

AN *IN VITRO* CHARACTERIZATION OF FUNCTIONAL INTERACTIONS BETWEEN PURIFIED
TELOMERE REPEAT BINDING FACTORS 1 AND 2 AND RAD51 RECOMBINASE

Brian D. Bower

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial
fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of
Genetics and Molecular Biology.

Chapel Hill
2014

Approved By

Jack D. Griffith

Shawn Ahmed

Michael B. Jarstfer

Dale A. Ramsden

Jeff J. Sekelsky

© 2014
Brian D. Bower
ALL RIGHTS RESERVED

ABSTRACT

BRIAN D BOWER: AN *IN VITRO* CHARACTERIZATION OF FUNCTIONAL INTERACTIONS BETWEEN PURIFIED TELOMERE REPEAT BINDING FACTORS 1 AND 2 AND RAD51 RECOMBINASE

(Under the direction of Jack D. Griffith)

A growing body of literature suggests that the homologous recombination/repair (HR) pathway cooperates with components of the shelterin complex to promote both telomere maintenance and non-telomeric HR. This may be due to the ability of both HR and shelterin proteins to promote strand invasion, wherein a single-stranded DNA (ssDNA) substrate base pairs with a homologous double-stranded DNA (dsDNA) template displacing a loop of ssDNA (D-loop). Rad51 recombinase catalyzes D-loop formation during HR, and telomere repeat-binding factor 2 (TRF2) catalyzes the formation of a telomeric D-loop that stabilizes a looped structure in telomeric DNA (t-loop) that may facilitate telomere protection. We have characterized this functional interaction *in vitro* using a fluorescent D-loop assay measuring the incorporation of Cy3-labeled 90 nucleotide telomeric and non-telomeric substrates into telomeric and non-telomeric plasmid templates. We report that pre-incubation of a telomeric template with TRF2 inhibits the ability of Rad51 to promote telomeric D-loop formation when pre-incubated with a telomeric substrate. This suggests Rad51 does not facilitate t-loop formation, and suggests a mechanism whereby TRF2 can inhibit HR at telomeres. We also report a TRF2 mutant lacking the dsDNA binding domain promotes Rad51-mediated non-telomeric D-loop formation, possibly explaining how TRF2 promotes non-telomeric HR. Finally, we report telomere repeat binding factor 1 (TRF1) promotes Rad51-mediated telomeric D-loop formation, which may facilitate HR-mediated replication fork restart and explain why TRF1 is required for efficient telomere replication.

To my grandmother, Elanor Shanz

ACKNOWLEDGEMENTS

I would like to thank:

My thesis advisor Jack Griffith for his unyielding support,
generosity and guidance throughout the course of my doctoral research

My family and friends for their love and support

My thesis committee for their thoughtful guidance and encouragement

TABLE OF CONTENTS

LIST OF FIGURES	ix
LIST OF TABLES	xii
LIST OF ABBREVIATIONS AND SYMBOLS	xiii
CHAPTER	PAGE
1 INTRODUCTION	1
Telomeric DNA	1
Telomere proteins	2
Homologous recombination/repair proteins.....	2
Chromosome end capping.....	3
Higher order telomere structure	4
DNA repair at the telomeres	5
Telomeric proteins and DNA repair	6
Scope of dissertation	7
REFERENCES	9
2 TRF2 AND TRF1 DIFFERENTIALLY MODULATE RAD51-MEDIATED DISPLACEMENT LOOP FORMATION IN VITRO.....	13
INTRODUCTION	13
MATERIALS AND METHODS.....	14
RESULTS	18
Fluorescent D-Loop Assay	18
TRF2 Inhibits Rad51-Mediated Telomeric But Not Non-Telomeric D-Loop Formation.....	19

TRF2ΔM Promotes Rad51-Mediated Telomeric But Not Non-Telomeric D-Loop	26
TRF2ΔB Inhibits Telomeric But Not Non-Telomeric Rad51-Mediated D-Loop Formation.....	28
TRF1 Promotes Rad51-Mediated Telomeric But Not Non-Telomeric D-Loop Formation.....	28
DISCUSSION.....	31
REFERENCES	35
3 BIOPHYSICAL AND ULTRASTRUCTURAL CHARACTERIZATION OF ADENO-ASSOCIATED VIRUS CAPSID UNCOATING AND GENOME RELEASE	38
INTRODUCTION	38
MATERIALS AND METHODS.....	39
RESULTS	46
Heat-induced DNA release depends on genome length.....	46
scAAV are more thermostable than ssAAV	50
Heat-induced exposure of VP1 is not dependent on genome size or complementarity	53
Tungsten-shadowing EM shows DNA secondary structure for dsDNA vectors	54
Molecular dynamics provides insight into intra-capsid genome organization.....	54
EM quantification of DNA release following thermal denaturation.....	58
Attempts at characterizing intra-capsid structure by EM.....	60
DISCUSSION	62
REFERENCES	66
4 A GUANOSINE-CENTRIC MECHANISM FOR RNA CHAPERONE FUNCTION	70
INTRODUCTION	70

MATERIALS AND METHODS.....	70
RESULTS & DISCUSSION.....	75
Time-resolved SHAPE analysis of MuLV RNA dimerization.....	77
Model-free clustering of nt-resolution kinetic dimerization profiles.....	79
Initial Interactions between NC and UP1 with MuLV monomer.....	82
Role of guanosine in RNA structure and RNA chaperone mechanism.....	83
REFERENCES.....	90
5 AN INVESTIGATION OF HETEROGENOUS RIBONUCELOPROTEIN A1 AND UNWINDING PROTEIN 1 DNA BINDING CHARACTERISTICS.....	92
INTRODUCTION.....	92
MATERIALS AND METHODS.....	93
RESULTS.....	95
hnRNP A1 & UP1 bind to ssDNA.....	95
hnRNP A1 bind preferentially to telomeric ssDNA.....	99
Possible deleterious effects of N-terminal tagging.....	103
DISCUSSION.....	104
REFERENCES.....	106
6 CONCLUSIONS AND FINAL THOUGHTS.....	108
Homologous recombination/repair in telomere maintenance & protection.....	108
Intra-capsid AAV genome organization is unamenable to EM characterization.....	110
REFERENCES.....	112

LIST OF FIGURES

FIGURE 1.1: Diagram of telomeric DNA	1
FIGURE 1.2: Telomere binding proteins.....	2
FIGURE 1.3: Representative telomere loops reported in the literature	5
FIGURE 2.S1: Proteins Used	19
FIGURE 2.1: TRF-mediated telomeric D-loop formation.....	20
FIGURE 2.2: TRF2 inhibits Rad51-mediated telomeric but not non-telomeric D-loop formation.....	21
FIGURE 2.S5: Area under the curve (AUC) calculation procedures	22
FIGURE 2.S2: TRF2-mediated D-loop formation requires telomeric homology while Rad51 does not.....	23
FIGURE 2.S3: TRF2 inhibits Rad51-mediated telomeric D-loop formation only when added early in D-loop reactions.....	25
FIGURE 2.3: TRF2 Δ M promotes Rad51-mediated non-telomeric but not telomeric D-loop formation	27
FIGURE 2.4: TRF2 Δ B promotes Rad51-mediated telomeric but not non-telomeric D-loop formation.....	29
FIGURE 2.5: TRF1 promotes Rad51-mediated telomeric but not non-telomeric D-loop formation.....	30
FIGURE 2.6: TRF1 and TRF2 differentially modulate Rad51-mediated telomeric and non-telomeric D-loop formation	34
FIGURE 3.1: Ultrastructural characterization of AAV capsid uncoating	47
FIGURE 3.2: Effect of packaged genome length on AAV capsid uncoating. ssAAV	48
FIGURE 3.3: Quantitative analysis of TEM images	49
FIGURE 3.4: Fluorimetric detection of AAV genome release under different conditions	51

FIGURE 3.5: Fluorimetric analysis of AAV capsid uncoating	
as a function of genome length and self-complementarity	52
FIGURE 3.6: Exposure of the VP1 N termini is not dependent	
on genome length or self-complementarity.....	53
FIGURE 3.7: Tungsten-shadowing EM of released genomes from ssCMV-GFP	55
FIGURE 3.8: Computational modeling of internal capsid	
pressure and genome organization.....	57
FIGURE 3.9: Characteristics of capsid adherence and	
genome release following thermal denaturation	59
FIGURE 3.10: EM analysis of end structure and genome length.....	61
FIGURE 3.11: Psoralen fails to crosslinks scAAV genomes in intact capsids	63
FIGURE 4.S1: Secondary structures of monomer and dimer states.....	76
FIGURE 4.1: Time-resolved SHAPE analysis of MuLV RNA dimerization.....	77
FIGURE 4.2: Model-free clustering of nucleotide-resolution	
kinetic profiles for dimerization.....	79
FIGURE 4.S2: Time-resolved SHAPE of the native sequence	
MuLV RNA dimerization at 600nM and 200 nM	80
FIGURE 4.3: Initial interactions between NC and UP1 with the MuLV monomer	82
FIGURE 4.4: Role of guanosine in RNA structure and	
mechanism of chaperone-mediated RNA folding.....	83
FIGURE 4.S3: Native and inosine-substituted RNA secondary structures	84
FIGURE 4.S4: Visualization of MuLV genomic RNA	
dimerization by non-denaturing electrophoresis.....	87
FIGURE 4.S5: Time-resolved SHAPE of UP1-mediated	
RNA dimerization of the native MuLV RNA.....	88

FIGURE 4.S6: Interactions between NC and UP1 with the MuLV monomer and dimer states.....	89
FIGURE 5.1: Electrophoretic mobility shift assay demonstrating hnRNP A1 ssDNA binding	96
FIGURE 5.2: hnRNP A1 binds to M13 ssDNA	98
FIGURE 5.3: Dynamic light scattering reveals mostly homogenous hnRNP A1	98
FIGURE 5.4: hnRNP A1 binds to strand displaced telomeric but not non-telomeric DNA	99
FIGURE 5.5: hnRNP A1 binds preferentially to one end of a template with telomeric and non-telomeric ssDNA on opposite sides of a linear template	100
FIGURE 5.6: Comparative binding of <i>E. coli</i> and hnRNP A1 to exposed telomeric ssDNA	101
FIGURE 5.7: Optimization of hnRNP A1 binding conditions	102
FIGURE 5.8: Optimization of UP1 binding conditions.....	103

LIST OF TABLES

TABLE 2.1: Properties of TRF2, TRF2 Δ B, TRF2 Δ M and TRF1 22

LIST OF ABBREVIATIONS AND SYMBOLS

5-FAM	5-Carboxyfluorescein
6-JOE	6-Carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein
Å	Angstrom
θ	Theta
κ	Kappa
Σ	Sigma
Δ	Delta
α	Alpha
AAV	Adeno-associated virus
AFM	Atomic force microscopy
AUC	Area under the curve
ATM	Ataxia telangiectasia mutated
bp	Base pair
BSA	Bovine serum albumin
BzCN	benzoyl cyanide
CBA	Chicken β Actin
CCD	Charge coupled device
cDNA	Coding DNA
CMV	Cytomegalovirus
Cy3	Cyanine 3
Δ	Delta
DNA	Deoxyribonucleic acid
D-loop	Displacement loop
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol

dsDNA	Double-stranded deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EF1a	Elongation factor 1 α
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
FANCD2	Fanconi anemia group D2 protein
GAG	Group specific antigen
GFP	Green fluorescent protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HJ	Holliday junction
hnRNP	Heterogenous ribonucleoprotein
HR	Homologous recombination/repair
HRP	Horseradish peroxidase
IRES	Internal ribosome entry site
ITR	Inverted terminal hairpin repeat
kbp	Kilobase pair
MD	Molecular dynamics
MRN	Mre11/Rad50/Nbs1 complex
MuLV	Moloney murine leukemia virus
MVM	Minute virus of mice
NC	Nucleoprotein capsid
NHEJ	Non-homologous end joining
nt	Nucleotide
PAL1	Palindromic sequence 1
PAL2	Palindromic sequence 2
PBS	Phosphate buffered saline

PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PTM	Post-translational modification
RNA	Ribonucleic acid
scAAV	Self-complimentary adeno associated virus
SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
SL1	Stem loop 1
SL2	Stem loop 2
ssAAV	Single-stranded adeno associated virus
ssDNA	Single-stranded deoxyribonucleic acid
ss/dsDNA	Single-stranded to double-stranded deoxyribonucleic acid
SDS	Sodium dodecyl sulfate
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline with Tween 20
TE	Tris-buffered EDTA solution
TEM	Transmission electron microscopy
t-loop	Telomere loop
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
TRF1	Telomere repeat-binding factor 1
TRF2	Telomere repeat-binding factor 2
TIN2	TRF1 interacting nuclear protein 2
UP1	Unwinding protein 1
vg	Viral genome
VP	Viral protein
WGA	Wheat germ agglutinin

WT

Wild type

wtAAV

Wild type adeno associated virus

promotes resection of the 5' end of the DSB to generate a 3' ssDNA tail vaguely similar in structure to the 3' tail of a telomere¹⁴. This ssDNA is rapidly bound by replication protein A (RPA)¹⁵, but is subsequently displaced by Rad51, the recruitment of which is promoted by breast cancer associated gene 2 (BRCA2) at the ss/dsDNA junction^{16, 17}. Rad51 then facilitates a process of homology search, whereby the ssDNA substrate interrogates available dsDNA for complimentary sequence¹⁸. Rad51 first forms a protein-mediated complex between the substrate ssDNA and complimentary dsDNA and thereafter promotes protein-independent base pairing of the substrate and template in an ATP dependent manner generating a displacement loop (D-loop) within the template¹⁸. Following displacement loop formation a variety of HR factors, including RecQ helicases such as Blooms (BLM)^{19, 20} and Werner (WRN)²⁰ helicase, promote the migration and expansion of the D-loop and the eventual formation of a Holliday Junction (HJ) wherein both strands of the damaged substrate are paired with homologous sequences. These holiday junctions can then be cleaved or 'resolved' by HJ resolvases, such as the SLX1/4 complex, leaving behind two single-strand breaks (SSBs), which are then ligated²¹.

Chromosome End Capping

As mentioned above, the ends of linear chromosomes superficially resemble the structure of a resected double-strand break²². However, cells must prevent the aberrant recognition of their telomeres as double-strand breaks. Failure to prevent this recognition may result in the initiation of cell cycle arrest, which if left unresolved can prompt apoptosis or senescence. Additionally, aberrant recognition of the telomeres as DNA damage can elicit aberrant repair of the telomeres as DNA damage. Aberrant repair of telomeres often results in the formation of a dicentric chromosome via joining of two telomeres²³. This presents an obstacle to chromosome segregation upon cell division, and can result in failed segregation or in chromosome breakage²⁴. To prevent these undesirable processes, human telomeres are protected by a number of processes. Most importantly, TRF2 and its accessory protein RAP1, have been reported to inhibit the activation of telomere-bound DNA-PK_{cs}, a critical step in NHEJ²⁵. Likewise, TRF2 has been reported to inhibit activation of the ataxia telangiectasia mutated (ATM) kinase, which is required for the

activation of MRNs exonucleolytic activities²⁶. However, the exact mechanism or mechanisms that mediate telomere protection or ‘capping’ *in vivo* remain only poorly understood and despite these processes it appears that unperturbed telomeres are transiently recognized as DNA damage during the cell cycle in spite of these processes²⁷.

