillating electromagnetic (EM) field, which transmits energies corresponding to their resonance frequencies, rotates the spins to the plane perpendicular to the static field (Figure 1.7). As they continue their precession in the transverse plane, a time-dependent signal that encodes the collective rotational frequencies of the spins is collected (Figure 1.7). Once processed by Fourier transformation, this signal provides a 'frequency map' that reflects the unique electronic surroundings of each spin (Figure 1.7).

[2] In the interests of clarity, nuclear magnetic resonance is described from a classical perspective. For a full quantum mechanical treatment of the basic NMR experiment, the reader is referred to (*187*, *188*).

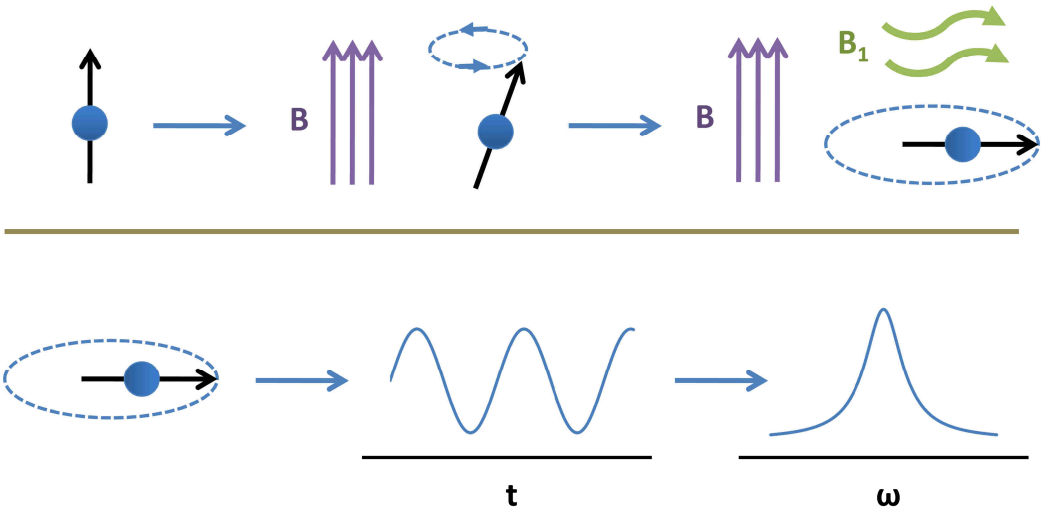


Figure 1.7. Nuclear magnetic resonance spectroscopy. A) Basic outline of the NMR experiment. Nuclear spins, represented by sphere and arrow, undergo precessive motion at a characteristic resonance frequency when placed in a static magnetic field (B). Application of an oscillating magnetic pulse (B_1) rotates the spin vector to the perpendicular plane, where it continues its precessional motion and induces an oscillating voltage signal. Fourier transform of this time-domain signal (t) generates a frequency spectrum (ω) with intensity centered at the spin's resonance frequency.

Without the presence of the excitatory EM field, transverse spin precession gradually returns to the original parallel precession extant prior to excitation. This return of excited spins to their ground energy state is known as 'relaxation' and provides a powerful means for accessing global and local molecular motions, as the timescales of motion influence the rapidity with which spins relax (*93*, *95-98*). This phenomenon will be discussed in more detail in a later section.

From the perspective of magnetic resonance, macromolecules are simply complex assemblies of nuclear spins. One-dimensional ^{1}H NMR exploits the high natural abundance of the ^{1}H isotope in macromolecules; however, the large number of protons present in a macromolecule (compared to a small molecule) often leads to extensive overlap within the resulting 1D ^{1}H spectrum. More advanced 2D (^{1}H, ^{15}N or ^{1}H, ^{13}C) or

3D (^{1}H, ^{13}C, ^{15}N) experiments, which allow greater resolution of these overlapping signals, require samples to be enriched with magnetically active isotopes (spin ½ nuclei -- ^{13}C, ^{15}N, or ^{31}P), which are less abundant in nature than is ^{1}H. Such enrichment is usually achieved by recombinant expression of the macromolecule in bacteria grown in minimal media supplemented with ^{15}N-ammonium chloride or ^{13}C-glucose as sole nitrogen and/or carbon sources (*99*).

The two-dimensional ^{15}N-^{1}H heteronuclear single-quantum coherence (HSQC) experiment has become a key tool for structural studies of proteins by NMR (*100*). For this particular experiment, the resonance frequencies of backbone amide ^{1}H and ^{15}N nuclei for each amino acid are correlated, forming a two-dimensional frequency map with signals corresponding to each ^{15}N-^{1}H pair from each residue, excepting prolines. Because the unique electronic surroundings of each nucleus dictate spin resonance frequencies, the distribution of signals in an HSQC spectrum serves as a unique 'fingerprint' for each protein. HSQC spectra provide valuable information on the foldedness of a protein, the proportion of residues localized to flexible regions of the polypeptide backbone, and the presence of different timescales of motion within different parts of a protein. Use of the HSQC becomes even more powerful when resonance assignments are available for the protein of interest, which correlate each ^{15}N-^{1}H frequency signal with its specific amino acid of origin. Standard assignment experiments and methodologies can be found in (*101*).

More importantly, the HSQC can be used to monitor changes to nuclear electronic environment upon introduction of a binding partner, variation in pH or temperature, or addition of a substrate. By tracking changes to peak position or intensity within the

spectrum, one can map an interaction surface on a protein of interest, calculate binding constants, or determine which regions of a protein are most forcefully impacted by changes in solution conditions. Of specific interest to the study of modular protein systems is the use of the HSQC to gain insight into the presence or absence of inter-domain contacts for an intact protein. This is usually achieved by comparison of signals in spectra acquired on individual modules relative to those of spectra acquired on the intact protein (*46*, *47*). Shifting or broadening of signals in spectra of the intact protein is considered indicative of inter-domain interaction.

Of note, protein mass has a direct bearing on the rapidity of NMR signal relaxation: the larger the protein or domain module, the faster signal relaxation. This effectively imposes a size limitation on systems that can be routinely studied by conventional HSQC (~35 kDa), as signals arising from macromolecules larger than this decay too quickly to be detected efficiently. In the past decade, the development of advanced NMR methodologies that address rapid relaxation has pushed this accessible size regime to 40-50 kDa, and creative experimental application of these techniques to select biological systems has granted insight into systems ranging from hundreds to thousands of kilodaltons (*102-105*).

Approaches to prolonging NMR signal lifetime or neutralizing mechanisms that induce signal relaxation rely on unique sample preparation strategies, specialized pulse sequences, or application of both. Deuterium enrichment (^{2}H) of macromolecular samples permits global reduction of the number of extraneous ^{1}H nuclei, which are potent inducers of signal relaxation (*106*). Since they exchange with solvent, the amide protons of interest are rendered NMR active by back-exchange of ^{2}H for ^{1}H during the course of

sample purification and preparation. Application of transverse relaxation optimized spectroscopy (TROSY) exploits destructive interference between different physical processes driving signal relaxation and selects for correlated ^{15}N-^{1}H signals with extended signal lifetime (*107*). Similarly, cross-correlated relaxation-enhanced polarization transfer (CRINEPT) takes advantage of interference between signal relaxation pathways to create signal correlation between ^{15}N-^{1}H pairs (*108*, *109*). Use of TROSY is most optimal for macromolecular masses ranging from 30-100 kDa; for CRINEPT, the optimal range extends to molecular masses in excess of 150 kDa.

NMR heteronuclear relaxation measurements

In addition to providing access to protein structure and biomolecular interactions, NMR is also an invaluable technique for probing global and local dynamics within macromolecules. Macromolecular motion can range from local fluctuations along the polypeptide backbone or side chains to global rotational diffusion of the macromolecule through solvent. For flexible, modular proteins, internal tumbling and hydrodynamic drag among linked domains is a key constraint driving conformational sampling of global architecture. Characterizing the speed and orientation of inter-domain motion can provide important insights into the conformational space accessible to a biomolecule and the kinds of substrate landscapes likely to exploit this conformational sampling.

As mentioned above, macromolecular motion can be monitored by probing the rate of NMR signal relaxation. This is possible because macromolecular motion causes oscillation of small magnetic fields formed by pairs of spins, which leads to de-excitation (relaxation) of spins that are excited by the applied EM field (*93*, *95*). The rate of signal

decay can then be translated into descriptors of global or local motion, such as the rotational diffusion tensor of a protein or amplitude and time-scale parameters for the fluctuations of amide bonds (*97, 98, 110*).

Relaxation can be monitored for any NMR active nucleus. For proteins, methods for monitoring ^{15}N heteronuclear relaxation are the most advanced, as these target NMR observable amide ^{15}N-^{1}H bonds along the polypeptide backbone and can be incorporated into a standard HSQC experimental framework. Experiments typically measure three types of relaxation: longitudinal or T_1 relaxation, transverse or T_2 relaxation, and cross-relaxation, which is indirectly monitored by measurement of NOEs. Details for each of these types of relaxation, as well as the quantum mechanical mechanisms driving relaxation can be found in associated reviews and references therein (*96-98, 110*).

Encoded within T_1, T_2, and NOE values for each HSQC signal are the relative frequencies of each motion executed or experienced by the associated amino acid residue. The distribution of these motional frequencies is referred to as the residue's spectral density function (SDF, $J(\omega)$) (Figure 1.8A). The spectral density function provides important information on the speed and amplitude of a residue's motion. In general, a high content of low frequencies within the SDF (Figure 1.8A) signals the presence of slow motions (timescales on the order of nanoseconds; protein rotational diffusion occurs on this timescale), while a more even distribution across both low and high frequencies reflects faster motion (timescales on the order of picoseconds; rapid amide bond fluctuations occur on this timescale) (Figure 1.8A).

Importantly, extensive work has gone into deriving mathematical relationships linking the value of the SDF at NMR resonance frequencies that drive nuclear excitation

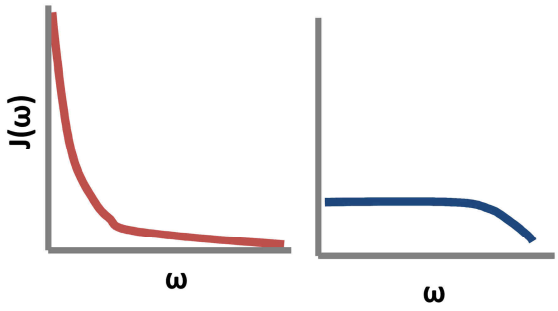
and transitions (i.e. $J(0)$, $J(\omega_N)$, $J(\omega_H)$, $J(\omega_H+\omega_N)$, $J(\omega_H-\omega_N)$) to parameters describing molecular motion, such as the principal components and orientation angles of the diffusion tensor (D_x, D_y, D_z, α, β, γ) (*111*, *112*). In the interest of completeness, equations describing the relationship between T_1, T_2, and NOE values and the spectral density function, as well as that between the spectral density function and isotropic diffusion parameters are provided (Figure 1.8B). In practice, direct calculation of diffusion tensor parameters from T_1, T_2, and NOE values is performed by software packages such as r2r1_diffusion (*111*), ROTDIF (*113*), and relax (*114*, *115*).

Rotational diffusion tensors provide two key pieces of information: (1) the speed of domain rotation and (2) the relative orientation of this rotation (i.e. if the domain is asymmetric in shape, does it rotate primarily about its long or short geometric axis?). In the case of flexible, modular proteins, examination of the rotational diffusion tensor for individual domains is particularly insightful, as similarities and differences among different diffusion tensors speak to the relative degree of inter-domain independence and hydrodynamic drag experienced by each domain module. Furthermore, changes to the rotational diffusion tensor imposed by binding of ligand (ssDNA or protein) can provide insight into architectural remodeling of inter-domain orientation.

In addition to descriptions of global rotational diffusion, relaxation data can also be used to provide insight into local fluctuations among amide ^{15}N-^{1}H bonds along the polypeptide backbone. Because bond motion is complicated and ^{15}N relaxation measurements provide only three observables for characterizing this motion[3], the most

[3] If ^{15}N relaxation measurements are made at more than one field strength, the number of available observables is increased. Even with tens of observables in hand, however, the motion of the system is still undersampled by several orders of magnitude.

A.



B.

$$\frac{1}{T_1}=\frac{d^2}{4}\left[6J(\omega_H+\omega_N)+J(\omega_H-\omega_N)+3J(\omega_N)\right]+c^2J(\omega_N)$$

$$\frac{1}{T^2}=\frac{d^2}{8}\left[6J(\omega_H+\omega_N)+J(\omega_H-\omega_N)+6J(\omega_H)+3J(\omega_N)+J(0)\right]+\frac{c^2}{2}\left[3J(\omega_N)+4J(0)\right]$$

$$\frac{\gamma_N}{\gamma_H}\left(\frac{1}{T_1}\right)(NOE-1)=\frac{d^2}{4}\left[6J(\omega_H+\omega_N)-J(\omega_H-\omega_N)\right]$$

where the dipolar constant, d, and the chemical shift anisotropy constant, c, are given as

$$d=\frac{\mu_0 h\gamma_N\gamma_H}{8\pi^2 r_{NH}^3} \qquad c=\frac{\Delta\sigma_x}{\sqrt{3}}$$

μ_0 – permeability of free space
h – Planck's constant
γ_N – magnetogyric ratio of nitrogen
γ_H – magnetogyric ratio of hydrogen
r_{NH} – length of amide NH bond

$$J(\omega)=\frac{6D_{iso}}{(6D_{iso})^2+\omega^2}$$

D_{iso} – rotational rate of diffusion

Figure 1.8. Investigating biomolecular motion. A) Theoretical spectral density functions for residues experiencing slow rotational diffusion exclusively (left, red) or diffusion coupled with rapid amide bond fluctuations (right, blue). B) Mathematical relationships between (1) rates of heteronuclear relaxtion (T_1, T_2, and NOE) and the spectral density function at proton and nitrogen transition frequencies, and (2) the spectral density function (J(ω)) and isotropic rotational diffusion tensor parameters (D_{iso}). Corresponding equations for anisotropic diffusion tensors can be found in (*111*).

common approach for describing bond motion is the Model-Free (MF) formalism of Lipari and Szabo (*116*, *117*). In this analytical approach, the authors define a generic amplitude term, the order parameter S^2, for describing spatial sampling of amide ^{15}N-^{1}H bond fluctuations. This is complemented by a generic timescale term, the internal correlation time τ_i (as opposed to the global correlation time describing rotational diffusion), for describing the time required for an amide ^{15}N-^{1}**H** bond to perform its spatial sampling. The formalism was eventually expanded to allow for description of certain residues using two motions: a fast motion (S^2_f, t_f) superimposed upon a slow motion (S^2_s, t_s) (*118*). As the slow motion approaches the amplitude and speed of the fast motion, the mathematical description reduces to the expressions for the single-motion framework.

The validity of the MF formalism relies upon complete independence of the global rotational diffusion of a protein and the local fluctuations along its polypeptide backbone (*116*, *117*). This can, unfortunately, present a problem for flexible, modular proteins as this assumption is not necessarily applicable for residues in flexible linkers. Furthermore, the rotational motion of each domain is not a sole function of Brownian diffusion through solvent, but is mutually influenced by the motion of its neighbors, causing time variation of diffusion tensor parameters. Calculation of order parameters and internal correlation times relies on an accurate assessment of the rotational diffusion tensor. The theoretical framework for making these calculations for diffusion tensors with constant rotational speed and orientation is well developed. However, a framework is lacking for performing the equivalent analysis in the context of speed variation and time-averaged asymmetry from the rotational diffusion tensor. A recently published

study addressing a two-site jump model for isotropic diffusion suggests that the field is beginning to address these issues (*119*). For now, MF analysis has the potential to be informative, but should not be the sole approach for gaining insight into local motion of flexible, modular proteins.

An alternative approach to characterizing local motion is spectral density mapping, where values of the spectral density function, J(ω), are calculated at NMR observable frequencies (0, ω_N, ω_H, $\omega_H+\omega_N$, $\omega_H-\omega_N$) for each amino acid residue and compared across the polypeptide backbone (*120, 121*). For relaxation measurements made at a single field strength, allowing for only three experimental observables, reduced spectral density mapping is employed (*122*), which requires the approximation $J(\omega_H) \sim J(\omega_H+\omega_N) \sim J(\omega_H-\omega_N)$. Examination of $J(\omega_H)$ values can broadly identify regions of the polypeptide backbone experiencing increased dynamic fluctuation at high frequencies.

Small-angle x-ray scattering

Over the past decade, small-angle x-ray scattering (SAXS) has proved an increasingly valuable tool for accessing global information on the architectural organization of macromolecules and their assemblies in a solution environment. Specific properties detected by x-ray scattering include macromolecular size and shape; the presence of foldedness, disorder, or aggregation; states of oligomerization; and conformational dynamics (*83*). Unlike diffraction from an ordered crystal matrix, scattering profiles are averaged across all molecular orientations present at random in solution. This has the effect of reducing the effective resolution of the spatial information encoded in the scattering curve, providing insight into the global molecular

envelope rather than details of internal structure. Even so, integrating a low resolution perspective of global architecture with high resolution domain or subunit structures can provide a comprehensive understanding of macromolecular assembly.

Acquisition of SAXS data is similar to that for x-ray diffraction (*123*). A monodisperse solution of the macromolecule or complex of interest is exposed to a high flux, coherent x-ray beam, which scatters from electron centers distributed throughout the macromolecule (Figure 1.9A). As the scattered x-rays travel forward to the detector, they undergo constructive and destructive interference to form a unique intensity pattern. Unlike the diffraction pattern from ordered crystal planes, though, scattering intensity is radially symmetric about the detector center, a consequence of random macromolecular orientation in solution, and decreases rapidly for scattering angles beyond 1° (or momentum transfer values beyond 0.3 $Å^{-1}$) (*83*, *84*, *88*). As a result, radial integration of the two-dimensional scattering pattern at discrete angular intervals from the detector center generates a macromolecule's scattering curve, I(q). Here, I(q) represents the integrated intensity for a given momentum transfer value, q, which is related to the corresponding scattering angle, 2θ, by $q = 4\pi \sin(\theta) / \lambda$, where λ is the corresponding x-ray wavelength (Figure 1.9A). For small-angle x-ray scattering studies reported in this research, scattering angles range from 0.1 to 1.1 degrees (q values of 0.01 to 0.322 $Å^{-1}$).

First-order analysis of the scattering curve, I(q), can provide important information on molecular size, aggregation state, and internal flexibility. Guinier transformation of I(q) (Figure 1.9B) allows calculation of the particle radius of gyration (R_g) within the low-q Guinier region of the curve (defined as $q*R_g < 1.3$). Visual inspection of the Guinier region can also reveal the presence of aggregation, as

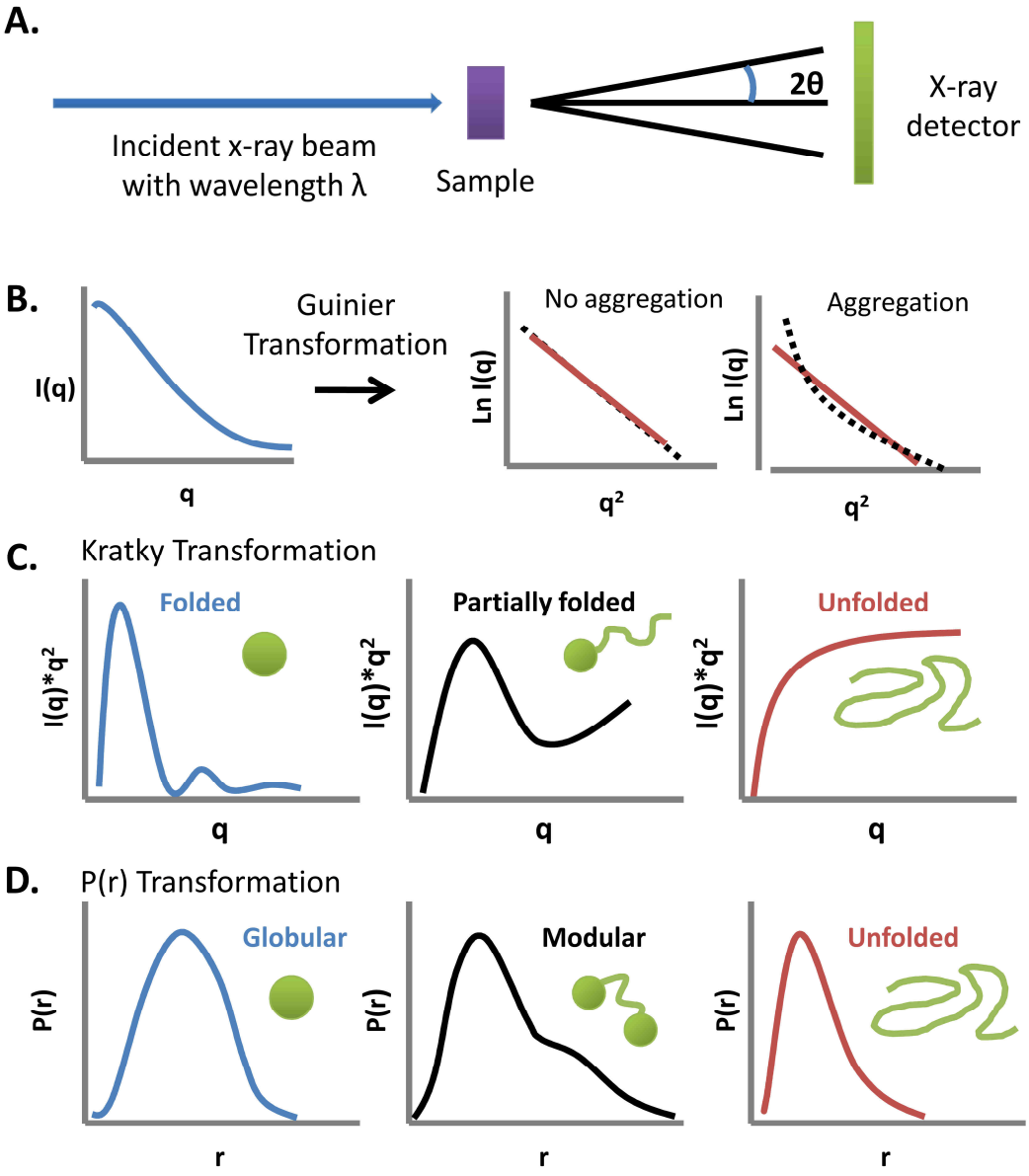


Figure 1.9. Small-angle x-ray scattering. A) Schematic of SAXS experimental set-up. B) Detection of aggregation by Guinier analysis. C) Assessment for internal flexibility and disorder using Kratky analysis. D) Pair distance distribution functions representative of different protein architectures. Adpated from Putnam, et al (*83*).

determined by non-linear behavior at low q values (Figure 1.9B). Kratky transformation of the scattering curve provides a visual assessment of internal flexibility within the

particle (Figure 1.9C). Particles with well-defined globular architecture will exhibit a parabolic peak that converges to baseline within the high-q region, while intrinsically disordered molecules will display a hyperbolic curve with non-convergent behavior at high q values. Macromolecules composed of a mixture of structured and flexible segments will demonstrate features from both scenarios: a well-defined parabolic peak that is upturned at high q values (Figure 1.9C) (*83*)

Direct insight into domain or subunit orientation, distance separation, and conformational mobility can be obtained by Fourier transformation of the scattering curve into real space to produce the pair-distribution function, P(r). Pair-distribution functions report on the frequency of inter-atomic distances within a macromolecule (*83, 88*). For compact, globular molecules, the P(r) function is manifest as a Gaussian-like distribution about the radius of gyration (Figure 1.9D). For architectures which deviate from spherical symmetry or possess multiple, flexibly linked components, this symmetric density of distances can vary to exhibit shoulders off the primary maximum, separate secondary maxima, or skewed asymmetry (Figure 1.9D) (*83*). Calculation of the P(r) function also provides an estimate of D_{max}, the maximum geometric distance present in the macromolecule or its conformational ensemble, in addition to a secondary estimate of R_g (*83, 88*). In the case of flexible, modular proteins, inter-domain dynamics results in a smoothing of secondary inter-domain distance peaks within the P(r) function and a tendency for the distribution to skew toward longer distances (Figure 1.9D) (*124*).

In recent years, methods for *ab initio* reconstruction of low resolution molecular envelopes directly from scattering curves have become computationally tractable (*125, 126*). Not only do such envelopes provide direct visualization of molecular shape, but

they also supply a three-dimensional framework for the docking of high resolution structures, if available (*127*). It is important to remember, however, that molecular envelopes represent the *averaged* molecular shape in solution. As with the P(r) distribution, heterogeneity among architectures or the presence of dynamics within a molecular population will cause smoothing of the molecular envelope (*124*). Furthermore, molecular envelope reconstruction is unable to distinguish the 'handedness' or shape chirality for asymmetric architectures, which can lead to ambiguous interpretation without complementary structural information. Even in these instances, though, molecular envelopes can provide at least first-order insights into the distribution of molecular domain architecture or changes to this architecture upon ligand binding (*84*, *88*).

Over the past five years, great strides have been made in extracting dynamics information from SAXS data by rigid-body modeling of macromolecular conformational ensembles (*89*, *128*). These techniques are particularly relevant for molecules possessing internal linkers or regions of disorder. The basic strategy involves initial simulation of a conformational population, usually 5,000 – 10,000 models, using rigid-body modeling methods that rely upon molecular dynamics calculations (Bilbo-MD) (*128*) or knowledge-based potentials (EOM) (*89*) to generate linker conformations. Theoretical scattering curves are calculated for each model and compared to the experimental scattering data. Random subsets of conformers are then selected and assessed for their ability to recapitulate the experimental scattering curve upon averaging. Model selection and subset organization are iteratively repeated and refined to yield a population of 'representative' conformers, whose collective theoretical scattering profile best matches

the experimental data. While ensemble selection methods can provide powerful insight into the most populated architectures of a macromolecular population, it is important to consider that the ensembles of flexibly linked and disordered molecules are a broad continuum of architectural states, rather than a limited series of discretely exchanging states. Advances in the theoretical framework describing these macromolecular systems and the ability to simulate them computationally will be required to extract a more complete description of conformational ensembles from SAXS data.

Small-angle neutron scattering

The advantages of high x-ray beam flux, and thus increased signal sensitivity, make SAXS the scattering methodology of choice for the majority of biomolecular systems. There are many instances, however, where high quality SAXS data cannot distinguish between competing models for the organization of multi-component assemblies within a molecular envelope. Small-angle neutron scattering (SANS) offers an alternative for dissecting the global scattering envelopes of multi-component assemblies through the use of contrast matching and contrast variation.

Where x-rays scatter from the electron clouds of atoms, neutrons scatter from atomic nuclei. In both cases, the strength of scattering is dependent upon the contrast between macromolecule and solvent (Figure 1.10A) (*83*). For x-rays, larger, compact electron-rich macromolecules are more distinct relative to the electron-diffuse solvent, and thus scatter more strongly. For neutrons, however, scattering contrast does not scale with the total number of atomic nuclei. Instead, neutrons scatter strongly only from certain kinds of atomic nuclei, among them ^{1}H and ^{2}H, which exhibit very distinct

scattering densities[5]. As such, neutron contrast for a protonated biomolecule in deuterated solvent is maximized due to the large difference in scattering density between ^{1}H and ^{2}H, producing a robust scattering signal. Conversely, neutron contrast is reduced (but not eliminated) when the same protonated biomolecule is studied in a protonated environment (*129*).

Equal scattering density between sample and solvent is known as a macromolecule's match point, where scattering signal from the macromolecule is indistinguishable from that of the solvent due to the lack of contrast. Notably, the scattering density of a solvent can be altered by adjusting its deuterium content; and matched solvent ^{1}H:^{2}H ratios have been calculated for the major classes of biomacromolecules with and without deuterium enrichment (protein, lipids, nucleic acids) (*129*). Contrast matching experiments exploit these differences in match points by acquisition of SANS data in a solvent matched to one component of a macromolecular complex (or vice versa) (Figure 1.10B) (*130*). The resultant scattering signal encodes spatial information exclusively from the non-matched portion of the complex in the context of its fully assembled state.

Contrast matching is optimally useful when shape information is desired for a single component of a complex (and when the match point of the component has been well established) (*130*). If shape information is sought for two or more components in the assembled state, contrast variation experiments can access this without the need for multiple sample preparations to target each component (*130*). With contrast variation, multiple neutron scattering profiles are acquired for the same sample across a range of

[5] The physical property that governs the strength of the interaction between neutrons and atomic nuclei is known as the coherent neutron scattering length density (SLD). Further description of the physics underlying neutron scattering can be found in (*189*).

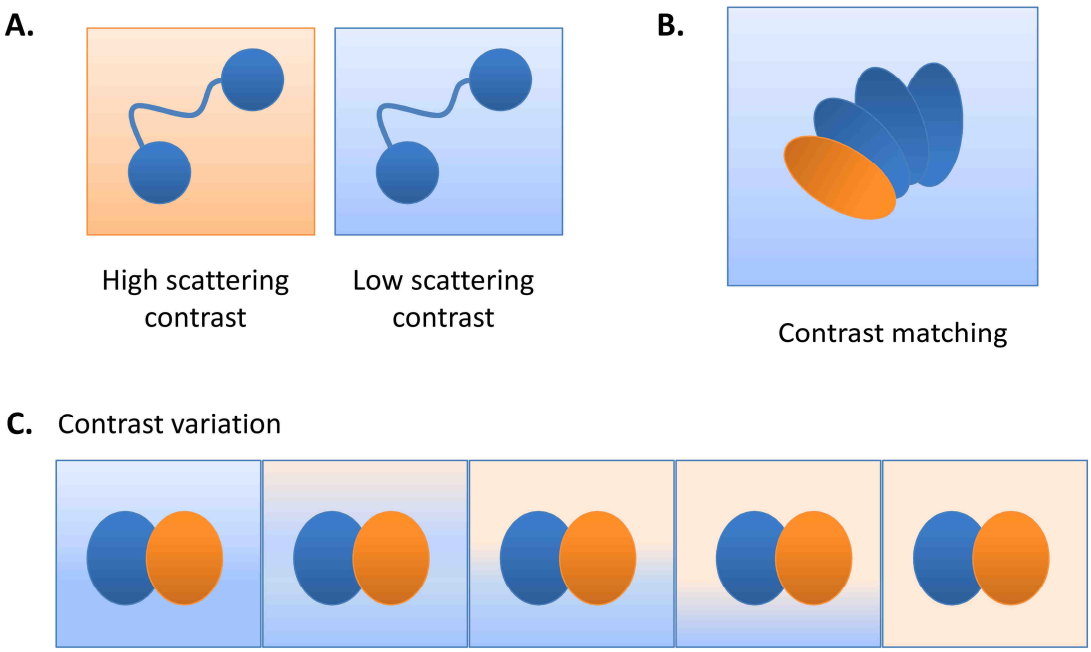


Figure 1.10. Small-angle neutron scattering. A) Schematic illustrating neutron contrast. Neutron contrast is maximized between protein and solvent when their respective scattering length densities differ as for a protonated protein in deutrated solvent (left). When protein and solvent have similar neutron scattering length densities (right), scattering contrast and the strength of the scattering signal are reduced. B) Contrast matching. Selective deuteration of one component and solvent matching of the remaining complex allows scattering detection exclusively from the deuterated component. C) Contrast variation. Neutron scattering profiles are acquired for the same sample across a range of H_2O:D_2O solvent mixtures. Individual scattering profiles for each component are subsequently deconvoluted from this series (*130*).

H_2O:D_2O solvent mixtures (typically 0%, 10%, 20%, 30%, 80%, and 100% D_2O) (Figure 1.10C). The match points of the complex and its components are extracted by comparison of sample scattering intensity across the contrast series, permitting deconvolution of the scattering function for each component of the complex in the assembled state (*131*). Conventional small-angle scattering analysis (Guinier, Kratky, and P(r) transformation) can proceed on SANS scattering curves for each assembled component, as well as the global scattering curve for the entire assembled complex.

More powerful still, scattering information from SANS can be combined with higher sensitivity SAXS data to guide *ab initio* reconstruction of the assembled complex with more accuracy (*132*).

From a practical perspective, SANS sample preparation and data acquisition present a number of technical challenges relative to SAXS. The relatively low flux of neutron beams requires 10-20 times the sample amount needed for SAXS and also demands greater sample stability at higher concentrations, as data collection can take hours, as opposed to the seconds needed for x-rays (*130*). The physics of the neutron beam entails measurement of additional calibration experiments, as well as greater mathematical intervention to ensure accurate reduction of the two-dimensional scattering pattern to a one-dimensional scattering curve (*133*). Despite these additional complications, with the appropriate expertise available SANS can provide unparalleled access to the explicit organization of macromolecular assemblies.

Experimental Overview

In an effort to advance a broader understanding of RPA function and its role in DNA processing, this dissertation project has probed the solution arrangement of modular domains within full-length RPA and the structural rearrangement of this inter-domain architecture as RPA negotiates the three interaction states of its DNA-binding trajectory. Chapter II addresses the conflicting views on RPA architecture through NMR studies on intact, $^{2}H,^{15}N$-enriched RPA and a number of multi-domain fragments. Correspondence between TROSY-HSQC spectra from RPA fragments and the spectrum acquired on full-length RPA indicate an absence of inter-domain contacts and favor a model for flexible

independence of RPA domains. This dynamic flexibility is maintained upon DNA binding for protein interaction domains 70N and 32C, but is lost for principal DNA-binding domains 70A and 70B as they engage the DNA substrate.

Chapter III describes a detailed biophysical study of internal inter-domain dynamics utilizing NMR ^{15}N relaxation measurements on select RPA tandem domain fragments and tracks how inter-domain motion is altered by DNA binding. These studies were initiated with the expertise and guidance of Dr. Kavita Dorai, a visiting professor from the Indian Institute of Science Education and Research in Mohali, India, and technical input from Dr. Markus Voehler of Vanderbilt's Biomolecular NMR Facility. Diffusion tensor analysis of relaxation data from the tandem RPA70AB fragment reveals a semi-independent rotational motion for each domain that is jointly correlated upon binding ssDNA substrate. A similar analysis on the multi-domain RPA70NAB construct reveals differential rotational motion for the closely linked 70A and 70B domains relative to the distantly connected 70N domain, highlighting the role of linker length in determining the unique solution behavior of domains executing distinct functions.

Chapter IV presents a global characterization of the DNA-binding core of RPA (RPA-DBC), employing SAXS to investigate architectural changes in each DNA-binding state and to track remodeling of the DNA binding cleft. SAXS sample preparation and data collection was carried out in collaboration with Dr. Susan Tsutakawa of the laboratory of Dr. John Tainer at the SIBYLS beamline (BL12.3.1) of the Advanced Light Source at Lawrence Berkeley National Laboratory. Dr. Tsutakawa also generously provided guidance in use of the software applications for the subsequent analysis of the scattering experiments. A series of SANS contrast variation experiments was also

collected on select complexes of RPA-DBC and ssDNA. While these experiments are not reported in this manuscript, they will be included in the published version of this work. SANS experiments were performed in collaboration with Dr. William Heller at the High Flux Isotope Reactor at Oak Ridge National Laboratory. Dr. Heller also provided invaluable assistance in navigating the reduction and analysis of SANS data, a new enterprise for this laboratory. The scattering results highlight the presence of inter-domain dynamics within the RPA DNA-binding core in the absence of DNA. RPA-DBC retains a portion of this flexibility as it undergoes a progressive compaction through the initial 10-nucleotide and intermediate 20-nucleotide DNA binding modes. The compaction is reversed, however, in the final 30-nucleotide binding mode, where the architecture is similar, but not identical, to that seen for the DNA-free state.

The broader implications of this research for RPA function and DNA processing are summarized in Chapter V. An integrated model for RPA solution architecture and its DNA-binding trajectory is proposed and evaluated in the context of the most recent literature on DNA processing. Future refinement of this view of RPA structural dynamics and a roadmap for probing the impact of protein interaction upon RPA architectural remodeling are outlined.

CHAPTER II

NMR ANALYSIS OF THE ARCHITECTURE AND FUNCTIONAL REMODELING OF A MODULAR MULTI-DOMAIN PROTEIN, RPA.[6]

The progression of DNA replication and repair requires the coordinated action of dynamic, multi-protein assemblies. We have previously proposed a critical role for proteins composed of multiple, flexibly attached domains in facilitating the action of these dynamic complexes (*24*). Because these proteins can undergo intra- and inter-domain rearrangements, they are able to interact optimally with the ever-changing substrate landscape present during DNA processing. RPA is a prototypical modular multi-domain DNA processing protein with flexible linkers of various lengths (Figure 2.1). The trimer core is a compact assembly of three OB-fold domains (RPA70C/32D/14) to which is appended the disordered RPA32N functional domain, the RPA32C winged-helix domain, and the tandem RPA70AB and the RPA70N OB-fold domains. Despite a wealth of information available on the structure and function of these domains, very few insights have been obtained about the architecture of intact RPA (*13*, *14*).

NMR spectroscopy in solution is a powerful tool for characterizing proteins under conditions that preserve intrinsic dynamic properties. The advent of TROSY, CRINEPT and related experimental approaches (*134*) has vastly increased the upper limit of

---

[6] This work has been published as a communication under Brosey, C.A., Chagot, M.-E., Ehrhardt, M., Pretto, D.I., Weiner, B.E., Chazin, W.J. (2009) NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA, *Journal of the American Chemical Society*, *131*, 6346-6347 (142).

molecular masses accessible to study by NMR. Examples range from the globular malate synthase (82 kDa) to the oligomeric GroEL-GroES complex (872 kDWa) to highly flexible domains from the ribosome (>2.5 MDa) (*102-105*). In the case of RPA (116 kDa) and many other multi-domain proteins, modularity and inter-domain flexibility are the critical properties that enable characterization of dynamic architectures by NMR.