Higher Order Telomere Structure

It has been discovered that telomeres possess a variety of unusual structures both at the level of primary sequence and the tertiary structure of the telomere nucleoprotein complex. The G-rich ssDNA tail can form an unusual secondary structure wherein four triguanosine tracts can form a quadruplex (G4) stabilized by a number of monovalent ions²⁸. This property is suspected to affect the replication of telomeric DNA as well²⁹. The C-rich complement of the G-rich strand of the telomeres is replicated discontinuously. As such, the G-rich strand is left transiently single-stranded, which may permit formation of G4 structures. These structures are likely to inhibit DNA replication directly by interfering with nucleotide addition and indirectly by promoting fork slipping. Accordingly, telomeres have been observed to be particularly difficult to replicate, and are prone to defects consistent with fork stalling and slipping³⁰. It has also been observed that telomeres from a variety of species often adopt a lariat or loop-like structure (T-loop)³¹⁻³⁴, which appear to be stabilized by base pairing between the ssDNA tail and internal telomeric sequence. T-loops appear to be subject to cleavage by HJ resolvases, such as the SLX1/4 complex³⁵, and resolution of T-loops is believed to result in the generation of extrachromosomal circular telomere DNAs (T-circles)³⁵. These T-circles are generated upon disruption of factors that may promote migration or dissociation of the D-loop or HJ that stabilizes the T-loop, such as RecQ helicases^{35, 36} or the regulator of telomere length 1 (RTEL1) helicase³⁵. T-circles are also observed to be highly abundant in cells with unusually long telomeres or telomeres that are maintained in a HR mediated and telomerase independent alternative pathway (ALT)³⁷. While it has been suggested that T-circles may play a role in lengthening telomeres, either by re-incorporation to the telomeres or by priming rolling-circle replication, these hypotheses have not been demonstrated in human cells *in vivo*³⁸.

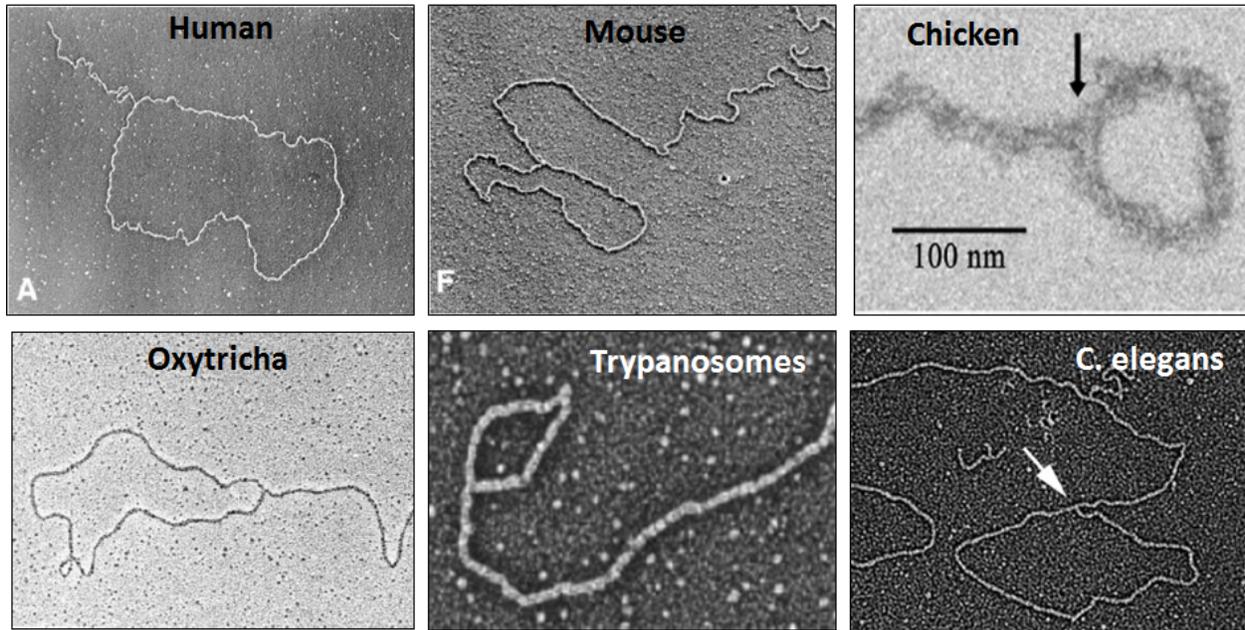


Figure 1.3: Representative telomere loops reported in the literature. T-loops have been observed at the telomeres of a variety of only-distantly related species by a variety of groups³¹⁻³⁴ using an assortment of techniques.

DNA Repair at the Telomeres

It has been paradoxically observed that telomere maintenance and protection require the activity of several DNA repair proteins. In contrast to HR, which is initiated by the MRN complex, NHEJ is initiated by binding of the Ku heterodimer (Ku 70 & Ku 80) and its associated DNA dependent protein kinase catalytic subunit (DNA-PK_{cs}). While TRF2 and Rap1 inhibit DNA-PK_{cs} activation at telomeres, it has been observed that both Ku and DNA-PK_{cs} are required for proper telomere maintenance. Disrupting the expression of Ku or DNA-PK_{cs} results in telomere shortening or aberrant telomere repair in a variety of mouse or human cell lines. This may be due to Ku's interaction with TRF2 or WRN helicase, which may facilitate telomere protection. Alternately, the binding of Ku to telomeres may inhibit alternative DSB repair pathways when activation of DNA-PK_{cs} is inhibited by the TRF2/Rap1 complex.

Likewise, it has been reported the BRCA2 recruits Rad51 to telomeres in a cell-cycle dependent manner, and disruption of the expression of either of these proteins results in telomere maintenance defects. It has been suggested that this defect may be due a telomere replication defect, rather than a

defect in telomere protection or capping. However, disrupting the expression of BRCA2 or Rad51 results in aberrant telomere repair even in quiescent and non-dividing cells, suggesting that the HR pathway is also required for telomere replication. It has been suggested that HR proteins including Rad51 may facilitate the formation of the T-loop structure. Immunodepletion of either TRF2 or Rad51 from nuclear extracts ablated the ability of those extracts to promote telomeric displacement loop formation; a requisite step in T-loop formation. This activity can be recovered upon addition of purified Rad51 or TRF2 to the immunodepleted extracts. However, there is evidence that TRF2 may also inhibit HR mediated processes at telomeres of humanized yeast strains⁵⁶.

While these sundry DNA repair proteins are known to play some important role in telomere biology, the nature of that role remains poorly understood. Our data suggest that HR processes at the telomeres are differentially regulated by TRF1 and TRF2. Literature suggests that this differential regulation may also be influenced by a variety of post-translational modifications and are mediated by several different protein domains. Many of these domains remain incompletely characterized. As such our understanding the actual role or roles of DNA repair proteins in telomere maintenance and biology remains cursory at best.

Telomeric Proteins and DNA Repair

Telomere proteins have also been observed to play a poorly understood role in DNA repair and maintenance both at telomeric and non-telomeric locations. The most well characterized proteins that have these paradoxical interactions are TRF1 and TRF2. TRF2 is rapidly phosphorylated in response to the induction of non-telomeric DSBs³⁹. Inhibition of this phosphorylation results in impaired break repair, suggesting that TRF2 does in fact promote repair⁴⁰. Supporting this hypothesis, it has been observed that TRF2 is rapidly recruited to sites of non-telomeric double-strand breaks in a manner that is dependent upon an N-terminal domain rich in basic residues⁴¹. While it has been suggested that this defect may arise due to impaired NHEJ, it may in fact be due to impaired HR. Overexpression of TRF inhibits NHEJ and promotes HR *in vivo*⁴¹. Likewise knockdown of TRF2 inhibits HR *in vivo*⁴¹. The data we present herein

suggests that the basic domain of TRF2 facilitates non-telomeric D-loop formation, a necessary initiating step in HR.

TRF1 has also been observed to facilitate telomere replication *in vivo*³⁰ despite previous reports that TRF1 actually inhibits replication of telomeric DNA *in vitro*⁴². It is plausible that *in vitro* TRF1 binding to telomeric DNA may pose an obstacle to DNA polymerase, which is alleviated by one or more factors *in vivo*. Supporting this hypothesis it has been observed that TRF1 is post-translationally modified (PTM) by replication complex associated proteins⁴³⁻⁴⁹. These PTMs reduce TRF1 binding affinity and may relieve the replication inhibition observed in an *in vitro* reconstituted replication reaction. Interestingly these PTMs are inhibited *in vivo* by TIN2⁴³ and a component of the Fanconi Anemia pathway, FANCD2⁴⁶. This careful modulation of TRF1 binding may only transiently remove TRF1 from near the replication fork. The data we present herein suggests that retaining TRF1 near the replication fork may be advantageous. Telomeric DNA is prone to fork stalling and slipping. Fork stalling event can be mitigated or reversed in an HR dependent manner, and we report that TRF1 promoted Rad51-mediated telomeric D-loop formation. This is a critical step in some models of HR mediated replication fork restart.

Scope of Dissertation

Presented here in is an investigation of functional interactions between TRF1, TRF2 and Rad51 *in vitro* (Chapter 2). This characterization will be the main focus of this dissertation, as this project constituted the bulk of my independent work in Dr. Griffith's laboratory. Published data characterizing genome organization and capsid stability of adeno-associated virus (Chapter 3)⁵⁰, and characterization of a guanosine centric mechanism of chaperone-mediated RNA folding (Chapter 4)⁵¹ are also described. Finally, unpublished data from a characterization of the DNA binding properties of hnRNP A1 and UP1 are described in Chapter 5. This projects are presented secondarily, as they constitute collaborative work on which I was not a first author, or work which did not yield data of sufficient quality for publication.

Chapter 2 describes our discovery that TRF1 and TRF2 differentially modulate Rad51 mediated telomeric and non-telomeric D-loop formation. The most significant finding of this chapter is likely our

discovery that TRF2 inhibits Rad51-mediated telomeric D-loop formation. This finding appears to contradict previous nuclear extract work by the Karlseder lab^{27, 52}, but may be supported by findings from Dr. Gilson's lab that TRF2 inhibits HR-mediated processed at humanized yeast telomeres⁵³. Our finding that TRF2ΔB promotes Rad51-mediated non-telomeric D-loop formation may explain how TRF2 can promote HR-mediated non-telomeric double-strand break repair⁴¹. Likewise, our finding that TRF1 promotes Rad51-mediated telomeric D-loop formation may suggest that TRF1 promotes telomeric replication fork restart, explaining how TRF1 may promote telomere replication *in vivo*³⁰ despite findings that TRF1 and TRF2 actually inhibit telomeric DNA replication *in vitro*⁴².

Chapter 3 describes a collaborative project between the Griffith and Asokan laboratories wherein a multidisciplinary approach was used to investigate the relative thermal stability of scAAV and ssAAV when loaded with genomes of varying lengths⁵⁰. I assisted with numerous EM examinations of AAV genome release following thermal denaturation. However, most of this data was not used in the final publication. Portions of this relevant unpublished data will be described in Chapter 3.

Chapter 4 describes a collaboration with the laboratory of Dr Kevin Weeks and his collaborators. We provided purified UP1 protein, which was used as a key control to determine whether the guanosine-centric mechanism of chaperone-mediated RNA folding they identified when using the MuLV NC protein was a conserved feature of RNA chaperones.

Chapter 5 describes EM studies aimed at examining telomeric DNA binding properties of hnRNP A1 and UP1. This project was not pursued due to our inability to find conditions in which hnRNP A1 bound in a conformation or structure that was amenable to EM examination.

REFERENCES

- [1] Mehle, C., Ljungberg, B., and Roos, G. (1994) Telomere shortening in renal cell carcinoma, *Cancer Res* 54, 236-241.
- [2] Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., Saito, M., and Ishikawa, F. (2009) RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway, *Mol Cell* 36, 193-206.
- [3] Chow, T. T., Zhao, Y., Mak, S. S., Shay, J. W., and Wright, W. E. (2012) Early and late steps in telomere overhang processing in normal human cells: the position of the final RNA primer drives telomere shortening, *Genes Dev* 26, 1167-1178.
- [4] Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2, *Nat Genet* 17, 231-235.
- [5] Nora, G. J., Buncher, N. A., and Opresko, P. L. (2010) Telomeric protein TRF2 protects Holliday junctions with telomeric arms from displacement by the Werner syndrome helicase, *Nucleic Acids Res* 38, 3984-3998.
- [6] Poulet, A., Buisson, R., Faivre-Moskalenko, C., Koelblen, M., Amiard, S., Montel, F., Cuesta-Lopez, S., Bornet, O., Guerlesquin, F., Godet, T., Moukhtar, J., Argoul, F., Declais, A. C., Lilley, D. M., Ip, S. C., West, S. C., Gilson, E., and Giraud-Panis, M. J. (2009) TRF2 promotes, remodels and protects telomeric Holliday junctions, *EMBO J* 28, 641-651.
- [7] Khan, S. J., Yanez, G., Seldeen, K., Wang, H., Lindsay, S. M., and Fletcher, T. M. (2007) Interactions of TRF2 with model telomeric ends, *Biochem Biophys Res Commun* 363, 44-50.
- [8] Fouche, N., Cesare, A. J., Willcox, S., Ozgur, S., Compton, S. A., and Griffith, J. D. (2006) The basic domain of TRF2 directs binding to DNA junctions irrespective of the presence of TTAGGG repeats, *J Biol Chem* 281, 37486-37495.
- [9] Li, B., Oestreich, S., and de Lange, T. (2000) Identification of human Rap1: implications for telomere evolution, *Cell* 101, 471-483.
- [10] Bae, N. S., and Baumann, P. (2007) A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends, *Mol Cell* 26, 323-334.
- [11] Ye, J. Z., Donigian, J. R., van Overbeek, M., Loayza, D., Luo, Y., Krutchinsky, A. N., Chait, B. T., and de Lange, T. (2004) TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres, *J Biol Chem* 279, 47264-47271.
- [12] Liu, D., Safari, A., O'Connor, M. S., Chan, D. W., Laegeler, A., Qin, J., and Songyang, Z. (2004) PTPN13 interacts with POT1 and regulates its localization to telomeres, *Nat Cell Biol* 6, 673-680.
- [13] Baumann, P., and Cech, T. R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans, *Science* 292, 1171-1175.
- [14] van den Bosch, M., Bree, R. T., and Lowndes, N. F. (2003) The MRN complex: coordinating and mediating the response to broken chromosomes, *EMBO Rep* 4, 844-849.

- [15] Robison, J. G., Lu, L., Dixon, K., and Bissler, J. J. (2005) DNA lesion-specific co-localization of the Mre11/Rad50/Nbs1 (MRN) complex and replication protein A (RPA) to repair foci, *J Biol Chem* 280, 12927-12934.
- [16] Liu, J., Doty, T., Gibson, B., and Heyer, W. D. (2010) Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA, *Nat Struct Mol Biol* 17, 1260-1262.
- [17] Jensen, R. B., Carreira, A., and Kowalczykowski, S. C. (2010) Purified human BRCA2 stimulates RAD51-mediated recombination, *Nature* 467, 678-683.
- [18] Chi, P., Van Komen, S., Sehorn, M. G., Sigurdsson, S., and Sung, P. (2006) Roles of ATP binding and ATP hydrolysis in human Rad51 recombinase function, *DNA Repair (Amst)* 5, 381-391.
- [19] Wu, L., Davies, S. L., Levitt, N. C., and Hickson, I. D. (2001) Potential role for the BLM helicase in recombinational repair via a conserved interaction with RAD51, *J Biol Chem* 276, 19375-19381.
- [20] Sakamoto, S., Nishikawa, K., Heo, S. J., Goto, M., Furuichi, Y., and Shimamoto, A. (2001) Werner helicase relocates into nuclear foci in response to DNA damaging agents and co-localizes with RPA and Rad51, *Genes Cells* 6, 421-430.
- [21] Wyatt, H. D., Sarbajna, S., Matos, J., and West, S. C. (2013) Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells, *Mol Cell* 52, 234-247.
- [22] Dewar, J. M., and Lydall, D. (2012) Similarities and differences between "uncapped" telomeres and DNA double-strand breaks, *Chromosoma* 121, 117-130.
- [23] van Steensel, B., Smogorzewska, A., and de Lange, T. (1998) TRF2 protects human telomeres from end-to-end fusions, *Cell* 92, 401-413.
- [24] Bertoni, L., Attolini, C., Tessera, L., Mucciolo, E., and Giulotto, E. (1994) Telomeric and nontelomeric (TTAGGG)_n sequences in gene amplification and chromosome stability, *Genomics* 24, 53-62.
- [25] Bombarde, O., Bobby, C., Gomez, D., Frit, P., Giraud-Panis, M. J., Gilson, E., Salles, B., and Calsou, P. (2010) TRF2/RAP1 and DNA-PK mediate a double protection against joining at telomeric ends, *EMBO J* 29, 1573-1584.
- [26] Karlseder, J., Hoke, K., Mirzoeva, O. K., Bakkenist, C., Kastan, M. B., Petrini, J. H., and de Lange, T. (2004) The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response, *PLoS Biol* 2, E240.
- [27] Verdun, R. E., Crabbe, L., Haggblom, C., and Karlseder, J. (2005) Functional human telomeres are recognized as DNA damage in G2 of the cell cycle, *Mol Cell* 20, 551-561.
- [28] Dapic, V., Abdomerovic, V., Marrington, R., Peberdy, J., Rodger, A., Trent, J. O., and Bates, P. J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides, *Nucleic Acids Res* 31, 2097-2107.
- [29] Lormand, J. D., Buncher, N., Murphy, C. T., Kaur, P., Lee, M. Y., Burgers, P., Wang, H., Kunkel, T. A., and Opresko, P. L. (2013) DNA polymerase delta stalls on telomeric lagging strand templates independently from G-quadruplex formation, *Nucleic Acids Res* 41, 10323-10333.

- [30] Sfeir, A., Kosiyatrakul, S. T., Hockemeyer, D., MacRae, S. L., Karlseder, J., Schildkraut, C. L., and de Lange, T. (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication, *Cell* 138, 90-103.
- [31] Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., and de Lange, T. (1999) Mammalian telomeres end in a large duplex loop, *Cell* 97, 503-514.
- [32] Murti, K. G., and Prescott, D. M. (1999) Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops, *Proc Natl Acad Sci U S A* 96, 14436-14439.
- [33] Munoz-Jordan, J. L., Cross, G. A., de Lange, T., and Griffith, J. D. (2001) t-loops at trypanosome telomeres, *EMBO J* 20, 579-588.
- [34] Nikitina, T., and Woodcock, C. L. (2004) Closed chromatin loops at the ends of chromosomes, *J Cell Biol* 166, 161-165.
- [35] Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I., Ding, H., and Boulton, S. J. (2012) RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity, *Cell* 149, 795-806.
- [36] Li, B., Jog, S. P., Reddy, S., and Comai, L. (2008) WRN controls formation of extrachromosomal telomeric circles and is required for TRF2DeltaB-mediated telomere shortening, *Mol Cell Biol* 28, 1892-1904.
- [37] Li, B., Reddy, S., and Comai, L. (2011) Depletion of Ku70/80 reduces the levels of extrachromosomal telomeric circles and inhibits proliferation of ALT cells, *Aging (Albany NY)* 3, 395-406.
- [38] Nosek, J., Rycovska, A., Makhov, A. M., Griffith, J. D., and Tomaska, L. (2005) Amplification of telomeric arrays via rolling-circle mechanism, *J Biol Chem* 280, 10840-10845.
- [39] Tanaka, H., Mendonca, M. S., Bradshaw, P. S., Hoelz, D. J., Malkas, L. H., Meyn, M. S., and Gilley, D. (2005) DNA damage-induced phosphorylation of the human telomere-associated protein TRF2, *Proc Natl Acad Sci U S A* 102, 15539-15544.
- [40] Huda, N., Tanaka, H., Mendonca, M. S., and Gilley, D. (2009) DNA damage-induced phosphorylation of TRF2 is required for the fast pathway of DNA double-strand break repair, *Mol Cell Biol* 29, 3597-3604.
- [41] Mao, Z., Seluanov, A., Jiang, Y., and Gorbunova, V. (2007) TRF2 is required for repair of nontelomeric DNA double-strand breaks by homologous recombination, *Proc Natl Acad Sci U S A* 104, 13068-13073.
- [42] Ohki, R., and Ishikawa, F. (2004) Telomere-bound TRF1 and TRF2 stall the replication fork at telomeric repeats, *Nucleic Acids Res* 32, 1627-1637.
- [43] Smith, S., and de Lange, T. (2000) Tankyrase promotes telomere elongation in human cells, *Curr Biol* 10, 1299-1302.
- [44] Ye, J. Z., and de Lange, T. (2004) TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex, *Nat Genet* 36, 618-623.

- [45] Beneke, S., Cohausz, O., Malanga, M., Boukamp, P., Althaus, F., and Burkle, A. (2008) Rapid regulation of telomere length is mediated by poly(ADP-ribose) polymerase-1, *Nucleic Acids Res* 36, 6309-6317.
- [46] Lyakhovich, A., Ramirez, M. J., Castellanos, A., Castella, M., Simons, A. M., Parvin, J. D., and Surralles, J. (2011) Fanconi anemia protein FANCD2 inhibits TRF1 polyADP-ribosylation through tankyrase1-dependent manner, *Genome Integr* 2, 4.
- [47] Dantzer, F., Giraud-Panis, M. J., Jaco, I., Ame, J. C., Schultz, I., Blasco, M., Koering, C. E., Gilson, E., Menissier-de Murcia, J., de Murcia, G., and Schreiber, V. (2004) Functional interaction between poly(ADP-Ribose) polymerase 2 (PARP-2) and TRF2: PARP activity negatively regulates TRF2, *Mol Cell Biol* 24, 1595-1607.
- [48] Simbulan-Rosenthal, C. M., Rosenthal, D. S., Boulares, A. H., Hickey, R. J., Malkas, L. H., Coll, J. M., and Smulson, M. E. (1998) Regulation of the expression or recruitment of components of the DNA synthesome by poly(ADP-ribose) polymerase, *Biochemistry* 37, 9363-9370.
- [49] Walker, J. R., and Zhu, X. D. (2012) Post-translational modifications of TRF1 and TRF2 and their roles in telomere maintenance, *Mech Ageing Dev* 133, 421-434.
- [50] Horowitz, E. D., Rahman, K. S., Bower, B. D., Dismuke, D. J., Falvo, M. R., Griffith, J. D., Harvey, S. C., and Asokan, A. (2013) Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release, *J Virol* 87, 2994-3002.
- [51] Grohman, J. K., Gorelick, R. J., Lickwar, C. R., Lieb, J. D., Bower, B. D., Znosko, B. M., and Weeks, K. M. (2013) A guanosine-centric mechanism for RNA chaperone function, *Science* 340, 190-195.
- [52] Verdun, R. E., and Karlseder, J. (2006) The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres, *Cell* 127, 709-720.
- [53] Saint-Leger, A., Koelblen, M., Civitelli, L., Bah, A., Djerbi, N., Giraud-Panis, M. J., Londono-Vallejo, A., Ascenzioni, F., and Gilson, E. (2014) The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres, *Cell Cycle* 13.
- [54] Kruger, A. C., Raarup, M. K., Nielsen, M. M., Kristensen, M., Besenbacher, F., Kjems, J., and Birkedal, V. (2010) Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level, *Eur Biophys J* 39, 1343-1350.
- [55] Flynn, R. L., Centore, R. C., O'Sullivan, R. J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J., and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA, *Nature* 471, 532-536.
- [56] Saint-Leger, A., Koelblen, M., Civitelli, L., Bah, A., Djerbi, N., Giraud-Panis, M. J., Londono-Vallejo, A., Ascenzioni, F., and Gilson, E. (2014) The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres, *Cell Cycle* 13.
- [57] Arat, N. O., and Griffith, J. D. (2012) Human Rap1 interacts directly with telomeric DNA and regulates TRF2 localization at the telomere, *J Biol Chem* 287, 41583-41594.

CHAPTER 2: TRF1 and TRF2 Differentially Modulate Rad51-Mediated Telomeric and Non-Telomeric Displacement Loop Formation *In Vitro*¹

INTRODUCTION

Mammalian telomeres consist of 5-15 kilobase pairs (kbp) of TTAGGG repeats that terminate in a 50-200 nucleotide (nt) single-stranded DNA (ssDNA) 3' tail. The telomere repeats and the single-stranded-to-double-stranded DNA (ss/dsDNA) junction provide a binding site for telomere-specific proteins that shelter telomeres from being recognized as DNA damage. While these shelterin proteins may directly inhibit DNA damage signaling,^{1, 2} the presence of a DNA loop at the end of the telomeres (t-loop) may also mediate telomere protection. One shelterin component, telomere repeat-binding factor 2 (TRF2), is required for t-loop formation *in vivo*,³ and can promote t-loop formation *in vitro*⁴ by facilitating a strand invasion reaction between the ssDNA tail and upstream dsDNA in a telomere. However, telomere protection also requires components of the homologous recombination/repair (HR) pathway, which may facilitate telomere replication or promote t-loop formation.

In vitro, telomeric replication forks are prone to slipping,⁵ and replication of telomeric DNA is inefficient⁶ and prone to defects consistent with fork stalling.⁷ *In vivo* fork stalling can be mitigated by proteins involved in the HR pathway.⁸ Accordingly, replication of telomeric DNA *in vivo* is sensitive to disruption of that pathway. The BRCA2 tumor suppressor recruits the Rad51-recombinase to telomeres during replication, and disrupting the expression of either of these proteins results in telomere shortening and fragility. These phenotypes are attenuated in cells possessing short telomeres and are exacerbated by chemical inhibition of DNA replication.⁹ As such it's likely that these defects are due in part to a telomere replication defect.

¹The following chapter describes work done in collaboration with Dr. Griffith. I materially contributed to all work described in this chapter⁴².

Disrupting the HR pathway in non-dividing cells results in aberrant telomere repair. Therefore, it is likely that the HR pathway also contributes to telomere protection in a replication-independent manner,⁹ possibly by promoting t-loop formation. Concordantly, both TRF2 and Rad51 are required for cell extracts to promote telomeric D-loop formation;¹⁰ a requisite step in t-loop formation. Interestingly, this relationship appears to be bi-directional. Overexpression of TRF2 promotes, while TRF2 knockdown inhibits HR *in vivo*.¹¹ While these observations suggest that TRF2 and HR cooperate functionally *in vivo*, this hypothesis contradicts these proteins' established *in vitro* activities. TRF2 induces positive supercoiling within telomeric dsDNA upon binding,¹² but Rad51 most efficiently promotes D-loop formation when acting upon negatively supercoiled dsDNA templates.¹³

To investigate functional interactions between shelterin proteins and the HR pathway, we undertook an *in vitro* characterization of the combined activities of purified proteins from these pathways. While the use of purified proteins permits an examination of their isolated functional interactions *in vitro*, such interactions may be affected by other proteins *in vivo*. The absence of such other proteins likely explains why the results of our assay contradict previous cell-extract based characterizations.¹⁰ We report that TRF2 inhibits the ability of Rad51 to promote telomeric D-loop formation, suggesting that Rad51 does not promote t-loop formation and elucidating a novel mechanism by which TRF2 inhibits aberrant DNA repair at the telomeres. In contrast, we report that TRF1 promotes Rad51-mediated telomeric D-loop formation, possibly explaining why TRF1 is required for efficient telomere replication. Finally, we report that a TRF2 mutant lacking the dsDNA binding domain was able to promote Rad51-mediated D-loop formation, suggesting that one or more TRF2 domains can positively modulate Rad51 activity and possibly explaining how TRF2 can facilitate HR.

MATERIALS AND METHODS

DNA Substrates, Templates and Competitors

A pBluescript derived plasmid containing a 103 bp telomeric tract (pBB: TTAGGG)₁₇T) was generated by conventional cloning via insertion of the BsmBI/BbsI fragment of pRST15⁴ into BsmBI cut

pRST15. A pBluescript derived plasmid containing a non-telomeric insert (pGL GAP) was generated as previously described.¹⁴ All plasmids were cultured in DH10B *E. coli* and purified using Qiagen Maxiprep kits. HPLC purified 5' Cy3 labeled G-rich telomeric 90 mer oligonucleotide (T90:[Cy3] (GGTTAG)₁₅), D1 oligonucleotide ([Cy3]AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA GTGAGGCACCTATCTCAGCGATCTGTCTATTT) and T3 promoter primer ([Cy3]ATTAACCCTCACTAAAGGA) and HPSF purified unlabeled T7 promoter primer (TAATACGACTCACTATAGGG) were ordered from Eurofins MWG Operon. A 255 bp Cy3-labeled PCR product was amplified from pBB using the 5' Cy3-labeled T3 and unlabeled T7 promoter primers and Q5 High Fidelity Polymerase (New England Biolabs) as per the manufacturer's instructions and purified using a DNA Clean & Concentrator-25 column (Zymo Research).

Proteins

Untagged Rad51 was expressed and purified as previous described¹⁵ from a pET-24 derived plasmid (EMD Millipore), which was generously provided by Dr. Richard Fishel (Ohio State University, USA). N-terminally hexahistidine tagged TRF2, TRF2ΔB, TRF2ΔM and TRF1 were purified from pTRC-HIS derived plasmids (Invitrogen) adapted from vectors generously provided by the laboratory of Dr. Eric Gilson¹² (University of Nice, France) or modified from vectors previously described.¹⁶ All TRF2 cDNAs were modified to include the Ala434 codon that is absent in HeLa derived TRF2 clones.¹⁷ Briefly, a pTRC-HIS plasmid was transformed into BL21(DE3)PlysS *E. coli* and serially passaged to inoculate 1 L of Terrific Broth (Sigma-Aldrich) containing 50 µg/ml ampicillin. The culture was grown to an OD of 0.6 at 595 nm, and protein expression was then induced via addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (Promega) for 4 h at 37°C. The cells were then recovered via centrifugation, washed with phosphate buffered saline, resuspended in 100 ml of buffer containing 20 mM HEPES at pH 7.5, 300 mM NaCl, 10% Glycerol, 0.5 mM DTT and 50 mM Imidazole supplemented with protease inhibitors (Roche) and then flash frozen in liquid nitrogen. The cells were then thawed and lysed via

sonication following addition of 1 mg/ml egg white lysozyme and 20 μ L of RQ1 DNase (Promega) and 20 μ L of RNase A (Sigma). The crude lysate was then centrifuged in an SW-41 Ti rotor at 41,000 RPM for 1.5 h. The supernatant was collected and serially purified over 1 ml HisTrap HP, HiTrap Heparin HP and HiTrap Q FF columns using an ÄKTApurifier FPLC (GE Bioscience). Rad51, TRF2, TRF2 Δ B and TRF2 Δ M protein were recovered in 20 mM HEPES at pH 7.5, 150 mM NaCl, 10% Glycerol and 0.5 mM DTT, while TRF1 was recovered in 20 mM HEPES at pH 7.5, 300 mM NaCl, 10% Glycerol, 0.5 mM DTT. These proteins were aliquoted, flash frozen with liquid nitrogen and stored at -80°C until use. Protein concentration was determined using a Biorad Protein Assay calibrated against a Bovine Gamma Globulin standard set (Biorad). For all proteins homogeneity was assessed as >90% by Coomassie staining of SDS-PAGE gels. Immediately prior to use in experiments TRF2, TRF2 Δ B, TRF2 Δ M and TRF1 were diluted to a final concentration of 4.25 μ M in buffer containing 19 mM HEPES-KOH, 203.8 mM NaCl, 1 mM CaCl, 1 mM ATP, 7% glycerol, and 0.7 mM DTT. All protein concentrations are reported as monomeric protein. Rad51 was purified to a concentration of 27.5 μ M and was used undiluted in all experiments. Fraction V Bovine Serum Albumin (Fisher) was diluted to 10 mg/ml in 20 mM potassium phosphate at pH 7.0, 50 mM NaCl, 5% Glycerol and 0.1 mM EDTA.