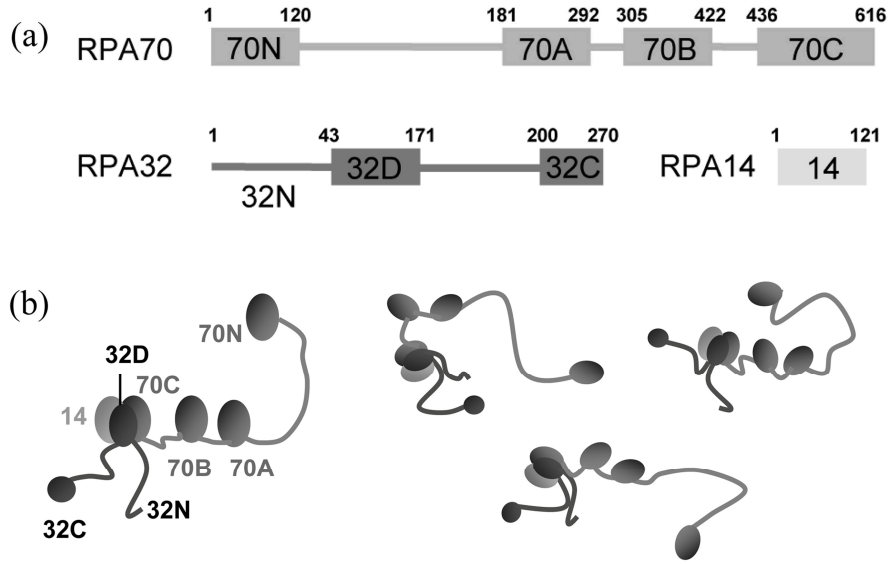


Figure 2.1. A) Domain organization of RPA. All domains are OB-folds, except for RPA32C, which is a winged-helix domain, and the disordered functional domain RPA32N. B) Illustration of 'structural dynamics' and time-dependent RPA architecture.

To illustrate the analytical framework, results are presented first for RPA70NAB (M_r 45.8 kDa), which has an asymmetric arrangement with a 70-residue N-A linker and a 10-residue A-B linker (Figure 2.1). The ^{15}N-^{1}H TROSY-HSQC spectrum of ^{15}N-enriched RPA70NAB reveals the presence of over 370 of the 400 expected signals from 422 residues (Figure 2.2A). The signals from each of the three domains appear in positions remarkably similar to those in NMR spectra of the three isolated domains (Figure 2.3). Thus, all three domains are structurally independent and resonance

assignments can be transferred from the isolated domains to RPA70NAB (*46*). NMR is highly sensitive to differences in the degree of inter-domain flexibility; the signals from the A and B domains are substantially weaker than the signals from the N domain, even though all three domains are approximately the same mass (Figure 2.2A). The differences arise from the fact that although the A and B domains are structurally independent, the short A-B tether partially restricts their motions, whereas the much longer N-A tether enables the N domain to tumble essentially freely in solution. The

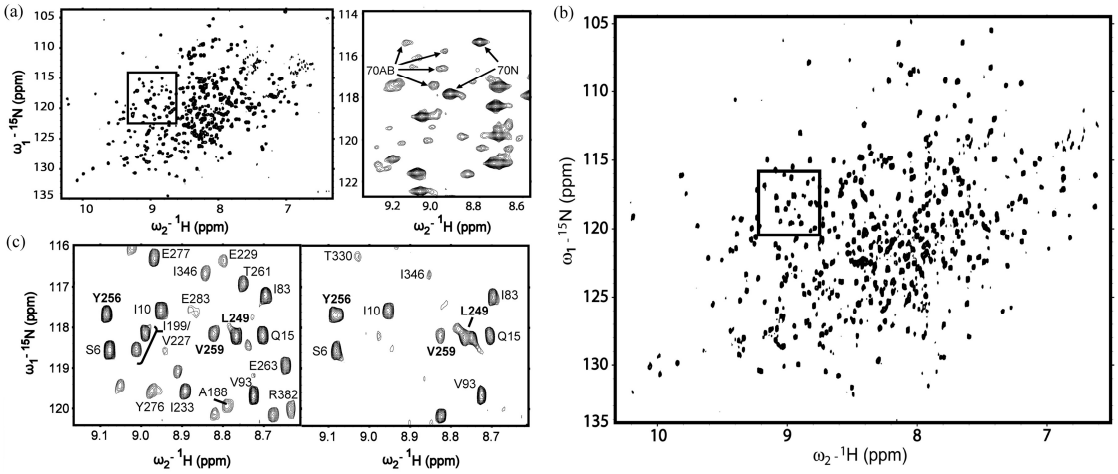


Figure 2.2. NMR Analysis of RPA. (a) ^{15}N-^{1}H TROSY-HSQC of ^{15}N-RPA70NAB recorded at 800 MHz, 25 ºC, and pH 6.0 (right) with expansion of the boxed region (left). (b) ^{15}N-^{1}H TROSY-HSQC of ^{2}H,^{15}N-RPA recorded at 800 MHz , 25 ºC, and pH 7.5. (c) Expanded views of the spectrum in (b) in the absence (left) and presence (right) of dT_{30}. Signals from RPA70 and RPA32 (bold) are labeled.

coupling of the A and B domains by the short tether is a critical factor for the ability to bind ssDNA with high affinity (*46*).

The same approach was applied to the analysis of full-length RPA. Figure 2.2B shows the remarkably high quality ^{15}N-^{1}H TROSY-HSQC spectrum of ^{2}H,^{15}N-enriched RPA. Over 350 of the 550 expected signals from the RPA70N, RPA70A, RPA70B, RP32C and RPA32N domains were identified in this spectrum. As in the case of

RPA70NAB, the signals appear in nearly identical positions as in the NMR spectra of the isolated domains, indicating the domains are structurally independent and enabling the transfer of resonance assignments directly to the intact protein (Figure 2.2C). Moreover, a clear hierarchy in the inter-domain dynamics was evident. The signals from RPA32N and RPA32C were very strong, indicating that these domains are nearly as flexible in the trimer as when isolated on their own. Lower intensity is observed for the signals from RPA70 N, A, and B domains, with N signals stronger than A and B as seen for RPA70NAB. In contrast, no signals were identified for the RPA70C/RPA32D/RPA14 trimer core. Although it has a relatively large mass (M_r 49.1 kDa), the trimer core on its own gives excellent spectra (Figure 2.4). The absence of signals in the intact protein is therefore attributable to slowing of its rate of tumbling due to the drag caused by the attachment of the five other domains. The ability to simultaneously probe five domains without interference from the trimer core in the TROSY-HSQC spectrum demonstrates the value of the dynamic hierarchy of different NMR experiments.

Having established a basis for analyzing RPA architecture, investigations were undertaken to characterize the remodeling of RPA structural dynamics upon binding ssDNA. Figure 2.2C shows a comparison of a region from ^{15}N-^{1}H TROSY-HSQC spectra of ^{2}H,^{15}N-enriched RPA obtained in the absence and presence of dT_{30}, which revealed three important observations. First, ssDNA has essentially no effect on RPA70N or RPA32C, showing directly that they play no role in the binding of ssDNA

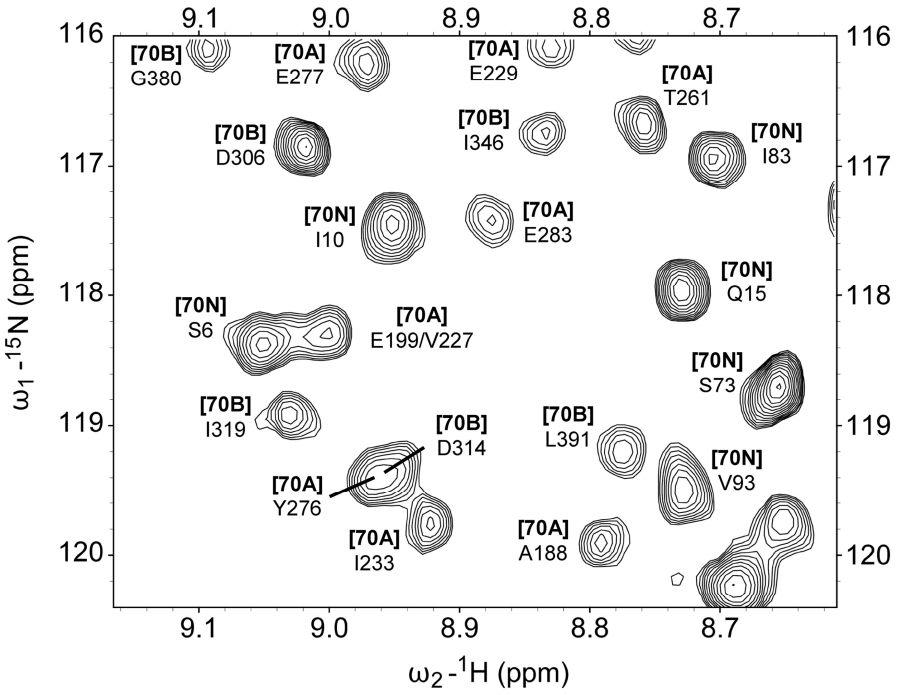


Figure 2.3. Identification of signals from individual domains within ^{15}N-RPA70NAB. Expanded view of ^{15}N-^{1}H TROSY-HSQC from ^{15}N-RPA70NAB recorded at 800 MHz, 25°C, and pH 6.0. Signals mapped to individual domains are labeled accordingly.

and remain available for functioning in the recruitment of other DNA processing factors. Second, the changes in signals of RPA70A and RPA70B upon binding of ssDNA were very similar to those observed when ssDNA is titrated into isolated RPA70AB. In addition, the signals from the A and B domains are seen to broaden upon binding of ssDNA, consistent with a tighter association of the tandem high affinity domains with the trimer core, which slows their rate of tumbling and increases the rate of relaxation. These results represent the first direct observation of the remodeling of RPA structural dynamics upon binding ssDNA and reflect DNA-induced alignment of RPA70AB with the trimer core.

The third important observation was that binding of ssDNA caused changes in the NMR signals of RPA32N, which also reflected remodeling of RPA. Comparison of NMR spectra for the intact protein and the isolated RPA32N domain revealed offsets

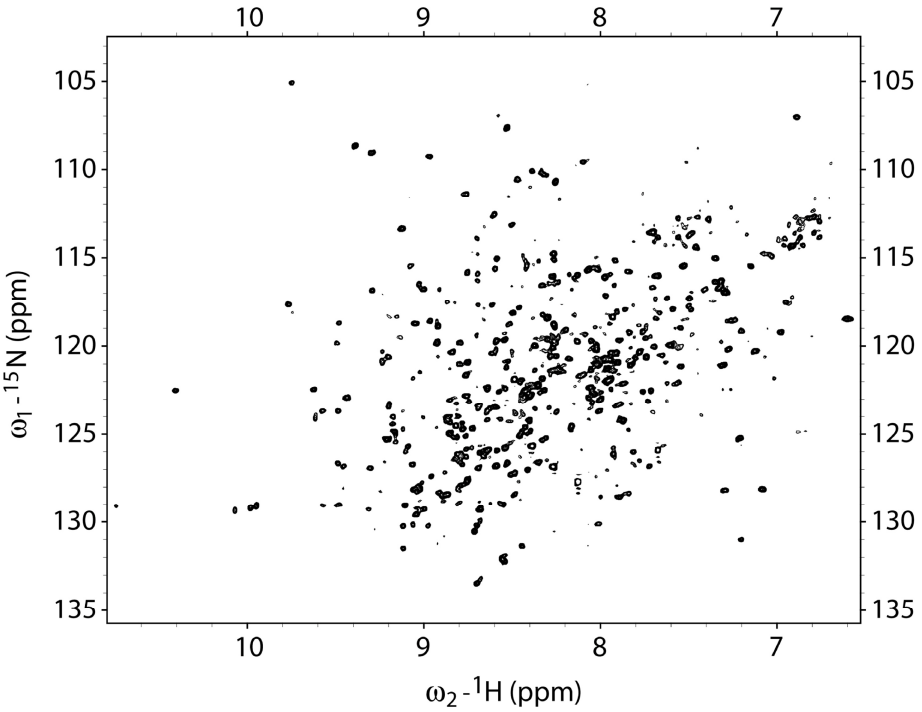


Figure 2.4. ^{15}N-^{1}H TROSY-HSQC of ^{2}H,^{15}N-RPA70C/32D/14 recorded at 800 MHz, 25 ºC, and pH 7.5.

in the intact protein until ssDNA is added (Figure 2.5). Since the 32N domain is the primary site for RPA phosphorylation, the transient interactions of RPA32N may explain why RPA actively involved in DNA processing (i.e. DNA-bound) can be efficiently phosphorylated by ambient cell-cycle machinery or DNA damage transducers. This hypothesis is consistent with previous reports that RPA associated with ssDNA is more accessible to kinase activity than the free protein (*78*, *135*, *136*). The change in availability of RPA32N may also help explain how damage-dependent phosphorylation

of RPA32N participates in redirecting processing of the DNA substrate from replication to repair (*137*, *138*).

The analysis of full-length RPA shows NMR can serve as an effective tool for evaluating the structural dynamics of challenging multi-domain proteins. While many obstacles remain to understanding the intricate choreography of DNA processing, we have demonstrated that NMR can contribute insight into the structural dynamics of the corresponding macromolecular machinery.

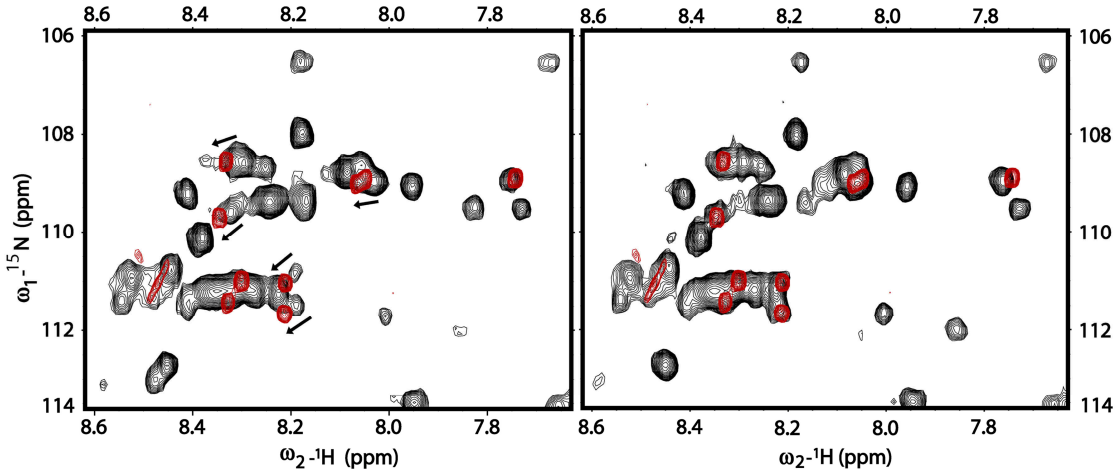


Figure 2.5. Release of RPA32N upon ssDNA binding. Overlay of ^{15}N-^{1}H TROSY-HSQC spectrum of free RPA32N (red) and intact ^{15}N-RPA (black) in the absence (left) and presence (right) of dT_{30}. Arrows indicate signals affected by ssDNA-binding.

Materials and Methods

Plasmid construction

Residues 1-422 of RPA70 were cloned into the pBG100 in-house expression vector (L. Mizoue, Center for Structural Biology, Vanderbilt University), introducing an H3C protease-cleavable N-terminal 6X-His tag. Residues 1-46 of RPA32 were cloned

into the pLM302 in-house expression vector, introducing an H3C protease-cleavable N-terminal 6XHis-MBP tag.

Protein production

All unlabeled RPA constructs were overexpressed in BL21(DE3) or Rosetta (DE3) cells at 37 °C and purified by Ni-NTA affinity chromatography (Sigma) using standard procedures. Fusion tags were left intact for all constructs, excepting RPA32N, whose His-MBP fusion tag was removed by H3C protease cleavage and re-passage over Ni-NTA resin. Full-length RPA heterotrimers were expressed overnight at 18ºC in BL21(DE3) p or BL21 Star cells and purified by Ni-NTA affinity chromatography, followed by heparin purification (GE Healthcare). Uniformly enriched ^{15}N and $^{2}H,^{15}N$ samples were prepared by growing bacteria overnight at room temperature (18ºC for full-length RPA) in a minimal medium containing 0.5 g/L $^{15}NH_4Cl$ (CIL, Inc.) in H_2O or $^{2}H_2O$, respectively.

NMR spectroscopy

NMR samples were concentrated to 0.1-0.2 mM in a buffer containing 30 mM HEPES, 100 mM NaCl, 5 mM βME, 10 μM $ZnCl_2$, and 200 mM arginine, pH 7.5. Spectra of full-length RPA and RPA70NAB were also acquired at pH 6.0 (30 mM NaCitrate substituted for HEPES) to reduce the effects of amide proton exchange in linker regions. Experiments were performed at 25 ºC using a Bruker *AVANCE* 800 NMR spectrometer equipped with a cryoprobe. Gradient–enhanced $^{15}N–^{1}H$ HSQC and TROSY–HSQC spectra were recorded with 2K complex points in the ^{1}H and 128

complex points in the ^{15}N dimension. NMR data were processed and analyzed using NMRPipe (*139*) and SPARKY v3.1 (*140*), respectively. Sequence specific assignments for RPA70NAB were obtained using published assignments for RPA70N and RPA70AB (*45*, *141*). Resonance assignments for RPA32C in intact RPA were made by direct transfer from isolated RPA32C (*48*). RPA32N signals were identified in the full-length protein by comparison with ^{15}N–^{1}H HSQC spectra from the isolated domain.

ssDNA titrations by NMR

Desalted ssDNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA) and used without further purification. Titrations were carried out by direct addition of stoichiometric dT_{30} to NMR samples. Sample dilution due to titration was less than 5%, and sample pH was confirmed after each experiment.

Acknowledgements

We gratefully acknowledge Markus Voehler, Alexey Bochkarev, Cheryl Arrowsmith and Ellen Fanning for valuable reagents and useful discussions, and support from the NIH (RO1 GM065484, T32 GM008320, PO1 CA092584, P30 CA068485, P50 ES000267).

CHAPTER III

STUDIES OF INTER-DOMAIN REMODELING WITHIN RPA70AB AND RPA70NAB UPON BINDING ssDNA BY NMR ^{15}N-RELAXATION

Introduction

As a central component of DNA processing assemblies, the eukaryotic ssDNA binding protein, Replication Protein A (RPA) utilizes a dynamic inter-domain organization to protect and organize ssDNA and coordinate a host of DNA processing factors requiring access to ssDNA substrates, enabling efficient adaptation to the changing substrate environments associated with genome maintenance and repair (*14*). A heterotrimer (M_r 70, 32, and 14 kDa), RPA contains a total of seven globular domains: three oligonucleotide/oligosaccharide binding (OB)-folds that associate non-covalently to form the trimeric core of the protein (70C/32D/14), an additional three flexibly linked OB-folds which form the N-terminus of RPA70 (70B, 70A, and 70N), and a flexibly linked winged helix domain at the C-terminus of RPA32 (32C) (*14*) (Figure 3.1). An eight region, the ~40-residue disordered N-terminus of RPA32 (32N), functions in cell-cycle and damage-dependent phosphorylation (*137*, *138*).

Binding of ssDNA is localized to the four central domains of RPA – 70A, 70B, 70C, and 32D – which contact ssDNA non-specifically from 5' to 3' (*28*, *52*). While the binding affinity of each domain is relatively weak (~10^{-6} M), the effect of close spatial proximity entailed by short linkers (10 and 15 residues for 70AB and 70BC, respectively) results in a collective binding affinity of ~10^{-9} M for the modular DNA-binding site (*46*,

56). Recruitment of other DNA processing proteins is facilitated primarily by domains 70N and 32C, though protein interaction with the principal DNA-binding domains 70A and 70B has also been described (*14*).

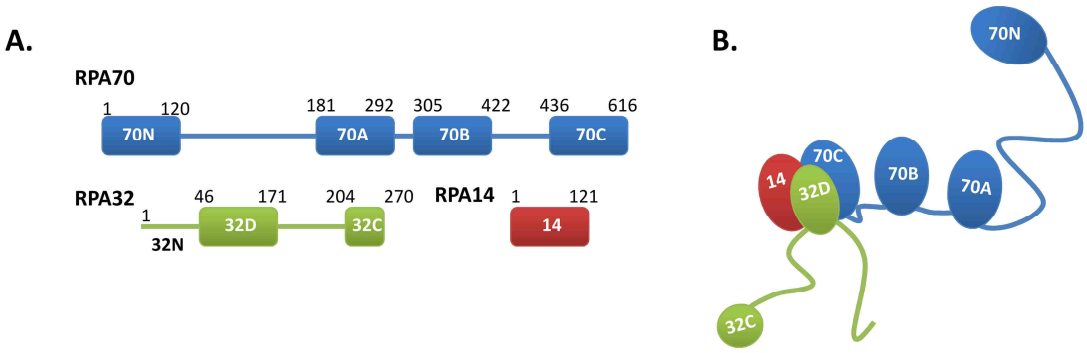


Figure 3.1. RPA domain organization and binding of ssDNA. A) Domain organization of RPA. B) Cartoon schematic of modular domain distribution in RPA. RPA70 (blue), RPA32 (green), RPA14 (red).

NMR studies of ^{2}H, ^{15}N-enriched full-length RPA in the absence and presence of ssDNA have revealed the existence of a dynamic, independent inter-domain architecture with little contact between globular domains (*142*). Subsequent studies of tandem domain fragments of the principal DNA-binding domains 70A and 70B and protein interaction domain 70N by SAXS have further identified the presence of multiple inter-domain orientations both in the absence and presence of ssDNA substrates (*143*). While these studies have demonstrated unequivocally that multiple inter-domain orientations exist for RPA in solution, the average spatial disposition or orientation of these inter-domain arrangements remains unknown.

To gain insight into the average inter-domain orientations experienced by RPA and extend the current understanding of RPA architecture, ^{15}N-relaxation parameters (T_1,

T_2, and NOE) have been measured on tandem domain fragments RPA70AB and RPA70NAB alone and RPA70AB bound to dT_{10} oligonucleotide. Derivation of individual rotational diffusion tensors for each domain from these relaxation parameters reveals semi-independent rotational motion for 70A and 70B that becomes largely correlated upon binding ssDNA. In comparison, rotational diffusion of the distantly linked 70N domain remains relatively unrestricted, allowing extensive orientational sampling. This differential rotational mobility correlates with and reinforces the corresponding functional needs of each domain, i.e. high affinity binding of ssDNA and interaction with other protein partners.

Materials and Methods

Materials

Plasmids for RPA70AB (pSV281) and RPA70NAB (pBG100) have been described previously (*46*, *142*). Both constructs contain N-terminal 6X-histidine fusion tags that are cleavable by TEV (RPA70AB) or H3C (RPA70NAB) proteases. TEV and H3C proteases were produced in-house. ssDNA substrates (dT_{10}) were purchased from Integrated DNA Technologies (IDT) with standard desalting purification and resuspended in sterile water.

Expression and purification of recombinant ^{15}N-RPA70AB and ^{15}N-RPA70NAB

Uniformly enriched ^{15}N-RPA70AB or ^{15}N-RPA70NAB were prepared by growing Rosetta(DE3) cells (Novagen, EMD Chemicals) at 37°C to OD_{600} 0.5-0.6 and expressing overnight at 25°C in a minimal medium containing 0.5 g/L $^{15}NH_4Cl$ (CIL,

Inc.). ^{15}N-enriched RPA70AB or RPA70NAB were purified using standard nickel affinity chromatography (Ni Sepharose 6 Fast Flow, Sigma) in a buffer containing 20 mM HEPES-KOH (pH 7.5), 500 mM NaCl, 5 mM βME, 10 μM $ZnCl_2$ with a linear elution gradient of 30-500 mM imidazole. To remove histidine fusion tags, relevant fractions were pooled and incubated with either TEV protease (RPA70AB) or H3C protease (RPA70NAB) during overnight dialysis at 4°C into 20 mM HEPES-KOH (pH 7.5), 5 mM βME, 10 μM $ZnCl_2$ with 100-200 mM NaCl. The protein was repassed over Ni affinity resin to remove free fusion tag and protease, then concentrated prior to loading on a HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated in 30 mM NaCitrate (pH 6.0), 100 mM NaCl, 5 mM βME, 10 μM $ZnCl_2$, 200 mM arginine. Relevant fractions were pooled, flash frozen in liquid nitrogen, and stored at -80°C.

Preparation of NMR samples

Proteins stocks of ^{15}N-RPA70AB or ^{15}N-RPA70NAB were thawed on ice and dialyzed into 30 mM NaCitrate, 100 mM NaCl, 5 mM βME, 10 μM $ZnCl_2$, pH 6.0 at 4°C, prior to concentration to 300 – 500 μM. For ^{15}N-relaxation samples bound to ssDNA, an equimolar amount of dT_{10} oligonucleotide was added directly to the protein concentrate.

ssDNA titration of RPA70AB by NMR

To aid in transferring assignments to the DNA-bound states of domains 70A and 70B, ^{15}N–^{1}H HSQC spectra were acquired on a titration series of RPA70AB bound to dT_{10}. Titrations were carried out by direct addition of dT_{10} oligonucleotide stock to

concentrated ^{15}N-RPA70AB for the molar ratios: 1:16, 1:12, 1:8, 1:4, 1:2, and 1:1. Samples were allowed to equilibrate 15 minutes prior to acquisition of each ^{15}N–^{1}H HSQC spectrum to ensure complete equilibrium binding. Total sample dilution due to the addition of ssDNA was less than 5%, and sample pH was confirmed at the end of the titration series.

NMR experiments

All NMR experiments were performed at 25 ºC using Bruker *AVANCE* 800 or 600 NMR spectrometers equipped with cryoprobes. Gradient–enhanced ^{15}N–^{1}H HSQC spectra were recorded with 1024 complex points in the ^{1}H and 128 complex points in the ^{15}N dimension. NMR data were processed using either Topspin 2.1 (Bruker Biospin) or NMRPipe (*139*) and analyzed with SPARKY v3.1 (*140*). Sequence specific assignments for RPA70AB and RPA70NAB were transferred from published assignments for RPA70N and RPA70AB (*45*, *141*). Assignments for domains 70A and 70B in complex with dT_{10} were confirmed via the ssDNA titration series described above, as well as from individual ssDNA titrations of 70A and 70B reported previously (*46*, *144*).

HSQC-based ^{15}N-relaxation measurements of T_1 and T_2 values were acquired using standard inverse detected pulse sequences (*145*, *146*), modified to include a gradient-enhanced water suppression scheme (*147*). T_1 values were measured for delays of 50, 100, 200 (x2), 300, 600, 1200, 2500, 4000 ms with an overall recovery delay of 5.0 – 6.0 s. T_2 values were measured for delays 17.3, 34.6, 51.8, 69.1 (x2), 86.4, 103.7, 138.2, 172.8, 207.4 ms with an overall recovery delay of 1.5 s. The ^{1}H-^{15}N NOE experiment (*145*) was acquired with a 3-second saturation period and interleaved

acquisition of saturated and non-saturated transients. Typical acquisition parameters were 128 (^{15}N) and 1024 (^{1}H) points in the ω_1 and ω_2 dimensions, respectively with 24 (T_1 and T_2) or 100-120 (NOE) transients collected for each t_1 increment.

Analysis of ^{15}N-relaxation data

Interleaved T_1, T_2, and NOE data were deconvoluted and processed in Topspin (Bruker Biospin) and imported into Sparky (*140*). T_1 and T_2 values for each residue were calculated by fitting a monoexponential decay to peak heights using Sparky's relaxation fitting feature and specifying 500 Monte Carlo simulations for error estimates. NOE values were calculated as the ratio of peak intensities for saturated and reference spectra (I_{sat}/I_{ref}). The NOE error was calculated according to

$$\sigma_{NOE} = NOE * \sqrt{\left(\frac{\sigma_{ref}}{I_{ref}}\right)^2 + \left(\frac{\sigma_{sat}}{I_{sat}}\right)^2} \qquad (1)$$

where the σ_{ref} and σ_{sat} are the root-mean-square (RMS) values of the noise floor for the reference and saturated spectra, respectively. These were calculated using an average of 4-5 measurements from the rm function in Sparky.

Ratios of T_1 and T_2 for calculation of the rotational diffusion tensor were generated in Excel, and errors were propagated as described for the NOE. Selection criteria for residues unaffected by fast motions or conformational exchange followed that of Tjandra et al (*111*). Essentially, residues were excluded from the diffusion tensor calculation where the NOE was less than 0.65 or for which

$$\frac{< T_2 > - T_2}{< T_2 >} - \frac{< T_1 > - T_1}{< T_1 >} > 1.5 * SD \qquad (2)$$

where $<T_2>$ and $<T_1>$ represent averages across all residues and SD represents the standard deviation among all residues for the difference expressed on the left side of the equation (*111*). Rotational diffusion tensors were calculated using the ModelFree script r2r1_diffusion (*111*, *148*) and ROTDIF (*113*). Structure files for specifying N-H bond vector orientations were derived from PDBs of crystal structures for 70AB in the absence (1FGU) and presence of ssDNA (1JMC) and the NMR solution structure for 70N (1EWI). The ellipsoid representations of each diffusion tensor were calculated using modified scripts in the program relax (*114*, *115*). PDB coordinates were also used in the calculation of theoretical rotational diffusion tensors for 70N and 70AB in the absence and presence of ssDNA using HYDRONMR (*149*, *150*).

Results

While previous NMR studies of tandem domain constructs from RPA have revealed a general lack of inter-domain contact (*46*, *47*, *142*), the average distribution of inter-domain orientation and how domain linking impacts this distribution remains unknown. To gain further insight into this aspect of RPA architecture, ^{15}N-relaxation parameters (T_1, T_2, and NOE) were acquired on the tandem domain constructs RPA70AB (±ssDNA) and RPA70NAB and used to calculate rotational diffusion tensors for the three globular domains. Since the speed and orientation of Brownian rotational diffusion are primarily influenced by macromolecular size and shape, the rotational diffusion tensor

can provide insight into the effective shape of a domain. For a domain tethered to a second rotating domain, the rotational diffusion tensor will also capture the orientational bias introduced by hydrodynamic drag from attachment of an additional rotating mass. Examination of the experimental rotational diffusion tensors of RPA domains, then, should reveal their average inter-domain rotational orientation in the context of inter-domain tethering, as well as the extent to which tethering influences their rotational motion.

Principal DNA-binding domains 70A and 70B exhibit semi-independent rotational diffusion in solution.

Visual inspection of amide relaxation parameters along the polypeptide backbone can provide first-order insight into domain mobility and reveal the presence of local flexibility within the backbone. For RPA70AB alone, Figure 3.2 reveals a slightly higher average T_1 for 70B versus that observed for 70A (Table 3.1), suggesting that the rotational motion of each domain is independent. Pronounced deviations from domain averages of T_1 and T_2 values occur in regions of high mobility, primarily the L_{12} (70A residues 212-219 and 70B residues 332-340) and L_{45} (70A residues 265-276 and 70B residues 383-390) DNA-binding loops of each domain (Chapter I), as well as the linker connecting 70A and 70B (residues 290-300), and the flexible C-terminus of 70B (residues 416-422) (Figure 3.2). A similar trend for these regions is observed from corresponding decreases in their NOE values (Figure 3.2), which are more sensitive to rapid internal motion along the polypeptide backbone.

Table 3.1. Average ^{15}N-relaxation parameters for RPA70AB at 800 MHz, 298 K.

	RPA70A[1]	RPA70B[2]	RPA70AB[3]
T_1 (s)	1.38 (0.11)	1.47 (0.13)	1.43 (0.12)
T_2 (s)	0.039 (0.002)	0.037 (0.002)	0.038 (0.002)
NOE	0.83 (0.04)	0.81 (0.04)	0.81 (0.04)

[1]The average over residues 187-287. Excludes DNA binding loops L_{12} (residues 212-219) and L_{45} (residues 265-276).
[2]The average over residues 300-413. Excludes DNA binding loops L_{12} (residues 332-340) and L_{45} (residues 383-390) and C-terminus (residues 416-422).
[3]The average over residues 187-413. Excludes DNA binding loops L_{12} and L_{45}, inter-domain linker (residues 290-300), and 70B C-terminus (residues 416-422).
Errors (in parentheses) are averaged across each range of residues.

Rotational diffusion tensors were calculated from T_1/T_2 ratios for 70A and 70B individually in the context of RPA70AB, as evaluating a single diffusion tensor for the entire tandem domain construct led to poor fits (data not shown). Since residues experiencing rapid local motion or conformational exchange can lead to under- or overestimation of rotational diffusion, only residues meeting the criteria of equation 2 and possessing an NOE > 0.65 were included in the calculation. In all, a total of 58 and 67 residues were used in the 70A and 70B tensor calculations, respectively. These data were sufficient to define the six parameters of a fully anisotropic diffusion tensor, and the relevant N-H vectors were evenly distributed throughout each domain, allowing for even orientational sampling (data not shown) (*151*).

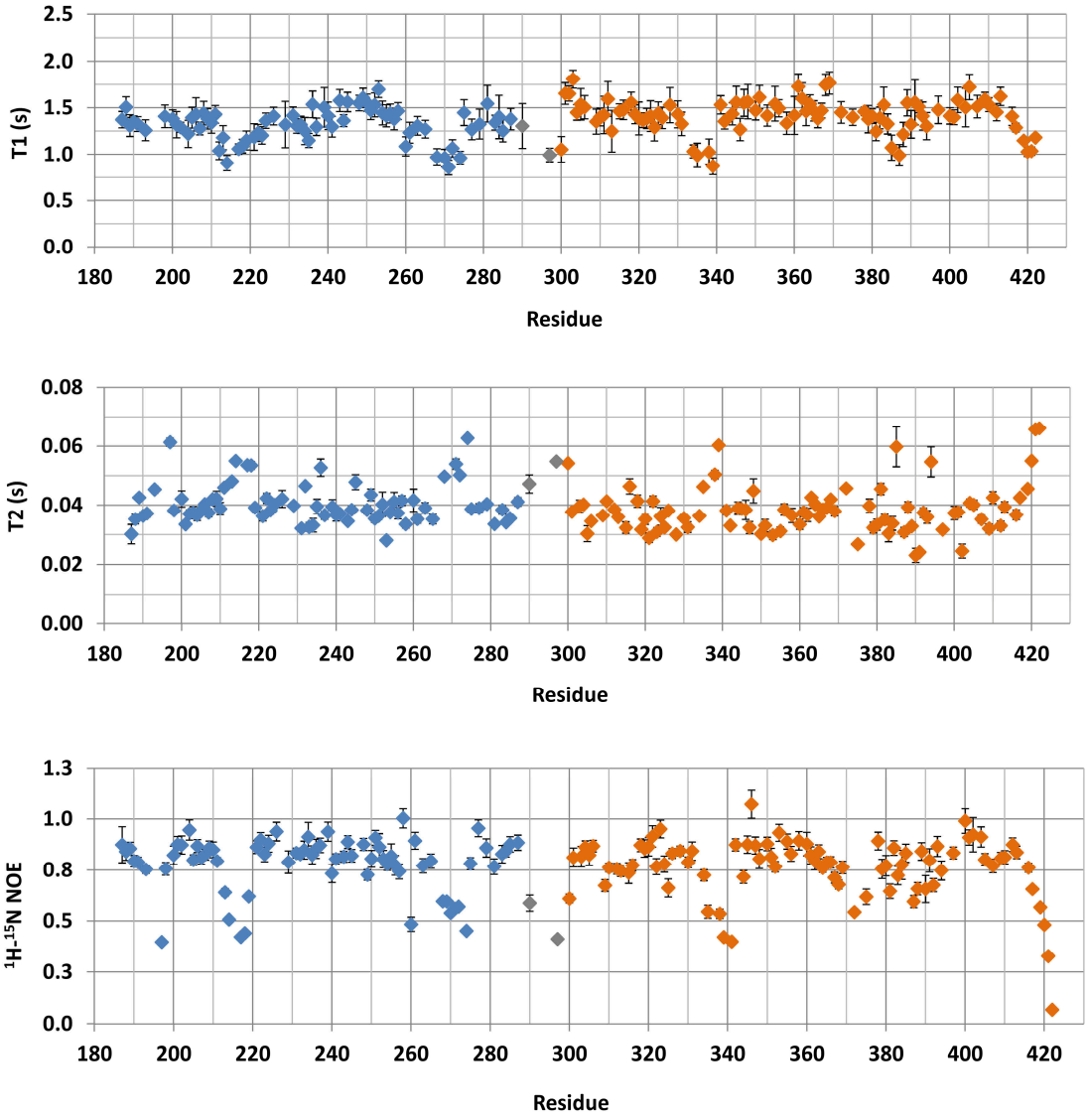
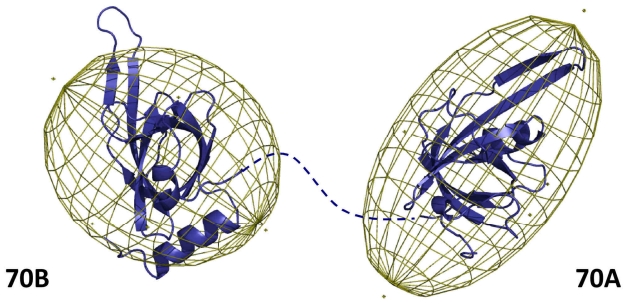


Figure 3.2. Backbone amide nitrogen (^{15}N) relaxation time constants (T_1 and T_2) and ^{1}H-^{15}N NOE values for RPA70AB at 800 MHz and 298 K. Residues for 70A are colored blue; residues for 70B, orange; and residues within the linker, gray. The average statistical error in the measurements was less than 10% for T_1 parameters and less than 5% for T_2 and ^{1}H-^{15}N NOE parameters.

Diffusion tensors were calculated assuming isotropic, axially symmetric, and fully anisotropic rotational motion and assessed for goodness-of-fit. Subsequent statistical testing indicated that the fit improved for both 70A and 70B upon selection of the axially

A.



B.

	Model	D_x (10^7 rad/s)	D_y (10^7 rad/s)	D_z (10^7 rad/s)	D_{iso} (10^7 rad/s)	τ_m (ns)	Anisotropy
70A	Isotropic				1.16	14.4	
	Axial	1.04 (0.04)	1.04 (0.04)	1.36 (0.07)	1.15	14.5	1.30 (0.05)
	Anisotropic	1.01 (0.20)	1.08 (0.11)	1.35 (0.26)	1.15	14.6	
70B	Isotropic				1.08	15.4	
	Axial	1.02 (0.04)	1.02 (0.04)	1.20 (0.11)	1.08	15.5	1.18 (0.09)
	Anisotropic	0.98 (0.09)	1.05 (0.10)	1.20 (0.12)	1.08	15.5	
	Model	**α**	**β**	**γ**	**X^2**	**F-statistic**	**p-value**
70A	Isotropic				1.92		
	Axial	71 (12)	163 (60)		1.47	6.76	5.9E-04
	Anisotropic	65 (72)	161 (13)	26 (118)	1.49	0.73	0.49
70B	Isotropic				1.78		
	Axial	178 (23)	55 (24)		1.66	2.54	0.064
	Anisotropic	179 (25)	54 (31)	67 (77)	1.67	0.71	0.49

Figure 3.3. Rotational diffusion tensor analysis for RPA70AB at 800 MHz and 298 K. A) Ellipsoid representations of the rotational diffusion tensor are superimposed upon inertial representations of domains 70A and 70B. The sizes of the ellipsoid major and minor axes are proportional to the corresponding rate of rotation about them, while the orientation represents the principal axis system of the diffusion frame. B) Rotational diffusion tensor parameters for 70A and 70B calculated in ROTDIF. D_x, D_y, and D_z represent rotational rates of diffusion about the principal axes of the diffusion tensor; D_{iso} is the effective isotropic rate of diffusion, $D_{iso} = \frac{1}{3}(D_x + D_y + D_z)$. τ_m is the rotational correlation time ($\tau_m = 1/6D_{iso}$). The Euler angles specifying the orientation of the principal diffusion axes with respect to the inertial frame (α, β, γ) are given in degrees. The model of best fit is highlighted in blue.