Displacement Loop Assay

For the displacement loop assay 2.4 μ M in nucleotides (nt) of the 5' Cy3 labeled telomeric 90 mer (26.67 nM Oligo) was incubated with no protein or 1,000-1,500 nM Rad51 at 37 °C for 10 min in a reaction buffer containing 5 mM HEPES-KOH at pH 7.5, 1 mM CaCl, 1 mM ATP, 0.8 mM DTT and 100 μ g/ml BSA. Simultaneously, the pBB plasmid, 35 μ M in base pairs (bp) or 10 nM plasmid was incubated with no protein or 100-500 nM TRF2, TRF2 Δ B, TRF2 Δ M or TRF1 at 37 °C for 10 min in reaction buffer and 100 μ g/ml BSA. Equal volumes of these reactions were then combined to give final Rad51 concentrations of 0 or 500-750 nM and a final concentration of 0-250 nM TRF2, TRF2 Δ B, TRF2 Δ M or TRF1. These reactions were incubated at 37 °C for 1 h then deproteinized via addition of 0.5% sodium dodecyl sulfate and 1 mg/ml Proteinase K (Ambion) and incubation at 37 °C for 15 min. Glycerol

loading buffer (5% Glycerol, 1.67 mM Tris, 0.17 mM EDTA, 0.017% SDS) was then added to 1X and the samples were separated for 30 min in a small-format 1% 1/2X TBE agarose gel at 100 V (6.67 V/cm) in a light-protected box in a 4 °C cold room. All figures are labeled with the final respective protein concentrations.

Electrophoretic Mobility Shift and Binding Competition Assay

To demonstrate binding via an electrophoretic mobility shift assay, 2.55 μ M in bp of the Cy3-labeled PCR product (10 nM product) was incubated with no protein or 100-500 nM TRF2, TRF2 Δ B, TRF2 Δ M or TRF1 at 37 °C for 10 min in reaction buffer supplemented with 100 μ g/ml BSA. To demonstrate binding specificity via a competition assay, an additional set of 500 nM reactions were performed in buffer containing no competitor or between a 1:1 (2.55 μ M bp) and 200:1 (510 μ M bp) excess of pGL GAP and then incubated at 37 °C for 25 min. To demonstrate that the induced supershifts were protein-mediated a 500 nM reaction containing no competitor was incubated for 10 min then deproteinized with SDS and proteinase K for 15 min. Glycerol loading buffer containing no SDS was then added to 1X and the samples were separated for 30 min in a small-format 1/2X TBE agarose gel at 100 V (6.67 V/cm) in a light-protected box in a 4 °C cold room. All figures are labeled with the final respective protein concentrations.

Imaging

All Cy3 labeled gel products were imaged using a Biorad Typhoon Scanner equipped with a 532 nM green laser module and a 580 nM bypass filter. Gels were imaged with a photomultiplier setting of 600 and a pixel size of 100 microns. All gels were imaged with a +3 mm focal plane setting. Gel image intensity was then adjusted using ImageQuant software (GE Life Sciences) and quantified using ImageJ software (NIH).

RESULTS

A Fluorescent TRF2 and Rad51-Mediated Displacement Loop Assay

To investigate functional interactions between Rad51 and TRF2 we developed a fluorescent displacement loop (D-loop) assay (Fig. 2.1A) adapted from previous TRF2 and Rad51 characterizations.^{12, 15} Untagged Rad51, and N-terminally hexahistidine tagged TRF1, TRF2 and TRF2 mutant proteins lacking either the N-terminal basic domain of TRF2 (TRF2 Δ B) or the C-terminal Myb domain of TRF2 (TRF2 Δ M) were purified from *E. coli* to >90% homogeneity (Supporting Information Fig. 2.S1). In this assay co-incubation of a Cy3-labeled telomeric ssDNA substrate (T90) with a dsDNA telomeric plasmid template (pBB) in the absence of any proteins resulted in low-to-undetectable levels (<0.5%) of spontaneous D-loop formation (Fig. 2.1B, C: Lane 1). In contrast, pre-incubation of the substrate with purified Rad51 protein prior to its addition to the template promoted D-loop formation in a Rad51-concentration dependent manner (Fig. 2.2A, B). Likewise, pre-incubation of the template with full length TRF2 protein prior to its addition to the substrate could promote D-loop formation across a discreet range of TRF2 concentrations (Fig. 2.1B). TRF2 Δ B exhibited only 47% of the activity of full-length TRF2 (Table 2.1), but this residual activity was similarly optimal across a narrow range of concentrations (Fig. 2.1B). In contrast, TRF2 Δ M and TRF1 respectively exhibited only 31% and 27% of the activity of full length TRF2 (Table 2.1), and were maximally active only at higher concentrations (Fig. 2.1C).

Rad51-mediated D-loop formation was observed to be homology driven. Rad51 could promote D-loop formation between telomeric substrates and templates, and non-telomeric substrates and templates, but not between a telomeric substrate and a non-telomeric template (Supporting Information Fig. 2.S2A, B). Under identical conditions, Rad51-mediated telomeric D-loop formation was 6.4 fold higher (Table 2.1) than non-telomeric D-loop formation (Supporting Information Fig. 2.S2). This is consistent with previous characterizations showing that the activity of Rad51 is enhanced on repetitive and GT rich substrates.^{18, 19} In contrast to Rad51, TRF2-mediated D-loop formation was observed to be

critically dependent upon telomeric homology. TRF2 could promote D-loop formation only between telomeric substrates and templates (Supporting Information Fig. 2.S2C, D).

TRF2 Inhibits Rad51-Mediated Telomeric But Not Non-Telomeric D-Loop Formation.

To test for functional interactions between TRF2 and Rad51, D-loop assay reactions were prepared in which the template was pre-incubated with either a fixed concentration of TRF2 or no protein, while the substrate was pre-incubated with one of several concentrations of Rad51 or no protein prior to the combination of the substrate and template reactions. Pre-incubation of a telomeric template with TRF2 reduced the ability of Rad51 to promote D-loop formation between the template and a homologous telomeric substrate by $52 \pm 5.1\%$ (Table 2.1; Fig. 2.2A, B). In contrast, TRF2 did not significantly inhibit Rad51-mediated non-telomeric D-loop formation (Table 2.1; Fig. 2.2C, D). Taken together, these data suggested that TRF2 differentially modulates Rad51-mediated telomeric and non-telomeric D-loop formation.

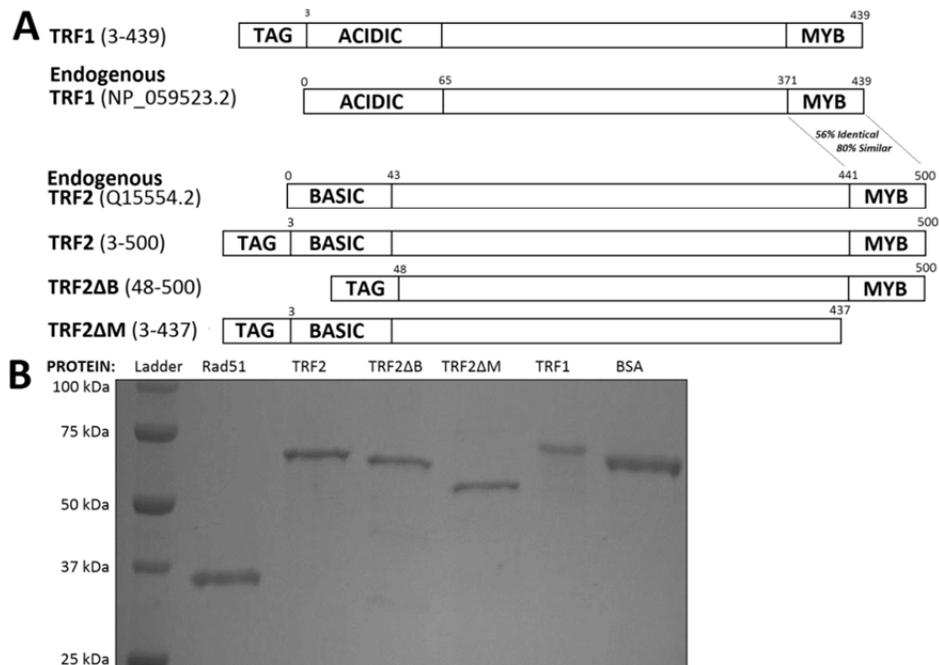


Figure 2.S1 Proteins Used. **(A)** Diagram and partial domain map of hexahistidine tagged TRF1, TRF2, TRF2ΔB and TRF2ΔM compared against endogenous TRF2 (Q15554.2) and endogenous TRF1 (NP_059523.2), partially adapted from⁴⁰. **(B)** Coomassie stained SDS-PAGE gel containing 2 μg of each of purified Rad51, TRF2, TRF2ΔB, TRF2ΔM, TRF1, and BSA.

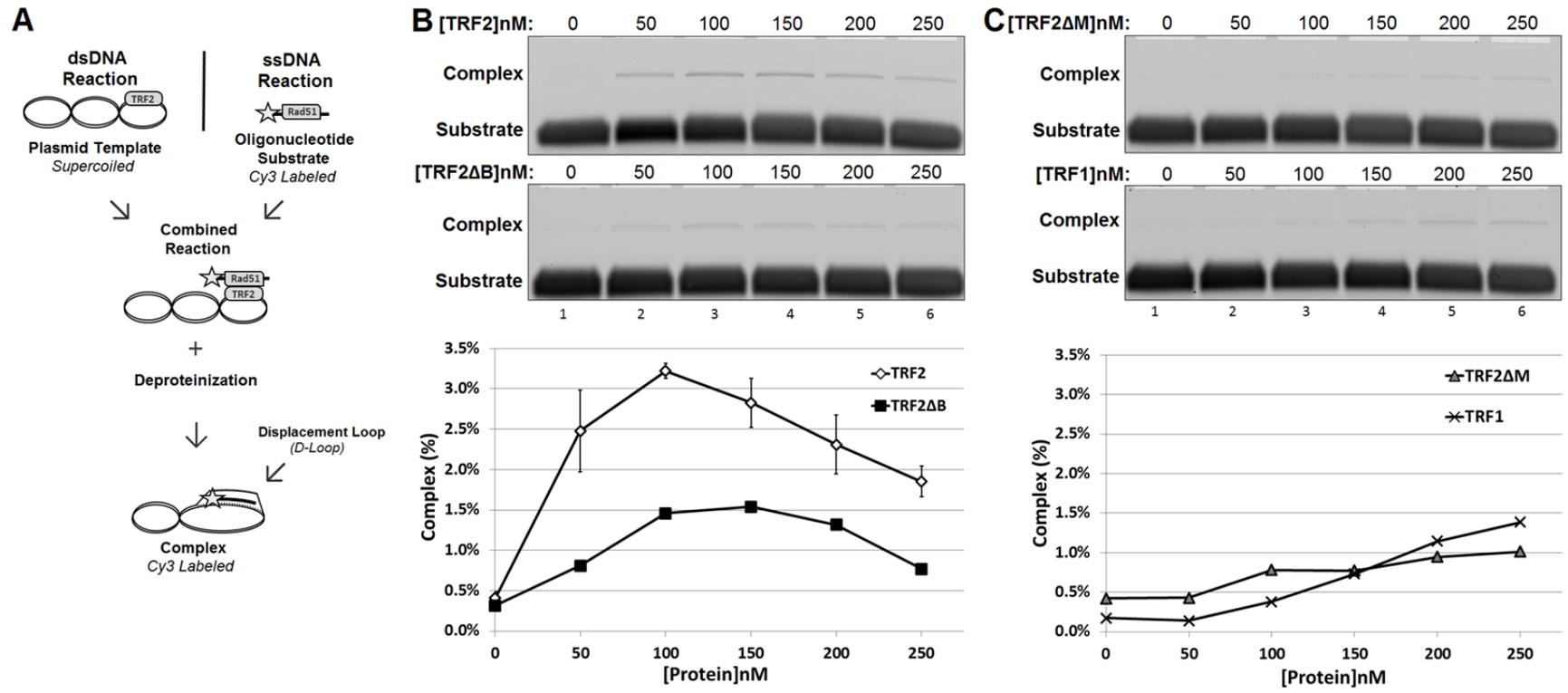


Figure 2.1 TRF-mediated telomeric D-loop formation. **(A)** Diagram of D-loop assay. **(B)** TRF2 and TRF2ΔB promote telomeric D-loop formation with an activity peak when included at a final concentration of between 100 nM (lane 3) and 150 nM (Lane 4) of protein. **(C)** TRF2ΔM and TRF1 promote telomeric D-loop formation only at higher concentrations.

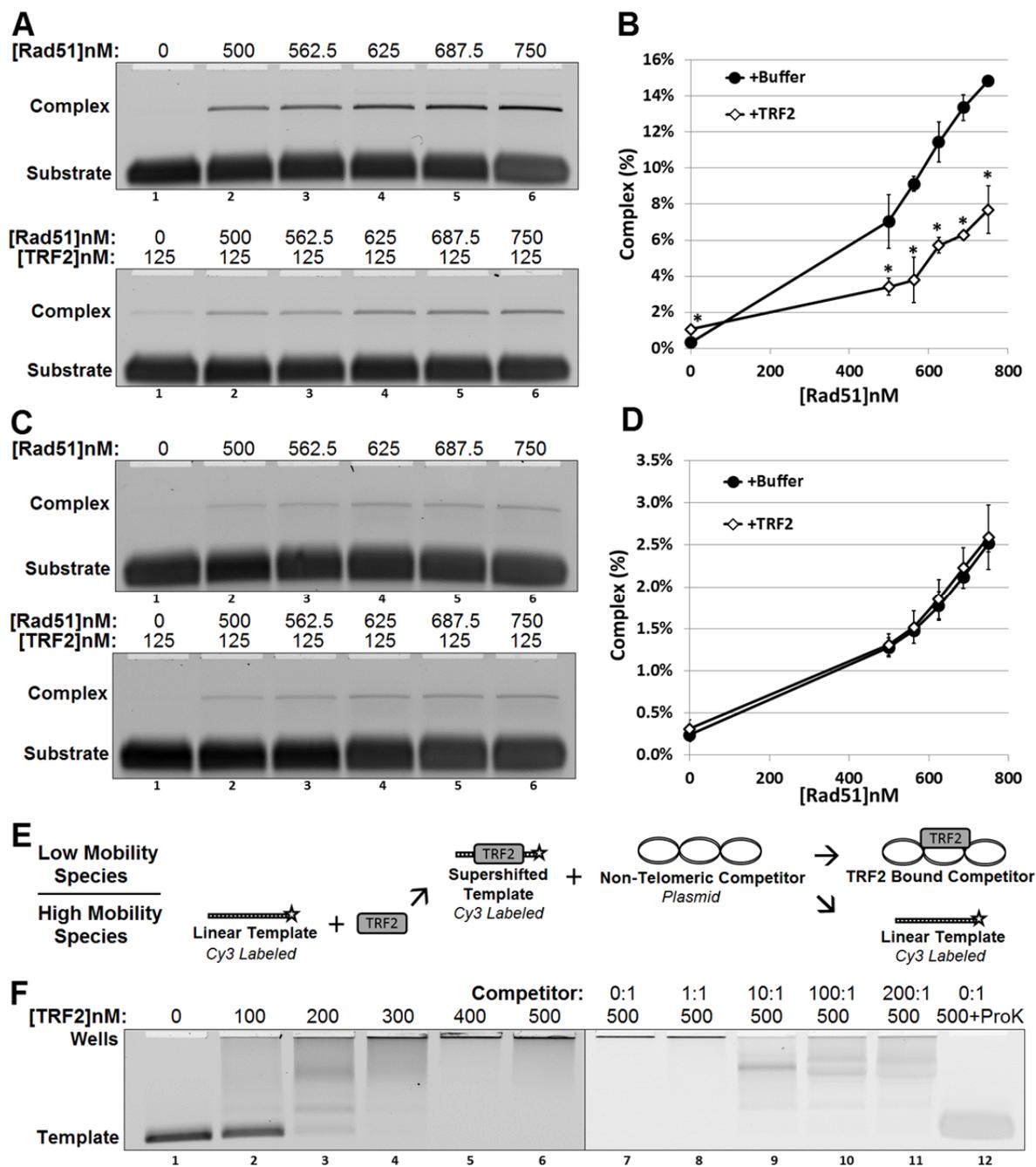


Figure 2.2 TRF2 inhibits Rad51-mediated telomeric but not non-telomeric D-loop formation. **(A)** Rad51 promotes telomeric D-loop formation in a concentration dependent manner that is inhibited by TRF2. **(B)** Quantification of data in (A). **(C)** Rad51 promotes non-telomeric D-loop formation in a concentration dependent manner that is not affected by TRF2. **(D)** Quantification of data in (C). **(E)** Diagram of DNA binding and competition assay. **(F)** TRF2 binding supershifts the template into the wells. This binding is specific and persists in the presence of high concentrations of non-telomeric competitor and is protein-mediated. Error bars shown 95% confidence interval, significant difference between +Buffer and +TRF2 (*), paired samples t-test $\alpha=0.05$ from three independent experiments.