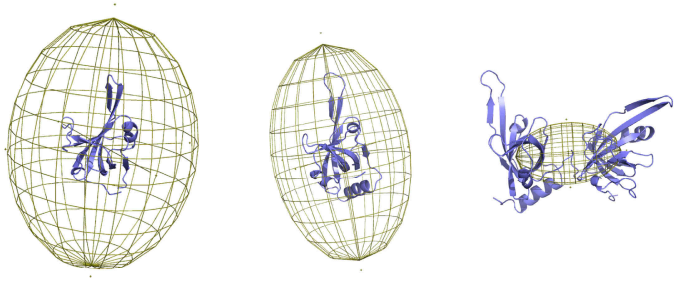
symmetric model over the isotropic model, but demonstrated negligible improvement upon application of the fully anisotropic model (Figure 3.3B).

Study of the axially symmetric rotational diffusion tensors of 70A and 70B reveals trends consistent with those initially observed for the ^{15}N relaxation parameters. The effective isotropic rotational diffusion rate for each domain differs slightly ($1.15x10^7$ rad/s for 70A and $1.08x10^7$ rad/s for 70B), indicating that 70A experiences more rapid tumbling in solution relative to 70B. More striking, though, is the difference in relative asymmetry and orientation of rotational tumbling for each domain (Figure 3.3). For 70A, the average orientation of rotational tumbling appears to correspond well with the domain's inertial frame ($\beta \sim 160°$ or a deviation of 20° from the inertial frame) and is favored about the long dimension of the domain, resulting in a 'twisting' motion about the point of linker attachment. For 70B, the preferred direction of rotational tumbling appears to be rotated ~55° with respect to the inertial frame of the domain, causing the domain to rotate preferentially at an angle to the DNA-binding cleft (Figure 3.3).

The disparity in rotational speed and orientation for each domain indicates that 70A and 70B experience independent rotational motion, even though they are tethered by a relatively short linker (10-residues). To determine if this rotational motion is consistent with that of the isolated domains, theoretical diffusion tensors were calculated from the corresponding PDB structures using HYDRONMR (*149*, *150*). Parameters for these theoretical rotational diffusion tensors and ellipsoid representations for each domain are reported in Figure 3.4. Notably, the predicted diffusion rates for each domain in isolation are higher than those measured experimentally, indicating that tethering of tandem

domains does introduce significant hydrodynamic drag to their respective rotational motion (c.f. D_{iso} 1.15x10^7 and 2.18x10^7 rad/s for 70A and D_{iso} 1.08x10^7 and 1.75x10^7 rad/s for 70B). Experimental diffusion rates are faster, though, than that predicted in the absence of any RPA70AB inter-domain motion (i.e. 70A and 70B remain fixed in space relative to each other) (Figure 3.4B). Interestingly, in contrast to the experimental results, the orientation of the theoretical diffusion tensors is aligned closely with the inertial tensor of each domain. While deviation between the inertial and experimental diffusion tensor for 70A is minimal, the difference is more pronounced for 70B (Figure 3.3), indicating that tethering introduces greater bias to the rotational tumbling of 70B than that of 70A. The theoretical diffusion tensor of the two domains fixed in tandem clearly deviates from all experimental orientations (Figure 3.4A, right panel).

A.



B.

Model	D_x (10^7 rad/s)	D_y (10^7 rad/s)	D_z (10^7 rad/s)	D_{iso} (10^7 rad/s)	τ_m (ns)	Anisotropy	α	β	γ
70A	1.92	2.00	2.61	2.18	7.6	1.3-1.4	49	169	41
70B	1.43	1.51	2.31	1.75	9.5	1.5-1.6	0	179	4
70AB	0.597	0.665	1.01	0.757	22.0	1.5-1.7	-34	176	-38

Figure 3.4. Theoretical rotational diffusion tensors for 70A and 70B individually and tethered. HYDRONMR calculations were specified for 800 MHz field strength and 298 K. A) Ellipsoid representations of the theoretical rotational diffusion tensor are superimposed upon inertial representations of domains 70A (left), 70B (middle), and 70AB (right). B) Theoretical rotational diffusion tensor parameters for 70A, 70B, and 70AB. Theoretical anisotropy is reported as a range from D_z/D_x and D_z/D_y. τ_m is the rotational correlation time ($\tau_m = 1/6D_{iso}$). Euler angles specifying the orientation of the principal diffusion axes with respect to the inertial frame (α, β, γ) are reported in degrees.

The rotational motion of 70A and 70B becomes correlated upon binding ssDNA.

To determine the impact of ssDNA binding on inter-domain orientation and rotational motion, ^{15}N-relaxation values were measured for RPA70AB bound to a dT_{10} oligonucleotide, and assignments were determined for the DNA-bound state by chemical shift perturbation assay (Materials and Methods). In contrast to the DNA-free state, the subtle difference in average T_1 and T_2 values for 70A versus 70B is now absent (Figure 3.5, Table 3.2). Moreover, there is a slight, but distinct increase, in the average T_1 for both domains in the DNA-bound state versus the DNA-free state, indicative of slower rotational motion for both domains. Deviations in T_1, T_2, and the NOE noted previously for the mobile DNA-binding loops L_{12} and L_{45} and the flexible linker connecting 70A and

Table 3.2. Average values for ^{15}N-relaxation parameters for RPA70AB bound to dT_{10} at 800 MHz and 298 K. Averages exclude regions of high mobility for RPA70AB in the absence of DNA for purposes of comparison (refer to Table 3.1). Errors (in parentheses) are averaged across each range of residues.

	RPA70A/dT_{10}	**RPA70B/dT_{10}**	**RPA70AB/dT_{10}**
T_1 (s)	1.83 (0.11)	1.86 (0.10)	1.85 (0.11)
T_2 (s)	0.037 (0.005)	0.038 (0.001)	0.038 (0.001)
NOE	0.83 (0.04)	0.83 (0.03)	0.83 (0.03)

70B are no longer present, but the distinctive trend noted in the C-terminus of 70B for the free protein remains (Figure 3.5). This loss of internal motion along the polypeptide backbone at L_{12} and L_{45} is consistent with reduced flexibility in the DNA-binding loops as they make contact with ssDNA substrate, while loss of flexibility in the inter-domain

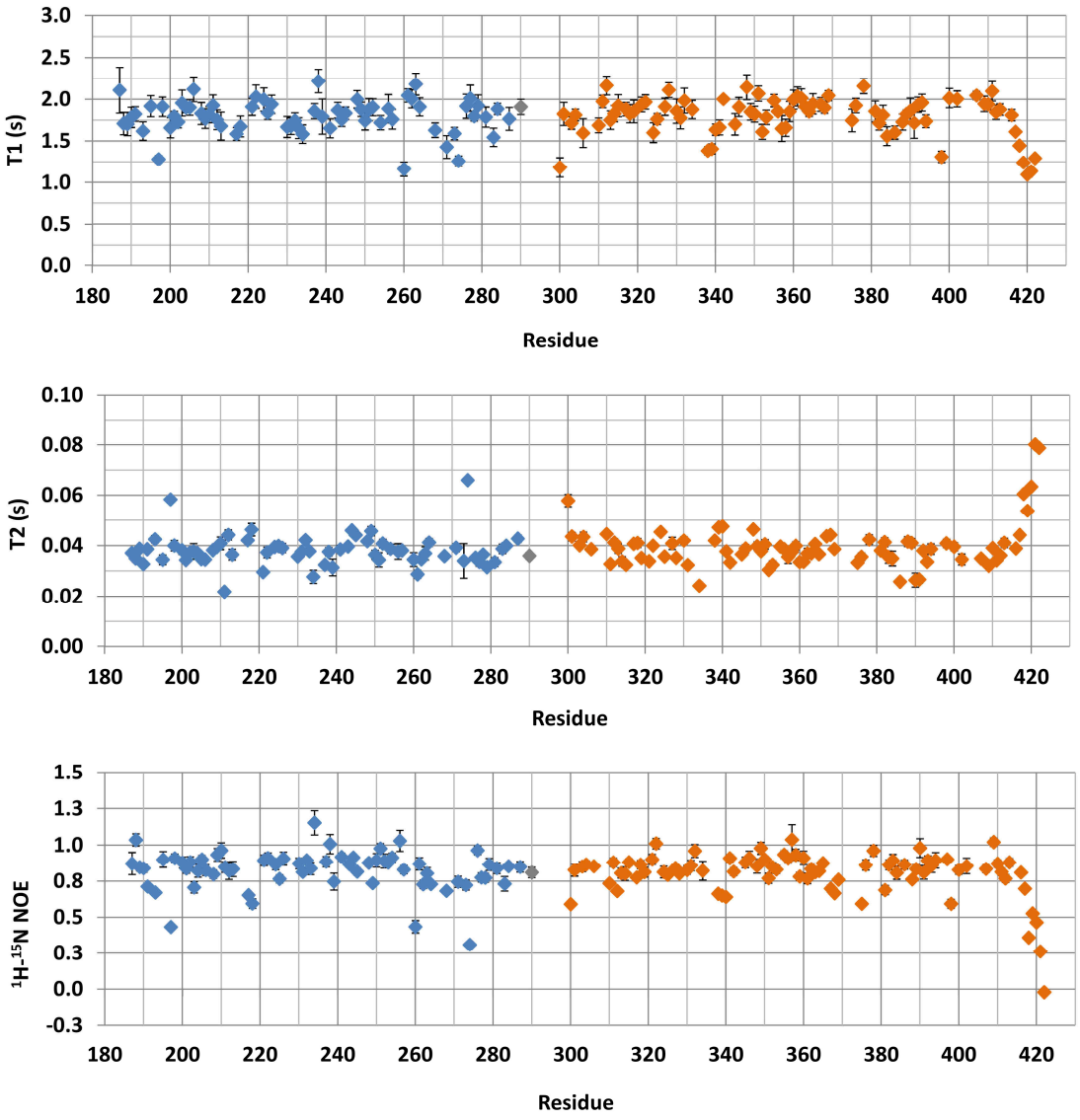
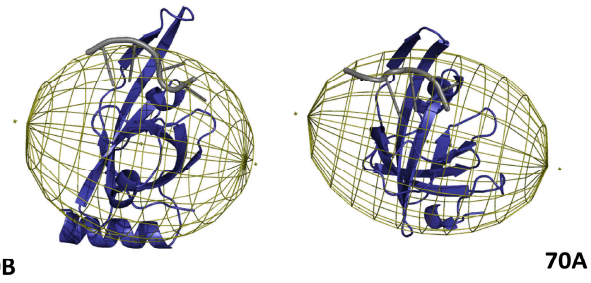


Figure 3.5. Backbone amide nitrogen (^{15}N) relaxation time constants (T_1 and T_2) and ^{1}H-^{15}N NOE values for RPA70AB bound to dT_{10} at 800 MHz and 298 K. Residues for 70A are colored blue; residues for 70B, orange; and residues within the linker, gray. The average statistical error in the measurements was 4-6%.

linker indicates restriction of inter-domain motion. Overall, these changes to the relaxation parameters are consistent with a broad decrease in the independence of global motion for 70A and 70B in the context of the RPA70AB construct.

A.



B.

	Model	D_x (10^7 rad/s)	D_y (10^7 rad/s)	D_z (10^7 rad/s)	D_{iso} (10^7 rad/s)	τ_m (ns)	Anisotropy
	Isotropic				0.99	16.8	
70A	Axial	0.92 (0.03)	0.92 (0.03)	1.15 (0.056)	1.00	16.7	1.24 (0.05)
	Anisotropic	0.89 (0.60)	0.96 (0.09)	1.14 (0.09)	1.00	16.7	
	Isotropic				0.97	17.2	
70B	Axial	0.91 (0.02)	0.91 (0.02)	1.17 (0.06)	1.00	16.7	1.28 (0.05)
	Anisotropic	0.88 (0.08)	0.94 (0.08)	1.17 (0.10)	1.00	16.7	
	Isotropic				0.98	17.0	
70AB	Axial	0.91 (0.02)	0.91 (0.01)	1.16 (0.04)	1.00	16.7	1.27 (0.04)
	Anisotropic	0.90 (0.12)	0.93 (0.10)	1.16 (0.23)	1.00	16.7	
	Model	**α**	**β**	**γ**	**χ^2**	**F-statistic**	**p-value**
	Isotropic				2.85		
70A	Axial	172 (17)	73 (16)		1.90	10.51	1.5E-05
	Anisotropic	170 (22)	75 (23)	122 (49)	1.92	0.77	0.47
	Isotropic				3.98		
70B	Axial	3 (12)	80 (9)		2.49	14.24	3.5E-07
	Anisotropic	4 (11)	79 (14)	130 (77)	2.52	0.57	0.57
	Isotropic				3.52		
70AB	Axial	116(9)	175(117)		2.23	25.20	9.3E-13
	Anisotropic	111 (101)	177 (9)	151 (114)	2.25	0.35	0.70

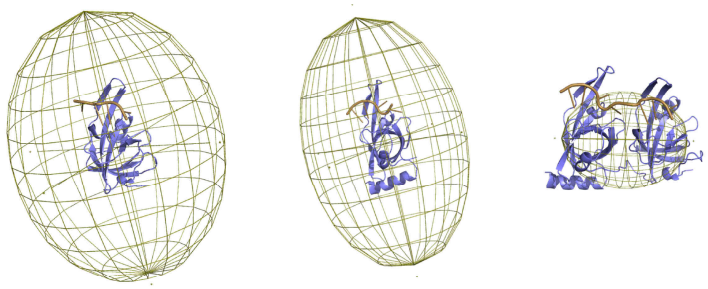
Figure 3.6. Rotational diffusion tensor analysis for RPA70AB bound to dT_{10} at 800 MHz and 298 K. A) Ellipsoid representations of the rotational diffusion tensor are superimposed upon inertial representations of domains 70A and 70B with ssDNA substrates. B) Rotational diffusion tensor parameters for 70A, 70B, and 70AB calculated in ROTDIF. D_x, D_y, and D_z represent rotational rates of diffusion about the principal axes of the diffusion tensor; D_{iso} is the effective isotropic rate of diffusion, $D_{iso} = \frac{1}{3}(D_x + D_y + D_z)$. τ_m is the rotational correlation time ($\tau_m = 1/6D_{iso}$). The Euler angles specifying the orientation of the principal diffusion axes with respect to the inertial frame (α, β, γ) are given in degrees.

As with the DNA-free state of RPA70AB, residues meeting the selection criteria described above were used to calculate rotational diffusion tensors for 70A (58 residues), 70B (67 residues), and 70AB (126 residues) (Figure 3.6). Axially symmetric diffusion remained the diffusion model of best fit for all cases. In the presence of a ssDNA substrate, the average speeds of rotational motion for 70A and 70B are now nearly identical (D_{iso} $1.00x10^7$ rad/s for both 70A and 70B) and match the effective rotational diffusion rate obtained when considering the two domains as a single diffusing body (D_{iso} $1.00x10^7$ rad/s for 70AB). Moreover, the orientation of rotational motion is now nearly perpendicular to the inertial frames of both 70A ($\beta = 73°$) and 70B ($\beta = 80°$), indicating that rotational tumbling occurs preferentially about an axis parallel to the DNA-binding cleft and ssDNA substrate (Figure 3.6). While comparison of chemical shifts between isolated and tandem 70A and 70B constructs upholds the absence of domain contact in the DNA-bound state (*46*), the high degree of correspondence between their respective diffusion tensors implies that their inter-domain motion is now well correlated.

To confirm that these experimental findings were consistent with synchronous rotational motion, theoretical diffusion tensors were calculated for isolated 70A, isolated 70B, and RPA70AB bound to ssDNA with HYDRONMR (Figure 3.7). Excellent correspondence was found between experimental diffusion tensors and those calculated for RPA70AB, whereas the diffusion rates calculated for isolated 70A and 70B clearly overestimated the rates of diffusion (Figure3.7B). The orientation of rotational motion from the theoretical calculations also agreed well with that derived from experimental ^{15}N-relaxation parameters (Figure 3.7A). Overall, these results support a realignment of

the average inter-domain orientation between 70A and 70B upon binding ssDNA, causing their respective rotational motions to lose independence and become correlated.

A.



B.

Model	D_x (10^7 rad/s)	D_y (10^7 rad/s)	D_z (10^7 rad/s)	D_{iso} (10^7 rad/s)	τ_m (ns)	Anisotropy	α	β	γ
70A/dC_4	2.05	2.10	2.66	2.27	7.3	1.3	74	176	-59
70B/dC_4	1.82	1.90	2.68	2.13	7.8	1.4	25	178	20
70AB/dC_8	0.873	0.981	1.22	1.02	22.8	1.2-1.4	59	177	60

Figure 3.7. Theoretical rotational diffusion tensors for 70A and 70B bound to ssDNA. individually and tethered. HYDRONMR calculations were specified for 800 MHz field strength and 298 K. A) Ellipsoid representations of the theoretical rotational diffusion tensor are superimposed upon inertial representations of domains 70A (left), 70B (middle) and 70AB (right) with ssDNA substrates. B) Theoretical rotational diffusion tensor parameters for 70A, 70B, and 70AB with ssDNA substrates. Theoretical anisotropy is reported as a range from D_z/D_x and D_z/D_y. τ_m is the rotational correlation time ($\tau_m = 1/6D_{iso}$). Euler angles specifying the orientation of the principal diffusion axes with respect to the inertial frame (α, β, γ) are reported in degrees.

The rotational motion of 70N is independent of 70A and 70B diffusion

As with the DNA-binding domains 70A and 70B, the protein interaction domain 70N remains free of inter-domain contact in the context of full-length RPA and other tandem domain constructs (*47*, *142*). Because of the comparatively long length of its connecting linker (60-70 residues), 70N is expected to exhibit greater rotational freedom than both 70A and 70B, though this has yet to be demonstrated experimentally. To gather further insight on the average inter-domain orientation of 70N when linked to 70A and 70B, as well as the degree of rotational freedom permitted by the 70NA linker, ^{15}N-

relaxation parameters were measured for the construct RPA70NAB, followed by extraction of individual diffusion tensor parameters for each domain.

Table 3.3. Average values for ^{15}N-relaxation parameters for RPA70NAB at 800 MHz and 298 K. Averages exclude regions of high mobility for RPA70AB in the absence of DNA (refer to Table 3.1) and 70N residues with ^{1}H-^{15}N NOE values three standard deviations below the average (residues 3, 21, 53, 90). Errors (in parentheses) are averaged across each range of residues.

	RPA70N	RPA70A	RPA70B	RPA70AB	RPA70NAB
T_1 (s)	1.20 (0.05)	1.58 (0.16)	1.63 (0.13)	1.61 (0.14)	1.48 (0.11)
T_2 (s)	0.049 (0.001)	0.038 (0.003)	0.046(0.004)	0.042 (0.004)	0.045 (0.003)
NOE	0.82 (0.03)	0.73 (0.08)	0.74 (0.06)	0.74 (0.07)	0.76 (0.06)

^{15}N-relaxation parameters for RPA70NAB are plotted in Figure 3.8 and summarized in Table 3.3. Assignments from 70N, 70A, and 70B were easily transferred (*152*); assignments for the 70NA linker were not available. Visual inspection of T_1 and T_2 plots reveals a marked difference in average values for 70N relative to 70A and 70B, indicating a high level of rotational independence for 70N. Notably, the level of noise and inter-residue variation appears to be greatly enhanced in domains 70A and 70B relative to 70N, which can be attributed to increased line broadening and decreased signal sensitivity for 70AB relative to 70N (Figure 3.8). Comparison of T_1 and T_2 values for 70AB residues in RPA70NAB to those measured for RPA70AB reveals a discernible increase in the average T_1 values for both domains, as well as an increase in the average T_2 for 70B (c.f. Tables 3.2 and 3.3). Examination of NOE values, which highlight local motion along the polypeptide backbone, reveal the characteristic deviations observed

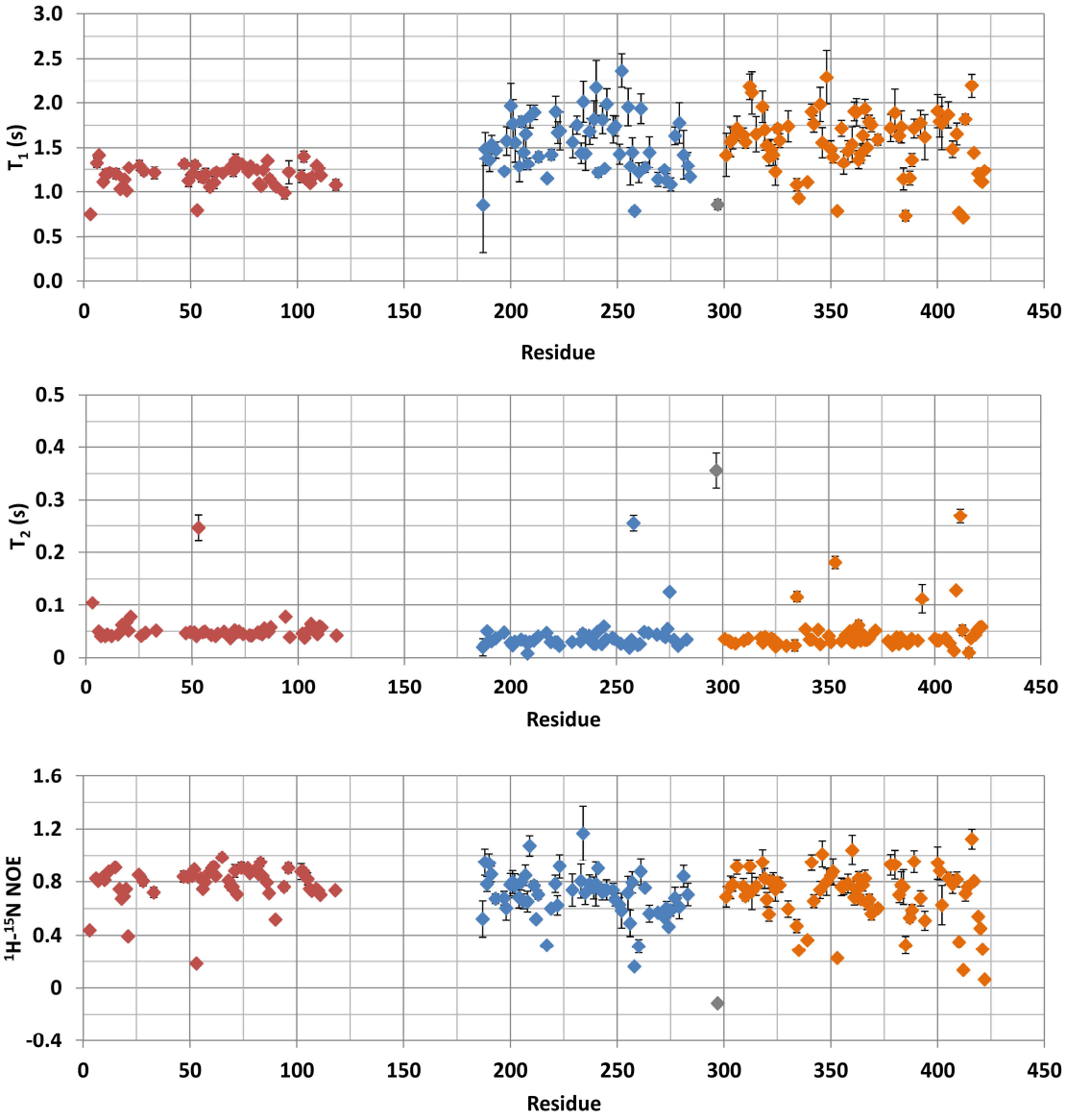
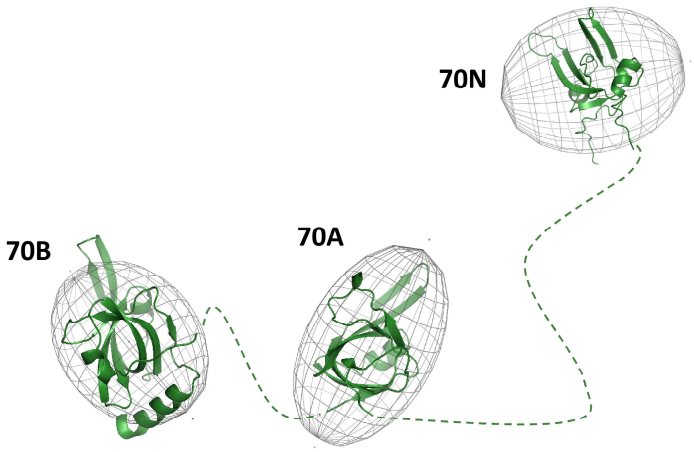


Figure 3.8. Backbone amide nitrogen (^{15}N) relaxation time constants (T_1 and T_2) and ^{1}H-^{15}N NOE values for RPA70NAB at 800 MHz and 298 K. Residues for 70N are colored red; residues for 70A, blue, residues for 70B, orange; and residues within the 70AB linker, gray. Assignments for residues in the 70NA linker were not available for these studies. The average statistical error in the measurements was less than 8% for all parameters considered across all domains. When parsed by domain, the statistical error was 3-4% for 70N and 8-10% for 70AB.

A.



B.

	Model	D_x (10^7 rad/s)	D_y (10^7 rad/s)	D_z (10^7 rad/s)	D_{iso} (10^7 rad/s)	τ_m (ns)	Anisotropy
	Isotropic				1.37	12.1	
70N	Axial	1.25 (0.07)	1.25 (0.07)	1.64 (0.17)	1.38	12.1	1.32
	Anisotropic	1.12 (0.17)	1.38 (0.23)	1.63 (0.22)	1.38	12.1	
	Isotropic				1.01	16.5	
70A	Axial	0.85 (0.06)	0.85 (0.06)	1.45 (0.75)	1.05	15.9	1.7
	Anisotropic	0.69 (0.31)	1.06 (0.43)	1.40 (0.46)	1.05	15.9	
	Isotropic				0.98	17.0	
70B	Axial	0.88 (0.07)	0.88 (0.07)	1.16 (0.13)	0.97	17.2	1.31
	Anisotropic	0.86 (0.24)	0.91 (0.26)	1.16 (0.30)	0.98	17.0	
	Model	**α**	**β**	**γ**	**χ^2**	**F-statistic**	**p-value**
	Isotropic				16.6		
70N	Axial	70 (20)	89 (23)		14.2	3.62	0.021
	Anisotropic	69 (39)	88 (23)	28 (29)	14.0	1.2	0.32
	Isotropic				20.0		
70A	Axial	135 (31)	166 (140)		17.9	2.46	0.079
	Anisotropic	110 (48)	171 (26)	3 (NaN)	17.9	9.3	0.41
	Isotropic				4.9		
70B	Axial	81 (35)	145 (52)		4.6	1.81	0.16
	Anisotropic	82 (53)	146 (40)	37 (NaN)	4.9	0.05	0.95

Figure 3.9. Rotational diffusion tensor analysis for RPA70NAB at 800 MHz and 298 K. A) Ellipsoid representations of the rotational diffusion tensor are superimposed upon inertial representations of domains 70N, 70A, and 70B. B) Rotational diffusion tensor parameters for 70N, 70A, and 70B calculated in ROTDIF. D_x, D_y, and D_z represent rotational rates of diffusion about the principal axes of the diffusion tensor; D_{iso} is the effective isotropic rate of diffusion, $D_{iso} = \frac{1}{3}(D_x + D_y + D_z)$. τ_m is the rotational correlation time ($\tau_m = 1/6D_{iso}$). The Euler angles specifying the orientation of the principal diffusion axes with respect to the inertial frame (α, β, γ) are given in degrees.

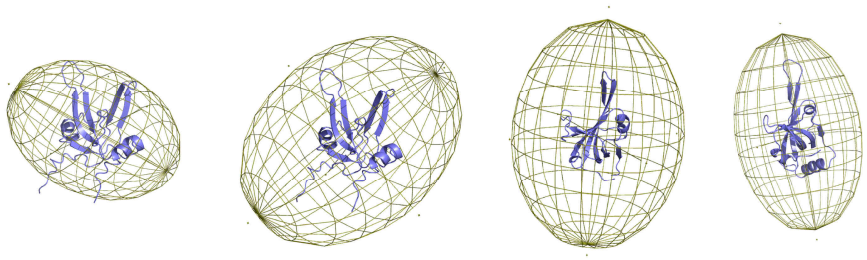
previously for 70A and 70B in the DNA-binding loops, inter-domain linker, and C-terminus. For 70N, low NOE values are observed in the L_{45} loops (residue 90); whereas we were unable to assess the trends for the 70N L_{12} loop**,** because the resonances were not well resolved in the original spectra (Figure 3.8).

While the rotational diffusion tensor for 70A in the context of RPA70NAB is very similar to its counterpart from the RPA70AB construct, the diffusion tensor for 70B is now more closely aligned with the inertial frame of the domain (c.f. Figures 3.3 and 3.9), suggesting that attachment of the 70NA linker affects the rotational motion of 70B without major perturbation to the rotational orientation of 70A. The speed of rotational diffusion for 70N is increased relative to that of 70A and 70B, consistent with its greater rotational freedom (Figure 3.9B). Curiously, the relative orientation of rotational tumbling for 70N differs markedly from that of 70A and 70B (Figure 3.9). In fact, the 70N diffusion frame is rotated back and away from the DNA/protein interaction cleft (β = 89°), nearly perpendicular, in comparison to the diffusion frames of 70A (β = 166°) and 70B (β = 145°) (Figure 3.9). This may be a consequence of the attachment point of the linker in each case, which is more centrally located for 70N, but is offset toward the back and front of the domain base for 70A and 70B, respectively.

Goodness-of-fit metrics for all diffusion tensors were globally much poorer for RPA70NAB than for RPA70AB with and without ssDNA substrate, despite the fact that the errors in the original ^{15}N T_1, T_2, and NOE measurements were comparable across all three sets of data (c.f. Table 3.3 to Tables 3.1 and 3.2). The RPA70NAB reduced χ^2 values were generally lower for 70B relative to 70A and 70N, suggesting that tethering by the 70NA linker may introduce greater complexity to the rotational trajectory of these

domains. As such, a high level of variance within the orientation of rotational tumbling may not be adequately described by diffusion tensor models that reflect motion associated with a fixed tumbling orientation.

A.



B.

Model	D_x (10^7 rad/s)	D_y (10^7 rad/s)	D_z (10^7 rad/s)	D_{iso} (10^7 rad/s)	τ_m (ns)	Anisotropy	α	β	γ
70N_trunc	1.747	1.956	2.422	2.04	8.2	1.2-1.4	23	120	-10
70N	1.200	1.423	1.842	1.49	11.2	1.3-1.5	-23	100	-31
70A	1.92	2.00	2.61	2.18	7.6	1.3-1.4	49	169	41
70B	1.43	1.51	2.31	1.75	9.5	1.5-1.6	0	179	4

Figure 3.10. Theoretical rotational diffusion tensors for 70N_trunc, 70N, 70A, and 70B. HYDRONMR calculations were specified for 800 MHz field strength and 298 K. A) Ellipsoid representations of the theoretical rotational diffusion tensor are superimposed upon inertial representations of domains (from left to right) 70N, 70N_trunc, 70A, and 70B. B) Theoretical rotational diffusion tensor parameters for 70N_trunc, 70N, 70A, and 70B. Theoretical anisotropy is reported as a range from D_z/D_x and D_z/D_y. τ_m is the rotational correlation time ($\tau_m = 1/6D_{iso}$). Euler angles specifying the orientation of the principal diffusion axes with respect to the inertial frame (α, β, γ) are reported in degrees.

To determine if the rotational motion of 70N mimiced the motion of the isolated domain, the theoretical rotational diffusion tensor for 70N alone was calculated in HYDRONMR (Figure 3.10). Diffusion tensors were calculated from two separate sets of 70N PDB coordinates. The first set included all residues determined as part of the original NMR solution structure (model 70N). The second set, however, excluded residues of the flexible N- and C-termini of the domain (residues 2-6 and 107-114) in an attempt to focus the calculation solely upon the globular core of the domain (model 70N_trunc). Interestingly, the theoretical diffusion parameters derived from the intact

70N model provided the best match to the experimental rotational diffusion data (Figure 3.10), indicating that the extended linker contributes to the hydrodynamic geometry of 70N rotational diffusion**.**

Discussion

Rotational motion in solution is exquisitely sensitive to macromolecular size and shape and is well positioned to reveal how domain linkage impacts inter-domain architecture. Here, rotational diffusion tensors have been determined for select RPA domains, the principal DNA-binding domains 70A and 70B and the protein interaction domain 70N, by measurement of ^{15}N-relaxation parameters on the tandem domain constructs RPA70AB and RPA70NAB. Analysis of rotational motion for domains 70A and 70B reveals semi-independent diffusion of each domain that is more restricted than that predicted for the free domains. Domain tumbling remains sufficiently unique, however, to allow independent diffusion frames for each domain. Rotational diffusion becomes highly correlated upon binding ssDNA, causing 70A and 70B to share identical diffusion properties and tumble as a single entity. Rotational diffusion of 70N is shown to be independent of that for both 70A and 70B, but its diffusion is clearly influenced by attachment of the 70NA linker.

Marked differences in the rate and orientation of domain rotational diffusion suggest that the primary factors driving the unique inter-domain dynamics of 70AB versus 70N and 70A versus 70B are the length of the associated inter-domain linkers and the geometry of linker attachment. The long, 60-70 residue linker connecting 70N to 70AB effectively isolates 70N from the hydrodynamic drag created by attachment of the 28-kDa 70A and 70B domains, allowing its rotational motion to approximate diffusion of

a free domain more closely. The short, 10-residue linker connecting 70A and 70B, however, maintains sufficient proximity between the domains that hydrodynamic drag becomes a key contribution to their diffusion and their relative inter-domain orientations.

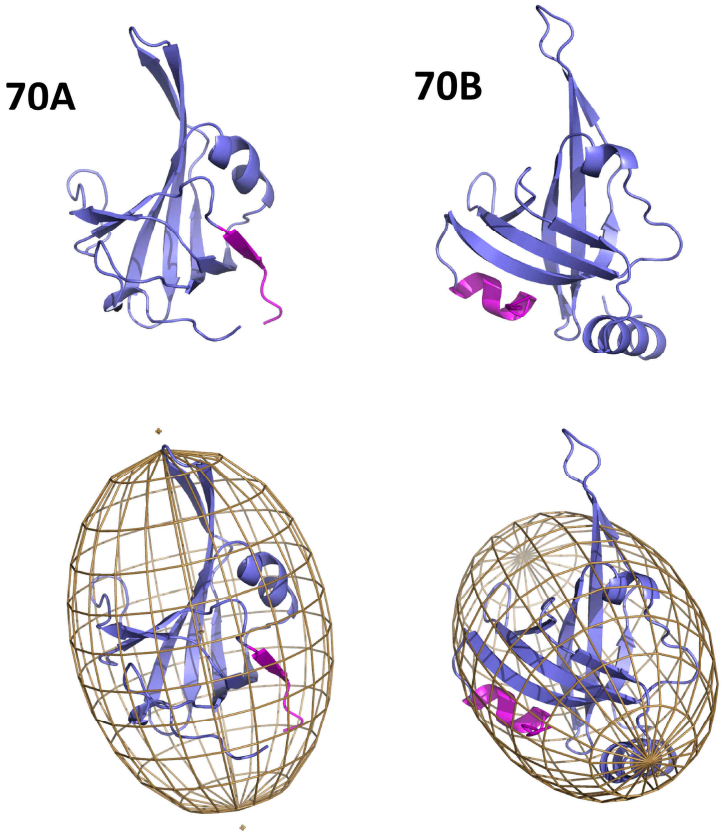


Figure 3.11. Linker attachment points for domains 70A (left) and 70B (right). Residues feeding into the 70AB linker are colored in magenta. Ellipsoid representations of the rotational diffusion tensor are superimposed upon inertial representations of each domain in the lower panel.

Interestingly, the orientation of 70A and 70B diffusion remains unique, despite the similar size and shape of these two domains. Examination of RPA70AB crystal structures, however, reveals that the two domains do differ in the attachment point of the linker. 70A is secured to the linker at the base of the domain on the face opposite the

DNA-binding cleft, while 70B is attached in an opposite manner at a point centered just below the DNA-binding cleft (Figure 3.11). It is likely that the asymmetry of these attachment points is the primary influence driving the unique direction of domain pivoting and may explain the differences observed in rotational orientation for 70A and 70B.

The rotational orientation of each domain and the geometry of their respective linker attachments are suggestive of how the two principal DNA-binding domains achieve efficient, polarized binding of ssDNA substrates. The rotational pivoting of 70B toward the front of its DNA-binding cleft optimally positions the domain to encounter a ssDNA substrate as soon as it is initially engaged by 70A. Meanwhile, the back-to-front attachment of the linker ensures that 70B binding occurs adjacent to the 3' side of 70A, preventing 70B from flipping around to bind at the 5' side of the domain or even in a back-to-back manner reminiscent of *E. coli* SSBs (Figure 3.12).

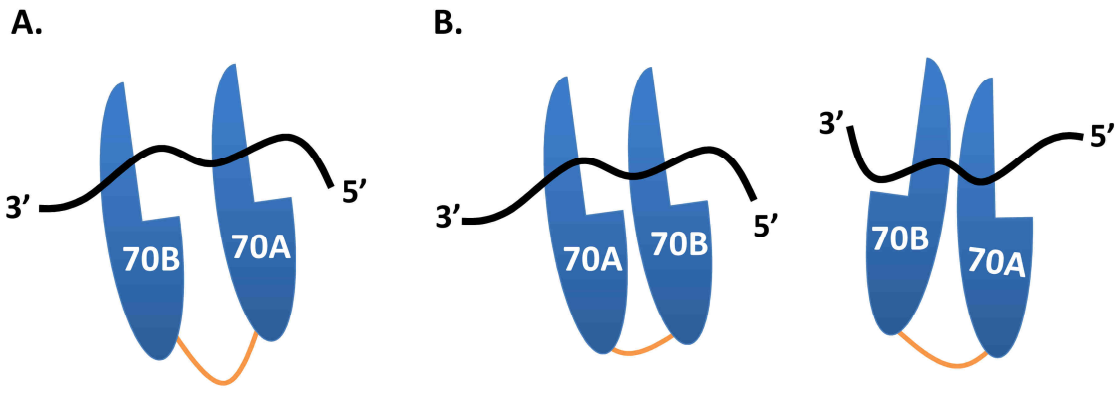


Figure 3.12. How rotational orientation and linker attachment potentially direct RPA70AB's known mode of DNA-binding. A) Canonical organization of domains 70A and 70B along ssDNA. B) Alternate organizations of domains 70A and 70B along ssDNA that are precluded by back-to-front linker attachment.