Table 2.1 Properties of TRF2, TRF2ΔB, TRF2ΔM and TRF1

Proteins	TRF-Induced Telomeric D-Loop Formation	Rad51-Mediated Telomeric D-loop Formation	Rad51-Mediated Non-Telomeric D-Loop Formation	Telomeric DNA Binding	Telomeric Binding Specificity	Migration in Agarose Gels
	AUC 0-250 nM % of TRF2	AUC 500-750 nM % Change From Buffer	AUC 500-750 nM % Change From Buffer	C _{1/2} , nM		
Buffer [†]	N/A	30.3 ± 0.4 0 ± 0.7%	4.7 ± 0.4 0 ± 7.6%	N/A	+	-
TRF2	6.0 ± 0.15 100 ± 2.4%	13.3 ± 0.8 * - 52 ± 5.1% *	4.7 ± 0.0 + 4 ± 2.2%	111 ± 8	+	-
TRF2ΔB	2.8 47%	22.1 ± 0.7 * - 31 ± 5.5% *	5.1 ± 0.7 + 5 ± 7.2%	257	+	-
TRF2ΔM	1.8 31%	29.0 ± 1.4 + 4 ± 4.0%	9.6 ± 0.4 * + 112 ± 13.0% *	319	-	+ / -
TRF1	1.6 27%	38.5 ± 1.9 * + 25 ± 1.0% *	5.4 ± 0.9 + 9 ± 5.2%	152	+	+

[†]Buffer data are averaged. Proteins are statistically compared against matched buffer controls.

TRF-induced D-loop formation calculated as area under the curve (AUC: Complex% × [TRF] nM) from Figure 1.

Rad51-mediated D-loop formation calculated as AUC (Complex% × [Rad51] nM) from Figures 2-5.

AUC calculation examples in Figure 2.S5.

C_{1/2} represents the concentration of TRF protein required to supershift 50% of template in EMSAs.

Errors shown are 95% confidence intervals from three independent experiments.

* denotes significant difference (p < 0.05) from buffer via two-tailed paired samples t-test.

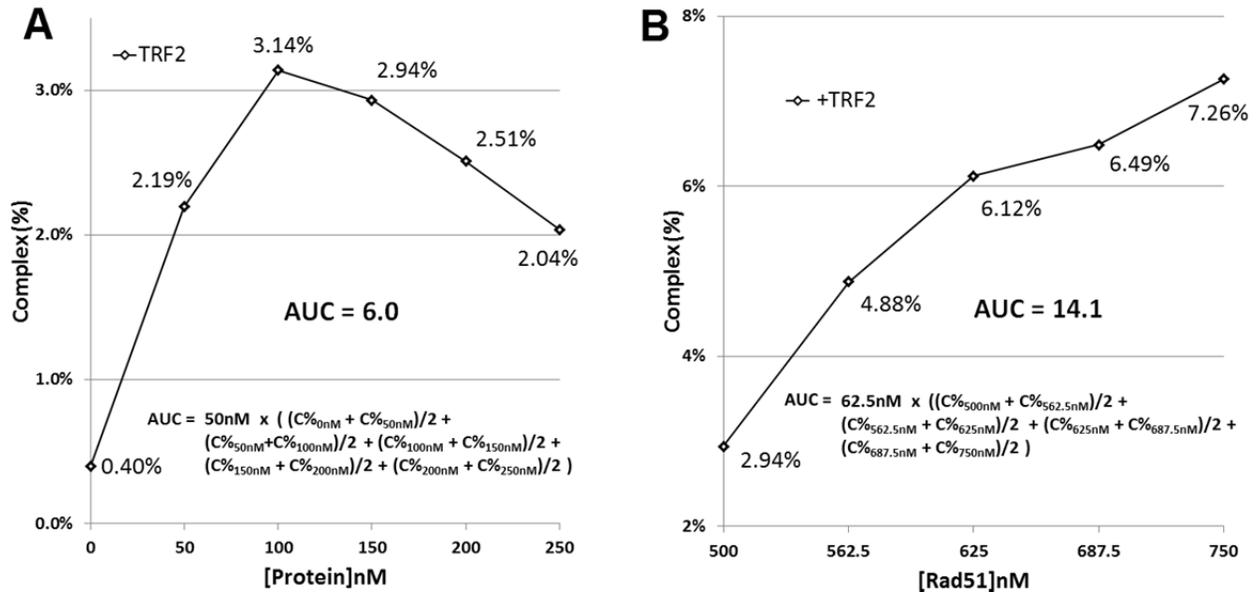


Figure 2.S5 Area under the curve (AUC) calculation procedures (C% = Complex %). **(A)** A representative activity trace from data shown in Figure 1C. AUC is calculated as (Complex% × [TRF2] nM) for all regions of the activity trace. **(B)** A representative activity trace from data shown in Figure 2A. AUC is calculated as (Complex% × [Rad51] nM) for 500-750 nM [Rad51].

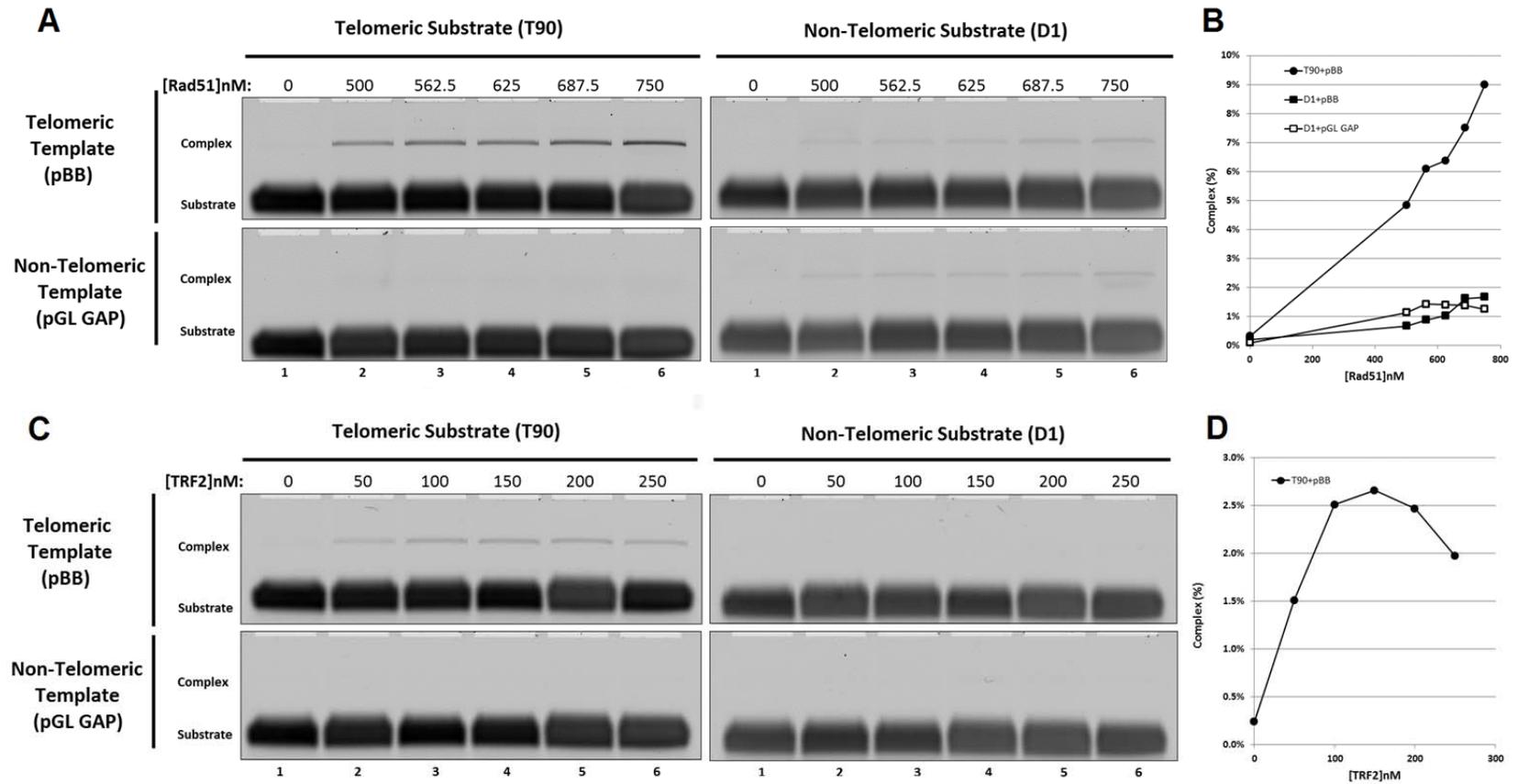


Figure 2.S2 TRF2-mediated D-loop formation requires telomeric homology while Rad51 does not. **(A)** Rad51 promotes telomeric D-loop formation between pBB and T90 (top left), but not between pGL GAP and T90. Rad51 can promote non-telomeric D-loop formation between D1 and both pGL GAP and pBB. **(B)** Quantification of (A). **(C)** TRF2 promotes D-loop formation only between T90 and pBB. **(C)** Quantification of (A). **(D)** Quantification of (C).

Rad51-mediated D-loop formation is a multi-step process initiated by Rad51 binding to ssDNA to form a nucleoprotein filament, which subsequently interrogates dsDNA for matching antisense sequence in a process known as ‘homology search’. In this process a Rad51-coated substrate initially forms a protein-mediated complex with a homologous template. Subsequently Rad51 promotes D-loop formation between the substrate and template.²⁰ To determine what step or steps of this process might be inhibited by TRF2 we performed several order of addition experiments.

We observed that the ability of TRF2 to inhibit Rad51-mediated telomeric D-loop formation was dependent upon addition of TRF2 early in the D-loop reaction (Fig. 2.S3). TRF2 could inhibit Rad51-mediated telomeric D-loop formation when pre-incubated with the telomeric template (T0) or when added to a combined reaction prior to D-loop formation (T0+10min). However, TRF2 could not inhibit Rad51-mediated D-loop formation if added after D-loop formation had already occurred (T0+3hrs). These observations suggested that TRF2 modulates Rad51-mediated D-loop formation via a passive mechanism, possibly by interfering with Rad51 filament formation, inhibiting homology search or by preventing subsequent D-loop formation.

We also observed that TRF2 could inhibit Rad51-mediated telomeric D-loop formation regardless of whether TRF2 was pre-incubated with the telomeric template or with the Rad51-coated substrate (Figure 2.S4). However, the degree of this inhibition was reduced when TRF2 was incubated with the substrate compared with when it was incubated with the template. This suggests that TRF2 does not inhibit Rad51 at the level of filament formation. Instead, it appears that the ability of TRF2 to inhibit Rad51-mediated telomeric D-loop formation is dependent upon the ability of TRF2 to bind to or modify the template.

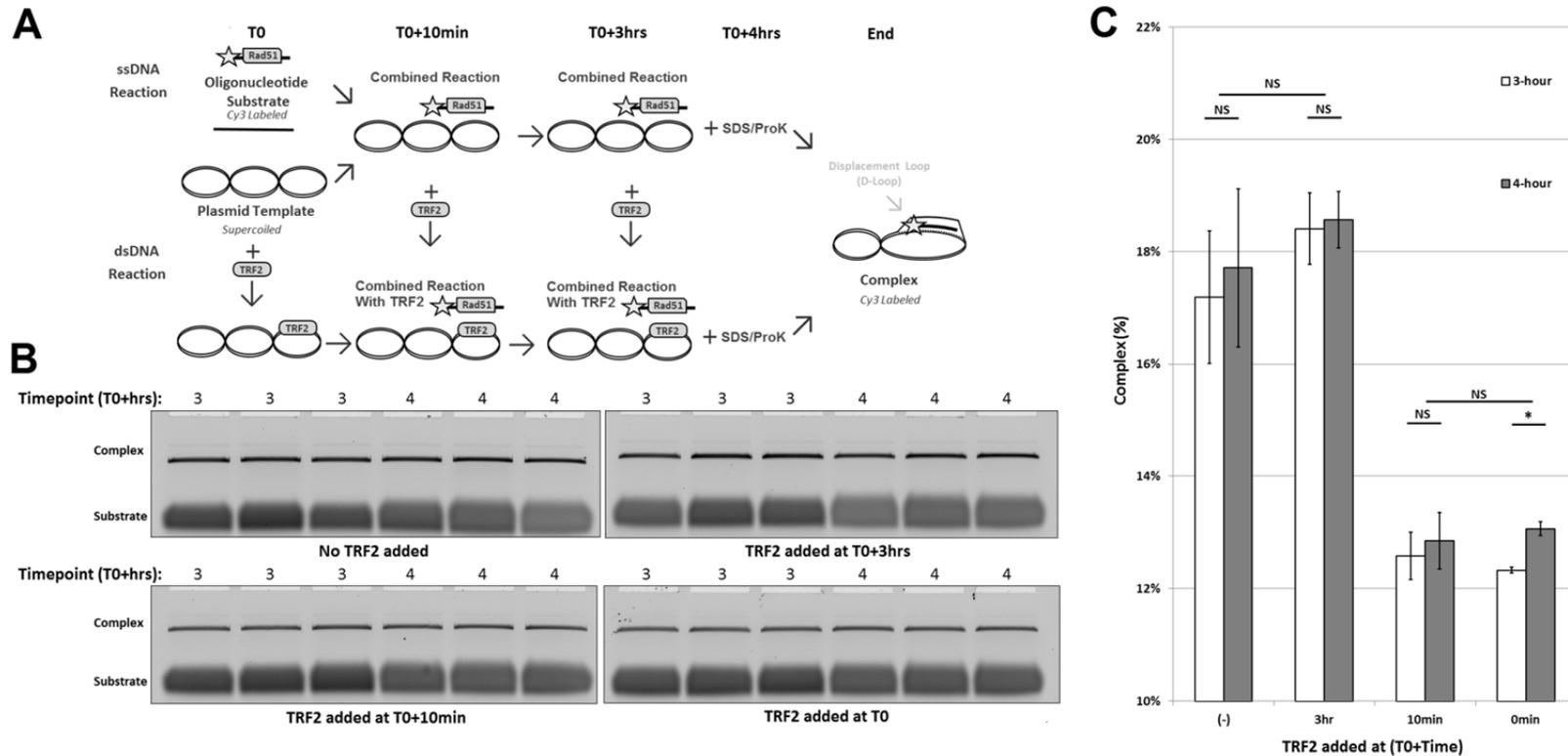


Figure 2.S3 TRF2 inhibits Rad51-mediated telomeric D-loop formation only when added early in D-loop reactions. **(A)** Diagram of order of addition experiment. **(B)** TRF2 inhibits Rad51-mediated telomeric D-loop formation when added early in the reactions (T0 and T0+10min vs no TRF2) but not when added late in the reaction (T0+3hrs vs no TRF2). **(C)** Quantification of (B), error bars shown 95% confidence intervals from three independent experiments. (*) significant difference between indicated samples, paired samples t-test $\alpha=0.05$. (NS) No significant difference.

To investigate whether the DNA binding activities of TRF2 mediate its ability to inhibit Rad51-mediated telomeric D-loop formation, we characterized the binding affinity and specificity of TRF2 using an electrophoretic mobility shift assay (EMSA) and a binding competition assay (Fig. 2.2E). Incubating a Cy3-labeled template containing a 103 base pair bp telomere tract with increasing concentrations of TRF2 resulted in a supershift of the template, consistent with stable TRF2 binding (Fig. 2.2F). The binding to TRF2 to the template was observed to be specific, and persisted even in the presence of high concentrations of non-telomeric competitor (Fig. 2.2F: lanes 8-11). Nearly all low-mobility species generated by TRF2 binding became trapped in the wells. This supershift was protein mediated, and could be disrupted by incubation with SDS and proteinase K (Fig. 2.2F: lane 12).

To further investigate possible mechanism by which TRF2 may inhibit Rad51-mediated telomeric D-loop formation we characterized the binding activity and the telomeric and non-telomeric Rad51-modulating activities of a variety of TRF2 mutant proteins and TRF1, a close homolog of TRF2.

TRF2 Δ M Promotes Rad51-Mediated Telomeric But Not Non-Telomeric D-Loop Formation.

The dsDNA binding activity of TRF2 is primarily directed by its C-terminal Myb domain. Deletion of this Myb domain reduces telomeric dsDNA binding affinity by a factor of 2.9 (Table 2.1), eliminates telomeric binding specificity and grossly alters DNA binding properties compared to full-length TRF2 (Fig. 2.3E).¹² Interestingly and in contrast to TRF2, TRF2 Δ M does not inhibit Rad51-mediated telomeric D-loop formation (Table 2.1; Fig. 2.3A, B). Also in contrast to TRF2, TRF2 Δ M was found to promote Rad51-mediated non-telomeric D-loop formation by 112 \pm 13.0% (Table 2.1; Fig. 2.3C, D). Taken together these observations suggest that the Myb domain of TRF2 both contributes to the ability of TRF2 to inhibit Rad51-mediated telomeric D-loop formation and suppresses the ability of TRF2 to promote Rad51-mediated non-telomeric D-loop formation.

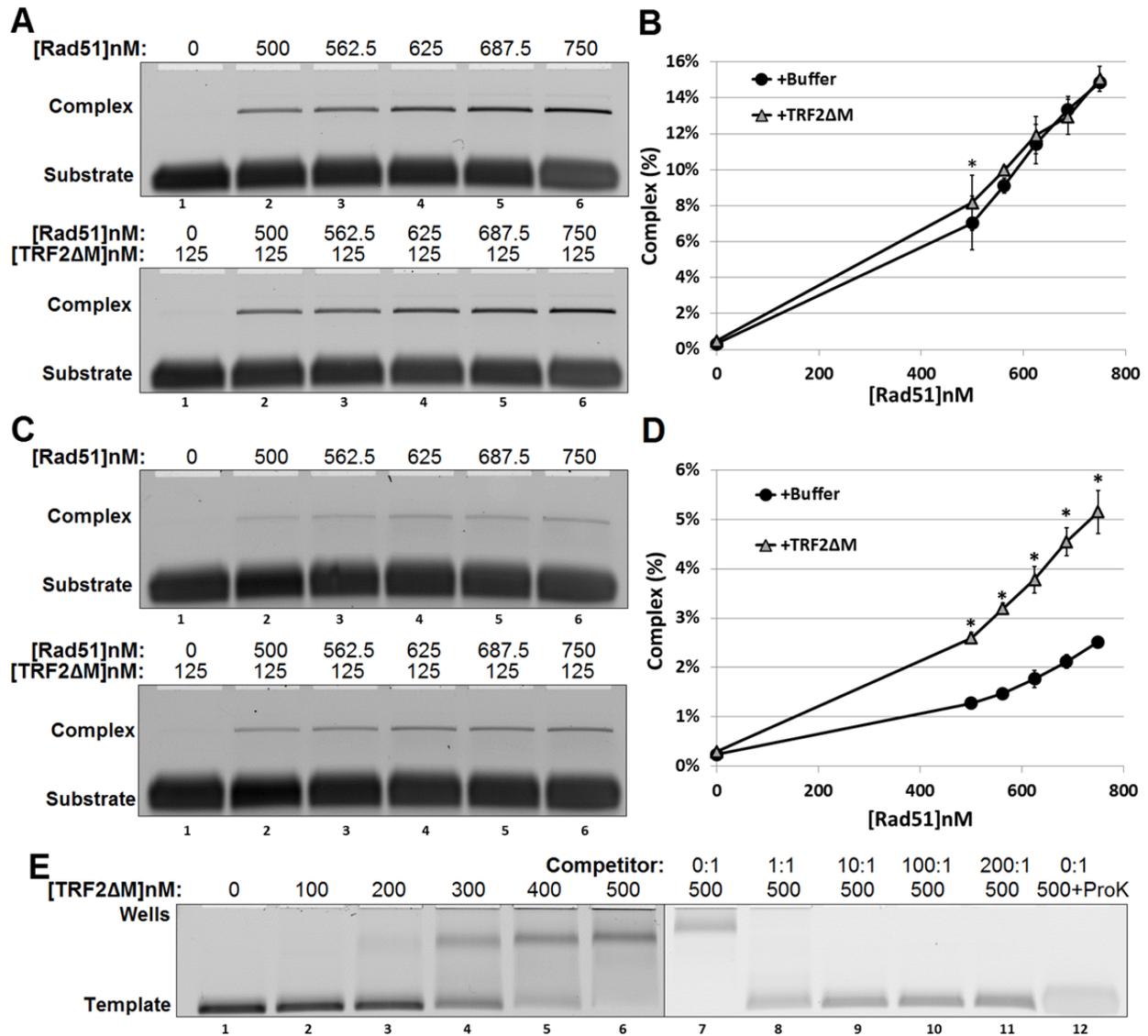


Figure 2.3 TRF2ΔM promotes Rad51-mediated non-telomeric but not telomeric D-loop formation. **(A)** Rad51 promotes telomeric D-loop formation in a concentration dependent manner that is not affected by TRF2ΔM. **(B)** Quantification of data in (A). **(C)** Rad51 promotes non-telomeric D-loop formation in a concentration dependent manner that is promoted by TRF2ΔM. **(D)** Quantification of data in (C). **(E)** TRF2ΔM binding supershifts the template into a lower mobility species and into the wells. This binding is non-specific and is disrupted by low concentrations of non-telomeric competitor and is protein-mediated. Error bars shown 95% confidence interval, significant difference between +Buffer and +TRF2ΔM (*), paired samples t-test $\alpha=0.05$ from three independent experiments.

TRF2 Δ B Inhibits Telomeric But Not Non-Telomeric Rad51-Mediated D-Loop Formation.

In addition to its Myb domain, TRF2 possesses an N-terminal domain rich in basic residues that has been implicated in directing the binding of TRF2 to ss/dsDNA junctions and unusual DNA structures.^{4, 21} This domain also promotes the annealing and migration of DNA joints in a manner not unlike that required during D-loop formation.²² To investigate whether the basic domain of TRF2 contributes to the ability of TRF2 to inhibit Rad51-mediated telomeric D-loop formation or the ability of TRF2 Δ M to promote Rad51-mediated non-telomeric D-loop formation we characterized the DNA binding affinity and specificity and Rad51-modulating activity of a TRF2 mutant protein lacking the basic domain of TRF2 (TRF2 Δ B).

Like TRF2 and in contrast to TRF2 Δ M, TRF2 Δ B was found to inhibit Rad51-mediated telomeric D-loop formation by 31 \pm 5.5% (Table 1; Fig. 2.4A, B), suggesting that the joint-binding activity of TRF2 is not required for TRF2 to inhibit Rad51-mediated telomeric D-loop formation. In contrast, TRF2 Δ B was not observed to affect Rad51-mediated non-telomeric D-loop formation (Table 2.1; Fig. 2.4C, D). Deletion of the basic domain resulted in an approximately 2.3-fold reduction in template binding affinity (Table 1) but did not reduce binding specificity (Fig. 2.4E: lanes 8-11) compared to full length TRF2. Like TRF2, TRF2 Δ B binding resulted in the template becoming trapped in the wells.

TRF1 Promotes Rad51-Mediated Telomeric But Not Non-Telomeric D-Loop Formation.

Our observation that TRF2 and TRF2 Δ B but not TRF2 Δ M could inhibit Rad51-mediated telomeric D-loop formation suggested that this inhibition could simply be due to Myb-domain directed dsDNA binding. To test this hypothesis we characterized the DNA binding and Rad51-modulating activity of TRF1, a TRF2 homolog with a highly similar Myb domain (Supporting Information Fig. 2.S1A).²³ Interestingly and in contrast to TRF2, TRF1 was found to promote Rad51-mediated telomeric D-loop formation by 25 \pm 1.0% (Table 2.1; Fig. 2.5A, B), suggesting that the ability of TRF2 to inhibit this process is not simply due to Myb domain binding. In contrast, TRF1 was found not to affect Rad51-mediated non-telomeric D-loop formation (Fig. 2.5C, D). Comparisons between TRF1 and TRF2 must

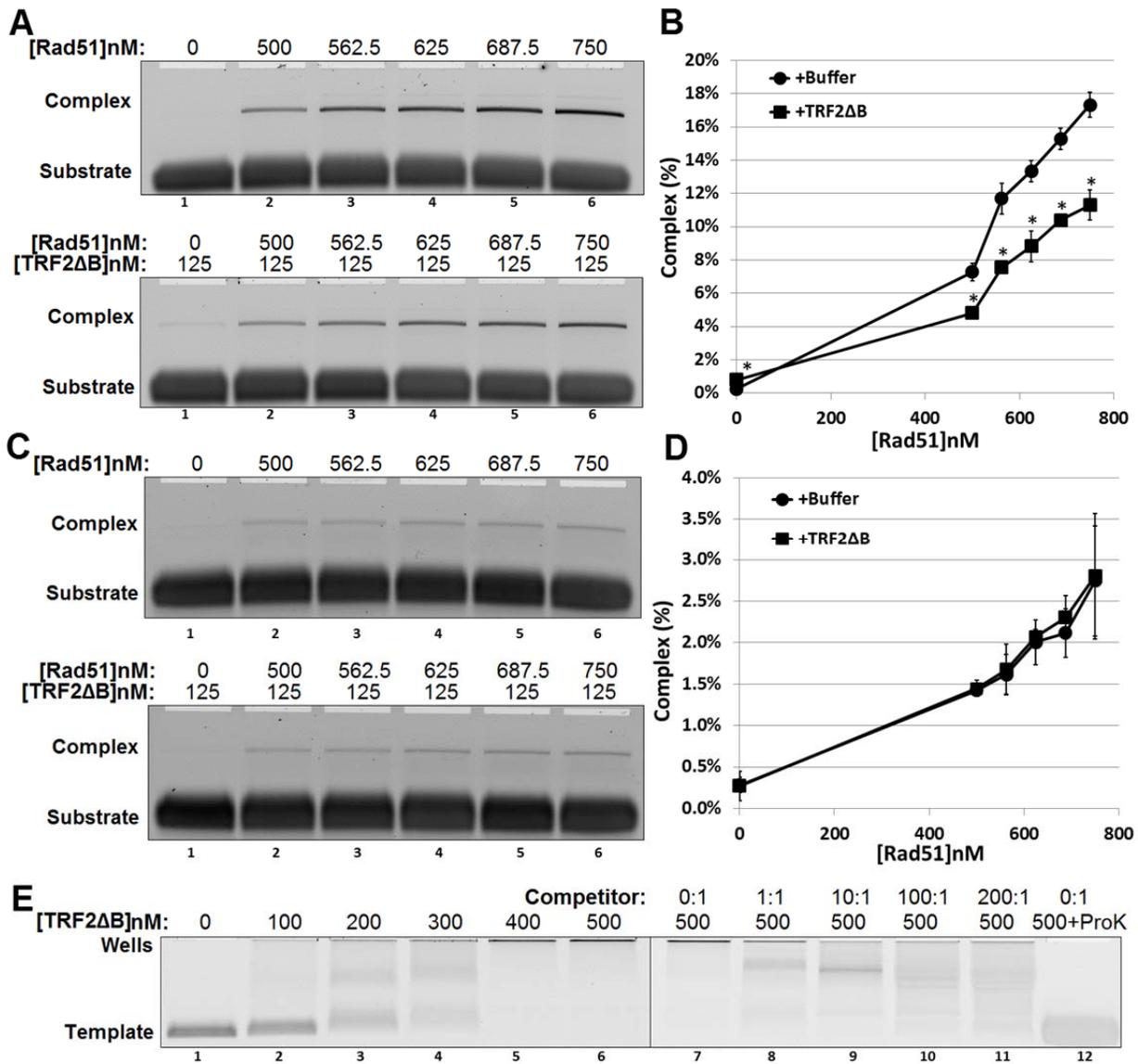


Figure 2.4 TRF2ΔB promotes Rad51-mediated telomeric but not non-telomeric D-loop formation. **(A)** Rad51 promotes telomeric D-loop formation in a concentration dependent manner that is promoted by TRF2ΔB. **(B)** Quantification of data in (A). **(C)** Rad51 promotes non-telomeric D-loop formation in a concentration dependent manner that is not affected by TRF2ΔB. **(D)** Quantification of data in (C). **(E)** TRF2ΔB binding supershifts the template into the wells. This binding is specific and persists in the presence of high concentrations of non-telomeric competitor and is protein-mediated. Error bars shown 95% confidence interval, significant difference between +Buffer and +TRF2ΔB (*), paired samples t-test $\alpha=0.05$ from three independent experiments.