Flexible, inter-domain organization is thought to be a critical aspect of RPA function. Understanding variation in the architecture of this inter-domain organization provides valuable insight into how RPA executes its function in binding ssDNA and engaging other DNA processing proteins. The close tethering of 70A and 70B results in an inter-domain arrangement that is only semi-independent in the DNA-free state. Partial restriction of this sampling of inter-domain orientations is expected to increase the rapidity of DNA binding by co-localizing the two domains in an orientation that is favorable for engaging ssDNA substrates, but still sufficiently flexible to allow the protein to adapt to the substrate as needed. In the case of protein interaction, though, the ability to access multiple, unrestricted domain orientations increases the likelihood of a successful binding interaction. The longer length of the 70NA linker (60-70 residues) enables such broad conformational sampling by allowing 70N sufficient rotational freedom to engage and recruit DNA processing partners representing a wide variety of architectures. Moreover, the extended linker also circumvents the steric obstruction that might arise were all RPA domains packed closely together, particularly if the target binding partner is presented in the context of a large multi-protein assembly.

From a biophysical point of view, the relationship between linker properties (length, geometry, amino acid content) and inter-domain dynamics is an emerging area of study that has been approached from both theoretical and emprirical perspectives (*153-155*). Much work remains, however, to provide a full account of the role of disordered linkers in directing inter-domain motion. For proteins such as RPA, this knowledge is critical to establishing how flexible inter-domain architectures function to promote the action of multi-protein machinery.

Acknowledgements

Dr. Kavita Dorai, a visiting professor to our lab from the Indian Institute of Science Education and Research in Mohali, India, and Dr. Markus Voehler of Vanderbilt's Biomolecular NMR Facility were instrumental in the initiation of this project and the initial piloting of pulse sequences used for the ^{15}N-relaxation experiments. Marie-Eve Chagot provided invaluable technical assistance in the preparation of ^{15}N-enriched stocks of RPA70AB used for a number of these experiments. Special acknowledgement is given to Dr. Sarah Soss Warnick for aid and guidance in the general analytical approach used for ^{15}N-relaxation data and application of the standard analysis software packages. Special thanks is given to Sonja Brooks, a former rotation student in our laboratory, who intrepidly analyzed the RPA70NAB ^{15}N-relaxation data. A final thank you is extended to Dr. Shibani Bhattacharya, a former post-doctoral fellow in our laboratory, now at the New York Center for Structural Biology, for providing suggestions and advice for troubleshooting ^{15}N-relaxation sequences during the start-up period of this project.

CHAPTER IV

INVESTIGATING REMODELING OF RPA ARCHITECTURE UPON BINDING ssDNA BY SMALL-ANGLE X-RAY SCATTERING

Introduction

As the primary ssDNA-binding protein in eukaryotes, Replication Protein A (RPA) plays a vital role in the organization and protection of single-stranded (ss) DNA during genome propagation and maintenance (reviewed in *12-14*). In addition to deflecting endonuclease activity and preventing DNA secondary structure formation, RPA acts as a platform for recruiting DNA processing factors and managing their access to ssDNA (*23, 156*). The ability to coordinate and participate in the rapid interchange of proteins at ssDNA templates is thought to arise in part from RPA's dynamic, modular architecture (*142*). While architectural remodeling enables RPA to adapt swiftly to an everchanging substrate landscape, specific structural models for how RPA engages ssDNA and coordinates protein access are currently unavailable.

As a modular heterotrimer, RPA's three subunits (RPA70, RPA32, RPA14) contain seven structured domains interconnected by flexible linkers (Figure 4.1). Three of these domains form the trimeric core of the protein (70C, 32D, 14), from which emanate the flexibly linked N-terminal domains of RPA70 (70B, 70A, 70N), as well as the disordered N-terminus and structured C-terminal domain of RPA32 (32N and 32C, respectively). All domains are OB-folds (oligosaccharide/oligonucleotide binding) with the exception of RPA32C (a winged helix domain) and have been structurally

characterized at high resolution (*39-45, 48, 49*). NMR experiments have indicated that these domains are structurally independent in the context of full-length RPA, resulting in a dynamic solution architecture (*143, 152*). The challenge of describing the average spatial disposition of this time-varying architecture and how it changes during the course of DNA processing has yet to be overcome.

Binding of ssDNA is mediated by the basic clefts of the four central OB-folds – 70A, 70B, 70C, and 32D, which engage ssDNA with 5' to 3' polarity, respectively, and occlude a site size of 30 nucleotides (Figure 4.1) *(28, 42)*. RPA binds ssDNA in three distinct binding modes, which correlate with different structural and functional states of the protein (*53, 54*). An initial 8-10 nucleotide binding mode involves the two principal DNA-binding domains 70A and 70B (Figure 4.1C). Progression through the subsequent ~20 and ~30 nucleotide binding modes is presumed to cause a shift in the overall architecture of RPA as DNA-binding domains of the trimer core become involved (first 70C for the 20-nucleotide binding mode, then 32D for the 30-nucleotide binding mode, Figure 4.1C). These binding modes are thought to represent pivot points at which other DNA processing proteins may act upon RPA to accelerate, hinder, or reverse progression through this DNA-binding trajectory and thus promote a particular DNA processing transaction.

X-ray diffraction, NMR, and scattering studies involving the principal DNA-binding domains 70A and 70B have provided early structural insight into the first step of this DNA-binding trajectory (*40, 42, 46, 143, 144*). Binding of a dC_8 substrate aligns and compacts 70A and 70B along the length of their DNA-binding clefts (*40, 42*). Even with its high affinity for ssDNA, though, the 70AB/dC_8 complex exhibits residual dynamics in

solution (*143*). The precise structural source of this motion remains to be determined, whether from fluctuations of the ssDNA substrate or torsional motion between 70A and 70B. The existence of motion within the DNA-bound complex, however, raises the possibility of a structural equilibrium that might be manipulated to enhance or reverse ssDNA binding.

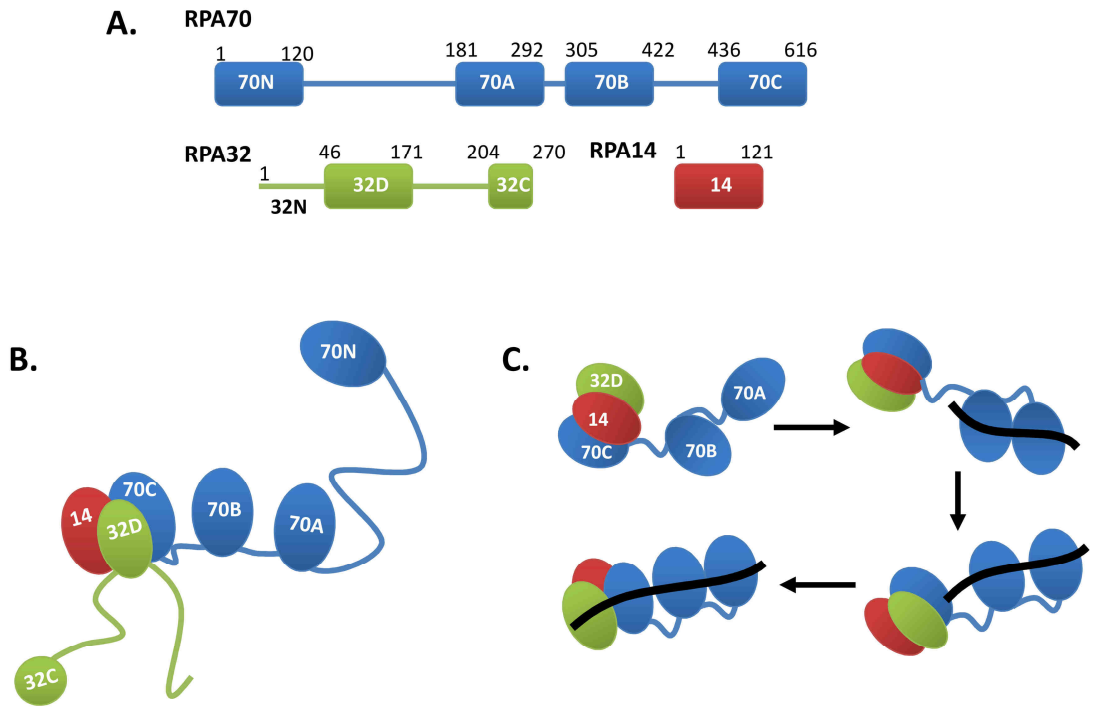


Figure 4.1. RPA domain organization and binding of ssDNA. A) Domain organization of RPA. B) Cartoon schematic of modular domain distribution in RPA. RPA70 (blue), RPA32 (green), RPA14 (red). C) RPA binds ssDNA in three discrete binding modes: (1) an initial 8-10 nucleotide mode that engages 70A and 70B, (2) an intermediate 12-23 nucleotide mode that includes 70A-70C, and (3) a final 28-30 nucleotide mode that includes all four DNA-binding domains (70A-70C, 32D).

In order to extend these initial insights to the full DNA-binding trajectory of RPA, small angle x-ray scattering (SAXS) experiments were conducted on the DNA-binding core of RPA (RPA-DBC) alone and bound to ssDNA substrates representative of RPA's

three binding modes (d10, d20, and d30). The use of SAXS allows for characterization of the global architecture and accompanying dynamics of each RPA binding mode under native solution conditions. Analysis of the low resolution spatial information gained from these experiments was extended by rigid body molecular dynamics simulations and computational modeling on RPA-DBC/ssDNA complexes. The results reveal that unbound RPA-DBC assumes multiple inter-domain orientations in solution, but the majority of the solution ensemble is only moderately extended. Binding of 10- and 20-nucleotide substrates results in a progressive compaction of the DNA-binding core, while interaction with the 30-nucleotide substrate restores RPA-DBC to an inter-domain arrangement more consistent with the DNA-free protein. Together, these findings suggest a dynamic, 'bi-modal' model for DNA-binding, where RPA is initially compacted to accommodate emerging ssDNA substrates, then resumes its default inter-domain distribution when 30-nucleotides become available. Elucidation of this fundamental structural trajectory provides an essential foundation for understanding the intricate interplay among RPA, ssDNA, and other DNA processing factors during DNA metabolism.

Materials and Methods

Materials

The pET15b vector containing RPA-DBC [$RPA70_{181-616}$/$RPA32_{43-171}$/14, or RPA70ABC/32D/14] was a kind gift of A. Bochkarev (*43*). Thrombin cleavable, 6X-His fusion tags precede N-termini of the 70ABC and 14 subunits. Active thrombin was purchased from CALBIOCHEM. All ssDNA substrates – dCCACCCCCCC,

dCCACCCCCCCCCCCCCCCC, and dCCACCCCCCCCCCCCCCCCCCCCCCCC (d10, d20, and d30, respectively), as well as fluorescently modified dC_{10}, dC_{20}, and dC_{30} oligonucleotides were purchased from Integrated DNA Technologies with standard desalting purification and were resuspended in sterile distilled water prior to use. Full-length RPA used in these studies was prepared as described (*152*) and provided as a kind gift by Dr. Mike Shell.

Expression and purification of recombinant RPA-DBC

Expression and purification of recombinant RPA-DBC have been described previously by the Bochkarev laboratory during its initial biochemical characterization (*43*). Further optimization of these protocols was required, however, in order to generate highly pure protein in quantities sufficient for SAXS. During the course of optimization, we found that RPA-DBC was highly toxic to *E. coli*, more so than the full-length protein (suggesting a beneficial steric effect from the presence of the extended protein interaction domains). Because of this, *E. coli* transformation efficiency proved to be very low (one or two colonies per 100 uL competent cells), and induced protein expression was highly inconsistent (one positive growth batch for every ten attempted).

The initial strategy to address RPA-DBC toxicity was to transfer the construct to a kanamycin-based vector with tighter control of its expression promoter (pBG100), thus circumventing the lability associated with ampicillin and the leaky expression characteristic of the pET15b promoter. Surprisingly, the transformation efficiency of this new vector proved far lower than that of the pET15b vector, and protein yield was extremely low for colonies which had successfully transformed. Attempting

transformation and expression with a variety of *E. coli* strains, media (LB, SOC, TB, Studier's minimal media), and expression conditions did not provide a ready a solution. We next considered that the order of subunits within the vector (RPA14, RPA32D, RPA70ABC) might be the source of this persistent toxicity. The subunit order had been designed previously to take advantage of robust expression from the first open reading frame (ORF) of the tricistronic vector, ensuring abundant production of the soluble subcomplex of RPA14 and RPA32 to aid in efficient solubilization of RPA70ABC. We postulated that reversing the order of the subunits in the reading frame (RPA70ABC, RPA14, RPA32D) would result in less efficient formation of the trimeric complex, and thus less toxicity from protein expression.

Implementation of this ORF order within the pBG100 vector appeared to resolve these problems, providing excellent transformation efficiency and consistent RPA-DBC expression. The overall yield of the protein was lower than expected, but was soon optimized to approach that of full-length RPA. SDS-PAGE of the purified protein, however, revealed differential staining intensity for RPA14 and RPA32D that suggested an altered stoichiometry between these subunits (2 RPA14: 1 RPA32D). Running an SDS-PAGE overlong caused the strongly staining RPA14 band to resolve into two separate species, raising the possibility of a co-purifying contaminant. This was subsequently confirmed by ESI-MS and proteomics analysis on purified protein, where the 16-kDa contaminant, previously assumed to be RPA32D, was identified as *E. coli* ferric uptake regulator (FUR). Extensive optimization of expression and purification was unable to prevent association of this protein with RPA-DBC during induced expression or to dislodge it once purified from bacteria.

In light of these developments, a renewed attempt was made to work with the original pET15b version of the construct, where strategies were developed to minimize time spent in liquid growth and stationary phases, the point at which ampicillin lability and leaky expression are most pronounced. This led to the current production protocol, which is as follows. Freshly transformed colonies of Rosetta(DE3) pLyS cells (Novagen, EMD Chemicals) were screened for robust expression at 37 °C. Successful colonies were replated overnight on solid LB media to generate high-density growth. Upon confirmation of RPA-DBC expression, cells were replated a second time to provide starter plates for liquid high-expression cultures. One-liter cultures were inoculated directly with cells scraped from the starter plates (1 plate/culture) and grown to OD_{600} 0.5-0.6 at 37 °C, followed by induction with 0.5 mM IPTG for 3 hours at 37 °C. Harvested cells were stored at -80 °C.

RPA-DBC was purified using standard nickel affinity chromatography (Ni Sepharose High Performace, GE Healthcare) in a buffer containing 20 mM HEPES-KOH (pH 7.5), 500 mM NaCl, 10 mM βME, 10 μM $ZnCl_2$ with a linear elution gradient of 40-500 mM imidazole. Relevant fractions were desalted into NaCl-free buffer (HiPrep 26/10 Desalting, GE Healthcare), followed by anion-exchange purification (SOURCE 15Q, GE Healthcare) in 20 mM HEPES-KOH (pH 7.5), 100 mM L-arginine, 10 mM βME, 10 μM $ZnCl_2$, 10% glycerol, with a linear elution gradient of 0-1.0 M NaCl. Target fractions were then concentrated and loaded onto a HiLoad 16/60 Superdex 200 column (GE Healthcare) equilibrated in 20 mM HEPES-KOH (pH 7.5), 200 mM NaCl, 10 mM βME, 10% glycerol. To remove histidine fusion tags, RPA-DBC was incubated with thrombin for two hours at room temperature, then loaded sequentially onto HiLoad

16/60 Superdex 75 and HiTrap Heparin HP columns (GE Healthcare) to remove thrombin and secondary cleavage products. Buffers were identical to those used previously for Superdex 200 and Source Q purification, respectively, with the exception of a 0 - 500 mM NaCl elution gradient for the heparin runs. As a final polishing step, the RPA-DBC pool was concentrated and passed through a Superdex 200 HR 10/30 column. Absence of ssDNA contamination was confirmed by assessment of the protein absorbance ratio at 260 and 280 nm. Average yield is approximately 1.4 mg per liter LB culture. Figure 4.2 provides a flow scheme and accompanying representative SDS-PAGE gels of RPA-DBC purification.

ESI-MS and proteomics analysis

Approximately 50 ug of purified RPA-DBC (prior to removal of histidine fusion tags) was analyzed by coupled HPLC and electrospray mass spectrometry to confirm the purity of the preparation. Proteomics analysis confirmed individual RPA subunit identity. Excision of individual subunit bands from a representative SDS-PAGE gel was followed by in-gel tryptic digestion and tandem mass spectrometry on extracted peptides. Bioinformatics analysis of MS/MS data was carried out in IDPICKER.

Fluorescence anisotropy ssDNA-binding assays

The ssDNA-binding activity of RPA-DBC was assessed by a rise in fluorescence anisotropy as increasing amounts of protein were added to polycytidine substrates labeled at their 5' ends with 6-carboxyfluorescein (5'-FAM-dC_{10}, -dC_{20}, -dC_{30}). Triplicate serial

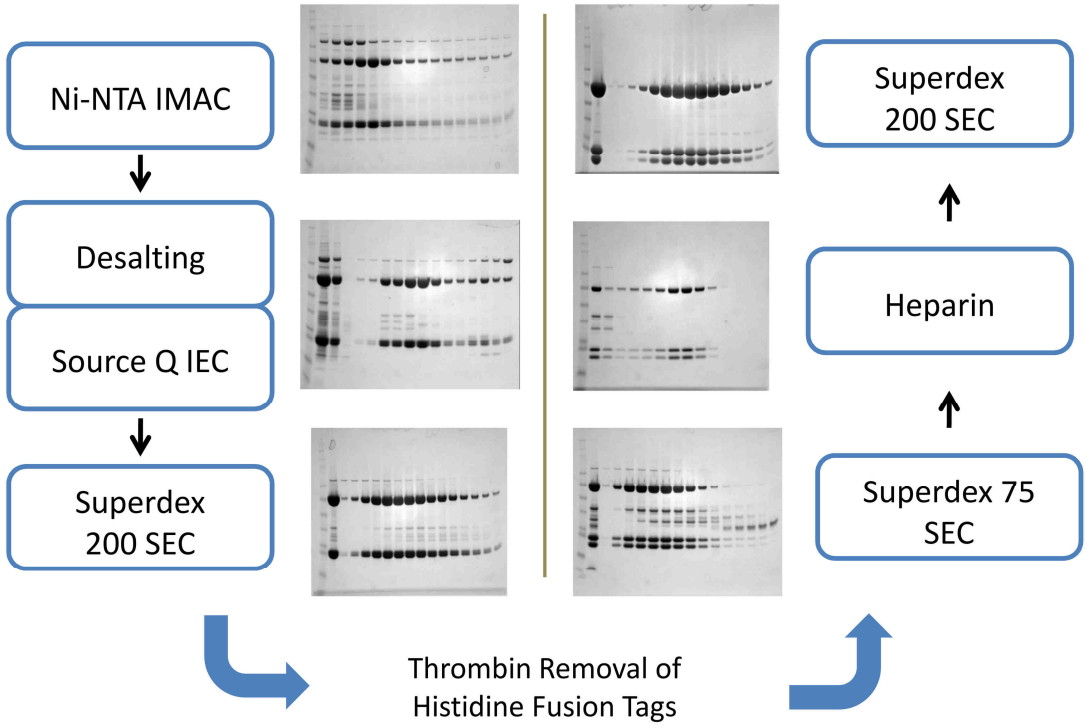


Figure 4.2. Flow chart for purification of RPA-DBC from *E. coli* and representative SDS-PAGE gels showing products of each step.

dilutions of protein (0 – 0.5 μM) were prepared in 384-well plates with 20 mM HEPES-KOH (pH 7.5), 200 mM NaCl, and 10 mM βME, then mixed with fluorescently-labeled ssDNA (final concentration 25 nM). Polarized fluorescent intensities were measured with a Spectramax M5 plate reader (Molecular Machines) at excitation and emission wavelengths of 492 nm and 520 nm, respectively, for 100 s (1 reading/s) and averaged. Dissociation constants (K_d) were calculated by fitting the data to a simple two-state binding model.

Size exclusion chromatography via multi-angle light scattering

Sample monodispersity was verified by multi-angle light scattering coupled with SEC (SEC-MALS). All experiments were carried out by in-line measurement of

ultraviolet absorbance at 280 nm (Agilent 1100 series, Agilent Technologies), static light scattering (DAWN HELEOS 8+, Wyatt Technology), and differential refractive index (Agilent 1200 series, Agilent Technologies). System calibration was performed on a BSA standard (Sigma), first resuspended, then twice exchanged into the SEC-MALS running buffer. RPA-DBC was combined with 1.2-1.5-fold excess ssDNA substrate and incubated on ice for 20-30 minutes. Samples (40 μL, 4-12 mg/mL) were then filtered and injected onto a Superdex 200 PC 3.2/30 column equilibrated in 20 mM HEPES-KOH (pH 7.5), 200 mM NaCl, 5 mM DTT, and 2% glycerol. Experiments were recorded and analyzed using ASTRA V (Wyatt Technology).

Preparation of RPA-DBC/ssDNA complexes for SAXS

Purified RPA-DBC was concentrated to 7.3 mg/mL, combined with 2-fold excess ssDNA substrate (d10, d20, or d30), and incubated on ice for 6-18 hours. To remove excess ssDNA, samples (300 uL) were injected onto a Superdex 200 HR 10/30 gel filtration column (GE Healthcare) equilibrated overnight in 20 mM HEPES-KOH (pH 7.5), 200 mM NaCl, 5 mM DTT, and 2% glycerol. Samples for RPA-DBC, RPA-DBC/d10, and RPA-DBC/d20 eluted as single peaks, consistent with their behavior observed by SEC-MALS (Figure 4.3). RPA-DBC/d30, however, produced a unique, second peak, which eluted after the primary peak. Bradford analysis and UV absorbance readings indicated the presence of both protein and DNA in this peak. As the primary peak of the elution profile remained intact, we chose to move forward with data collection on this sample series.

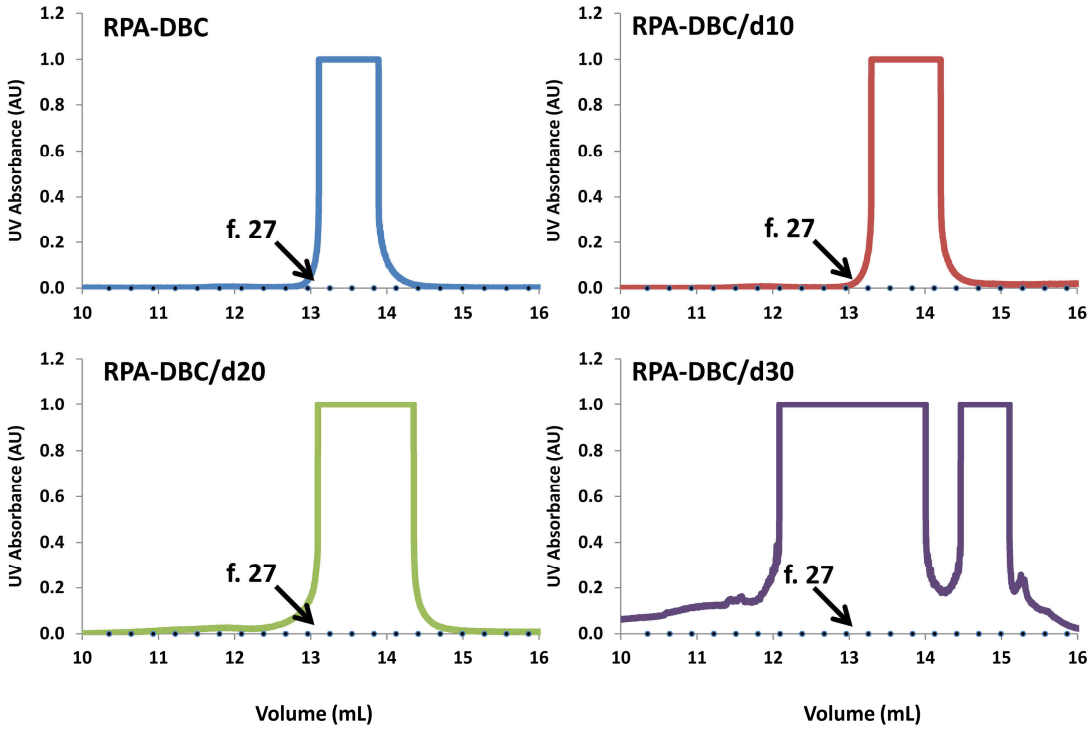


Figure 4.3. Preparative Superdex 200 gel filtration profiles for RPA-DBC before SAXS data acquisition. The arrow labeled 'f.27' is the position of fraction 27.

Coincident with the unexpected elution profile for RPA-DBC/d30 was the absence of excess, unbound ssDNA peaks for all DNA-bound samples, suggesting underestimation of the original protein or ssDNA substrate concentration during sample preparation. Reassessment of ssDNA stock concentrations indicated that the original addition of ssDNA had been on the order of 1.2-1.5-fold excess. Reassessment of protein concentration was not possible after completion of the SEC runs. A_{280} readings from the peak eluting fraction for RPA-DBC (2.4 mg/mL), however, were consistent with a starting concentration of 12 mg/mL, assuming the standard 1:5 dilution factor estimated by the SIBYLS beamline staff for the Superdex 200 column. This concentration is also consistent with the total amount of protein originally shipped for the experiments. Assuming 12 mg/mL as an upper limit, the minimum amount of RPA-DBC loaded with

ssDNA would be on the order of 70-90% prior to separation by gel filtration with no free ssDNA to spare.

In light of the unexpected SEC results, the decision was made to pursue data collection with full partitioning of each elution profile. Conservative fractionation (290 uL fractions) allowed the 2-3 mL eluted peak widths to be divided into portions comprising at most 10-15% of the total peak. If in the most extreme case, 70-90% of an RPA-DBC sample is DNA-bound, one would expect the most central portion of the peak to contain homogeneously DNA-bound RPA-DBC, which is readily extracted by the small volume of fractionation.

Small-angle x-ray scattering data collection

SAXS data was collected at the SIBYLS beamline 12.3.1 at the Advanced Light Source, Lawrence Berkeley National Laboratory. Scattering measurements were performed on 20 μL samples at 15 °C using a Hamilton robot for loading samples from a 96-well plate into a helium-purged sample chamber. Data were collected on the original gel filtration fractions from each SEC run, as well as concentration series from fractions sampled from each primary SEC elution peak and the anomalous secondary RPA-DBC/d30 peak (~2-8-fold concentration). Fractions prior to the void volume and concentrator eluates were used for buffer subtraction.

SAXS experiments were acquired using an X-ray beam from a multilayer monochromator of 12 keV (λ 1.3 Å) covering the following momentum transfer range: $0.011\ Å^{-1} < q < 0.322\ Å^{-1}$, where q is defined as $q = 4\pi \sin(\theta/2)/\lambda$ with scattering angle θ/2 and wavelength λ (Å). The multilayer monochromator provides increased X-ray flux,

allowing stronger signals for lower protein concentrations. Sequential exposures (0.5, 0.5, 2, 5, and 0.5 s) were taken, and data were monitored for radiation-dependent aggregation. All SAXS data were collected using the MarCCD 165 detector in fast frame transfer mode and reduced via normalization to the incident beam intensity. Buffer scattering was subtracted from protein scattering. This was followed by azimuthal averaging to obtain the intensity $I(q)$ versus q scattering plot visualized by xmgrace. The data were analyzed using PRIMUS (Primary Analysis & Manipulations with Small Angle Scattering Data) version 3.0 from ATSAS 2.3 (*157*), from which Guinier, Kratky, $P(r)$, and CRYSOL analyses were generated.

Computational modeling

Individual *ab initio* molecular envelopes were calculated in GASBOR (*125*); ten to fifteen GASBOR runs were subsequently averaged and filtered using the DAMAVER suite (*158*). The model of RPA-DBC was constructed using PDB coordinates from the X-ray crystal structures of RPA70AB/dC_8 (entry 1JMC) and RPA70C/32D/14 (entry 1L1O). Missing loops in 70C and 32D, as well as the 70B/70C linker, were built manually in PyMol (*159*), then joined and optimized with the Modeller 9v4 interface in Chimera (*160*, *161*). From these starting coordinates, conformational ensembles were generated by rigid body molecular dynamics simulations with BILBO-MD in the presence (R_g 30-50 Å) and absence of R_g restraints (*162*).

To create RPA-DBC/d20 and RPA-DBC/d30 models, docked 70C/dC_4 and 32D/dC_4 structures were generated in HADDOCK, relying upon 'active' residues and protein/DNA base-stacking restraints defined from the RPA70AB/dC_8 crystal structure

(*163*, *164*). DNA-bound 70C and 32D domains were substituted within the original RPA-DBC model and intervening sections of ssDNA were manually added in PyMol (*159*). Subsequent R_g and P(r) calculations for select RPA-DBC models were generated with GNOM from theoretical scattering curves simulated with CRYSOL or the FoXS server (*165*). All molecular graphics were prepared using PyMol (*159*).

Results and Discussion

Preparation and validation of RPA DNA-binding core (RPA-DBC)

Flexible, modular proteins offer many complications to the interpretation of scattering profiles, as their architecture is not fixed in space, but instead varies with time (*124*, *143*, *166*). In particular, the presence of continuous variation in long-range distances can overshadow more subtle architectural changes occurring at closer proximity (*83*). For full-length RPA, the extended, flexible linkers associated with the dedicated protein interaction domains 70N and 32C, as well as the disordered N-terminus of RPA32 (32N), offer the potential to dilute the electron pair distance distribution and obscure changes in the DNA-binding core of RPA. In an effort to simplify the scattering analysis and amplify the architectural features of interest, the decision was made to focus on a construct containing only the DNA-binding core – RPA-DBC (RPA70ABC/32D/14; Figure 4.1). This construct presents only three structural modules for subsequent pair distance distribution analysis – 70A, 70B, and the heterotrimeric core of RPA (70C/32D/14) – with relatively short intervening linkers (10 and 15 residues for 70AB and 70BC, respectively). The architecture of the DNA-binding core in isolation is anticipated to be unchanged from that of the full-length protein, as structural

independence of the excluded domains relative to the DNA-binding core has been established previously, both in the absence and presence of ssDNA (*143*, *152*).

High-yield expression and purification of RPA-DBC were successfully optimized for *E. coli* (Figures 4.2 and 4.4A) (*43*). Subunit masses and identity for the purified protein were validated by electrospray mass spectrometry (ESI-MS) and proteomics analysis, respectively (Figures 4.4B and 4.4). Fluorescence anisotropy binding assays verified similar ssDNA binding activity for RPA-DBC relative to full-length RPA (Figure 4.4C), and the measured binding affinities are comparable to those in the literature (0.9 nM to 0.028 nM; (*55*, *167*)).

Sample homogeneity is critical for scattering studies, as even minor amounts of sample aggregation can strongly distort the scattering (*84*, *123*). Accordingly, monodispersity of RPA-DBC was assessed by SEC-MALS (Superdex 200 PC 3.2/30), both in the absence of ssDNA and bound to ssDNA substrates representative of RPA's three binding modes (10-, 20-, and 30-nucleotides). In order to facilitate consistent placement of RPA-DBC at the 5' end of each substrate, polycytidine sequences were engineered to include a single adenine at position 3 (i.e. d-CCACCCCCCC, d-CCACCCCCCCCCCCCCCCCC, d-CCACCCCCCCCCCCCCCCCCCCCCCCCC (d10, d20, d30, respectively)), as previous studies have indicated preferential binding of

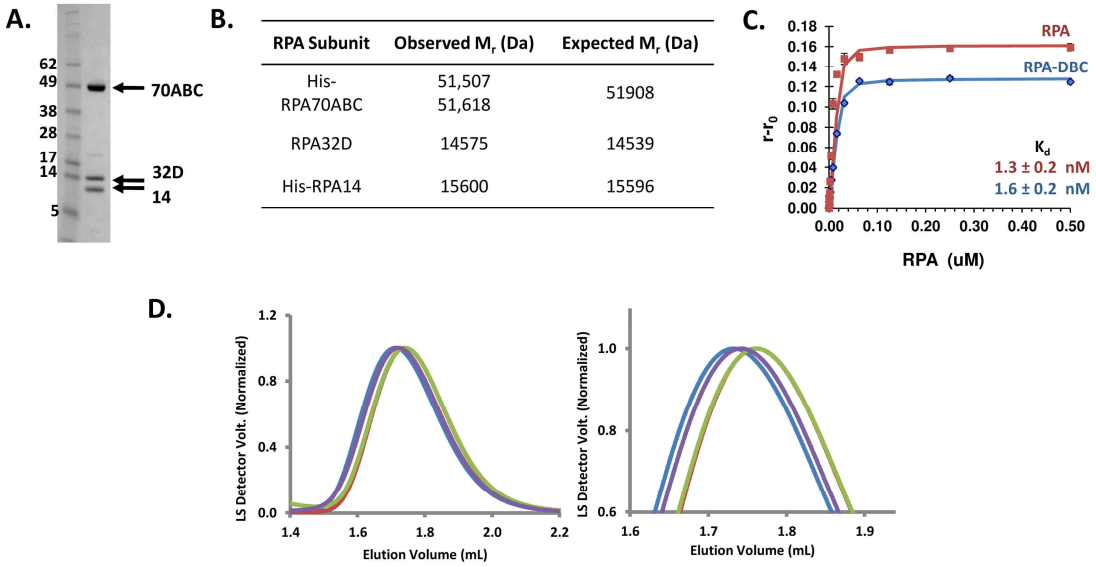


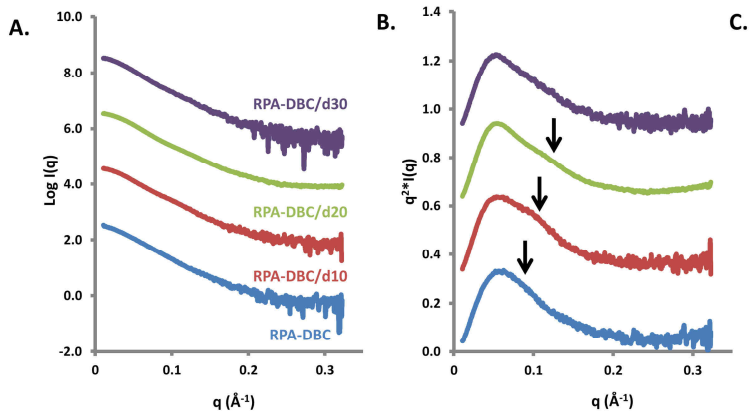
Figure 4.4. Preparation and validation of RPA-DBC. A) SDS-PAGE of purified RPA-DBC in 20 mM HEPES-KOH (pH 7.5), 200 mM NaCl, 10 mM βME, and 5% glycerol. B) ESI-MS analysis of RPA-DBC subunits. C) Fluorescence anisotropy measurement of binding affinity of full-length RPA (■) and RPA-DBC (●) for 5'-FAM-dC$_{30}$. K_d values represent the average of three measurements. D) Elution profiles from SEC-MALS on RPA-DBC alone (blue) and bound to d10 (red), d20 (green), and d30 (purple) substrates.

the 70A domain to d-CCAC (*143*). Monodisperse elution profiles were obtained for all DNA-free and DNA-bound species (Figures 4.4D) with only small amounts of concentration-dependent aggregation eliminated in the void volume. As expected, RPA-DBC in complex with d10 and d20 substrates elute later than the free protein (Figure 4.4D), consistent with the presence of a more compact architecture induced by the binding of ssDNA. Notably, this trend is reversed for the 30-nucleotide binding mode, where the delay in elution, though later, is more consistent with that of the DNA-free protein (Figure 4.4D).

Small-angle x-ray scattering on RPA-DBC alone and bound to ssDNA substrates

Small-angle x-ray scattering profiles were acquired on RPA-DBC in the absence of ssDNA and bound to representative substrates for each binding mode (Figure 4.5A), following a final SEC purification to eliminate trace aggregation and excess ssDNA substrate (see Materials and Methods). Initial assessment of concentration-dependent aggregation and radiation damage by visual inspection of scattering curves underscores the unique solution behavior of each sample. In the absence of ssDNA, RPA-DBC remains highly sensitive to concentration-dependent aggregation (in excess of 2.5 mg/mL) and radiation exposure (greater than 0.5 s) (Figure 4.6A). RPA-DBC/d10 avoids such aggregation at higher concentrations, but retains radiation sensitivity. In marked contrast, RPA-DBC/d20 complexes remain resilient to both aggregation and radiation damage at all concentrations and exposures tested. Unfortunately, an unusual SEC elution profile, contaminated background subtraction, and mostly aggregated concentration series were encountered in the analysis for RPA-DBC/d30, so results for this sample series are inconclusive. Broadly speaking, though, progressive DNA-binding appears to improve RPA-DBC stability and ultimately protects against radiation damage.

Following visual assessment, the most representative scattering curves were selected for further analysis. A radius of gyration (R_g) for each sample was derived from Guinier analysis (Figures 4.5C and 4.6B). Comparison of R_g values across all samples reveals progressive compaction of RPA-DBC upon interaction with d10, then d20



Sample	Rg (Å)
RPA-DBC	38.8 ± 0.2
RPA-DBC/d10	36.3 ± 0.1
RPA-DBC/d20	35.0 ± 0.04
RPA-DBC/d30	38.6 ± 0.2

Figure 4.5. RPA-DBC SAXS data collection. A) Azimuthally averaged scattering profiles (I(q) v. q) for RPA-DBC alone and bound to d10, d20, and d30 substrates. The scattering curve for RPA-DBC/d20 was acquired at four times the concentration of the other samples, which is why the noise at high q values is less in comparison. B) Kratky transformations of scattering data. C) Summary of RPA-DBC R_g values derived from Guinier analysis.

substrates. The change in R_g observed upon binding d10 (2.5 Å) is in excellent agreement with that originally observed from scattering studies of the two principal DNA-binding domains, RPA70AB, bound to an 8-nucleotide substrate (2.2 Å) (*143*). Binding of RPA-DBC to d20 results in an additional decrease in the R_g of 1.3 Å (Figure 4.5C). Interaction with d30, however, restores the R_g to a value nearly identical to that of the DNA-free state, suggesting that RPA's 30-nucleotide binding architecture mirrors the extended state of the DNA-free protein, consistent with the trend initially detected by SEC-MALS.