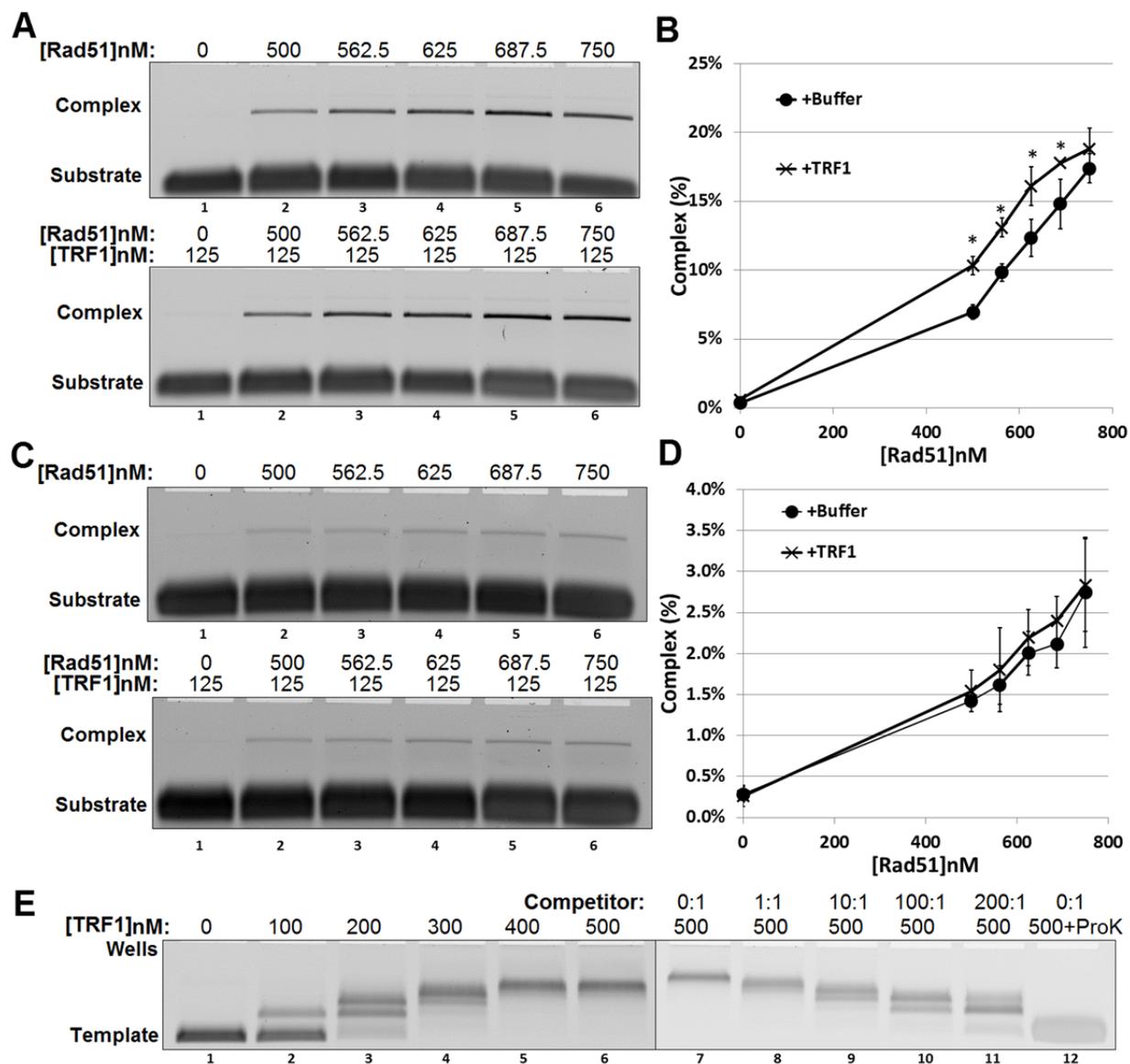


Figure 2.5 TRF1 promotes Rad51-mediated telomeric but not non-telomeric D-loop formation. **(A)** Rad51 promotes telomeric D-loop formation in a concentration dependent manner that is promoted by TRF1. **(B)** Quantification of data in (A). **(C)** Rad51 promotes non-telomeric D-loop formation in a concentration dependent manner that is not affected by TRF1. **(D)** Quantification of data in (C). **(E)** TRF1 binding supershifts the template into several low-mobility species. This binding is specific and persists in the presence of high concentrations of non-telomeric competitor and is protein-mediated. Error bars shown 95% confidence interval, significant difference between +Buffer and +TRF1 (*), paired samples t-test $\alpha=0.05$ from three independent experiments.

be made with caution, as despite possessing comparable DNA binding affinity and telomeric sequence specificity (Table 2.1) their binding behavior is otherwise grossly different when examined in an EMSA. Whereas TRF2 binding shifts a telomeric template into the wells (Fig. 2.2F), TRF1 binding shifts the species into increasingly larger complexes as the TRF1 concentration is increased (Fig. 2.5E). This behavior is perhaps consistent with previous observations that while TRF2 binds to telomeric dsDNA as a large oligomeric structure, TRF1 binds as a smaller complex.^{24, 25} Likewise, this property may be consistent with observations that TRF2 can promote the formation of unusual DNA structures and induce topological changes within telomeric DNA to a greater degree than TRF1.^{12, 26}

DISCUSSION

The results of this study suggest a model whereby TRF1 and TRF2 differentially regulate Rad51-mediated telomeric and non-telomeric D-loop formation. This would promote efficient telomeric DNA replication and non-telomeric HR while inhibiting aberrant HR at the telomeres. TRF1 promotes Rad51-mediated telomeric D-loop formation, which may facilitate replication fork restart and explain why TRF1 is required for efficient telomere replication. In contrast, TRF2 potentially inhibits Rad51-mediated telomeric D-loop formation, providing yet another mechanism by which TRF2 can inhibit DNA repair at telomeres. Finally, TRF2 Δ M promotes Rad51-mediated D-loop formation, providing insight into how TRF2 may contribute to HR. Our findings are generally in good agreement with previous characterizations, and what contradictions exist are likely due to methodological differences.

Data from our *in vitro* characterization must be compared with other *in vivo* and genetic characterizations with caution. The activities of TRF1, TRF2, and Rad51 are modulated *in vivo* by a variety of other proteins, including other shelterin and HR proteins, which are absent from our assay. Additionally, the templates and substrates used in our assay are necessarily different from their *in vivo* analogs. The 3' tails of telomeres are adjacent to a ss/dsDNA junction, which were absent from the substrates used in this work. Furthermore, telomeres are several kbp in length, whereas our template

possesses only a 103 bp telomeric tract. Finally, while the topology of telomeric DNA *in vivo* is unknown, the templates used in our assay were negatively supercoiled.

While we observed that TRF2 inhibits Rad51, it has previously been reported that TRF2 and Rad51 appear to exhibit functional cooperation. Immunodepletion of TRF2 or Rad51 from cell extracts ablates the ability of those extracts to promote telomeric D-loop formation.¹⁰ Moreover, supplementation of such immunodepleted extracts with purified Rad51 or TRF2 can restore telomeric D-loop formation.¹⁰²⁷ However, the presence of factors *in vivo* that are absent from our *in vitro* characterization may affect the activities of TRF2 and Rad51.

Although TRF1 and TRF2 can be found at telomeres throughout the cell cycle and TRF1 promotes efficient telomeric replication,⁷ TRF1 and TRF2 inhibit DNA replication *in vitro*.⁶ However, TRF1 and TRF2 binding are inhibited by post-translational modifications (PTMs), some of which are conferred by replication-complex associated proteins.²⁸⁻³⁴

These proteins may facilitate replication by transiently removing TRF1 and TRF2 from telomeric DNA near the replication fork.³⁰ Interestingly, PTMs that reduce TRF1 binding are inhibited *in vivo* by another shelterin protein, TIN2,²⁸ and by FANCD2, a component of the Fanconi anemia pathway.³¹ Comparable PTMs of TRF2 are not likewise inhibited. However, these TRF2 PTMs likely disrupt TRF2 dimerization, which would be expected to abrogate Myb domain binding but that may not affect basic domain binding. As TRF1 can promote and TRF2 can inhibit Rad51-mediated telomeric D-loop formation, the depletion of TRF2 but not TRF1 from DNA near the replication fork may facilitate HR-mediated fork restart within the telomeres (Fig. 2.6A). This process may also be facilitated by basic-domain mediated recruitment of TRF2 to regressed forks, where it can both protect the nascent Holliday junction (HJ) from HJ resolvases²² and recruit RecQ helicases that can promote fork migration.^{35, 36} Likewise, the presence of TRF2 on telomeric dsDNA away from the fork may prevent HR-mediated strand invasion reactions and protect the telomeres from aberrant repair (Fig. 2.6B).

The role of TRF2 in the HR pathway, especially in non-telomeric contexts, is not well understood. TRF2 is rapidly recruited to genomic dsDNA breaks (DSBs) and this recruitment requires

the basic domain of TRF2 but not its Myb-domain³⁷ and can occur in an ATM deficient background. Additionally, TRF2 is phosphorylated by ATM³⁸ in response to DNA damage,³⁹ and mutations that disrupt TRF2 phosphorylation inhibit DNA repair.⁴⁰ While it has been suggested that this DNA repair defect may be due to impaired non-homologous end joining (NHEJ),³⁹ the defect may also be due to impaired HR. Overexpression of TRF2 and TRF2 Δ M promote HR *in vivo*.¹¹ Likewise, knockdown of TRF2 inhibits HR but not NHEJ *in vivo*.¹¹ Our finding that TRF2 Δ M can promote Rad51-mediated D-loop formation suggests a novel mechanism by which TRF2 can promote HR (Fig. 2.6C). Upon induction of a DSB, TRF2 may undergo basic-domain mediated recruitment to the site of damage. TRF2 may then help recruit proteins such as the Mre11/Rad50/Nbs1 (MRN) complex,⁴¹ which promotes end resection in preparation for HR. Following end resection, Rad51 binding and homology search, the basic domain of TRF2 may facilitate Rad51-mediated D-loop formation by promoting the opening of the template dsDNA in a manner similar to a Rad51 accessory protein, Rad54.¹³

This model of the interaction of TRF1, TRF2 and Rad51 provides insight into both telomere biology and the HR pathway. Previous characterizations suggested that TRF2 and Rad51 cooperate to promote telomeric D-loop and possibly t-loop formation *in vivo*, despite apparent incompatibilities in the *in vitro* activities of these proteins. Our finding that TRF2 inhibits Rad51-mediated telomeric D-loop formation suggests that Rad51 does not contribute to t-loop formation, or that this inhibition must be alleviated by additional factors *in vivo*. While it has previously been reported that TRF1 is required for efficient telomere replication, this requirement seems at odds with other reports that TRF1 can inhibit telomere replication. Our finding that TRF1 promotes Rad51-mediated telomeric D-loop formation suggests that TRF1 may facilitate telomere replication by promoting HR-mediated replication fork restart. Finally, our observation that TRF2 Δ M can promote Rad51-mediated non-telomeric D-loop formation may explain how TRF2 can promote HR *in vivo*.

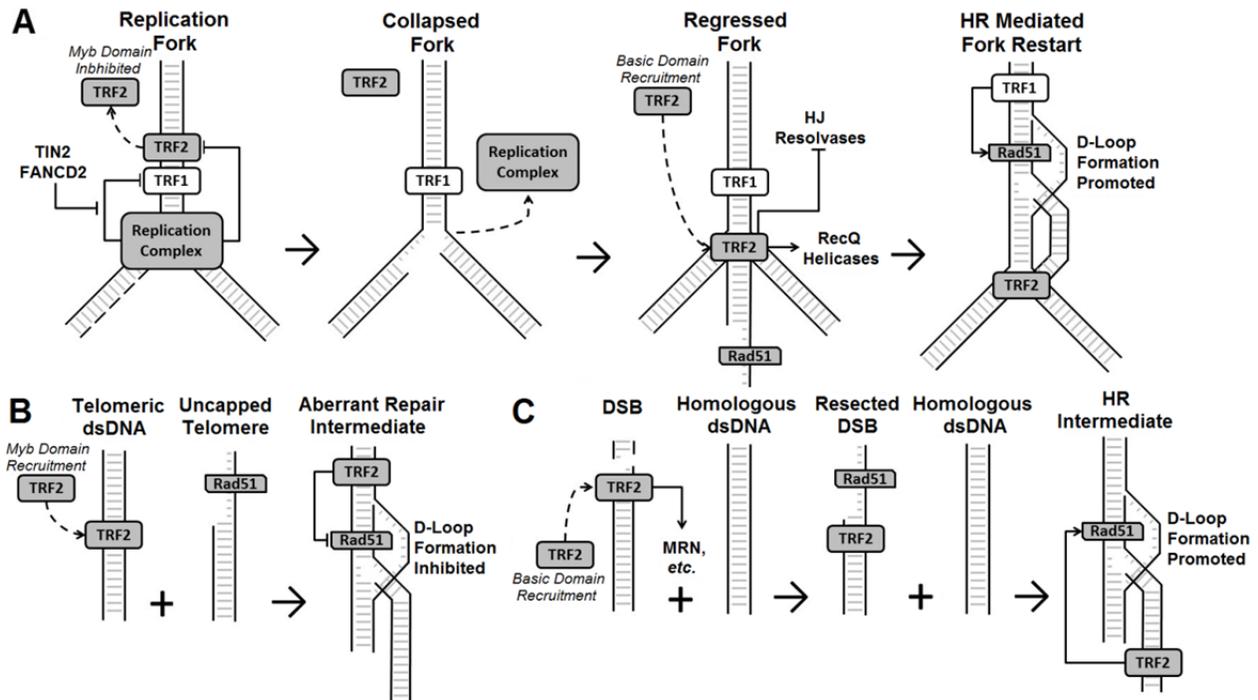


Figure 2.6 TRF1 and TRF2 differentially modulate Rad51-mediated telomeric and non-telomeric D-loop formation. **(A)** Post translational modifications may deplete TRF2 but not TRF1 from telomeric DNA near a replication fork, possibly by inhibiting TRF2 Myb domain binding. Following fork collapse, basic domain directed TRF2 binding can protect regressed forks from cleavage and recruit factors that promote fork migration. Finally, TRF1 can promote D-loop formation away from the fork and thereby promote HR mediated fork restart. **(B)** TRF2 inhibits Rad51-mediated telomeric D-loop formation, which may prevent aberrant repair processes at the telomeres. **(C)** TRF2 is recruited to DSBs, where it may promote recruitment of enzymes that promote end resection. Afterwards, the basic domain of TRF2 may promote Rad51-mediated D-loop formation and thereby promote HR.

REFERENCES

- [1] Bombarde, O., Bobby, C., Gomez, D., Frit, P., Giraud-Panis, M. J., Gilson, E., Salles, B., and Calsou, P. (2010) TRF2/RAP1 and DNA-PK mediate a double protection against joining at telomeric ends, *EMBO J* 29, 1573-1584.
- [2] Karlseder, J., Hoke, K., Mirzoeva, O. K., Bakkenist, C., Kastan, M. B., Petrini, J. H., and de Lange, T. (2004) The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response, *PLoS Biol* 2, E240.
- [3] Doksan, Y., Wu, J. Y., de Lange, T., and Zhuang, X. (2013) Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation, *Cell* 155, 345-356.
- [4] Stansel, R. M., de Lange, T., and Griffith, J. D. (2001) T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang, *EMBO J* 20, 5532-5540.
- [5] Fouche, N., Ozgur, S., Roy, D., and Griffith, J. D. (2006) Replication fork regression in repetitive DNAs, *Nucleic Acids Res* 34, 6044-6050.
- [6] Ohki, R., and Ishikawa, F. (2004) Telomere-bound TRF1 and TRF2 stall the replication fork at telomeric repeats, *Nucleic Acids Res* 32, 1627-1637.
- [7] Sfeir, A., Kosiyatrakul, S. T., Hockemeyer, D., MacRae, S. L., Karlseder, J., Schildkraut, C. L., and de Lange, T. (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication, *Cell* 138, 90-103.
- [8] North, J. A., Amunugama, R., Klajner, M., Bruns, A. N., Poirier, M. G., and Fishel, R. (2013) ATP-dependent nucleosome unwrapping catalyzed by human RAD51, *Nucleic Acids Res* 41, 7302-7312.
- [9] Badie, S., Escandell, J. M., Bouwman, P., Carlos, A. R., Thanasoula, M., Gallardo, M. M., Suram, A., Jaco, I., Benitez, J., Herbig, U., Blasco, M. A., Jonkers, J., and Tarsounas, M. (2010) BRCA2 acts as a RAD51 loader to facilitate telomere replication and capping, *Nat Struct Mol Biol* 17, 1461-1469.
- [10] Verdun, R. E., and Karlseder, J. (2006) The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres, *Cell* 127, 709-720.
- [11] Mao, Z., Seluanov, A., Jiang, Y., and Gorbunova, V. (2007) TRF2 is required for repair of nontelomeric DNA double-strand breaks by homologous recombination, *Proc Natl Acad Sci U S A* 104, 13068-13073.
- [12] Amiard, S., Doudeau, M., Pinte, S., Poulet, A., Lenain, C., Faivre-Moskalenko, C., Angelov, D., Hug, N., Vindigni, A., Bouvet, P., Paoletti, J., Gilson, E., and Giraud-Panis, M. J. (2007) A topological mechanism for TRF2-enhanced strand invasion, *Nat Struct Mol Biol* 14, 147-154.
- [13] Sigurdsson, S., Van Komen, S., Petukhova, G., and Sung, P. (2002) Homologous DNA pairing by human recombination factors Rad51 and Rad54, *J Biol Chem* 277, 42790-42794.
- [14] Subramanian, D., and Griffith, J. D. (2005) p53 Monitors replication fork regression by binding to "chickenfoot" intermediates, *J Biol Chem* 280, 42568-42572.