Kratky transformation of the data (Figure 4.5B) yields broad profiles typical of a multi-domain protein (*83*, *124*), and the progressive redistribution of the primary maximum at q = 0.1 $Å^{-1}$ points toward the presence of discrete conformational changes within RPA-DBC for each increase in substrate length. Curiously, high-q divergence of

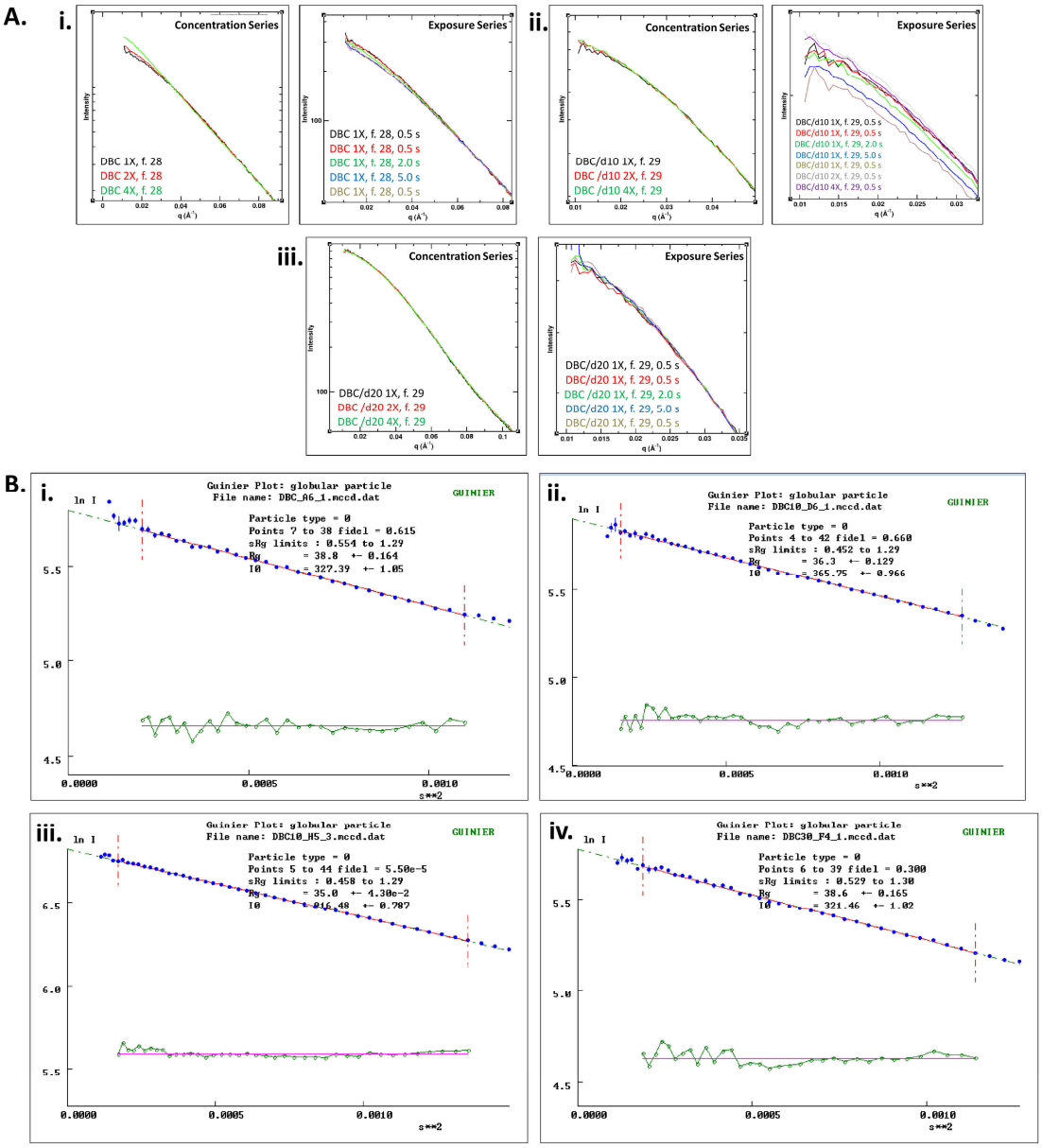


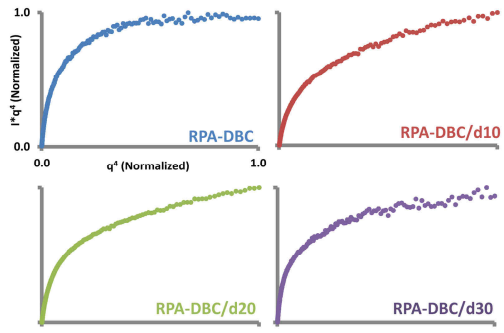
Figure 4.6. Processing and analysis of SAXS data. A) Visual inspection of I(q) versus q scattering curves in Grace. RPA-DBC (i), RPA-DBC/d10 (ii), and RPA-DBC/d20 (iii). Grace plots constructed by Dr. Susan Tsutakawa (Tainer Laboratory, Lawerence Berkeley National Laboratories). B) Guinier analysis of RPA-DBC scattering data. RPA-DBC (i), RPA-DBC/d10 (ii), RPA-DBC/d20 (iii), and RPA-DBC/d30 (iv).

the Kratky plots, a hallmark for partial flexibility in proteins (*88*), is quite subtle for RPA-DBC in the absence of ssDNA and remains unaffected for subsequent DNA-bound complexes (Figure 4.5B). Lower sensitivity within the high q-region of the scattering curve, as well as the lower proportion of flexible linkers relative to the remainder of the protein (~5% of total residues), may hinder detection of internal flexibility by Kratky transformation.

Recent theoretical work has advocated a more sensitive assessment of flexibility by application of the Porod-Debye fourth power law (*168*). In this approach, non-asymptotic behavior within the Porod region of the transformed scattering curve serves as a benchmark for the diffuseness of a biomolecule's electron density contrast, and thus an indicator of high internal density or diffuse flexibility. If the exact mass of the biomolecule is known, qualitative evaluation of the transformed scattering curve can be supplemented by calculations of the packing density ($d_{protein}$) from the Porod volume (V_p).

In the absence of ssDNA, Porod-Debye transformation of the RPA-DBC scattering curve results in a distinct asymptote, indicating well-defined electron density contrast between protein and solvent (Figure 4.7A). Subsequent examination of the packing density verifies that the protein mass is not unilaterally condensed in a single globular mass, as the $d_{protein}$ of 1.08 g/mL is approximately 18% lower than that predicted for a uniform mass distribution, 1.32 g/mL (Figure 4.7B). This is consistent with the discrete inter-domain disposition of RPA-DBC, where the majority of scattering centers fall within structured domains, but flexible linkers preclude confinement of the overall architecture to a single volume.

A.



B.

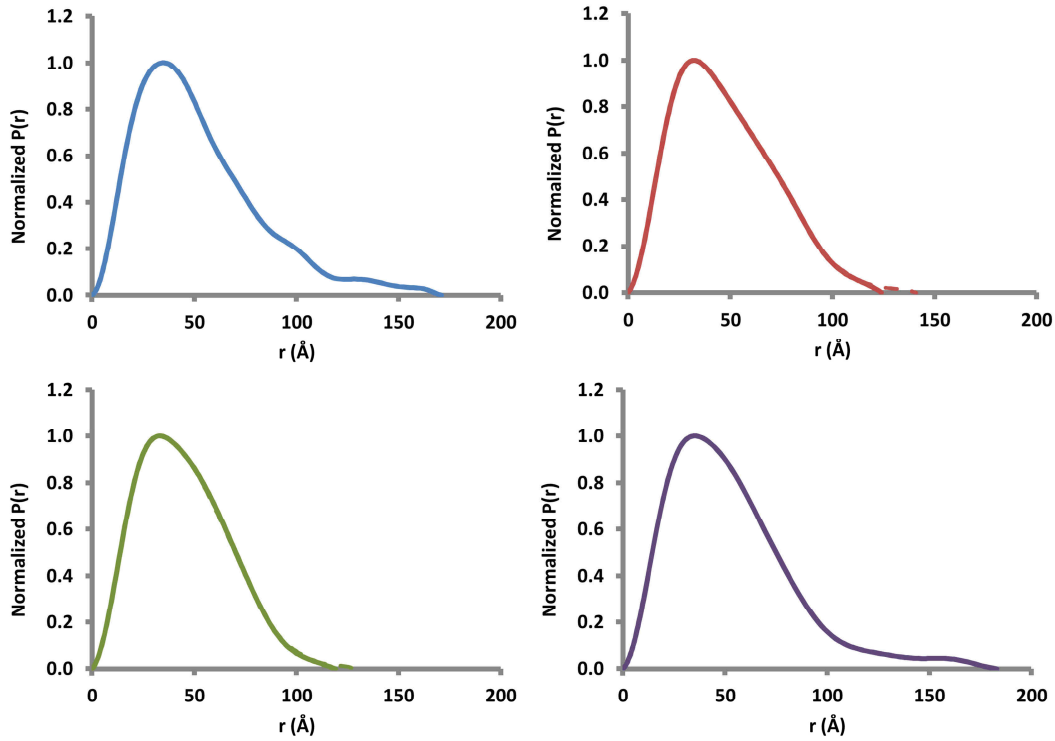
Sample	Porod Volume (Å^3)	$d_{protein}$ (g/mL)	Predicted $d_{protein}$ (g/mL)
RPA-DBC	120,000	1.08	1.32
RPA-DBC/d10	138,000	0.97	1.33
RPA-DBC/d20	134,000	1.03	1.34
RPA-DBC/d30	148,000	0.97	1.34

Figure 4.7. RPA-DBC Porod-Debye analysis. A) Normalized Porod-Debye plots [I(q)*q^4 versus q^4]. B) Summary of Porod-Debye statistics.

When RPA-DBC is bound to ssDNA substrates, however, the asymptote is lost, signifying increased diffuseness in the electron density contrast (Figure 4.7A). This is precisely the reverse, though, of what is anticipated – that is, inter-domain compaction upon DNA-binding is expected to provide an even higher concentration of electron density and thus retention of the Porod-Debye plateau and a general increase in particle packing density. The extended conformation of the bound ssDNA may provide the simplest explanation of the diffuse electron density, though the accompanying increase in Porod volume exceeds that predicted for direct addition of RPA-DBC and ssDNA volumes. Further theoretical and experimental analyses based upon the differential scattering contrasts of protein and DNA will likely be required to resolve this.

Electron pair distance distributions, P(r), were generated for each scattering curve via iterative determination of the maximum dimension D_{max} using GNOM (Figure 4.8). For RPA-DBC/d10 and RPA-DBC/d20 complexes, D_{max} iteration generated two successful D_{max} values, here designated as primary and secondary D_{max}; the essential

A.



B.

Sample	Guinier R_g (Å)	Real-Space R_g (Å)	D_{max} (Å)
RPA-DBC	38.8 (0.2)	42.1	171
RPA-DBC/d10	36.3 (0.1)	37.5 / 38.1	124 / 141
RPA-DBC/d20	35.0 (0.04)	35.3 / 35.6	119 / 128
RPA-DBC/d30	38.6 (0.2)	42.9	183

Figure 4.8. RPA-DBC P(r) distribution analysis. A) Normalized RPA-DBC P(r) distributions. B) Summary of P(r) statistics. Errors for Guinier-derived R_g values are indicated in parentheses.

shape of the distributions were identical regardless of the D_{max} chosen. The P(r) distribution for RPA-DBC exhibits a broad, highly extended tail, characteristic of a flexible, modular protein (*88*, *124*) (Figure 4.8A). Comparison of normalized P(r) distributions from the DNA-bound complexes of RPA-DBC again reveals progressive

compaction of the protein upon binding d10 and d20 substrates, evidenced by loss of the distribution's extended tail and decrease in D_{max} (Figure 4.8A). Binding of d30 restores the original span in distance; however, the distribution retains unique features within the intermediate distance range (50 Å < r < 100 Å) (Figure 4.8A). Examination of the corresponding distance range for RPA-DBC/d10 and RPA-DBC/d20 complexes also highlights density features unique to each complex, providing clear indication of distinct structural changes for each DNA-binding state, coordinate with changes in the breadth of the global molecular envelope.

Inter-domain organization within RPA-DBC

Collectively, comparisons of R_g values, Porod-Debye plots, and P(r) distributions from RPA-DBC alone and bound to ssDNA substrates point toward a general compaction of RPA-DBC architecture upon binding d10 and d20, but release of this compaction upon binding d30, resulting in a final DNA-bound conformation more consistent with the DNA-free protein. To derive more explicit information on specific inter-domain re-arrangements occurring within RPA-DBC during these binding events, a series of molecular envelopes were generated *ab initio* from optimized P(r) distributions using GASBOR, followed by envelope averaging with DAMAVER. With the expectation of intrinsic inter-domain flexibility for RPA-DBC (that is potentially retained upon binding ssDNA), molecular envelopes should be viewed as averaged representations of inter-domain distribution, rather than definitive reconstructions of fixed architectures (*124*, *143*). This is supported by the relatively poor goodness-of-fit values (5.0 – 6.7)

generated from comparison of experimental P(r) distributions to individual GASBOR envelopes.

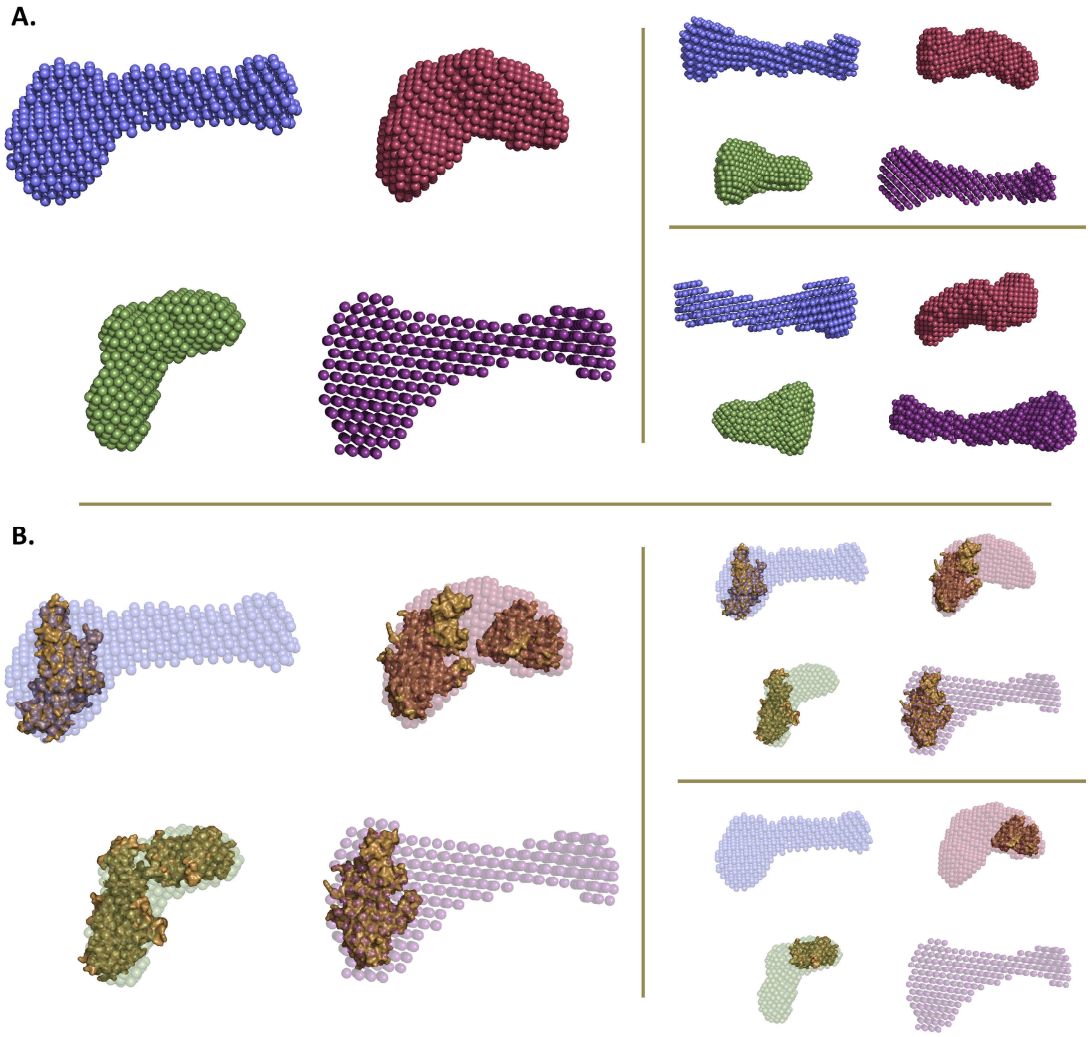


Figure 4.9. Averaged RPA-DBC molecular envelopes. A) GASBOR averaged molecular envelopes. The views to the right are a 90° rotation from the bottom (above), followed by a 180° rotation from the left (below. RPA-DBC (blue), d10 (red), d20 (green), d30 (purple). B) Manual comparison and docking of 70C/32D/14 (1L1O, above right) and 70AB/dC_8 (1JMC, below right) crystal coordinates with scattering envelopes.

The averaged RPA-DBC envelope reveals the presence of a comparatively linear molecular volume, anchored by a primary lobe of density that remains constant across all

DNA-bound complexes (Figure 4.9A). The arm of density which extends from this anchoring lobe is noticeably condensed in the RPA-DBC/d10 and RPA-DBC/d20 complexes, but is again lengthened for RPA-DBC/d30, albeit with a slightly different profile relative to the DNA-free state (Figure 4.9A). Because the extended arm of density is the primary portion of the envelope impacted upon introduction of d10, it is tentatively assigned as the location of principal DNA-binding domains 70A and 70B, the primary domains involved in the 10-nucleotide binding mode. The relatively larger size of the anchoring lobe and the consistency of its shape for all DNA-binding states are suggestive of RPA's trimer core.

To determine the feasibility of this inter-domain arrangement, molecular envelopes from previous scattering studies of RPA70AB (*143*), as well as published crystal structures of the trimer core (RPA70C/32D/14) (*43*) and RPA70AB bound to a dC_8 substrate (*40*) were manually docked into RPA-DBC molecular envelopes (Figure 4.9B, 4.10). The RPA trimer core, as represented by the crystal structure, can be accommodated by the anchoring lobe for all cases (Figure 4.9B) and clearly exceeds the volumes of the remaining densities, which are attributed to 70A and 70B. Correspondence between the trimer core and the molecular envelopes, while comparable, is not entirely optimal, particularly for the domain 70C, which possesses a zinc-ribbon motif that tends to protrude beyond the envelope (Figure 4.9B). Superposition of the bilobal RPA70AB scattering envelope upon the extended density of RPA-DBC *sans* ssDNA shows excellent correspondence between the two, indicating that structural averaging from the dynamic inter-domain motion of RPA70AB is preserved in the context of the full DNA-binding core (Figure 4.10A) (*143*). By analogy, the intervening

density between 70B and the trimer core region is consistent with dynamic motion between these two modules.

Direct translation of the more compact molecular envelope from RPA70AB/dC_8 to those of RPA-DBC bound to ssDNA is not quite as straightforward. Overlays with RPA-DBC/d10 and RPA-DBC/d20 show localized areas of mismatched density, while RPA-DBC/d30 lacks all correspondence with the smaller envelope (Figure 4.10B). Interestingly, substitution of the RPA70AB/dC_8 crystal structure within RPA-DBC/d20 shows marked improvement in fitting of the density (Figure 4.10B). In solution, RPA70AB/dC_8 retains residual inter-domain dynamics (*143*). Equivalent residual inter-domain motion would be expected for RPA-DBC/d10; for this case, it is possible that the 2-nucleotide increase in substrate length (10 versus 8) might be sufficient to yield a slightly different molecular envelope. The engagement of three DNA-binding domains in the RPA-DBC/d20 complex, however, would effectively tether the 70B domain to its substrate, which would explain this envelope's superior correspondence to the fixed crystal structure. For RPA-DBC/d30, failure to accommodate either crystal structure or scattering envelope suggests that the 70A and 70B domains experience a different averaged conformation on their substrate. Their inter-domain motion, if present, is not identical to the DNA-free protein, as the DNA-free RPA70AB scattering envelope fails to overlay the RPA-DBC/d30 envelope effectively (Figure 4.10C).

In comparing RPA-DBC molecular envelopes with and without ssDNA, the pronounced compaction of RPA-DBC/d10 relative to RPA-DBC *sans* ssDNA becomes evident (Figure 4.9A). In fact, the RPA-DBC/d10 envelope more closely resembles the general shape of RPA-DBC/d20, rather than that of free RPA-DBC, particularly when

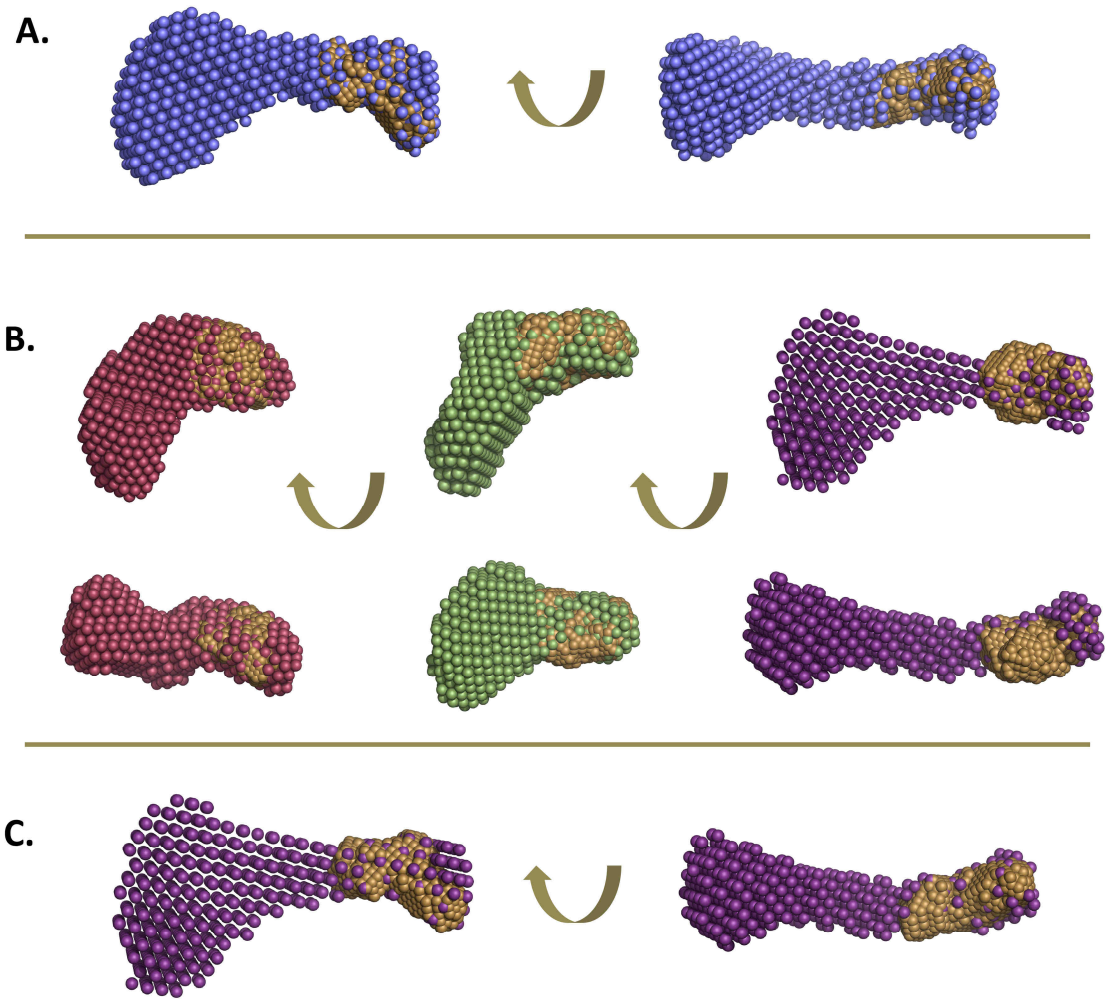


Figure 4.10. Comparison of RPA-DBC and RPA70AB molecular envelopes. A) Overlay of RPA-DBC (blue) and RPA70AB (gold) envelopes in the absence of ssDNA. B) Overlay of DNA-bound RPA-DBC (d10 in red, d20 in green, and d30 in purple) and RPA70AB/dC_8 (gold) envelopes. C) Overlay of DNA-free RPA70AB envelope (gold) with RPA-DBC/d30 (purple).

comparing the density in between the trimer core and 70B (even though the linker is expected to remain extended upon binding d10). It is possible that this excessive compaction simply reflects the altered hydrodynamic state of 70A and 70B when jointly bound to ssDNA, where the increase in effective mass and hydration for a 70AB/d10 module produces greater viscous drag relative to the trimer core, and hence a less

extended range of inter-domain orientations. Insights into whether this molecular scattering volume is consistent with the ensemble-averaged orientation of RPA-DBC/d10 can be obtained from computational simulations (see below).

Apart from this, comparison of RPA-DBC molecular envelopes verifies that the initial compaction for RPA's first binding mode involves binding and alignment of the principal DNA-binding domains, 70A and 70B. Progression to the 20-nucleotide binding mode causes maximum compaction of the DNA-binding core with direct apposition of 70AB and the trimer core. The structural rearrangements underlying transition to the 30-nucleotide binding mode, however, are less intuitive and may involve redistribution of the DNA-binding modules across the span of the ssDNA substrate, increased dynamic spatial averaging within the DNA-binding core, or some combination of the two that would result in an averaged, extended envelope similar to that of the DNA-free state.

While docking studies support the proposed arrangement (and rearrangement) of domains within RPA-DBC scattering volumes, the location of the DNA-binding site remains unresolved – whether it lies on the convex or concave face of the molecular envelope (Figure 4.9). Taking RPA-DBC/d20 as an example, a convex DNA-binding site would allow for a contiguous path from 70AB to the trimer core with moderate curvature of the ssDNA substrate. Such an inter-domain arrangement, however, forces an entropically unfavorable steric proximity between the 70B and 70C domains (Figure 4.11A). A concave DNA-binding site would obviate the steric strain between 70B and 70C (Figure 4.11B), but would destroy the continuity of the binding channel, forcing the ssDNA substrate either to kink back to make sequential contact with all the DNA-binding

sites or to skip binding of the 70C domain in favor of maintaining a less strained conformation.

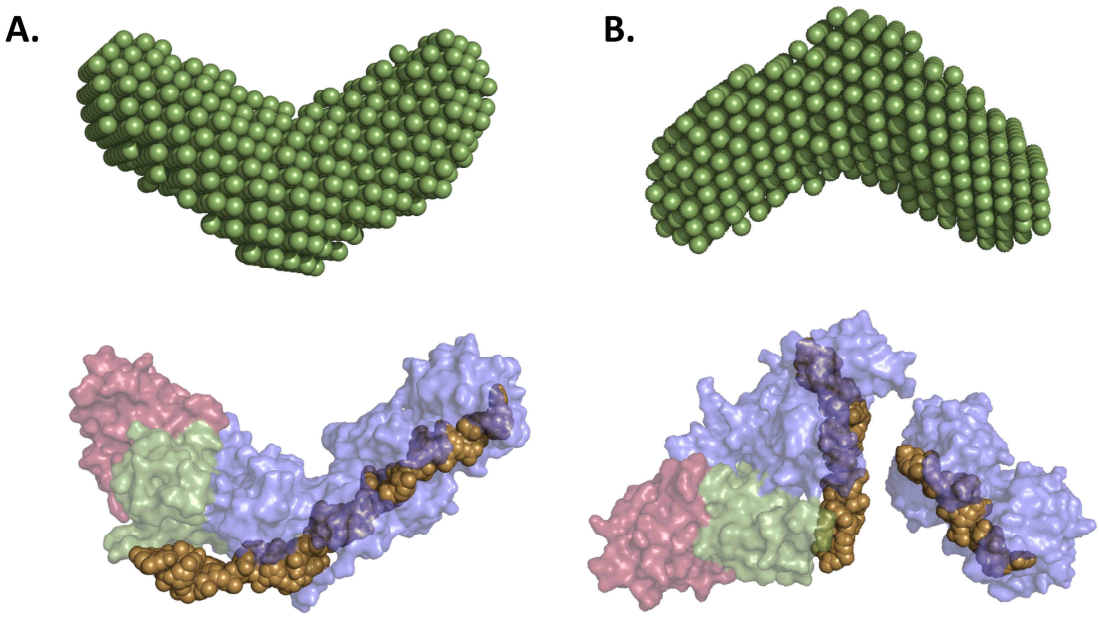


Figure 4.11. Potential RPA-DBC inter-domain orientations assuming a convex (A) or concave (B) ssDNA-binding site. RPA70 domains (blue), RPA32D (green), RPA14 (red), ssDNA (gold).

RPA-DBC ensemble modeling indicates limited extension in the absence of DNA

While analysis of ensemble-averaged RPA-DBC molecular envelopes allows insight into specific inter-domain re-arrangements accompanying each DNA-binding mode, how the path of ssDNA evolves across these transitions and whether DNA substrates are stationary or fluctuating within the binding cleft remains undetermined. Moreover, inter-domain dynamics within RPA-DBC, both in the presence and absence of ssDNA, are expected to play a critical role in the plasticity and organization of the modular DNA-binding cleft during binding and release of ssDNA substrates, particularly

in the context of DNA processing assemblies. Density averaging within the molecular envelope, unfortunately, precludes access to this specific dynamic information.

Computational simulations provide a valuable avenue for exploring the conformational sampling capacity of flexible, modular proteins (*89*, *124*, *162*). For the purposes of these studies, RPA-DBC molecular dynamics (MD) simulations were run using BILBO-MD, a rigid body MD program that uses the CHARMM forcefield (*162*). As the BILBO-MD algorithm is currently unable to accommodate nucleic acids, analysis was limited to RPA-DBC and a 'mock' version of RPA-DBC/d10, where domains 70A and 70B were defined as a single rigid body in the orientation of the dC_8 co-crystal structure. A starting model of RPA-DBC was constructed from published 70AB and trimer core crystal structures (*40*, *43*); linkers and missing loops were added using the Modeller extension of Chimera (*160*, *161*). From this model, 6600 confomers were generated under a constrained R_g range of 30-50 Å for RPA-DBC *sans* ssDNA, as well as RPA-DBC/d10 with the modified rigid body boundaries described above.

Plots of individual R_g and D_{max} values extracted from each conformer versus goodness-of-fit to the experimental data are displayed in Figure 4.12. While R_g distributions exhibit best-fit minima consistent with the experimentally derived values (Figure 4.12), D_{max} values are systematically underestimated (c.f. 135 Å simulated D_{max} to 171 Å experimental D_{max} for RPA-DBC). Review of the CHARMM simulation trajectories from select R_g categories revealed that R_g restraints imposed upon the volume explored by RPA-DBC forced compaction (R_g 30 Å) or extension (R_g 50 Å) of the protein. Simulation trajectories for RPA-DBC also highlighted different domain

sampling properties, depending upon the choice of reference module (domain held fixed in space during the simulation). To ensure adequate exploration of conformational space

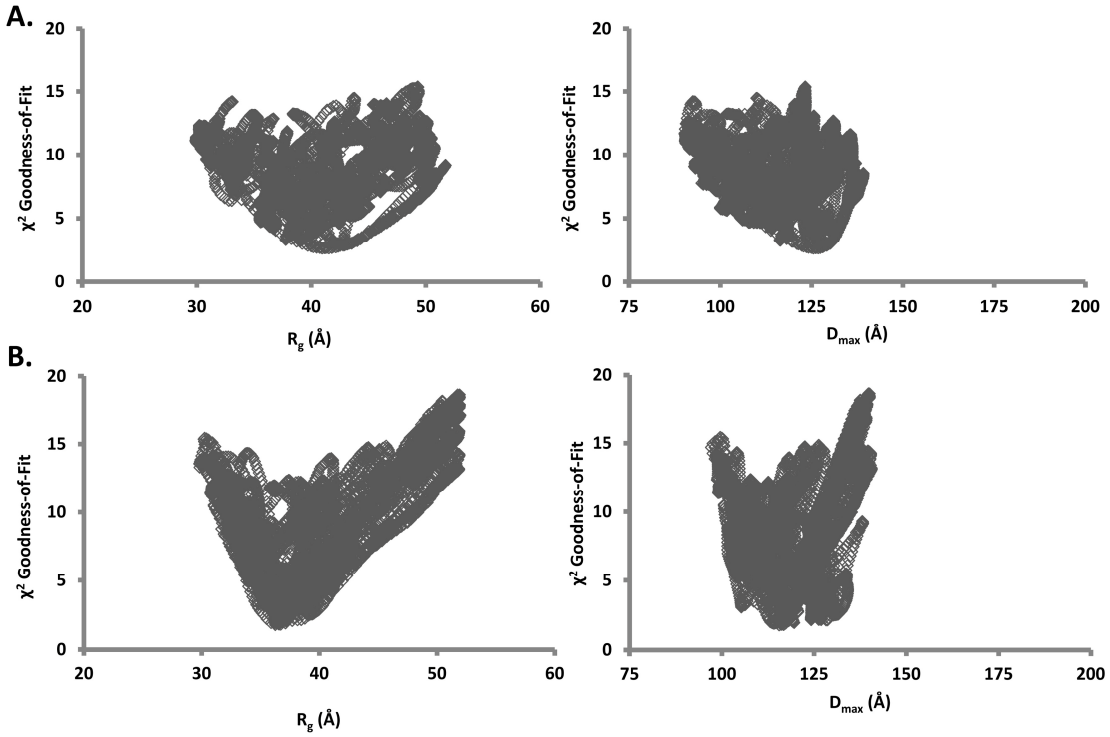


Figure 4.12. Summary of RPA-DBC BILBO-MD ensembles in the presence of R_g restraints. A) Plots of χ^2 goodness-of-fit versus R_g and D_{max} for RPA-DBC. B) Plots of χ^2 goodness-of-fit versus R_g and D_{max} for mock RPA-DBC/d10.

without steric interference from domain clashing, the central 70B domain was selected as the reference module for subsequent RPA-DBC simulations.

The simulations were repeated, this time in the absence of R_g restraints, generating a total of 15,000 conformers each for RPA-DBC and RPA-DBC/d10. Plots of unrestrained R_g and D_{max} values versus experimental goodness-of-fit provide a more

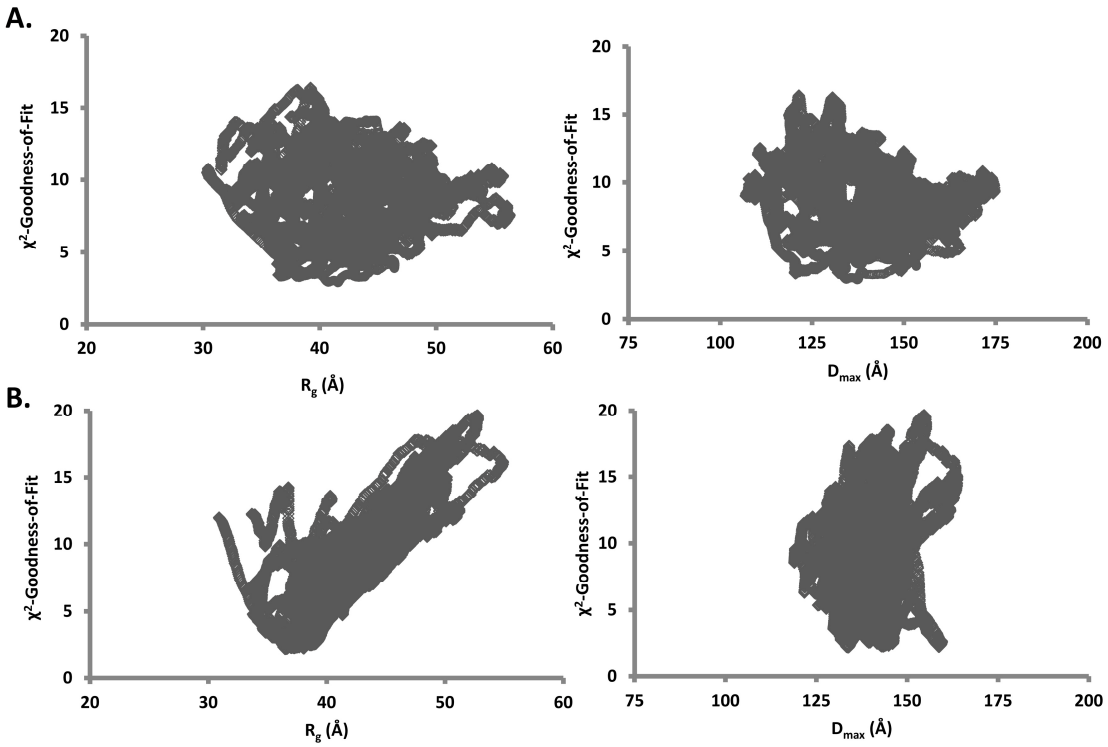


Figure 4.13. Summary of RPA-DBC BILBO-MD ensembles in the absence of R_g restraints. A) Plots of χ^2 goodness-of-fit versus R_g and D_{max} for RPA-DBC. B) Plots of χ^2 goodness-of-fit versus R_g and D_{max} for mock RPA-DBC/d10.

complete representation of RPA-DBC conformational space (Figure 4.13). While small gaps throughout these distributions suggest the potential for additional conformational sampling, symmetric distributions of maximum x, y, and z distances from each set of simulations confirm that the conformational sampling is sufficient to permit comparison to the experimental data (data not shown). For RPA-DBC *sans* ssDNA, the absence of a well-defined minimum in the goodness-of-fit plots for R_g and D_{max} indicates that multiple inter-domain orientations are capable of describing the experimental scattering data equally well (Figure 4.13), consistent with the P(r) and molecular envelope analysis. The presence of a broad minimum in the goodness-of-fit plots from RPA-DBC/d10, however, suggests that the average range of inter-domain orientations consistent with the

experimental scattering data is now more limited, again reflective of the P(r) and molecular envelope analysis and consistent with compaction upon binding ssDNA.

Intriguingly, the average radius of gyration for both simulated RPA-DBC ensembles is appreciably greater than the experimental values (Figure 4.14A). In fact, for both RPA-DBC and RPA-DBC/d10, experimental R_g values fall at the sparsely populated, lower end of each simulated R_g range (Figure 4.14B). Thus, the simulated population distribution generated by BILBO-MD is not consistent with the actual values in solution. This points toward inaccuracies in the simulation, which likely arise from the high temperature applied during the trajectory and a lack of consideration for molecular hydration. The greatest D_{max} from the simulated RPA-DBC ensemble, however, remains consistent with the experimental value (Figure 4.14A), suggesting that this fully extended conformation is sampled in solution. The greatest D_{max} from RPA-DBC/d10, though, is approximately 20-40 Å greater than the experimental value (Figure 4.14A), again pointing to simulation inaccuracy. When compared to theoretical D_{max} values from the trimer core (D_{max} 93 Å) and 70AB/dC_8 (D_{max} 60 Å), the average D_{max} values for both simulated RPA-DBC ensembles (139-140 Å) suggest that the most frequent arrangement of 70A, 70B, and trimer core modules within the simulated ensemble is not necessarily linear as indicated by the *ab initio* envelopes, but rather arced.

To gain insight into the conformational architectures, selected BILBO-MD conformers close to the experimental R_g and D_{max} were extracted and compared (Figure 4.15). As expected, RPA-DBC conformers representative of the experimental R_g had an arced, rather than a linear inter-domain distribution, contrary to the *ab initio* molecular envelope (Figure 4.15A, left panel). This curvature is lost, though, for

A.

Sample	Experimental R_g (Å)	Simulated Average R_g (Å)	Experimental D_{max} (Å)	Simulated Maximum D_{max} (Å)	Simulated Average D_{max} (Å)
RPA-DBC	38.8	42.3	171	174.7	139.1
RPA-DBC/d10	36.3	43.2	124 / 141	163.8	140.1

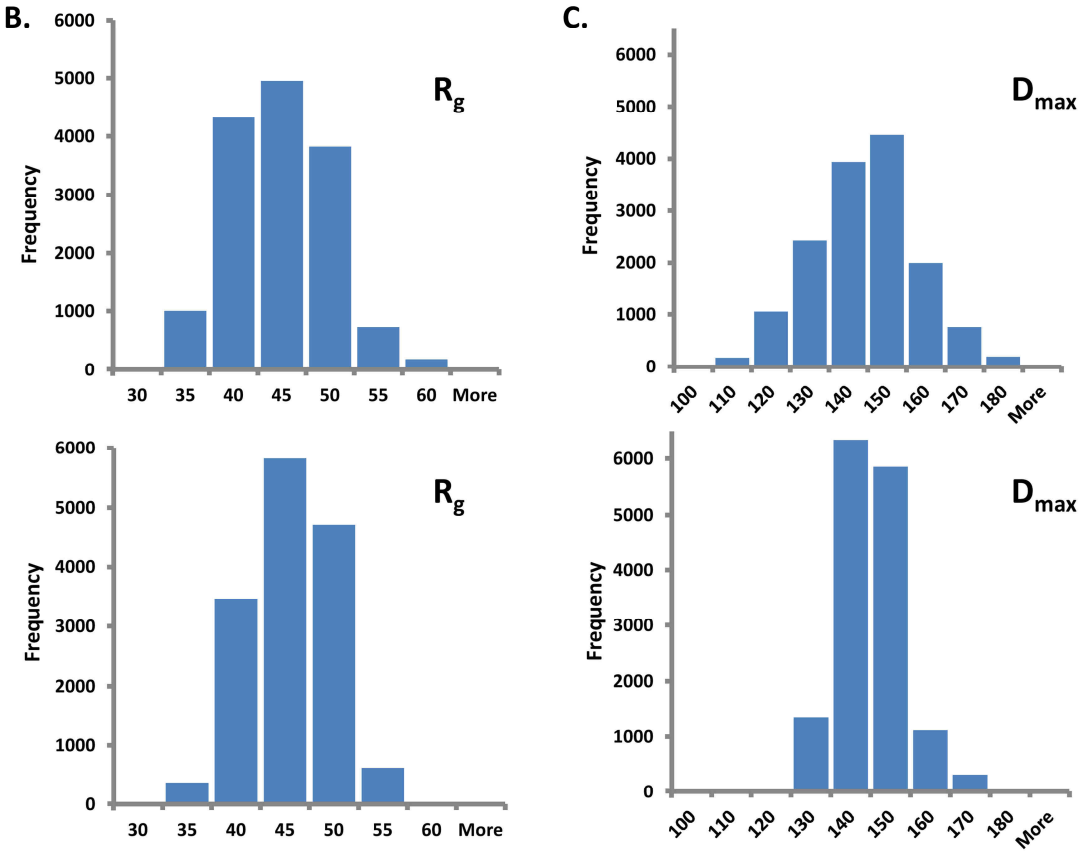


Figure 4.14. Comparison of parameters from simulated BILBO-MD ensembles to experimental scattering parameters. A) Summary of simulated and experimental statistics. B) Frequency distributions of R_g for simulated RPA-DBC (top) and RPA-DBC/d10 (bottom) ensembles. C) Frequency distributions of D_{max} for simulated RPA-DBC (top) and RPA-DBC/d10 (bottom) ensembles.

conformers consistent with the experimental D_{max}, where all domains are at their maximum, linear extension (Figure 4.15A, right panel). A similar pattern is observed for

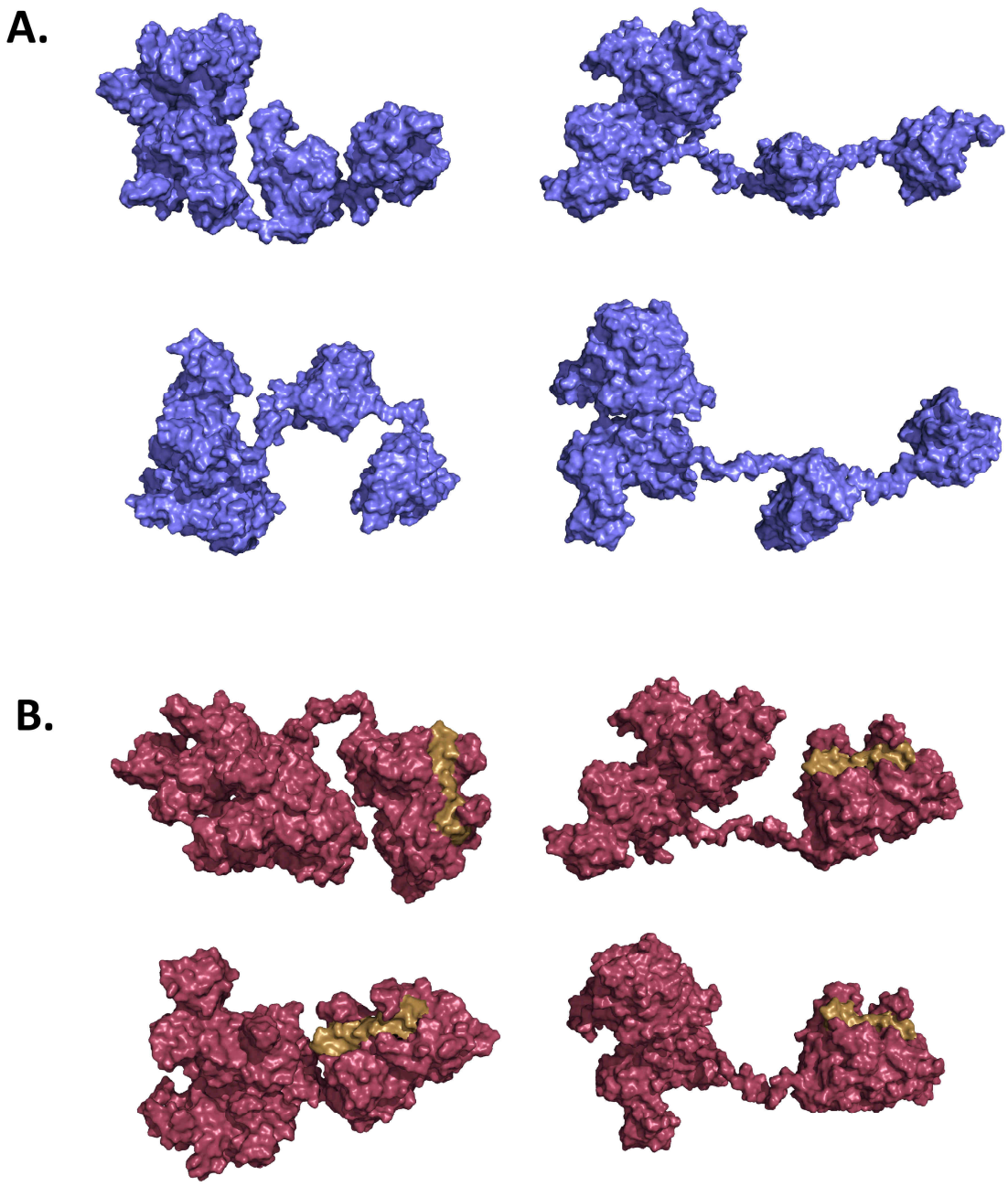


Figure 4.15. Representative BILBO-MD conformers for experimental R_g (left panels) and D_{max} (right panels) values for RPA-DBC (A) and mock RPA-DBC/d10 (B).

RPA-DBC/d10, where conformers consistent with the experimental R_g reflect a closer inter-domain organization (Figure 4.15B, left panel), as opposed to the more open

arrangement associated with conformers displaying the experimental D_{max} (Figure 4.15B, right panel).

As mentioned above, discrepancies between average experimental and simulated parameters point toward inaccuracies in the simulation. It would seem that BILBO-MD is able to sample the breadth of accessible inter-domain orientations, but is unable to account for the relative abundance of these orientations within a population. This partly arises from running the simulation under high temperatures (1500K), which can exaggerate the sampling of extended inter-domain orientations that may be less accessible under physiological temperatures. The rigid-body protocol employed by BILBO-MD may also favor oversampling of more extended conformations by overlooking the effects of hydrodynamic viscosity. Specifically, the program relies upon designation of a stationary, reference domain, from which neighboring 'moving' domains are iteratively rearranged after MD simulation of inter-domain linkers (*128*). As such, conformational effects imposed by hydrodynamic drag from rotational diffusion of the protein through solvent (including semi-independent diffusion from *all* mobile domains) are neglected. Contributions from transient inter-domain attraction or repulsion driven by electrostatic, van der Waals, or hydration effects are ignored as well. In light of this, determining whether the inter-domain arrangements associated with the experimental RPA-DBC R_g values (Figure 4.15) truly represent the most abundant orientation in solution will require advanced molecular dynamics approaches that apply more realistic solvent and temperature parameters.

While an accurate description of the full RPA-DBC solution ensemble is not accessible through BILBO-MD, the simulations do provide important information on

inter-domain distribution that is unavailable from examining averaged molecular envelopes. It is clear that multiple RPA-DBC conformations are present in solution; however, the current evidence suggests that the most abundant conformations exhibit limited inter-domain extension. A similar conclusion applies to simulations of RPA-DBC/d10. In light of this, the pronounced compaction of the RPA-DBC/d10 molecular envelope relative to RPA-DBC, noted above (Figure 4.9A), may simply reflect the conformational predisposition of the RPA-DBC ensemble. As a result, the greater extension in the molecular envelope of RPA-DBC relative to those of RPA-DBC/d10 and RPA-DBC/d20 may arise from ensemble averaging over a more diverse distribution of inter-domain orientations, rather than an architecture that is predominantly extended in solution.

RPA-DBC assumes a close, curved inter-domain orientation in the presence of ssDNA

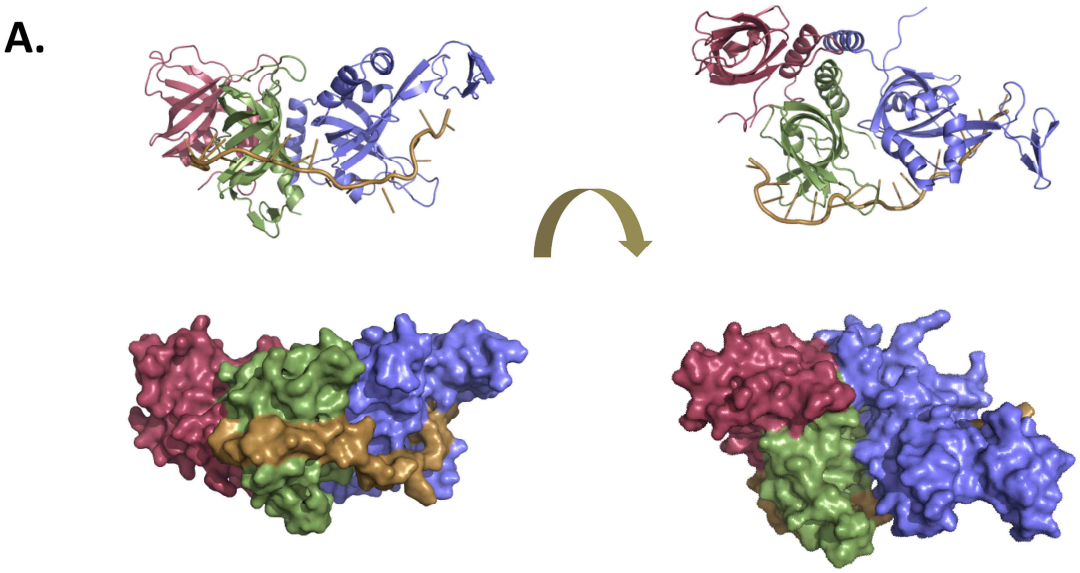
Evidence for limited inter-domain extension within the RPA-DBC solution ensemble would indicate a similar inter-domain distribution for RPA-DBC/d30 that is not readily detectable by molecular envelope comparison. As BILBO-MD is unable to accommodate nucleic acids, select models of RPA-DBC in complex with d20 and d30 substrates were built by hand to provide insight into the correspondence between specific inter-domain distributions and experimental parameters derived from the scattering data (R_g, D_{max}, P(r)). Complexes of 70C and 32D bound to dC_4 substrates were generated in HADDOCK (*163*, *164*), based upon restraints defined from critical protein-DNA base-stacking interactions in the 70AB/dC_8 crystal structure. Docked domains were overlaid

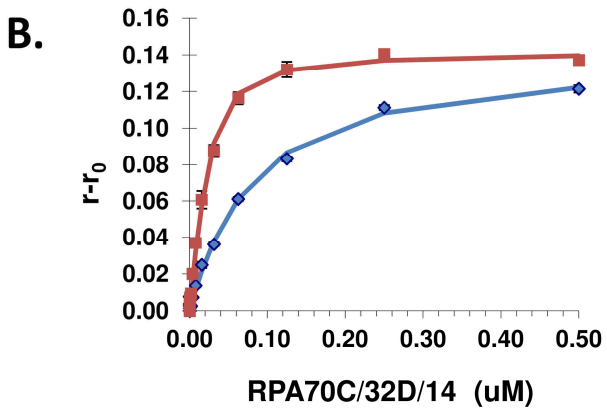
with the original RPA-DBC model (see above) in Pymol and Chimera, and intervening segments of the ssDNA substrate were added manually.

Prior to comparison with the experimental data, assembly of the models led to a number of observations. The first of these regards the minimum and maximum number of intervening nucleotides accommodated by 70A/70B and 70B/70C. For both cases, a minimum of two intervening nucleotides is required to prevent steric overlap between domains (the 70AB/dC_8 crystal structure displays this as well) (*40*). Separation of 70A and 70B by more than 4 nucleotides leads to overextension of the 70AB linker, whereas approximately 6-8 intervening nucleotides led to full extension between 70B and 70C.

A second observation considers the substrate capacity of the trimer core, where 7 bases are required to bridge the 70C and 32D binding sites on adjacent faces of the trimer core (Figure 4.16). With three nucleotides engaging each DNA-binding domain, this suggests that at least 13 nucleotides are required to fill the trimer core completely. Measurements of RPA70C/32D/14 affinity for 10- and 20-nucleotide substrates by fluorescence polarization are consistent with this, as the trimer core's affinity for the 20-nucleotide substrate is greatly increased relative to that for the 10-nucleotide substrate (Figure 4.16). Altogether, these measurements would predict a minimum substrate length of 23 nucleotides (8 for 70AB, 2 for 70B/70C, 13 for trimer core) in order to occupy all RPA-DBC DNA-binding domains completely. This substrate length is consistent with the 32D domain remaining unoccupied with a 20-nucleotide substrate, as has traditionally been defined for the intermediate binding mode. For a 30-nucleotide substrate, though, interaction with this minimum substrate length could result in RPA-DBC binding at 8 slightly different locations along the length of the 30-nucleotide substrate. Consequently,

future experiments examining the final binding mode should consider limiting the substrate to 23-24 nucleotides in order to reduce potential variations in RPA-DBC's binding register.





Substrate	K_d (nM)
5'-FAM-dC_{10}	144 ± 6
5'-FAM-dC_{20}	29 ± 3

Figure 4.16. A) Model of RPA70C/32D/14 bound to ssDNA substrate. B) Fluorescence anisotropy binding curves for RPA70C/32D/14 titrations with 5'-FAM-dC_{10} (●) and 5'-FAM-dC_{20} (■). K_d values represent the average of three measurements.

Three models were generated for RPA-DBC/d20 with the intent of examining the role of inter-domain spacing and linear domain arrangement in complementing the

scattering data (Figure 4.17). The first model (panel A) considers the influence of steric freedom between the modules of the DNA-binding site by incorporating six nucleotides between 70AB and the trimer core. This model is meant to test whether the entropically favorable steric freedom gained by allowing this extra inter-modular spacing is still

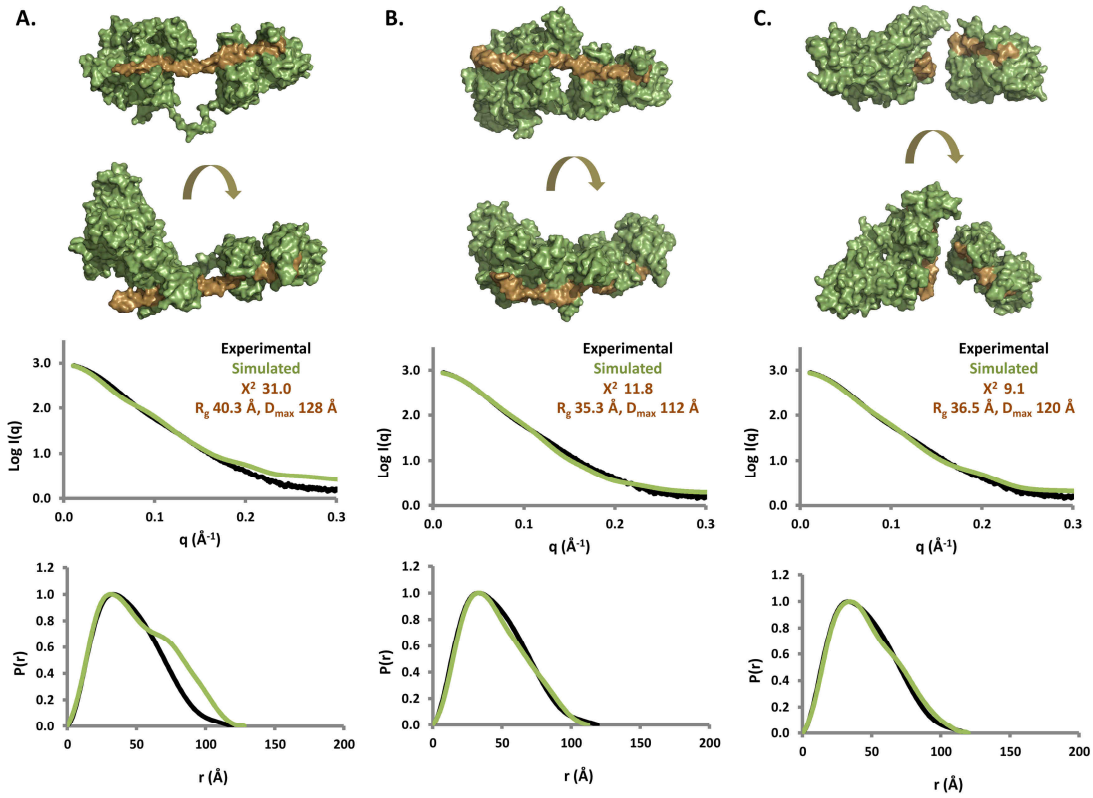


Figure 4.17. Modeling of RPA-DBC/d20 complexes with overlays of experimental (black) and theoretical (green) I(q) curves and P(r) distributions. A) Model 1 (extended inter-domain orientation). B) Model 2 (linear inter-domain orientation). C) Model 3 (arced inter-domain orientation).

consistent with the compaction observed in the *ab initio* envelope and P(r) distribution (Figure 4.17A, the 32D domain remains unoccupied in this case). The simulated scattering function from this model differs considerably from the experimental data (χ^2 31.0, Figure 4.17A), while the R_g (~40.3 Å) clearly exceeds that derived by Guinier

analysis (34.9 Å). Moreover, the experimental P(r) distribution lacks the secondary shoulder of the simulated distribution, which arises from spatial separation of 70AB and the trimer core. Overall, it would appear that an extended model is not consistent with the experimental data, suggesting that proximity between 70AB and the trimer core are favored in the intermediate binding mode.

The second and third models examine whether a proximate arrangement of domains within the RPA-DBC/d20 complex are aligned linearly or as a curved arc, evaluating the potential for curvature in the modular DNA-binding cleft as noted in the Bilbo-MD simulations. The second model presents a relatively linear arrangement of the central domains (Figure 4.17B), while the third model attempts to accommodate the sharply curved arrangement indicated by the RPA-DBC/d20 molecular envelope (Figure 4.17C). Here, R_g values from both models are consistent with the experimental data (R_g 35.3 and 36.5 Å, respectively) and their χ^2 goodness-of-fit values are relatively similar (compare 11.8 to 9.1). While these two models more closely approximate the experimental P(r) distribution than the model with broader inter-domain spacing (Figure 4.17A), the relatively high χ^2 values indicate that neither is a particularly good match to the data. As such, determination of the precise geometric arrangement of domains (linear versus arced) clearly warrants further investigation by more advanced modeling approaches than those presented here. Consideration of dynamics and ensemble-based modeling strategies will also likely prove insightful.

In the case of RPA-DBC/d30, two models were generated with the intent of examining the inter-domain proximity on the longer substrate (Figure 4.18). The first considered maximum spread between 70A/70B and 70B/70C with 4 and 6 bases,

respectively, incorporated between these domains (Figure 4.18A). The second model retained the extended placement between 70B and 70C, but substituted the minimum two-base separation between 70A and 70B that is observed in the crystal structure (Figure 4.18B). R_g values for both models were 43.5 Å (fully extended) and 41.1 Å (partially extended), which is considerably higher than the experimental value of 38.6 Å. Moreover, the associated P(r) distributions possess distinct secondary shoulders that are not present in the experimental distribution (Figure 4.18), representing closer association between 70AB and the trimer core than the modeled inter-domain spacing. The D_{max} values of the models (144 and 131 Å, respectively), however, fall well short of the experimental D_{max} of 183 Å. As with the DNA-free protein, the majority of RPA-DBC/d30 molecules in solution appear to adopt a less extended state than might be expected from a well-spaced, linear alignment of the DNA-binding domains along the 30-nucleotide substrate. The existence of a high experimental D_{max}, though, indicates the presence of a population distribution that includes highly extended orientations within the ensemble. Interestingly, the intra-domain features in the RPA-DBC/d30 P(r) distribution ($r < 100$ Å) remain very similar to those of RPA-DBC/d20 (Figure 4.17), suggesting that the interdomain orientations in the two binding modes may be similar.

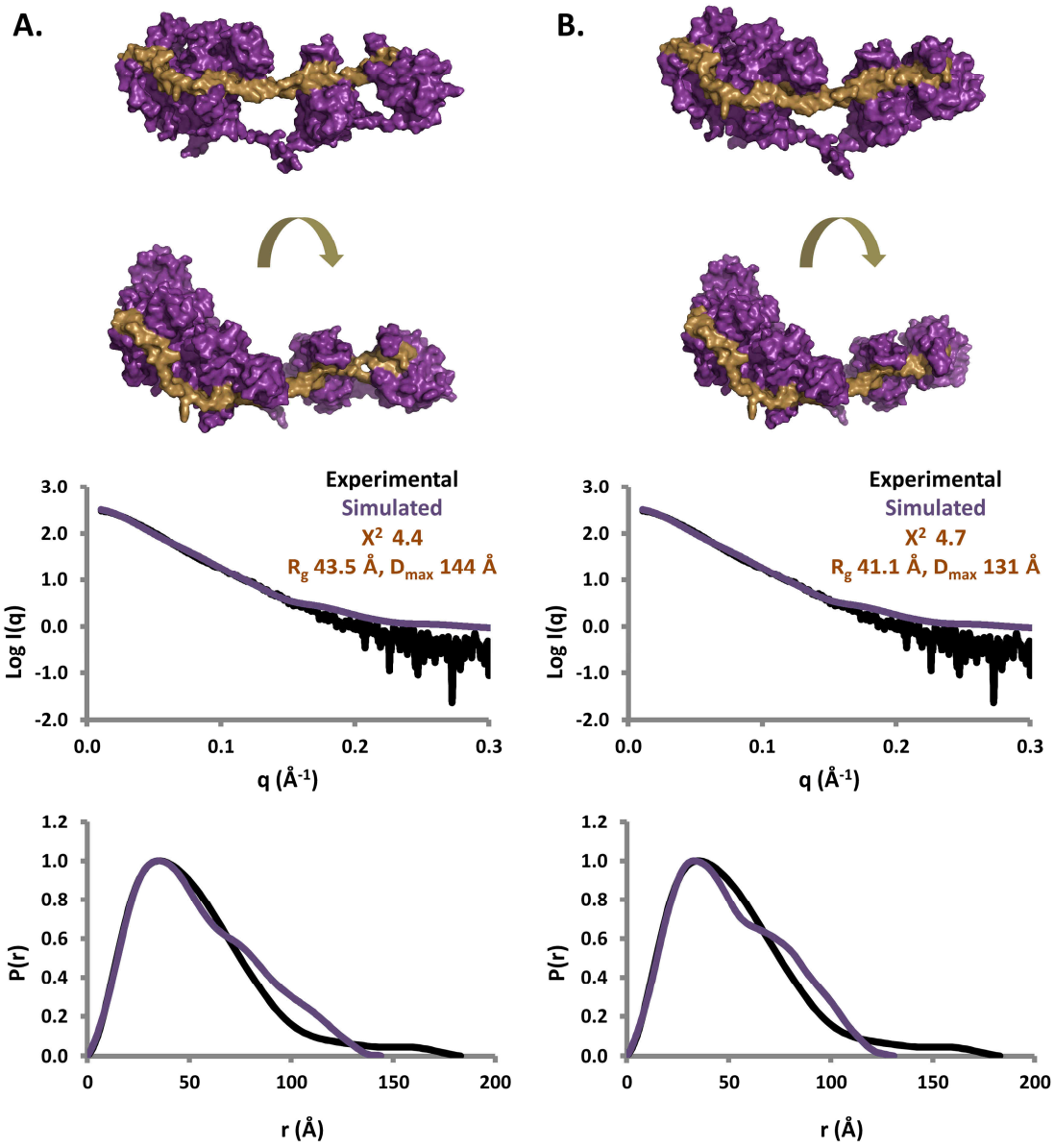


Figure 4.18. Modeling of RPA-DBC/d30 complexes with overlays of experimental (black) and theoretical (purple) I(q) curves and P(r) distributions. A) Model 1 (maximum 70A/70B and 70B/70C spacing). B) Model 2 (maximum 70B/70C spacing).

RPA-DBC follows a dynamic 'bi-modal' mechanism of DNA-binding

Comparisons between RPA-DBC models generated by Bilbo-MD and manual methods and their associated P(r) distributions were used to correlate specific inter-domain orientations with unique features of the experimental P(r) distributions. The results are summarized in Figures 4.19 and 4.20.

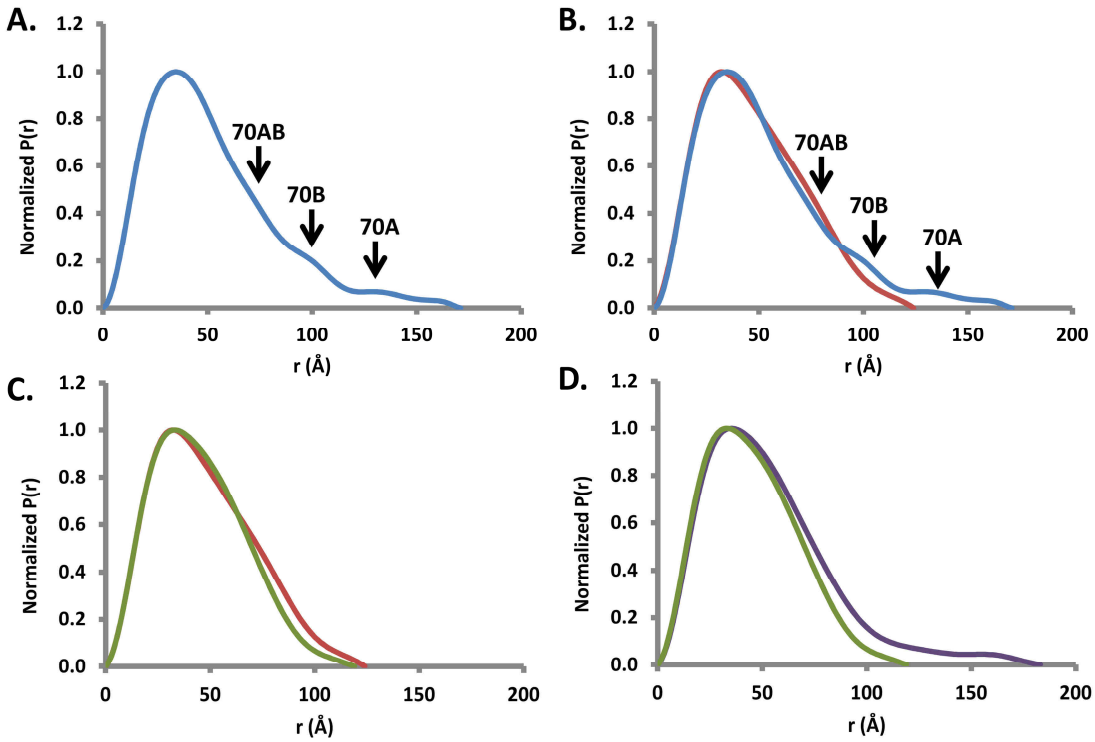


Figure 4.19. Integrating RPA-DBC modeling and experimental P(r) distributions. A) RPA-DBC P(r) distribution. Arrows indicate distances between the given module and the trimer core. As an example, the local 70A feature centered at 130 Å represents all distances between 70A and the trimer core. B) Comparison of RPA-DBC (blue) and RPA-DBC/d10 (red). C) Comparison of RPA-DBC/d10 (red) and RPA-DBC/d20 (green). D) Comparison of RPA-DBC/d20 (green) and RPA-DBC/d30 (purple).

In the absence of ssDNA, the P(r) profile for RPA-DBC contains a primary maximum representative of all intra-domain distances and three 'shoulders' that represent distances between the trimer core and 70A, 70B, and 70AB (Figure 4.19A). In the case

of the ‘70AB’ shoulder, the 70A and 70B domains are sufficiently close to be considered a single module, while the ‘70A’ and ‘70B’ shoulders capture distances associated with the spatial separation of these domains. When RPA-DBC binds d10, the discrete shoulders associated with 70A and 70B disappear, while the shoulder corresponding to 70AB increases slightly, indicating that DNA-binding has resulted in substantial reduction in the population of structures where the two domains are far apart (Figure 4.19B). For the 20-nucleotide binding mode, the 70AB shoulder shifts in slightly and merges with the primary maximum (Figure 4.19C). This correlates with the maximum compaction of RPA-DBC when bound to d20. In the case of RPA-DBC/d30, the continuity between the 70AB shoulder and the primary maximum remains unchanged; however, there is now an increase in long-range distances in excess of 100 Å. (Figure 4.19D). This long-range portion of the P(r) distribution lacks the features associated with the spatial separation of 70A and 70B that was observed in the absence of ssDNA (Figure 4.19). However, the intra-domain features of the RPA-DBC/d30 distribution ($r < 100$ Å) remain similar to those of the RPA-DBC/d20 distribution. The final 30-nucleotide binding state, then, appears to encompass a mixture of inter-domain orientations either reminiscent of the compact 20-nucleotide binding mode or exhibiting more extended orientations.

Taken together, the changes in inter-domain orientation that accompany RPA-DBC through its three binding modes suggest that the structural remodeling of RPA-DBC is ‘bi-modal,’ that is there are two primary architectural transitions taking place during DNA-binding. The first transition is the compaction of RPA-DBC’s inter-domain architecture during the first two binding states. Binding of 8-10 nucleotide substrates

draws the 70A and 70B domains together, where progression to a 20-nucleotide substrate positions the trimer core close to the 70AB domains (Figure 4.20). The second transition involves a dynamic redistribution of domains for some subset of the RPA-DBC solution ensemble, converting the compact architecture of the intermediate binding mode to a more extended, final binding state. In the subsequent mixture of conformers, the trimer core and 70AB domains can occupy inter-domain orientations where they are farther apart from each other than in the compact d20 state (Figure 4.20).

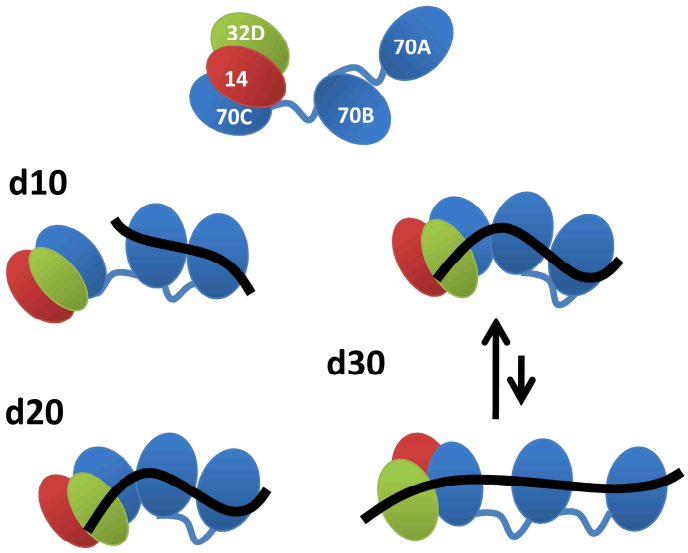


Figure 4.20. Proposed model for RPA-DBC DNA-binding trajectory. Binding of d10 and d20 substrates results in progressive compaction of the DNA-binding core through alignment of 70A and 70B (d10), then 70AB and trimer core (d20). Binding of the d30 substrate allows RPA-DBC to rearrange upon the ssDNA.

The question of how exactly ssDNA threads its way through RPA remains. Does the substrate follow the convex or concave surface of the arced scattering envelope? While additional SAXS experiments on intermediate ssDNA substrates may provide

insight into this, a full delineation of the DNA density can be obtained from small-angle neutron scattering (SANS), which can provide sufficient scattering contrast to distinguish protein and DNA density.

Conclusions

These studies have provided the first structural perspective on RPA's DNA-binding trajectory. A combination of small-angle x-ray scattering and computational simulation have established that the DNA-binding core of RPA (RPA-DBC) exists as a mixture of inter-domain orientations, where the most abundant of these orientations assumes a close, curved inter-domain arrangement. This ensemble of inter-domain orientations is progressively compacted and restricted upon binding 10- and 20-nucleotide substrates, but partially resumes a mixed inter-domain distribution upon binding 30-nucleotide substrates. The compaction associated with the first two binding modes arises primarily from alignment and apposition of 70A and 70B domains (d10), followed by alignment and apposition of 70AB and the trimer core (d20). The subsequent return to a mixture of inter-domain orientations in the final binding state (d30) may arise from variations in the distribution of DNA-binding domains along the substrate, greater dynamic fluctuations between the substrate and the DNA-binding core, or some combination of the two.

Overall, two primary architectural transitions define RPA's DNA-binding trajectory – inter-domain compaction, followed by a return to a mixed population of compact and extended states. The mechanistic implications of this 'bi-modal' understanding of RPA remodeling upon binding ssDNA will be discussed in fuller detail

in the next chapter, as well as strategies for determining how this fundamental structural pathway is integrated into DNA processing transactions.

Acknowledgements

The pET15b plasmid for RPA-DBC was a kind gift of Alexey Bochkarev. Dr. Miaw-Sheue Tsai of the EMB Core of the Structural Biology of DNA Repair (SBDR) program supervised construction of the two versions of RPA-DBC in the pBG100 vector series. Norie Sugitani, a research assistant in our laboratory, collected the ESI-MS data on RPA-DBC, while Dr. W. Hayes McDonald of the Mass Spectometry Research Center at Vanderbilt performed the proteomics analysis. Full-length RPA used in the fluorescence anisotropy studies was a kind gift of Dr. Mike Shell, a post-doctoral fellow in our laboratory. Dr. Susan Tsutakawa of the Tainer laboratory at LBNL has been an invaluable collaborator with our laboratory for the collection of SAXS data and discussions of their analysis and interpretation. Dr. Rob Rambo of the Tainer laboratory at LBNL has also generously provided help and advice for applying the Porod-Debye analysis to this data.

CHAPTER V

DISCUSSION AND FUTURE DIRECTIONS

Summary of this work

'Adaptable architecture' is an emerging paradigm in the field of DNA processing biology, a theme applicable to individual DNA processing proteins and elaborate protein assemblies alike. This research project has focused upon investigating and describing the dynamic architecture of one of the most fundamental participants in eukaryotic DNA processing, RPA. Key discoveries and themes from this work are summarized below.

RPA possesses an 'independent,' dynamic inter-domain architecture

Application of advanced NMR methodologies has permitted unprecedented access to the solution architecture of full-length RPA, resolving an implicit debate between competing views of RPA quaternary structure ('compact' versus 'independent'). TROSY-HSQC experiments have revealed an absence of inter-domain contact among RPA's five structured modules, counter to the original 'compact' architecture indicated by STEM (*78*). Not only do individual modules remain free of physical contact, but their diffusive rotational motion remains independent as well, as witnessed by variation in signal linewidths within the TROSY-HSQC. This dynamic architecture is maintained even in the context of DNA binding, as the structure and motion of protein interaction domains 70N and 32C are unaffected by RPA interaction with a 30-nucleotide substrate, compared to DNA binding domains 70A and 70B. This structural autonomy of domains

dedicated to DNA binding versus protein interaction is well correlated with RPA's role as a recruitment platform and coordinator for DNA processing.

Domain linkage and DNA binding impose motional correlation on RPA domain dynamics

Variation in rotational motion among RPA domains appears to be driven primarily by linker length, as revealed by heteronuclear relaxation experiments on RPA tandem domain fragments, RPA70AB and RPA70NAB. Comparison of theoretical rotational diffusion tensors for free domains to those of their linked counterparts establishes that rotational motion of the linked domains is not completely free, but rather depends upon the degree of tethering imposed by the interconnecting linker. Rotational tumbling of the DNA binding domains 70A and 70B is seen to become highly correlated upon binding a ssDNA substrate. For RPA70NAB, 70N maintains a higher degree of diffusive autonomy compared to 70A and 70B, consistent with their respective roles in engaging protein and ssDNA substrates, as well as the observation of functional autonomy through structural independence noted for full-length RPA.

RPA architecture condenses, then extends, upon passing through its DNA-binding trajectory.

The structural adaptability of RPA's dynamic inter-domain architecture is reflected in the structural progression of the protein's DNA-binding trajectory. Small-angle scattering analysis of the initial 8-10 and intermediate 12-23 nucleotide binding modes has uncovered a progressive compaction of inter-domain architecture (first 70A and 70B, then 70AB and the trimer core) as the DNA-binding core attempts to

accommodate these comparatively short substrates. Advancement to the final 28-30 nucleotide binding state prompts release of this inter-domain compaction, and the DNA-binding core reverts to a dynamic mixture of compact and open inter-domain distributions along the ssDNA. The evolving contour of the DNA-binding core over the course of this trajectory, whether convex or concave, remains to be resolved by future SANS analysis. Nonetheless, the scattering profiles provide the first comprehensive view of inter-domain rearrangement as RPA binds ssDNA and confirm the existence of residual motion within the DNA-binding core even when bound to its substrate.

Implication of the results

Re-envisioning the three binding modes of RPA

The ability to correlate a particular architecture with each binding mode of RPA provides tremendous mechanistic insight into how RPA adapts to and engages ssDNA substrates. As expected, the initial binding mode (10 nt) engages and aligns domains 70A and 70B, while the trimer core remains dynamically flexible. The intermediate binding mode (20 nt) results in close proximity of the trimer core and 70AB, generating a condensed, bilobal, inter-domain architecture. The final binding mode (30 nt) is associated with release of the steric constraint observed for the intermediate state, suggesting that a more complicated assortment of open and compact inter-domain orientations typifies the default DNA-bound architecture.

Results of computational modeling indicate that binding of ssDNA with a minimum number of nucleotides to occupy all four DNA-binding domains should leave little room for redistribution of domains along the ssDNA (Chapter IV). The presence of

extended distributions of domains in the final complex, then, suggests that the established occluded site size of 30 nucleotides (*55*, *56*) is not the minimum interaction footprint for all four DNA-binding domains, which is estimated at 23-24 nucleotides from computational models and fluorescence studies (*82*). In this view, the 'final' binding mode, as defined by full occupation of all four DNA-binding domains, can encompass at least two distinct architectures—condensed or extended—contingent upon available substrate length. This architectural dichotomy may explain the ambiguity within the literature in defining nucleotide ranges for RPA's intermediate and final binding modes. The 12-23 nucleotide span invoked for the intermediate binding mode (*54*) encompasses a range in which ssDNA initially makes contact with 70C, but must then traverse a third of the trimer core to reach and occupy 32D (*17* and Chapter IV). The condensed, bilobal architecture of the intermediate state is expected to be present throughout this range. Conflicting ranges of 23-27 nt (*54*) or 28-30 nt (*14*, *55*) for the final interaction state define the minimum site requirement for full occupation of the DNA-binding core (23-24 nt) and the best-fit model for extended inter-domain orientation on ssDNA (28-30 nt).

In light of this, accurate delineation of RPA's three DNA binding states should require consideration of the architectures of each interaction mode, in addition to domain involvement and nucleotide length range. For the initial binding mode, participation of domains 70A and 70B correlates well with the corresponding scattering profile and a site size of 8-10 nucleotides (Figure 5.1). In the case of the intermediate state, additional engagement of the trimer core through 70C, and ultimately 32D, generates a condensed architecture over the corresponding length range of 12-24 nucleotides (Figure 5.1). With all four DNA-binding domains now occupied, however, the final binding mode is best

described as an ‘architecture in transition’ across the corresponding length range of 24-30 nucleotides, as the redistribution of DNA-binding domains progressively shifts the core architecture to include a significant population of extended states (Figure 5.1).

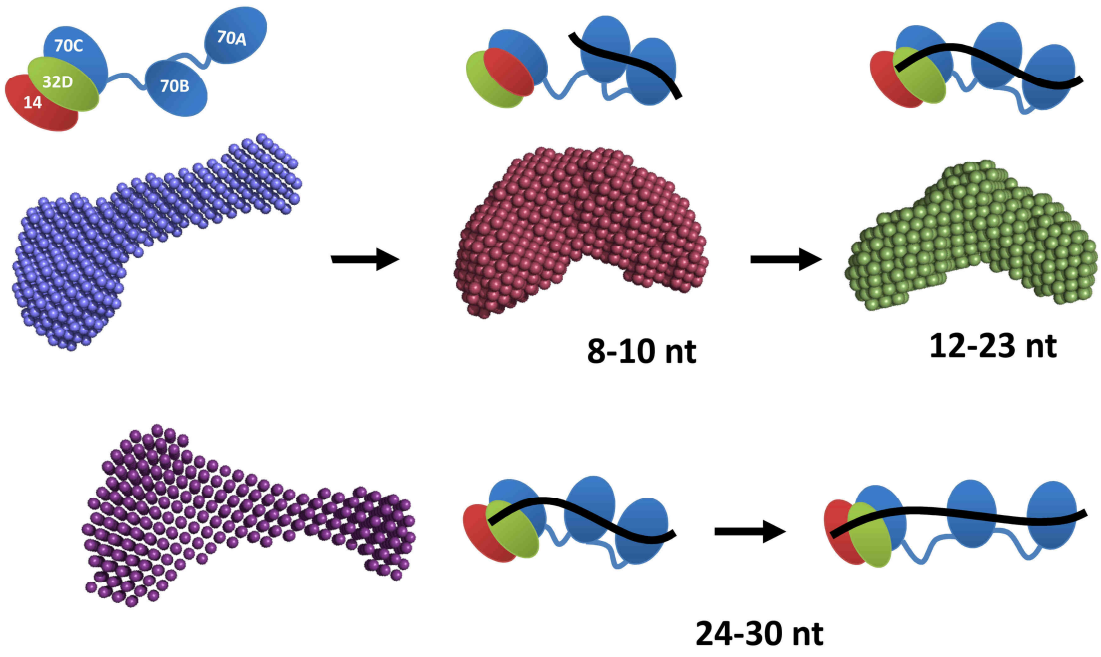


Figure 5.1. Architectural model for RPA’s DNA-binding trajectory. Cartoon diagrams of each binding mode are shown next to their corresponding scattering envelope from Chapter IV. Prior to binding ssDNA, RPA-DBC possesses a slightly extended inter-domain orientation in solution. The initial binding state involves substrates 8-10 nucleotides in length and engages domains 70A and 70B. The intermediate binding state encompasses substrates 12-23 nucleotides in length and is characterized by a condensed architecture as the ssDNA proceeds to engage trimer core domains 70C and 32D. The final binding state of 24-30 nucleotides is characterized by a redistribution of DNA-binding domains, with a transition from the compact architecture of the intermediate binding mode to the extended architecture associated with the final occluded site size of 30 nucleotides.

The available scattering data driving this model represent snapshots of very specific points along this architectural trajectory, and additional experimental verification is warranted to test its validity and as well as its relevance *in vivo* (*vide infra*). Even so,

this long-sought structural view of RPA's DNA binding-trajectory now allows the formulation of key questions addressing how these structural transitions drive DNA processing. Of particular interest is how other DNA processing factors might take advantage of these architectural changes to facilitate or, more importantly, to disrupt binding of ssDNA by RPA.

Integrating protein intervention into the DNA-binding trajectory of RPA

The ability to accelerate or interrupt the binding of RPA to its ssDNA substrate is critical to the efficient progression of DNA processing machinery. Studies investigating the role of protein interaction in manipulating DNA-binding by RPA have focused primarily upon the relevance of specific domain-domain contacts (SV40 Tag-OBD/70AB and Tag-OBD/32C, Chapter I). The availability of a structural model for RPA's DNA-binding trajectory now permits these studies to be placed into a more global architectural context. Among the reports discussed in Chapter I was the specific interaction between the principal DNA-binding domains 70A and 70B and the origin-binding domain of SV40 large T-antigen (Tag-OBD), an interaction shown to stimulate RPA's ability to bind ssDNA, purportedly through pre-alignment of the DNA-binding channel (*29*). Of particular interest was the ability of Tag to form a ternary complex with RPA bound to a dT_8 substrate, but subsequent loss of this interaction for substrates representative of intermediate (dT_{15}) and final binding modes (dT_{30}).

Examination of the core architectures of the initial and intermediate binding modes of RPA suggests that steric effects are a primary contributor to the disruption of the Tag-OBD/70AB interaction. Manual superposition of the known structure of

monomeric Tag-OBD with its approximate binding site on RPA70AB in the context of RPA-DBC/d10 and RPA-DBC/d20 complexes places Tag-OBD in close proximity to the trimer core (Figure 5.2). The higher-order assembly of active large T-antigen as a hexamer would be expected to create steric clash, more so with the compact architecture of RPA-DBC/d20. The progressive rearrangement of RPA's inter-domain distribution, then, may be responsible for freeing the DNA-binding core from large T-antigen as RPA finalizes contact with the ssDNA. As such, the loading mechanism would be intrinsically primed to disengage, allowing for 'hand-off' of RPA to the next protein partner.

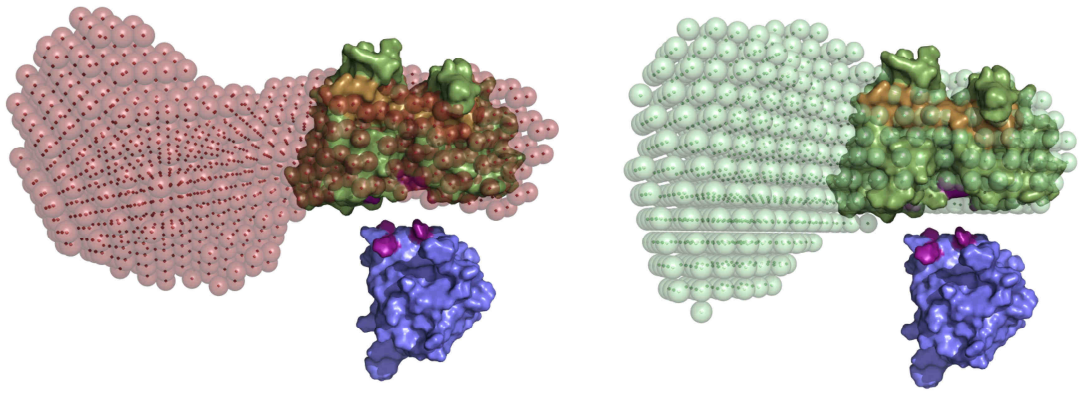


Figure 5.2. Superposition of hypothetical 70AB/Tag-OBD complex with RPA-DBC scattering envelopes. Overlay of hypothetical binding complex between Tag-OBD and RPA70AB with scattering envelopes from RPA-DBC/d10 (left) and RPA-DBC/d20 (right). Residues facilitating protein-protein contact are highlighted in magenta.

Initiation of primer synthesis for the SV40 system of replication relies upon contact between Tag-OBD and domain 32C, an interaction hypothesized to trigger displacement of RPA from ssDNA (*30*). Displacement is proposed to occur through the exertion of tension on the linker between 32C and the final, labile DNA-binding domain,

32D, exposing enough of the ssDNA template to enable polymerase α/primase to access the ssDNA and subsequently displace the remainder of the RPA DNA-binding core (*30*). Subsequent studies from this group and that of Dr. Ellen Fanning have since determined that primase itself is also capable of interacting with 32C (*170*) and also the tandem domains, 70N, 70A, and 70B, which may also facilitate displacement of RPA from ssDNA, though this has yet to be investigated experimentally.

The structural models of RPA's DNA-binding core reveal no obvious obstacle to preventing displacement of the trimer core via tension exerted upon the 32D/32C linker. The high degree of flexibility between the trimer core and domains 70A and 70B in the initial binding mode indicate, however, that tension leveraged through the 32D/32C linker alone may not be sufficient to remove the entire DNA-binding core, unless it is sufficient to propagate to the linker between domains 70C and 70B. It is perhaps at this point that interaction between 70AB and primase is relevant for full removal of RPA from ssDNA. This remains purely speculative, though. Full elucidation of the structural mechanism behind protein-driven displacement of RPA will require a clearer understanding of the full range of protein-protein contacts involved, as well as the intrinsic dynamics present within RPA's DNA-bound state. Research strategies to pursue this are outlined in more detail in the next section.

The role of RPA structural dynamics in DNA processing

Dynamic inter-domain architectures are optimally suited to accommodate the shifting substrate environments of DNA metabolism. Not only is such modular organization useful for adapting to the unique geometries of binding partners, but flexible

linking also ensures the proximity of distinct biochemical functions (*24*). The findings noted above illustrate how DNA binding and protein interaction functions are structurally integrated in the context of full-length RPA through the flexible linking of their cognate domains. Such an inter-domain organization specifies how RPA can simultaneously organize and coordinate both ssDNA and protein participants during DNA processing transactions.

Well characterized examples of this can be found with the DNA damage response machinery, where stretches of RPA bound to ssDNA serve to signal checkpoint activation. Recruitment of ambient checkpoint proteins ATRIP, RAD9, and MRE11 to DNA-bound RPA is facilitated by their interaction with the 70N domain (*63*, *64*), which is now shown to remain unconstrained in the DNA-bound state (*143*, *152*). Assembly of the nucleotide excision repair complex serves as another example of the independent coordination of DNA and protein substrates by RPA. In this particular case, the polarity of RPA binding to ssDNA determines and maintains the polarity of excision nuclease recruitment and action (*4*, *28*).

In addition to an ability to manage separate DNA and protein substrates, RPA is also able to adapt directly to variation in the binding environment of ssDNA, through architectural versatility of its DNA-binding core. Compression and release of the DNA-binding core over the course of the DNA-binding trajectory indicates that RPA can tailor its inter-domain orientation to the available length of the ssDNA substrate. Such 'custom' interaction ensures that RPA can shield ssDNA, regardless of context, and provides an essential mechanism for efficiently sequestering ssDNA that falls below the standard occluded site size of 30 nucleotides. The binding of ssDNA substrates appears

to outweigh the steric strain imposed by a condensed architecture, as the core assumes a highly compacted inter-domain arrangement when a ssDNA substrate of minimal length is available.

Precisely how the architectural plasticity of RPA's DNA binding core impacts DNA metabolism has not been directly captured by standard biochemical approaches as yet. At best, this structural flexibility has been detected indirectly by *in vitro* UV cross-linking studies on RPA and primer-template substrates (*53*) or RPA and replicating SV40 DNA (*171*, *172*). In the study of primer-template substrates, the efficiency of subunit cross-linking relied upon substrate length, where RPA70 was more effectively captured in the context of shorter substrates (13-14 nt) and RPA32 in the context of longer substrates (19-31 nt) (*53*). For replicating SV40 DNA, cross-linking efficiency was initially optimal for RPA32 with emerging RNA-DNA primers, then favored for RPA70 with subsequent primer extension (*171*, *172*). A clearer understanding of how RPA's DNA-binding core is rearranged during active cellular DNA processing will require the creative combination of methods that can access this dynamic molecular environment.

Understanding and accessing adaptable protein architectures through dynamics

The explosion of 'adaptable architecture' as a general theme for protein structures has emphasized the need for an integrated understanding of protein function as a product of both architecture and dynamics. In the case of RPA, a time-varying, modular architecture is critical for participation in the diverse substrate landscapes of DNA metabolism. The challenge of accessing and characterizing such dynamic architectures remains an active area of theoretical and experimental development. While approaches

such as SAXS, SANS, and NMR relaxation are becoming more mainstream, their routine application to many important biological systems is awaited.

The description of RPA's dynamic, modular architecture represents an important advance in the field of DNA processing; it is also significant for making use of these integrated methodologies. As such, this work can serve as a reference for similar experimental approaches to other modular SSBs and OB-fold proteins. The fundamental quandary of high resolution domain information in the absence of a knowledge of global architecture persists for nearly all SSBs and countless DNA processing proteins, among them important targets such as nucleotide excision repair factors XPA and XPC, homologous repair factor BRCA2, and the telomere-binding proteins Pot1-Sten1.

Future Directions

Research is a self-evolving path, where interesting results inspire continuing questions. This project has fundamentally advanced our understanding of RPA's intrinsic dynamic architecture and how this architecture is impacted by the binding of ssDNA. The architectural models generated from these findings now await further testing and refinement, while new questions focusing upon the biological implications of these results are ripe for exploration. Outlines of three key areas for future investigation are described in detail below.

The first segment focuses upon refining our view of RPA 'structural dynamics' by providing a more quantitative description of the spatial volume explored by the full-length protein. This line of inquiry would be expected to yield the average spatial disposition and orientation of individual domains with respect to RPA's trimer core, and thus a more complete picture of the general volume occupied by RPA. This information

would set the stage for incorporating full-length RPA into models for multi-protein DNA processing assemblies.

The second section concentrates upon validation and refinement of the model for DNA-binding developed from the scattering experiments reported here (Figure 5.1). Key questions include resolving the exact placement of the DNA-binding core within the scattering envelope, examining the specific orientations and dynamics of individual domains across the three binding modes, and investigating the architectures of 'transition' points between each binding mode. As referred to above, this information would provide a more detailed platform for understanding the fundamental structural transitions which accompany DNA processing transactions.

The final section seeks to address the challenging biological problem of how protein intervention results in displacement of RPA from ssDNA. Using minimal components of the SV40 system of replication, primary areas of focus would include identification of RPA domains necessary for displacement and description of architectural changes within the DNA-binding core generated by manipulation of these target domains. By understanding the architectural hallmarks of a displaced DNA-binding core, we can begin to integrate this 'displacement' mechanism into the progression of specific DNA processing pathways, including DNA replication, damage response, and repair.

Extending and detailing the current view of RPA structural dynamics

TROSY-HSQC experiments have proved invaluable in establishing a flexible, independent inter-domain architecture for RPA. Measurements of rotational diffusion

tensors have provided complementary information regarding average inter-domain orientations for domains 70N, 70A, and 70B, and scattering data has highlighted a moderately extended architecture for the DNA-binding core (70ABC/32D/14). Nonetheless, critical questions remain on many of the details of RPA's modular, flexible architecture. Average domain orientations have yet to be established in the context of the full-length protein, the accessible spatial distances sampled by each domain remain undefined, and details of the global architecture of the full-length protein are unknown. As with the project reported here, a combination of NMR and SAXS on full-length RPA and multi-domain fragments is proposed to explore these questions.

Establishing average domain orientation in full-length RPA

NMR heteronuclear relaxation experiments (Chapter III) have the capacity to establish the rotational orientation of a domain of interest. Full-length RPA, however, presents a number of technical challenges for measurement of ^{15}N relaxation. Primary among these is the requirement for an extended recovery delay in the measurement of large T_1 values, a consequence of increased hydrodynamic drag on domains closely coupled to the 50-kDa trimer core (i.e. 70B and 70A). As a result, the acquisition time required for running relaxation experiments on the full-length protein using standard methods is impractical with regards to sample stability or shared user demands on the instrumentation.

An alternative source of information on inter-domain orientation and motion can be found in measurement of ^{15}N-^{1}H residual dipolar couplings (RDCs) and subsequent calculation of the averaged molecular alignment tensor for each domain. Rather than

reporting directly upon rotational motion as the diffusion tensor does, the molecular alignment tensor provides an assessment of how molecular tumbling is orientationally biased upon introduction of alignment media to the sample solvent. The orientation of this tumbling bias, in turn, indirectly reports upon shape and rotational freedom in the context of alignment. Theoretical details of RDCs, their measurement, and calculation of the molecular alignment tensor are detailed in a number of reviews (*173*, *174*).

As the structures of RPA's domains are already well established by high resolution methods, the power of RDCs lies in their ability to reveal how flexible tethering affects domain motion. Differences between the theoretical alignment tensor predicted for a domain in isolation and that measured for the domain in the context of full-length RPA will highlight the rotational bias created by flexible attachment to the rest of the protein. Collectively, then, the alignment tensors will provide a view of the average inter-domain orientation for domains 70N, 70A, 70B, and 32C in the context of full-length RPA. As signals from the trimer core are not detected in TROSY spectra of full-length ^{2}H, ^{15}N-RPA, other NMR experiments or examination of alternative RPA constructs, such as RPA-DBC, will be required to obtain this information.

Full determination of an alignment tensor with its five parameters (D_a, R, α, β, γ) will require acquisition of five separate sets of ^{15}N-^{1}H RDCs under different alignment conditions. As optimization of sample alignment is very much an empirical undertaking, alignment media should be established initially with ^{15}N-enriched RPA prior to investigation of a deuterated sample. Since the analytical approach relies upon molecular alignment created by steric obstruction of rotational tumbling, alignment media that act

sterically, such as bicelles and strained polyacrylamide gels, should be preferentially explored.

Interpretation of RDC data for RPA will pose a significant challenge. Initial analysis will focus upon comparing and contrasting alignment tensors calculated for each domain from the measured sets of ^{15}N-^{1}H RDCs. This domain-by-domain approach is preferred for proteins like RPA, since the time-varying architecture of the full-length protein does not present a fixed molecular frame for globally referencing all ^{15}N-^{1}H RDCs. Rather, the flexible independence of each domain causes the ordering environment of the alignment media to impact each domain's rotational tumbling differently. As a result, each domain alignment tensor should possess a unique magnitude and orientation and will represent the averaged alignment experienced by the domain as it dynamically samples multiple orientations in solution (*174*, *175*). Differences in the alignment tensors across the domains of RPA are expected to be reflective of the differences in domain shape and tethering environments (i.e. domains experiencing highly anisotropic motion, whether from shape or tethering by short linkers, will exhibit a greater magnitude in the alignment tensor).

The magnitude and orientation of domain alignment tensors calculated in the context of the full-length protein will differ, though, from those determined for each domain in isolation, as tethering by inter-domain linkers will increase the anisotropy of rotational tumbling in the context of alignment. Comparison of domain alignment tensors measured in the context of full-length RPA and those calculated for each domain in isolation should reveal how linkage impacts this orientational sampling for each domain, and thus provide insight into the average inter-domain orientation experienced in

the context of full-length RPA. Validation of these average domain orientations will require ensemble methods, such as molecular dynamics simulations on the full-length protein, to provide a sufficiently large population of RPA structures for assessing the average theoretical orientation and alignment of each domain. A similar approach has been described by Maciejewski and colleagues for a three-domain derivative of human factor H (FH1-3), with the exception that ensemble generation was driven by simulated annealing in Xplor-NIH, rather than molecular dynamics trajectories (*176*).

Determining average inter-domain distances in full-length RPA

Paramagnetic relaxation enhancement (PRE) refers to the ability of a paramagnetic center, such as a metal ion or organic radical, to accelerate the rate at which an excited NMR signal returns to its equilibrium state. Association of a paramagnetic center with a biomolecule, then, results in broadening of NMR signals. Since the degree of relaxation enhancement depends upon the proximity of the paramagnetic center according to $\frac{1}{r^6}$, where r is the intervening distance, PREs provide an effective means for determining the relative distances between a known paramagnetic center and those surrounding nuclei that are within a range of 12-35 Å (*177*). In the instance where the paramagnetic center is situated on a separate, flexibly linked domain, however, PREs will reflect an averaged distance between the paramagnetic agent and nuclei located on the opposing domain, weighted according to $\frac{1}{r^6}$.

Incorporation of a paramagnetic center within RPA may be readily achieved by cobalt substitution at the zinc ribbon of 70C, establishing the trimer core as a central point of reference for the remaining flexibly linked domains. The use of metal ions as

paramagnetic probes for proteins has been well documented by the laboratory of Lewis Kay for the case of copper ions (Cu^{2+}) (*178*). Application of a similar strategy with cobalt will require knowledge of the corresponding electronic *g*-factor and the correlation time associated with its electronic spin relaxation, values which are already established or which can be measured by NMR experiments (*179*).

Qualitative information on those domains that approach the trimer core within the range of 12-35 Å can be readily obtained by comparison of TROSY-HSQC signal intensities acquired from the diagmagnetic versus the paramagnetic state. Quantitative interpretation of average distances calculated from PREs (weighted by $\frac{1}{r^6}$), though, will require an understanding of the upper and lower inter-domain distances accessible to each domain within the spatial range of the paramagnetic effect. Rigid-body modeling engines, such as BilboMD and EOM can provide estimates of this distance range, while small-angle scattering can provide experimental validation. Once these distance ranges are available, though, care must be exercised in the interpretation of the distance average, as the $\frac{1}{r^6}$ dependence of paramagnetic phenomena causes shorter distances to have an exaggerated influence on the inter-domain average. As described above for the RDC measurements, ensemble analysis of inter-domain distances from molecular dynamics simulations on full-length RPA should be helpful in exploring the impact of this bias on the PRE measurements. Once these PRE-based average inter-domain distances have been determined and properly interpreted, their incorporation into the emerging structural model for full-length RPA will provide important information on the average spatial distribution of domains 70N, 70A, 70B, and 32C with respect to the trimer core.

Probing global architecture of the RPA conformational ensemble with SAXS

While RDCs and PREs measured by NMR can provide detailed information on *average* inter-domain orientations and distances, they are limited in their ability to provide information on the *range* of domain orientations and distances experienced by the entire solution ensemble of full-length RPA. Insight into this range of accessible inter-domain distances and orientations is critical to understanding the degree of inter-domain motion present in the full-length protein, as well as the array of architectural forms assumed by RPA. Small-angle x-ray scattering (SAXS) is the principal experimental technique of choice for accessing global spatial information from a protein ensemble in solution and is capable of capturing *all* distances present within the population. A combination of SAXS and advanced MD rigid-body modeling is proposed to explore the range of inter-domain architectures accessible to full-length RPA and to determine which of these architectures occur most frequently in solution.

Interpretation of scattering curves and molecular envelopes from dynamic, multi-domain proteins presents a number of complications, as the experimental data simultaneously encodes multiple inter-domain arrangements. Moreover, the behavior of scattering curves for modular protein systems has only recently received systematic documentation through computational simulation studies (*89, 180*) and still lacks much of the theory that underlies the current understanding of more static, globular proteins (*125, 181*). Because of this, pursuit of SAXS on full-length RPA is also an opportunity to determine how inter-domain motion impacts the experimental scattering curve and models derived from this data.

Acquisition of SAXS data on full-length RPA would follow the methods detailed in Chapter IV, followed by standard data analysis (Guinier, Kratky, pair-distance transformations, etc.). Spatial parameters derived from the scattering data (R_g, D_{max}, P(r) distribution) would then be available for comparison to spatial parameters derived via NMR (i.e. PREs). Bilbo-MD could be used to simulate an initial conformational ensemble for comparison with the experimental scattering data; the limitations detailed in Chapter IV, however, suggest that a full rigid-body molecular dynamics (MD) or Brownian dynamics (BD) simulation would be a better resource for generating a simulated population for full-length RPA. The subsequent simulated ensemble and analysis of the average, simulated SAXS curve from this ensemble would be informative in determining the range of theoretically possible RPA architectures, as well as whether Brownian rotational motion alone or both Brownian motion and viscosity effects can account for the spatial and orientational properties measured experimentally by NMR and SAXS.

Validating the relative abundance of specific inter-domain arrangements within a solution population for any protein still remains a challenge. For RPA, experimental information from NMR, SAXS, and MD/BD can lend support to the relative frequency of general categories of inter-domain orientations. As an example, this might include resolving whether all inter-domain linkers are predominantly extended in solution or possess the more condensed persistence length of a random coil polymer, based on assessment of PRE-derived distances and SAXS R_g and D_{max} values, as well as the relative frequency of these respective conformational arrangements within MD/BD trajectories. The assessment of definitive conformational ensembles for flexible, modular

proteins continues to remain an active area for methods development in structural biology.

Integration of NMR, SAXS, and MD/BD data into a model of full-length RPA will provide a comprehensive view of RPA's general inter-domain disposition, as well as the diversity of architectures accessible to the protein in solution. This structural model will provide an invaluable foundation for understanding the role of RPA architecture in the context of multi-protein assemblies and how this architecture evolves during the course of different DNA processing events.

Refining the model of RPA's DNA-binding trajectory

Small-angle scattering has provided the first structural model of how RPA architecture accommodates different substrate lengths as it follows its DNA-binding trajectory. Nonetheless, key questions remain regarding RPA's DNA-binding pathway. The precise location of the DNA-binding core remains ambiguous, as does the relative orientations of specific domains within the DNA-binding core. Moreover, the inter-domain architectures associated with 'transition' points between binding modes remain a critical area of interest for understanding the initial 'capture' of 70C by ssDNA and changes to the DNA-binding core as it transitions from its minimal binding site size of 24 nucleotides to its occluded site size of 30 nucleotides.

Delineating the path of DNA through RPA with SANS

Studies from small-angle x-ray scattering reveal two possible orientations for the site of DNA binding: the smooth convex face of the RPA-DBC molecular envelopes or

the sharply kinked concave face of the opposite side (Chapter IV, Figure 4.8). The convex arrangement is expected to introduce marked steric strain on domains 70AB and the trimer core, while the concave arrangement would accommodate less physical contact between these two sections of the DNA-binding core. The use of small-angle neutron scattering (SANS), specifically contrast variation experiments, is proposed to resolve the ambiguity in the placement of the DNA-binding channel.

The principle and rationale for neutron contrast variation is discussed in Chapter I. Here, neutron scattering profiles will be collected on RPA-DBC/ssDNA complexes in buffers consisting of 0%, 10%, 20%, 30%, 80%, and 100% D_2O. As protein and ssDNA possess different neutron scattering properties, well-established mathematical algorithms can be implemented to extract separate scattering curves for RPA-DBC and its ssDNA substrate. Analysis can then proceed as for conventional x-ray scattering, allowing derivation of P(r) functions and reconstruction of molecular envelopes. Ideally, reconstruction of the protein scattering envelope should highlight the channel occupied by scattering density from the DNA. As the scattering studies have revealed domain spreading in the case of the final 30 nucleotide substrates, the ssDNA substrates for SANS will be designed to occupy the DNA-binding core without allowing excess room for spreading of domains (24 nucleotides), as well as to capture the features of the architecture of the final, occluded site size (30 nucleotides).

Examining inter-domain re-orientation during RPA DNA-binding

While scattering studies have provided an unprecedented view of global changes in the DNA-binding core as RPA traverses its DNA-binding trajectory, the specific

behavior of individual domains during DNA-binding has relied upon data from smaller fragments (RPA70AB, Chapter III) or modeling studies comparing specific inter-domain arrangements to the experimental scattering data (Chapter IV). As for full-length RPA, the use of ^{15}N-^{1}H residual dipolar couplings (RDCs) measured by NMR and comparison of the resulting experimental and theoretical alignment tensors is proposed to track changes in inter-domain orientation for the DNA-binding core bound to 10-, 20-, and 30-nucleotide substrates.

These experiments present a number of technical challenges, foremost among them the size limitations imposed by the 78 kDa DNA-binding core, RPA-DBC. Deuterium enrichment has shown promising results in a pilot study of this construct at 800 MHz, where strong NMR signals were observed for 70A and 70B, and slightly broader signals from the trimer core. Signal strength is maintained upon introduction of a 10-nucleotide substrate, but is diminished for a 30-nucleotide substrate. Additional optimization at higher field strength (900 MHz), as well as higher temperatures (in excess of 298 K), would allow for improved signal sensitivity for the final 30-nucleotide binding mode. The use of alternative NMR experiments (CRINEPT/CRIPT) for higher molecular weight systems would also be an option for accessing this complex (*108*, *109*). Progressive changes to the magnitude and orientation of domain alignment tensors with each mode of DNA binding will highlight specific architectural changes imposed by DNA-binding, as well as the presence of residual inter-domain motion in the intermediate and final binding modes.

Mapping the architecture of 'transition' states between DNA binding modes

While SAXS has illuminated the characteristic architectures for each DNA binding mode of RPA, the updated structural model for RPA's DNA-binding trajectory raises new questions about the architectural transitions among binding modes. Specifically, what are the inter-domain rearrangements that accompany the DNA-binding core as it accommodates substrate lengths that bridge the initial and intermediate binding modes (10-13 nucleotides) and the intermediate and final binding modes (24-30 nucleotides)? As with the previous studies, SAXS is proposed as the primary approach to examine changes to the global profile of the DNA-binding core of RPA when bound to substrates that are 10-13 nucleotides in length for the first transition and 24-30 nucleotides in length for the second transition.

The molecular envelopes derived for the first transition would be expected to show a progressive compaction of 70AB and trimer core density, as the substrate becomes sufficiently long to capture and tether 70C. It is possible though that the weak binding affinity of 70C may in fact result in residual motion of the trimer core, preventing transition to the intermediate mode until the substrate is long enough to at least partially engage 32D as well. The molecular envelopes for the second transition from the compact intermediate binding mode to the final binding mode are expected to show a progressive lengthening of the scattering envelope up to the occluded site size of 30 nucleotides. To characterize the presence and degree of residual motion between 70AB and the trimer core during these transitions, collection and analysis of ^{15}N-^{1}H RDCs are again proposed. The extent to which inter-domain motion remains in these transition states will be

important to determining how labile these architectures are to disruption of DNA-binding.

Completion of these experiments will serve to validate the original model of DNA binding proposed by this research and to provide a more in-depth understanding of how the DNA-binding core of RPA is rearranged to accommodate different ssDNA substrates. As with studies on the full-length protein, this knowledge will be critical to furthering our understanding of how RPA captures and comes off ssDNA substrates and specifically how other DNA processing proteins can target these different architectural states to promote or hinder DNA binding.

Elucidating a 'displacement' mechanism for RPA driven by protein intervention

Paramount to understanding the role of RPA in DNA metabolism is an awareness of how other DNA processing proteins dislodge RPA from its nucleic acid substrate in order to gain access to ssDNA templates. As mentioned above, examination of the structural pathway for DNA-binding by RPA-DBC does not provide a clear answer to this issue. This final section is devoted to identification of specific domains required for the displacement of RPA from ssDNA substrates and the architectural changes to the DNA-binding core that accompany manipulation of these domains.

Development of an RPA 'displacement' assay

The design of this assay would be similar to that used to detect loss of interaction between RPA and SV40 large T-antigen (Tag) as RPA transitions from its initial to intermediate and final binding modes (*29*). Here, researchers linked RPA to resin via

antibody coupling and pre-incubated the bound RPA with Tag to form a complex. Subsequent washing of the RPA-Tag complexes with increasing concentrations of dT_8, dT_{15}, and dT_{30} oligonucleotide substrates resulted in no, partial, or complete loss of Tag, respectively, from the resin, as RPA architecturally 'remodeled' when transitioning from its initial to final binding mode. A similar strategy is proposed for detecting displacement of RPA from ssDNA substrates. A 30-nucleotide substrate would be linked to resin using a biotin/streptavidin system and pre-incubated with full-length RPA. As there is precedent for the involvement of Tag in displacing RPA from ssDNA in the SV40 system of replication (*30*), RPA/ssDNA complexes would initially be incubated with increasing concentrations of Tag and retention of RPA on the ssDNA resin would be monitored. Since polymerase α/primase (pol-prim) may also be required for displacement, it would also be tested individually and in combination with Tag for the ability to disrupt DNA binding by RPA. Length of the DNA substrate and RPA/DNA stoichiometry would also be parameters open for optimization. Once the necessary combination of 'displacement proteins' is identified, investigation of specific RPA domains required for the disruption of DNA binding would follow.

Identification of specific RPA domains required for displacement from ssDNA.

Building upon the displacement assay established in the preceding section, the experimental system would be modified to test the ability of Tag and/or pol-prim to disrupt DNA binding for various fragments of RPA. Initial studies would examine displacement of RPA deletion mutants lacking either or both 70N and 32C domains. Follow-up studies would focus upon the involvement of the DNA-binding core in this

process by repeating the assay with the DNA-binding core alone, RPA-DBC, and its two primary modules: RPA70AB and RPA70C/32D/14 (trimer core) (substrate length would be modified to fit the DNA-binding sites of these smaller constructs). Based upon known domain interactions between Tag and RPA (and pol-prim and RPA), these studies are expected to reveal 32C as necessary for the efficient disruption of DNA binding. A possible secondary target may also emerge for domains 70A and 70B. Previous studies have already established residues critical for interaction between 32C and 70AB (*29*, *30*), and charge-reversal mutants would be available to validate the role of direct protein interaction in triggering RPA displacement.

Delineating the mechanism of RPA displacement by NMR

Once specific domains involved in RPA displacement have been identified, an RPA construct representing a 'minimal displacement system' (likely the DNA-binding core, RPA-DBC, with domain 32C attached) will be positioned for more detailed analysis by NMR. Initial chemical shift perturbation experiments will focus upon recreating the results initially demonstrated by biochemical assay, where unlabeled Tag (or pol-prim) is titrated into an $^{2}H,^{15}N$-enriched sample of the target RPA construct pre-loaded with ssDNA substrate. A reversal or re-appearance in RPA chemical shifts to positions characteristic of the DNA-free state would indicate successful displacement from the ssDNA.

Examination of chemical shift perturbation data should also reveal other key aspects of the displacement process. For example, is binding of the protein partner strong enough to be observed during displacement, and does the interaction continue after

ssDNA has been released? Are target RPA domains bound in sequence (32C first, then 70AB) or simultaneously? As interaction between RPA and Tag has been studied extensively by NMR, characteristic chemical shift perturbations associated with binding between these two proteins are readily available, allowing direct assessment of this question. Even without a prior knowledge of these signal perturbations, though, as would be the case for pol-prim, one would still be able to detect the presence, order, and persistence of binding by comparing signal positions to those of RPA's DNA-free state.

Chemical shift perturbation data should also reveal the site within the DNA-binding channel at which displacement originates. Are signals from 32D and 70C targeted first, as we would anticipate, or is the interaction uniformly disrupted across the entire channel? What are the timescales of the displacement reaction? Does the mechanism rely upon the intrinsically weak binding equilibria of 32D and 70C to facilitate eventual removal of the remaining core, or does it force a far more rapid removal of the DNA-binding channel by other means?

Repetition of the titration assay with charge-reversal mutations incorporated into target protein interaction sites on the RPA construct would serve to validate these findings by NMR, as well as to parse which domain interactions contribute to specific aspects of the displacement mechanism. For example, is displacement merely weakened or completely compromised by charge reversal in just one protein interaction site? Is there a particular portion of the DNA-binding core affected by loss of a given protein interaction? Is displacement completely inhibited by applying charge reversal to all interaction sites, or is there a hidden contact site yet undiscovered?

The primary challenge of these experiments arises from the large sizes of RPA's protein partners, Tag (~540 kDa hexamer) and pol-prim (~350 kDa), which may cause signal broadening upon association with RPA. Sub-stochiometric titrations at high magnetic field strength (800-900 MHz), in addition to deuterium enrichment of the RPA construct, would be used to circumvent this. Should signal broadening be particularly severe, information could still be obtained by use of CRINEPT/CRIPT, or alternatively by study of RPA charge reversal mutants that weaken the interaction with Tag or pol-prim, but retain the capacity for partial displacement.

Characterizing the architecture of a 'displaced' RPA DNA-binding core

The ultimate goal for this work is the development of a structural model for displacement of RPA from ssDNA. Observations from the chemical shift perturbation assays described above, as well as information provided by more detailed structural studies of full-length RPA and the DNA-binding core bound to ssDNA, will provide much of the foundation for this model. Nonetheless, the ability to observe a 'displaced' architecture for the DNA-binding core directly would be an invaluable benchmark for understanding this process.

Whether intermediates representative of the 'displaced' architectural state of RPA exist in solution long enough to be captured by conventional NMR is currently unknown. In the event that their presence is transient, recently developed methodology for studying invisible, 'excited' protein conformations by NMR has been made available from the laboratory of Lewis Kay (*182-184*) and holds great promise for permitting access to these intermediate RPA architectures. A common example of this approach combines NMR

relaxation dispersion and residual dipolar coupling measurements to determine relative domain alignments of abundant 'ground' state protein conformations and less populated 'excited' conformations (*182-184*). The success of this methodology, though, relies upon the presence of a very slow transition between ground and excited states, on the order of microseconds or milliseconds, such that each state presents a distinct population in solution. Should RPA displacement occur on these slow timescales, whether spontaneously or by intervention from Tag or pol-prim, application of the appropriate pulse schemes should provide information on the rate of displacement, as well as the relative inter-domain alignments of displacement intermediates. In the event that Tag or pol-prim accelerate RPA displacement to faster timescales, it may be possible to turn to charge-neutralizing or reversal mutants to hinder interaction with RPA, and thus reduce the speed and efficiency of displacement to the target timescale range.

The characterization of invisible, excited states is still relatively new and has yet to be applied to a protein that possesses the architectural complexities of RPA. Experimental optimization is likely to be the rule rather than the exception in this case, though strategies for accommodating large proteins, particularly segmental labeling and methyl TROSY methodology, are becoming increasingly more common (*185*, *186*) Nonetheless, this approach offers the possibility of accessing intermediate architectural states of RPA that may be impossible to capture by more conventional structural methodologies and that are crucial to our understanding of a fundamental process in DNA metabolism – the removal of RPA from its ssDNA substrates.

Projection for future research on RPA

Future study of RPA architecture and function is essential to furthering our current understanding of the molecular coordination involved in DNA replication, damage response, and repair events. The research plan outlined above seeks to extend this knowledge by first providing a more detailed view of RPA 'structural dynamics' through mapping accessible and average inter-domain distances and orientations of the full-length protein. The plan moves on to refinement of the current architectural model for RPA's DNA-binding trajectory and how the DNA-binding core is remodeled by interaction with ssDNA. The final section investigates the mechanism by which protein intervention disrupts DNA-binding by RPA and the corresponding RPA inter-domain architectures associated with displacement. Pursuit of these three research focuses should provide critical insights into how RPA organizes and participates in the wide diversity of DNA processing transactions performed by the cell.

Concluding Remarks

Our perspective on how cellular genomes are propagated and maintained has advanced swiftly in the 25 years since the original discovery of RPA. With landmark advances in structural biology, particularly in the areas of NMR spectroscopy and small-angle scattering, we are now in an era where the architectural landscape of DNA metabolism is just coming into view. The choreographed assembly and interchange of multi-protein complexes on DNA substrates has emerged as a focal point for rapid cellular adaptation to the shifting substrate environments of DNA replication, damage response, and repair. We are now in a position to map the molecular mechanisms of

these diverse biochemical transactions in terms of the spatial and temporal evolution of their multi-protein DNA processing machines.

The scope of this dissertation research has sought to characterize the architectural diversity of one of the most critical DNA processing proteins, the human ssDNA-binding protein, RPA. In addition to providing our first view of the global disposition of RPA's inter-domain organization, this project has also examined how this protein's dynamic quaternary structure is refashioned upon binding ssDNA. From this we have gained a structural perspective on one of the most pervasive molecular actions in DNA processing – the protection and organization of ssDNA. These findings are an essential prerequisite to understanding how RPA coordinates access to ssDNA templates and regulates progression of DNA processing events.

For the future, there lies the need for further refinement and development of the structural models to emerge from this work. More importantly, though, the stage is now set for examining how the basic architectural changes that accompany RPA's DNA-binding trajectory are influenced by and propagated to the greater DNA processing machinery. It is hoped that the scope of RPA research in the next quarter century will prove as fruitful and interesting as for the one past.

BIBLIOGRAPHY

1. Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid, *Nature 71*, 737-738.

2. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kaçmaz, K., and Linn, S. (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints, *Annual Review of Biochemistry 73*, 39-85.

3. Hanawalt, P. C., and Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises, *Nature Review. Molecular Cell Biology 9*, 958-970.

4. Reardon, J. T., and Sancar, A. (2005) Nucleotide excision repair, *Progress in Nucleic Acid Research and Molecular Biology 79*, 183-235.

5. Baute, J., and Depicker, A. (2008) Base excision repair and its role in maintaining genome stability, *Critical Reviews in Biochemistry and Molecular Biology 43*, 239-276.

6. Jiricny, J. (2006) The multifaceted mismatch-repair system, *Nature Reviews: Molecular Cell Biology 7*, 335-346.

7. Lieber, M. R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway, *Annual Review of Biochemistry 79*, 181-211.

8. Lumir Krejci , Ling Chen , Stephen Van Komen, P. S. and A. T. (2003) Mending the break: Two DNA double-strand break repair machines in eukaryotes, *Progress in Nucleic Acid Research and Molecular Biology 74*, 159-201.

9. Muniandy, P. A., Liu, J., Majumdar, A., Liu, S.-ting, and Seidman, M. M. (2010) DNA interstrand crosslink repair in mammalian cells: step by step, *Critical Reviews in Biochemistry and Molecular Biology 45*, 23-49.

10. Ehmsen, K. T., and Heyer, W.-D. (2008) Biochemistry of meiotic recombination: formation, processing, and resolution of recombination intermediates., *Genome Dynamics and Stability 3*, 91-164.

11. Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., and Keck, J. L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes, *Critical Reviews in Biochemistry and Molecular Biology 43*, 289-318.

12. Wold, M. S. (1997) Replication protein A: A heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism, *Annual Review of Biochemistry*. *66*, 61-92.

13. Iftode, C., Daniely, Y., and Borowiec, J. A. (1999) Replication protein A (RPA): The eukaryotic SSB, *Critical Reviews in Biochemistry and Molecular Biology 34*, 141-180.

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

15. Richard, D. J., Bolderson, E., and Khanna, K. K. (2009) Multiple human single-stranded DNA binding proteins function in genome maintenance: structural, biochemical, and functional analysis, *Critical Reviews in Biochemistry and Molecular Biology 44*, 98-116.

16. Lohman, T. M., and Ferrari, M. E. (1994) Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities, *Annual Review of Biochemistry 63*, 527-570.

17. Murzin, A. G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences, *The EMBO Journal 12*, 861-867.

18. Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S. (2003) Nucleic acid recognition by OB-fold proteins, *Annual Review of Biophysics and Biomolecular Structure 32*, 115-133.

19. Flynn, R. L., and Zou, L. (2010) Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians, *Critical Reviews in Biochemistry and Molecular Biology 45*, 266-275.

20. Bochkarev, A., and Bochkareva, E. (2004) From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold, *Current Opinion in Structural Biology 14*, 36-42.

21. Yang, C., Curth, U., Urbanke, C., and Kang, C. (1997) Crystal structure of human mitochondrial single-stranded DNA binding protein at 2.4 Å resolution, *Nature Structural Biology 4*, 153-157.

22. Richard, D. J., Bolderson, E., Cubeddu, L., Wadsworth, R. I. M., Savage, K., Sharma, G. G., Nicolette, M. L., Tsvetanov, S., McIlwraith, M. J., Pandita, R. K., Takeda, S., Hay, R. T., Gautier, J., West, S. C., Paull, T. T., Pandita, T. K., White, M. F., and Khanna, K. K. (2008) Single-stranded DNA-binding protein hSSB1 is critical for genomic stability, *Nature 453*, 677-681.

23. Yuzhakov, A., Kelman, Z., Hurwitz, J., and O'Donnell, M. (1999) Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme, *The EMBO Journal 18*, 6189-6199.

24. Stauffer, M. E., and Chazin, W. J. (2004) Structural mechanisms of DNA replication, repair, and recombination, *The Journal of Biological Chemistry 279*, 30915-30918.

25. Ahn, B., Lee, J. W., Jung, H., Beck, G., and Bohr, V. A. (2009) Mechanism of Werner DNA helicase: POT1 and RPA stimulates WRN to unwind beyond gaps in the translocating strand, *PloS One 4*, e4673.

26. Brosh, R. M., Li, J. L., Kenny, M. K., Karow, J. K., Cooper, M. P., Kureekattil, R. P., Hickson, I. D., and Bohr, V. A. (2000) Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity, *The Journal of Biological Chemistry 275*, 23500-23508.

27. Kenny, M. K., Lee, S. H., and Hurwitz, J. (1989) Multiple functions of human single-stranded-DNA binding protein in Simian Virus 40 DNA replication: Single-strand stabilization and stimulation of DNA polymerases alpha and delta, *Proceedings of the National Academy of Sciences, U.S.A. 86*, 9757-9761.

28. de Laat, W. L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1998) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair, *Genes & Development 12*, 2598-2609.

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

30. Arunkumar, A. I., Klimovich, V., Jiang, X., Ott, R. D., Mizoue, L., Fanning, E., and Chazin, W. J. (2005) Insights into hRPA32 C-terminal domain--mediated assembly of the simian virus 40 replisome, *Nature Structural & Molecular Biology*. *12*, 332-339.

31. Fairman, M. P., and Stillman, B. (1988) Cellular factors required for multiple stages of SV40 DNA replication in vitro, *The EMBO Journal 7*, 1211-1218.

32. Wobbe, C. R., Weissbach, L., Borowiec, J. A., Dean, F. B., Murakami, Y., Bullock, P., and Hurwitz, J. (1987) Replication of simian virus 40 origin-containing DNA in vitro with purified proteins, *Proceedings of the National Academy of Sciences, U.S.A. 84*, 1834-1838.

33. Richter, A., Sapp, M., and Knippers, R. (1986) Are single-strand-specific DNA binding proteins needed for mammalian DNA replication?, *Trends in Biochemical Sciences 11*, 283.

34. Moore, S. P. (1991) The human homologous pairing protein HPP-1 is specifically stimulated by the cognate single-stranded binding protein hRP-A, *Proceedings of the National Academy of Sciences, U.S.A. 88*, 9067-9071.

35. O'Sullivan, R. J., and Karlseder, J. (2010) Telomeres: protecting chromosomes against genome instability, *Nature Reviews: Molecular Cell Biology 11*, 171-181.

36. Miller, S., Moses, K., Jayaraman, L., and Prives, C. (1997) Complex formation between p53 and replication protein A inhibits the sequence-specific DNA binding of p53 and is regulated by single- stranded DNA, *Molecular and Cellular Biology 17*, 2194-2201.

37. Mer, G., Bochkarev, A., Chazin, W. J., and Edwards, A. M. (2000) Three-dimensional structure and function of replication protein A, *Cold Spring Harbor Symposia 65*, 193-200.

38. Gomes, X. V., Henricksen, L. A., and Wold, M. S. (1996) Proteolytic mapping of human replication protein A: evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA, *Biochemistry 35*, 5586-5595.

39. Richard A. Pfuetzner, Alexey Bochkarev, Lori Frappier, and Alex M. Edwards. (1997) Replication Protein A. Characterization and crystallization of the dna binding domain, *Journal of Biological Chemistry 272*, 430-434.

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

41. Alexey Bochkarev, Elena Bochkareva, Lori Frappier, and Aled M. Edwards. (1999) The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding, *The EMBO Journal 18*, 4498-4504.

42. Bochkareva, E., Belegu, V., Korolev, S., and Bochkarev, A. (2001) Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding, *The EMBO Journal 20*, 612-618.

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

44. Jacobs, D. M., Lipton, A. S., Isern, N. G., Daughdrill, G. W., Lowry, D. F., Gomes, X., and Wold, M. S. (1999) 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 14*, 321-331.

45. Bochkareva, E., Kaustov, L., Ayed, A., Yi, G.-S., Lu, Y., Pineda-Lucena, A., Liao, J. C. C., Okorokov, A. L., Milner, J., Arrowsmith, C. H., and Bochkarev, A. (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A, *Proceedings of the National Academy of Sciences, U.S.A. 102*, 15412-15417.

46. Arunkumar, A. I., Stauffer, M. E., Bochkareva, E., Bochkarev, A., and Chazin, W. J. (2003) Independent and coordinated functions of replication protein A tandem high affinity single-stranded DNA binding domains, *The Journal of Biological Chemistry 278*, 41077-41082.

47. Daughdrill, G. W., Ackerman, J., Isern, N. G., Botuyan, M. V., Arrowsmith, C. H., Wold, M. S., and Lowery, D. F. (2001) The weak interdomain coupling observed in the 70 kDa subunit of human replication protein A is unaffected by ssDNA binding, *Nucleic Acids Research 29*, 3270-3276.

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

49. Deng, X., Habel, J. E., Kabaleeswaran, V., Snell, E. H., Wold, M. S., and Borgstahl, G. E. O. (2007) Structure of the full-length human RPA32/14 complex gives insight into the mechanism of DNA binding and complex formation, *Journal of Molecular Biology 49*, 865-876.

50. Henricksen, L., Umbricht, C., and Wold, M. (1994) Recombinant replication protein A: expression, complex formation, and functional characterization, *Journal of Biological Chemistry 269*, 11121-11132.

51. Stigger, E., Dean, F. B., Hurwitz, J., and Lee, S. H. (1994) Reconstitution of functional human single-stranded DNA-binding protein from individual subunits expressed by recombinant baculoviruses, *Proceedings of the National Academy of Sciences, U.S.A. 91*, 579-583.

52. Iftode, C., and Borowiec, J. A. (2000) 5‘ → 3‘ Molecular polarity of human replication protein A (hRPA) binding to pseudo-origin DNA substrates, *Biochemistry 39*, 11970-11981.

53. Lavrik, O. I., Kolpashchikov, D. M., Weisshart, K., Nasheuer, H.-P., Khodyreva, S. N., and Favre, A. (1999) RPA subunit arrangement near the 3’-end of the primer

is modulated by the length of the template strand and cooperative protein interactions, *Nucleic Acids Research 27*, 4235-4240.

54. Bastin-Shanower, S. A., and Brill, S. J. (2001) Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding, *The Journal of Biological Chemistry 276*, 36446-36453.

55. Kim, C., Snyder, R. O., and Wold, M. S. (1992) Binding properties of replication protein A from human and yeast cells., *Molecular and Cellular Biology 12*, 3050-3059.

56. Kim, C., Paulus, B. F., and Wold, M. S. (1994) Interactions of human replication protein A with oligonucleotides, *Biochemistry*. American Chemical Society *33*, 14197-14206.

57. Brill, S. J., and Bastin-Shanower, S. (1998) Identification and Characterization of the Fourth Single-Stranded-DNA Binding Domain of Replication Protein A, *Molecular and Cellular Biology 18*, 7225-7234.

58. Sibenaller, Z. A., Sorensen, B. R., and Wold, M. S. (1998) The 32- and 14-kilodalton subunits of replication protein A are responsible for species-specific interactions with single-stranded DNA, *Biochemistry 37*, 12496-12506.

59. Kim, C., and Wold, M. S. (1995) Recombinant human replication protein A binds to polynucleotides with low cooperativity, *Biochemistry 34*, 2058-2064.

60. Philipova, D., Mullen, J. R., Maniar, H. S., Lu, J., Gu, C., and Brill, S. J. (1996) A hierarchy of SSB protomers in replication protein A, *Genes & Development 10*, 2222-2233.

61. Bochkareva, E., Frappier, L., Edwards, A. M., and Bochkarev, A. (1998) The RPA32 Subunit of Human Replication Protein A Contains a Single-stranded DNA-binding Domain, *Journal of Biological Chemistry 273*, 3932-3936.

62. Zhou, H.-X. (2001) The Affinity-Enhancing Roles of Flexible Linkers in Two-Domain DNA-Binding Proteins, *Biochemistry 40*, 15069-15073.

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

64. Xu, X., Vaithiyalingam, S., Glick, G. G., Mordes, D. A., Chazin, W. J., and Cortez, D. (2008) The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling., *Molecular and Cellular Biology 28*, 7345-7353.

65. Wang, Y., Putnam, C. D., Kane, M. F., Zhang, W., Edelmann, L., Russell, R., Carrión, D. V., Chin, L., Kucherlapati, R., Kolodner, R. D., and Edelmann, W. (2005) Mutation in Rpa1 results in defective DNA double-strand break repair, chromosomal instability and cancer in mice, *Nature Genetics 37*, 750-755.

66. Brill, S. J., and Stillman, B. (1991) Replication factor-A from Saccharomyces cerevisiae is encoded by three essential genes coordinately expressed at S phase, *Genes & Development 5*, 1589-1600.

67. Chen, C., Umezu, K., and Kolodner, R. D. (1998) Chromosomal rearrangements occur in S. cerevisiae rfa1 mutator mutants due to mutagenic lesions processed by double-strand-break repair, *Molecular Cell 2*, 9-22.

68. Chen, C., and Kolodner, R. D. (1999) Gross chromosomal rearrangements in Saccharomyces cerevisiae replication and recombination defective mutants, *Nature Genetics 23*, 81-85.

69. Hass, C. S., Gakhar, L., and Wold, M. S. (2010) Functional characterization of a cancer causing mutation in human replication protein A, *Molecular Cancer Research 8*, 1017-1026.

70. Chen, L., Neubauer, A., Kurisu, W., Waldman, F. M., Ljung, B. M., Goodson, W., Goldman, E. S., Moore, D., Balazs, M., and Liu, E. (1991) Loss of heterozygosity on the short arm of chromosome 17 is associated with high proliferative capacity and DNA aneuploidy in primary human breast cancer, *Proceedings of the National Academy of Sciences, U.S.A 88*, 3847-3851.

71. Coles, C., Thompson, A. M., Elder, P. A., Cohen, B. B., Mackenzie, I. M., Cranston, G., Mackay, J., Macdonald, M., Steel, C. M., Chetty, U., Nakamura, Y., and Hoyheim, B. (1990) Evidence implicating at least two genes on chromosome 17p in breast carcinogenesis, *The Lancet 336*, 761-763.

72. Andrews, B. J., and Turchi, J. J. (2004) Development of a high-throughput screen for inhibitors of replication protein A and its role in nucleotide excision repair, *Molecular Cancer Therapeutics 3*, 385-391.

73. Bouley, J., Pionneau, C., Varinot, J., Biard, D., Genestie, C., Antoine, M., Coulet, F., Stern, M.-H., Stoppa-Lyonnet, D., and Soubrier, F. (2010) Proteomic analysis of BRCA1-depleted cell line reveals a putative role for replication protein A2 up-regulation in BRCA1 breast tumor development, *Proteomics--Clinical Applications 4*, 489-498.

74. Shuck, S. C., and Turchi, J. J. (2010) Targeted inhibition of Replication Protein A reveals cytotoxic activity, synergy with chemotherapeutic DNA-damaging agents, and insight into cellular function, *Cancer Research 70*, 3189-3198.

75. Anciano Granadillo, V. J., Earley, J. N., Shuck, S. C., Georgiadis, M. M., Fitch, R. W., and Turchi, J. J. (2010) Targeting the OB-Folds of Replication Protein A with Small Molecules, *Journal of Nucleic Acids 2010*, 304035.

76. Shuker, S. B., Hajduk, P. J., Meadows, R. P., and Fesik, S. W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR, *Science 274*, 1531-1534.

77. Glanzera, J. G., Liu, S., and Oakley, G. G. (2011) Small molecule inhibitor of the RPA70 N-terminal protein interaction domain discovered using in silico and in vitro methods, *Bioorganic and Medicinal Chemistry 19*, 2589-2595.

78. Blackwell, L., Borowiec, J., and Masrangelo, I. (1996) Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase, *Molecular and Cellular Biology 16*, 4798-4807.

79. Iftode, C., and Borowiec, J. (1997) Denaturation of the simian virus 40 origin of replication mediated by human replication protein A, *Molecular and Cellular Biology 17*, 3876-3883.

80. Pestryakov, P. E., Khlimankov, D. Y., Bochkareva, E., Bochkarev, A., and Lavrik, O. I. (2004) Human replication protein A (RPA) binds a primer-template junction in the absence of its major ssDNA-binding domains, *Nucleic Acids Research 32*, 1894-1903.

81. Blackwell, L. J., and Borowiec, J. A. (1994) Human replication protein A binds single-stranded DNA in two distinct complexes, *Molecular and Cellular Biology 14*, 3993-4001.

82. Cai, L., Roginskaya, M., Qu, Y., Yang, Z., Xu, Y., and Zou, Y. (2007) Structural characterization of human RPA sequential binding to single-stranded DNA using ssDNA as a molecular ruler, *Biochemistry 46*, 8226-8233.

83. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations, and assemblies in solution, *Quarterly Reviews of Biophysics 40*, 191-285.

84. Rambo, R. P., and Tainer, J. A. (2010) Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering, *Current Opinion in Structural Biology 20*, 128-137.

85. Perry, J. J. P., Cotner-Gohara, E., Ellenberger, T., and Tainer, J. A. (2010) Structural dynamics in DNA damage signaling and repair., *Current Opinion in Structural Biology 20*, 283-294.

86. Blackledge, M. (2010) Mapping the conformational mobility of multidomain proteins, *Biophysical Journal 98*, 2043-2044.

87. Markwick, P. R. L., Bouvignies, G., Salmon, L., McCammon, J. A., Nilges, M., and Blackledge, M. (2009) Toward a unified representation of protein structural dynamics in solution, *Journal of the American Chemical Society 131*, 16968-16975.

88. Tsutakawa, S. E., Hura, G. L., Frankel, K. A., Cooper, P. K., and Tainer, J. A. (2006) Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography, *Journal of Structural Biology 158*, 214-221.

89. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering, *Journal of the American Chemical Society 129*, 5656-5664.

90. Keeler, J. (2005) Understanding NMR Spectroscopy. John Wiley & Sons Ltd., Chichester, England.

91. Bloch, F., Hansen, W. W., and Packard, M. (1946) Nuclear Induction, *Physical Review 69*, 127.

92. Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946) Resonance Absorption by Nuclear Magnetic Moments in a Solid, *Physical Review 69*, 37-38.

93. Bloch, F. (1946) Nuclear Induction, *Physical Review 70*, 460-474.

94. Bloch, F., Hansen, W. W., and Packard, M. (1946) The Nuclear Induction Experiment, *Physical Review 70*, 474-485.

95. Solomon, I. (1955) Relaxation Processes in a System of Two Spins, *Physical Review 99*, 559-565.

96. Bruschweiler, R., and Case, D. A. (1994) Characterization of biomolecular structure and dynamics by NMR cross relaxation, *Progress in Nuclear Magnetic Resonance Spectroscopy 26*, 27-58.

97. Luginbuhl, P., and Wuthrich, K. (2002) Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules, *Progress in Nuclear Magnetic Resonance Spectroscopy 40*, 199-247.

98. Korzhnev, D. M., Billeter, M., Arseniev, A. S., and Orekhov, V. Y. (2001) NMR studies of Brownian tumbling and internal motions in proteins, *Progress in Nuclear Magnetic Resonance Spectroscopy 38*, 197-266.

99. Li, M. X., Corson, D. C., and Sykes, B. D. (2002) Structure determination by NMR: Isotope labeling, *Methods in Molecular Biology 173*, 255-65.

100. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, *Chemical Physics Letters 69*, 185-189.

101. Rule, G. S., and Hitchens, T. K. (2006) Fundamentals of Protein NMR Spectroscopy, pp 277-312. Springer, Dordrecht, The Netherlands.

102. Tugarinov, V., and Kay, L. E. (2003) Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G, *Journal of Molecular Biology 327*, 1121-1133.

103. Horst, R., Bertelsen, E. B., Fiaux, J., Wider, G., Horwich, A. L., and Wüthrich, K. (2005) Direct NMR observation of a substrate protein bound to the chaperonin GroEL, *Proceedings of the National Academy of Sciences, U.S.A. 102*, 12748-12753.

104. Helgstrand, M., Mandava, C. S., Mulder, F. A. A., Liljas, A., Sanyal, S., and Akke, M. (2007) The ribosomal stalk binds to translation factors IF2, EF-Tu, EF-G and RF3 via a conserved region of the L12 C-terminal domain, *Journal of Molecular Biology 365*, 468-79.

105. Hsu, S.-T. D., Fucini, P., Cabrita, L. D., Launay, H., Dobson, C. M., and Christodoulou, J. (2007) Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy, *Proceedings of the National Academy of Sciences, U.S.A. 104*, 16516-16521.

106. Matthews, S. (2004) Perdeuteration/site-specific protonation approaches for high-molecular-weight proteins, *Methods in Molecular Biology 278*, 35-45.

107. Pervushin, K., Roland Riek, Gerhard Wider, and Kurt Wüthrich. (1997) 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, U.S.A. 94*, 12366-12371.

108. Riek, R., Wider, G., Pervushin, K., and Üthrich, K. (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules, *Proceedings of the National Academy of Sciences, U.S.A. 96*, 4918-4923.

109. Riek, R., Pervushin, K., and Wüthrich, K. (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution, *Trends in Biochemical Sciences 25*, 462-468.

110. Palmer, A. G. (2004) NMR characterization of the dynamics of biomacromolecules, *Chemical Reviews 104*, 3623-3640.

111. Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995) Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation, *Journal of the American Chemical Society 117*, 12562-12566.

112. Fushman, D., Xu, R., and Cowburn, D. (1999) Direct determination of changes of interdomain orientation on ligation: use of the orientational dependence of 15N NMR relaxation in Abl SH(32), *Biochemistry 38*, 10225-10230.

113. Walker, O., Varadan, R., and Fushman, D. (2004) Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from 15N relaxation data using computer program ROTDIF, *Journal of Magnetic Resonance 168*, 336-345.

114. d'Auvergne, E. J. and Gooley, P. R. (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces, *Journal of Biomolecular NMR 40*, 107-119.

115. d'Auvergne, E. J. and Gooley, P. R. (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor, *Journal of Biomolecular NMR 40*, 121-133.

116. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *Journal of the American Chemical Society 104*, 4546-4559.

117. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, *Journal of the American Chemical Society 104*, 4559-4570.

118. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins, *Journal of the American Chemical Society 112*, 4989-4991.

119. Wong, V., Case, D. A., and Szabo, A. (2009) Influence of the coupling of interdomain and overall motions on NMR relaxation., *Proceedings of the National Academy of Sciences,U.S.A. 106*, 11016-11021.

120. Peng, J. W., and Wagner, G. (1992) Mapping of spectral density functions using heteronuclear NMR relaxation measurements, *Journal of Magnetic Resonance 98*, 308-332.

121. Peng, J. W., and Wagner, G. (1992) Mapping of the spectral densities of nitrogen-hydrogen bond motions in Eglin c using heteronuclear relaxation experiments, *Biochemistry 31*, 8571-8586.

122. Farrow, N., Zhang, O., Szabo, A., Torchia, D., and Kay, L. (1995) Spectral density function mapping using 15N relaxation data exclusively, *Journal of Biomolecular NMR 6*, 153-162.

123. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., Tsutakawa, S. E., Jenney, F. E., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S.-J., Scott, J. W., Dillard, B. D., Adams, M. W. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS), *Nature Methods 6*, 606-612.

124. Bernadó, P. (2010) Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering, *European Biophysics Journal 39*, 769-780.

125. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) Determination of domain structure of proteins from x-ray solution scattering, *Biophysical Journal 80*, 2946-2953.

126. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *Journal of Applied Crystallography 42*, 342-346.

127. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data., *Biophysical Journal 89*, 1237-1250.

128. Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering, *General Physiology and Biophysics 28*, 174-180.

129. Jacrot, B. (1976) The study of biological structures by neutron scattering from solution, *Reports in Progress in Physics 39*, 911-953.

130. Whitten, A. E., and Trewhella, J. (2009) Small-angle scattering and neutron contrast variation for studying bio-molecular complexes, *Methods in Molecular Biology 544*, 307-323.

131. Whitten, A. E., Cai, S., and Trewhella, J. (2008) MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies, *Journal of Applied Crystallography 41*, 222-226.

132. Petoukhov, M. V., and Svergun, D. I. (2006) Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution., *European Biophysics Journal 35*, 567-576.

133. Grillo, I. (2008) Soft Matter Characterization: Small-Angle Neutron Scattering and Applications in Soft Condensed Matter (Borsali, R., and Pecora, R., Eds.), pp 723-782. Springer Netherlands, Dordrecht.

134. Riek, R., Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wüthrich, K. (2002) Solution NMR techniques for large molecular and supramolecular structures, *Journal of the American Chemical Society 124*, 12144-12153.

135. Fotedar, R., and Roberts, J. M. (1992) Cell cycle regulated phosphorylation of RPA-32 occurs within the replication initiation complex, *The EMBO Journal 11*, 2177-2187.

136. Fang, F., and Newport, J. (1993) Distinct roles of cdk2 and cdc2 in RP-A phosphorylation during the cell cycle, *Journal of Cell Science 106*, 983-994.

137. Binz, S. K., Sheehan, A. M., and Wold, M. S. (2004) Replication Protein A phosphorylation and the cellular response to DNA damage, *DNA Repair 3*, 1015-1024.

138. Zou, Y., Liu, Y., Wu, X., and Shell, S. M. (2006) Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses., *Journal of Cellular Physiology 208*, 267-273.

139. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *Journal of Biomolecular NMR 6*, 227-293.

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

141. Bhattacharya, S., Arunkumar, A. I., Sullivan, S. L., Botuyan, M.-V., Arrowsmith, C. H., and Chazin, W. J. (2004) 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 28*, 195-196.

142. Brosey, C. A., Chagot, M.-E., Ehrhardt, M., Pretto, D. I., Weiner, B. E., and Chazin, W. J. (2009) NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA., *Journal of the American Chemical Society 131*, 6346-6347.

143. Pretto, D. I., Tsutakawa, S., Brosey, C. A., Castillo, A., Chagot, M.-E., Smith, J. A., Tainer, J. A., and Chazin, W. J. (2010) Structural dynamics and single-stranded

DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering, *Biochemistry 49*, 2880-2889.

144. Bhattacharya, S., Botuyan, M.-V., Hsu, F., Shan, X., Arunkumar, A. I., Arrowsmith, C. H., Edwards, A. M., and Chazin, W. J. (2002) Characterization of binding-induced changes in dynamics suggests a model for sequence-nonspecific binding of ssDNA by replication protein A, *Protein Science 11*, 2316-2325.

145. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry 28*, 8972-8979.

146. Skelton, J. N., Palmer, A. G., Akke, M., Kordel, J., Rance, M., and Chazin, W. J. (1993) Practical aspects of two-dimensional proton-detected 15N spin relaxation measurements, *Journal of Magnetic Resonance, Series B 102*, 253-264.

147. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (2002) Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity, *Journal of Magnetic Resonance 102*, 241-245.

148. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme, *Journal of Molecular Biology 246*, 144-163.

149. García De La Torre, J., Huertas, M. L., and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations, *Journal of Magnetic Resonance 147*, 138-146.

150. Bernadó, P., García De La Torre, J., and Pons, M. (2002) Interpretation of 15N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR, *Journal of Biomolecular NMR 23*, 139-150.

151. Fushman, D., Ghose, R., and Cowburn, D. (2000) The effect of finite sampling on the determination of orientational properties: A theoretical treatment with application to interatomic vectors in proteins, *Journal of the American Chemical Society 122*, 10640-10649.

152. Brosey, C. A., Chagot, M.-E., Ehrhardt, M., Pretto, D. I., Weiner, B. E., and Chazin, W. J. (2009) NMR analysis of the architecture and functional remodeling of a modular multidomain protein, RPA., *Journal of the American Chemical Society*. American Chemical Society *131*, 6346-7.

153. García de la Torre, J., Pérez Sánchez, H. E., Ortega, A., Hernández, J. G., Fernandes, M. X., Díaz, F. G., and López Martínez, M. C. (2003) Calculation of

the solution properties of flexible macromolecules: methods and applications., *European Biophysics Journal 32*, 477-486.

154. Bae, S.-H., Dyson, H. J., and Wright, P. E. (2009) Prediction of the rotational tumbling time for proteins with disordered segments, *Journal of the American Chemical Society 131*, 6814-6821.

155. Walsh, J. D., Meier, K., Ishima, R., and Gronenborn, A. M. (2010) NMR studies on domain diffusion and alignment in modular GB1 repeats, *Biophysical Journal 99*, 2636-2646.

156. Kowalczykowski, S. C. (2000) Some assembly required, *Nature structural biology 7*, 1087-1089.

157. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis, *Journal of Applied Crystallography 40*, s223-s228.

158. Svergun, V. V. V. and D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering, *Journal of Applied Crystallography 36*, 860-864.

159. The PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC.

160. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis., *Journal of Computational Chemistry 25*, 1605-1612.

161. Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures., *Protein Science 9*, 1753-1773.

162. Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering, *General Physiology and Biophysics 28*, 174-189.

163. Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information., *Journal of the American Chemical Society 125*, 1731-1737.

164. Dijk, M. van, and Bonvin, A. M. J. J. (2010) Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance, *Nucleic Acids Research 38*, 5634-5647.

165. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles, *Nucleic Acids Research 38*, W540-544.

166. Bernadó, P., Modig, K., Grela, P., Svergun, D. I., Tchorzewski, M., Pons, M., and Akke, M. (2010) Structure and Dynamics of Ribosomal Protein L12: An Ensemble Model Based on SAXS and NMR Relaxation., *Biophysical Journal 98*, 2374-2382.

167. Patrick, S. M., and Turchi, J. J. (2001) Stopped-flow kinetic analysis of replication protein A-binding DNA: damage recognition and affinity for single-stranded DNA reveal differential contributions of k(on) and k(off) rate constants, *The Journal of Biological Chemistry 276*, 22630-22637.

168. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and instrinsically unstructured biological macromolecules by SAS using the porod-debye law, *Biopolymers 95*, 559-571.

169. Binz, S. K., and Wold, M. S. (2008) Regulatory functions of the N-terminal domain of the 70-kDa subunit of replication protein A (RPA), *The Journal of Biological Chemistry 283*, 21559-21570.

170. Vaithiyalingam, S., Warren, E. M., Eichman, B. F., and Chazin, W. J. (2010) 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, U.S.A. 107*, 13684-13689.

171. Mass, G., Nethanel, T., and Kaufmann, G. (1998) The Middle Subunit of Replication Protein A Contacts Growing RNA-DNA Primers in Replicating Simian Virus 40 Chromosomes, *Molecular and Cellular Biology 18*, 6399-6407.

172. Mass, G., Nethanel, T., Lavrik, O. I., Wold, M. S., and Kaufmann, G. (2001) Replication protein A modulates its interface with the primed DNA template during RNA-DNA primer elongation in replicating SV40 chromosomes, *Nucleic Acids Research 29*, 3892-3899.

173. Lipsitz, R. S., and Tjandra, N. (2004) Residual dipolar couplings in NMR structure analysis, *Annual Review of Biophysics and Biomolecular Structure 33*, 387-413.

174. Tolman, J. R., and Al-Hashimi, H. M. (2003) NMR studies of biomolecular dynamics and structural plasticity using residual dipolar couplings, *Annual Reports on NMR Spectroscopy 51*, 105-166.

175. Fischer, M. W., Losonczi, J. A., Weaver, J. L., and Prestegard, J. H. (1999) Domain orientation and dynamics in multidomain proteins from residual dipolar couplings, *Biochemistry 38*, 9013-9022.

176. Maciejewski, M. M., Tjandra, N., and Barlow, P. N. (2011) Estimation of inter-domain flexibility of N-terminus of factor H using residual dipolar couplings, *Biochemistry 50*¸ 8138-8149.

177. Battiste, J. L., and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data, *Biochemistry 39*, 5355-5365.

178. Donaldson, L. W., Skrynnikov, N. R., Choy, W.-Y., Muhandiram, D. R., Sarkar, B., Forman-Kay, J. D., and Kay, L. E. (2001) Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy, *Journal of the American Chemical Society 123*, 9843-9847.

179. Iwahara, J., Tang, C., and Marius Clore, G. (2007) Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules, *Journal of Magnetic Resonance 184*, 185-195.

180. Heller, W. T. (2005) Influence of multiple well defined conformations on small-angle scattering of proteins in solution, *Acta crystallographica. Section D 61*, 33-44.

181. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophysical Journal 76*, 2879-2886.

182. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E., and Kay, L. E. (2007) Measurement of bond vector orientations in invisible excited states of proteins, *Proceedings of the National Academy of Sciences, U.S.A. 104*, 18473–18477.

183. Hansen, D. F., Vallurupalli, P., and Kay, L. E. (2008) Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states, *Journal of Biomolecular NMR 41*, 113-120.

184. Baldwin, A. J., and Kay, L. E. (2009) NMR spectroscopy brings invisible protein states into focus, *Nature Chemical Biology 5*, 808-814.

185. Muona, M., Aranko, A. S., Raulinaitis, V., and Iwaï, H. (2010) Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro, *Nature Protocols 5*, 574-587.

186. Sprangers, R., and Kay, L. E. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR, *Nature 445*, 618-622.

187. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., I., Rance, M., and Skelton, N. J. (2007) Protein NMR Spectroscopy: Principles and Practice 2nd ed., pp 29-113. Elsevier Academic Press, Burlington.

188. Rule, G. S., and Hitchens, T. K. (2006) Fundamentals of Protein NMR Spectroscopy, pp 89-134. Springer, Dordrecht, The Netherlands.

189. Bacon, G. E., and Lonsdale, K. (1953) Neutron diffraction, *Reports on Progress in Physics 16*, 1-61.