- [15] Amunugama, R., He, Y., Willcox, S., Forties, R. A., Shim, K. S., Bundschuh, R., Luo, Y., Griffith, J., and Fishel, R. (2012) RAD51 protein ATP cap regulates nucleoprotein filament stability, *J Biol Chem* 287, 8724-8736.
- [16] Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A., and de Lange, T. (2002) DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2, *Curr Biol* 12, 1635-1644.
- [17] Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2, *Nat Genet* 17, 231-235.
- [18] Biet, E., Sun, J., and Dutreix, M. (1999) Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure, *Nucleic Acids Res* 27, 596-600.
- [19] Tracy, R. B., Baumohl, J. K., and Kowalczykowski, S. C. (1997) The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences, *Genes Dev* 11, 3423-3431.
- [20] Forget, A. L., and Kowalczykowski, S. C. (2010) Single-molecule imaging brings Rad51 nucleoprotein filaments into focus, *Trends Cell Biol* 20, 269-276.
- [21] Fouche, N., Cesare, A. J., Willcox, S., Ozgur, S., Compton, S. A., and Griffith, J. D. (2006) The basic domain of TRF2 directs binding to DNA junctions irrespective of the presence of TTAGGG repeats, *J Biol Chem* 281, 37486-37495.
- [22] Poulet, A., Buisson, R., Faivre-Moskalenko, C., Koelblen, M., Amiard, S., Montel, F., Cuesta-Lopez, S., Bornet, O., Guerlesquin, F., Godet, T., Moukhtar, J., Argoul, F., Declais, A. C., Lilley, D. M., Ip, S. C., West, S. C., Gilson, E., and Giraud-Panis, M. J. (2009) TRF2 promotes, remodels and protects telomeric Holliday junctions, *EMBO J* 28, 641-651.
- [23] Court, R., Chapman, L., Fairall, L., and Rhodes, D. (2005) How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures, *EMBO Rep* 6, 39-45.
- [24] Poulet, A., Pisano, S., Faivre-Moskalenko, C., Pei, B., Tauran, Y., Haftek-Terreau, Z., Brunet, F., Le Bihan, Y. V., Ledu, M. H., Montel, F., Hugo, N., Amiard, S., Argoul, F., Chaboud, A., Gilson, E., and Giraud-Panis, M. J. (2012) The N-terminal domains of TRF1 and TRF2 regulate their ability to condense telomeric DNA, *Nucleic Acids Res* 40, 2566-2576.
- [25] Khan, S. J., Yanez, G., Seldeen, K., Wang, H., Lindsay, S. M., and Fletcher, T. M. (2007) Interactions of TRF2 with model telomeric ends, *Biochem Biophys Res Commun* 363, 44-50.
- [26] Yoshimura, S. H., Maruyama, H., Ishikawa, F., Ohki, R., and Takeyasu, K. (2004) Molecular mechanisms of DNA end-loop formation by TRF2, *Genes Cells* 9, 205-218.
- [27] Verdun, R. E., Crabbe, L., Haggblom, C., and Karlseder, J. (2005) Functional human telomeres are recognized as DNA damage in G2 of the cell cycle, *Mol Cell* 20, 551-561.
- [28] Smith, S., and de Lange, T. (2000) Tankyrase promotes telomere elongation in human cells, *Curr Biol* 10, 1299-1302.

- [29] Ye, J. Z., and de Lange, T. (2004) TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex, *Nat Genet* 36, 618-623.
- [30] Beneke, S., Cohausz, O., Malanga, M., Boukamp, P., Althaus, F., and Burkle, A. (2008) Rapid regulation of telomere length is mediated by poly(ADP-ribose) polymerase-1, *Nucleic Acids Res* 36, 6309-6317.
- [31] Lyakhovich, A., Ramirez, M. J., Castellanos, A., Castella, M., Simons, A. M., Parvin, J. D., and Surrallés, J. (2011) Fanconi anemia protein FANCD2 inhibits TRF1 polyADP-ribosylation through tankyrase1-dependent manner, *Genome Integr* 2, 4.
- [32] Dantzer, F., Giraud-Panis, M. J., Jaco, I., Ame, J. C., Schultz, I., Blasco, M., Koering, C. E., Gilson, E., Menissier-de Murcia, J., de Murcia, G., and Schreiber, V. (2004) Functional interaction between poly(ADP-Ribose) polymerase 2 (PARP-2) and TRF2: PARP activity negatively regulates TRF2, *Mol Cell Biol* 24, 1595-1607.
- [33] Simbulan-Rosenthal, C. M., Rosenthal, D. S., Boulares, A. H., Hickey, R. J., Malkas, L. H., Coll, J. M., and Smulson, M. E. (1998) Regulation of the expression or recruitment of components of the DNA synthesome by poly(ADP-ribose) polymerase, *Biochemistry* 37, 9363-9370.
- [34] Walker, J. R., and Zhu, X. D. (2012) Post-translational modifications of TRF1 and TRF2 and their roles in telomere maintenance, *Mech Ageing Dev* 133, 421-434.
- [35] Machwe, A., Xiao, L., and Orren, D. K. (2004) TRF2 recruits the Werner syndrome (WRN) exonuclease for processing of telomeric DNA, *Oncogene* 23, 149-156.
- [36] Opresko, P. L., von Kobbe, C., Laine, J. P., Harrigan, J., Hickson, I. D., and Bohr, V. A. (2002) Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases, *J Biol Chem* 277, 41110-41119.
- [37] Bradshaw, P. S., Stavropoulos, D. J., and Meyn, M. S. (2005) Human telomeric protein TRF2 associates with genomic double-strand breaks as an early response to DNA damage, *Nat Genet* 37, 193-197.
- [38] Tanaka, H., Mendonca, M. S., Bradshaw, P. S., Hoelz, D. J., Malkas, L. H., Meyn, M. S., and Gilley, D. (2005) DNA damage-induced phosphorylation of the human telomere-associated protein TRF2, *Proc Natl Acad Sci U S A* 102, 15539-15544.
- [39] Huda, N., Abe, S., Gu, L., Mendonca, M. S., Mohanty, S., and Gilley, D. (2012) Recruitment of TRF2 to laser-induced DNA damage sites, *Free Radic Biol Med* 53, 1192-1197.
- [40] Huda, N., Tanaka, H., Mendonca, M. S., and Gilley, D. (2009) DNA damage-induced phosphorylation of TRF2 is required for the fast pathway of DNA double-strand break repair, *Mol Cell Biol* 29, 3597-3604.
- [41] Zhu, X. D., Kuster, B., Mann, M., Petrini, J. H., and de Lange, T. (2000) Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres, *Nat Genet* 25, 347-352.
- [42] Bower, B. D., and Griffith, J. D. (2014) TRF1 and TRF2 Differentially Modulate Rad51-Mediated Telomeric and Nontelomeric Displacement Loop Formation in Vitro, *Biochemistry* 53, 5485-5495.

Chapter 3: Biophysical and Ultrastructural Characterization of Adeno-Associated Virus Capsid Uncoating and Genome Release¹

INTRODUCTION

Adeno-associated virus (AAV) is a small (25 nm) nonenveloped virus belonging to the family *Parvoviridae* and genus *Dependovirus*. The AAV capsid packages a single-stranded (ssDNA) genome approximately 4.7 kb in length¹. The wild-type genome consists of two open reading frames flanked by two inverted terminal hairpin repeats (ITRs). The ITRs, which are 145 nucleotides each, are the only *cis* element in the AAV genome required for successful packaging^{2,3}. The AAV capsid is composed of 60 (T = 1) viral protein subunits VP1, VP2, and VP3, in approximately the ratio 1:1:10. The three different subunits are generated from overlapping reading frames and interact within the capsid through the common VP3 subunit region. The largest VP1 subunit is known to possess a phospholipase A2 domain required for infectivity⁴. Because of its broad tropism, lack of pathogenicity, and flexibility in genome content, AAV has become a promising candidate for therapeutic gene transfer applications. In the past 2 decades, AAV has been utilized as a gene transfer vector in a number of phase I and phase II clinical trials treating various genetic diseases⁵.

Different AAV serotypes infect cells by engaging a variety of cell surface glycans and coreceptors, followed by endocytic uptake^{4,6,7}. Viral particles are then thought to escape from the endosome and translocate to the nucleus, where the ssDNA genome is released and undergoes second-strand synthesis. Engineered AAV genomes containing a mutant 3' ITR have been shown to package dimeric, self-complementary DNA (scDNA)⁸. Such scAAV vectors have the advantage of escaping ssDNA degradation⁹ and bypassing second-strand synthesis, which is a rate-limiting step preceding

¹The following chapter describes work done in collaboration with Dr. Eric Horowitz, Dr. Shefaet Rahman, Dr. David Dismuke, Dr. Michael Falvo, Dr. Jack Griffith, Dr. Stephen Harvey, and Dr. Aravind Asokan. I significantly contributed to EM analysis of AAV genome release via thermal denaturation⁴⁷.

transgene expression by AAV vectors^{10,11}. These features have been shown to enable rapid onset of transgene expression by scAAV vectors.

The packaging capacity of ssAAV and scAAV vectors has been extensively studied¹²⁻¹⁷. However, little is known about the consequences of packaging subgenomic-length DNA or self-complementary genomes for AAV capsid uncoating. Previously, studies with minute virus of mice (MVM) packaging subgenomic-length DNA have demonstrated that such defective particles do not release their genomes *in vitro*¹⁸. Along with the observation that the MVM genome reinforces and increases the stiffness of MVM capsids¹⁹, these studies suggest that the viral genome exerts an internal pressure on the walls of the capsid, which is critical for proper capsid uncoating. Thermal analysis has previously been used to understand large conformational changes in AAV capsids such as exposure of the buried phospholipase A2 domain during infection^{18,20-22}. In the current study, we utilized atomic force microscopy (AFM), electron microscopy (EM), and fluorescence-based assays to characterize the biophysical and ultrastructural properties of different AAV vectors and the impact of thermal stimuli on capsid uncoating/genome release. In addition, we utilized computational modeling and molecular dynamics (MD) to further understand the potential organization and architecture of DNA packaged within ssAAV and scAAV vectors.

MATERIALS AND METHODS

Chemicals and reagents

Mica was purchased (Ted Pella Inc.) and cleaved immediately prior to use. Four-hundred-mesh copper Formvar carbon-coated TEM grids (Ted Pella Inc.) were made hydrophilic by glow discharge immediately before use. Uranyl acetate (Ted Pella Inc.) was used as received. Uranyl acetate solutions were centrifuged for several minutes prior to use to remove any precipitate or aggregates. SYBR gold (Invitrogen Life Sciences) was first diluted in water immediately prior to use. Tris-HCl and spermidine (ThermoFisher) were utilized as received.

Virus production and purification

All viral vectors were generated at the UNC Vector Core by iodixanol gradient ultracentrifugation followed by ion-exchange chromatography and vector genome titers determined by dot blot assay as well as verified by quantitative PCR (qPCR) as described previously²³. Single-stranded AAV serotype 2 vectors packaging different transgene cassettes were as follows: wild-type AAV genome containing *rep* and *cap* genes (4.7 kb)²⁴, chicken β -actin (CBA) promoter-driven firefly luciferase (4.1 kb)²⁵, CBA promoter-driven tdTomato (3.8 kb)²⁵, cytomegalovirus (CMV) promoter-driven firefly luciferase (3.6 kb)²⁶, CMV promoter-driven green fluorescent protein (GFP) (3.4 kb)²³, and an EF1a promoter-driven mCherry with an internal ribosome entry site (IRES) signal followed by WGA-Cre (5.3 kb)²⁷. Self-complementary AAV2 vectors utilized in the study were as follows: CMV promoter-driven green fluorescent protein with lambda phage genome stuffer DNA (5.0 kb)²³, CMV promoter-driven green fluorescent protein (4.6 kb)²⁸, and CBA hybrid promoter-driven green fluorescent protein (4.1 kb)²⁸.

AFM

Freshly cleaved mica was treated with 50 μ l of polylysine (50 μ g/ml) for 2 h at 37°C in a humid chamber. Mica was then washed three times with distilled water (dH₂O) and dried. Viral particles (1×10^{10} vector genomes [vg]) in $1\times$ phosphate-buffered saline (PBS) were heated in a thermocycler for 30 min to 37 or 65°C, followed by immediate chilling to 4°C. AAV was then deposited on the treated mica surface (10 μ l) in a humid chamber for 45 min at room temperature. Mica pieces were then washed three times with dH₂O and gently air dried. Atomic force microscopy (AFM) images were acquired using an Asylum MFP3D atomic force microscope. All images were collected in air using the tapping mode. Images were then processed using the MFP3D software.

TEM

Viral particles (2×10^{10} vg) in $1\times$ PBS were heated in a thermocycler for 30 min to 37, 55, or 65°C, followed by immediate chilling to 4°C. Vectors were then adsorbed onto 400- μ m mesh carbon-

coated copper transmission electron microscopy (TEM) grids for 2 min. After washing twice with 0.2- μm -filtered dH_2O , the grids were stained with freshly prepared 2% uranyl acetate for 30 s. After drying, grids were imaged using a Zeiss LEO 910 transmission electron microscope. Ten to 20 images of each grid were captured in random locations to get an accurate sampling of the viral particles. Each image was then analyzed using ImageJ by counting the number of full particles based on their distinct morphologies and intensity of uranyl acetate staining. Particle counts for each virus and experiment ranged from 395 to 5,317 total particles. In all cases, the average and standard deviation for each temperature were determined using the percentage of full particles from each image. Empty particles determined by their differential staining²⁹ were excluded from counting due to the presence of fragmented capsids and diverse morphology.

Fluorometric assay for detection of thermally induced genome release

Viral vectors ($1 \times 10^{10}\text{vg}$) in $1\times$ PBS with 25 μM SYBR gold were heated using a Roche LightCycler 480. Samples were heated in a stepwise manner (2°C steps) from 37°C to 95°C . Viruses were held at each temperature for 5 min prior to measuring fluorescence. Each sample was subjected to thermal treatment in triplicate. In addition to a dye-only control, each virus was heated to 95°C prior to addition of SYBR gold to determine whether any reversible fluorescence changes originated from melting of DNA secondary structures. Data were normalized to baseline spectra obtained prior to and after complete thermal transition to obtain the ratio of uncoated particles to intact particles. Melting temperature (T_m) values were defined as the temperature at which uncoated and intact viral particles were at 50% each (uncoated particle ratio = 0.5). For subjecting samples to increasing osmotic pressure, wild-type AAV in PBS (pH 7.2) was supplemented with 5%, 10%, 20%, or 30% polyethylene glycol (PEG) 8000 (Sigma). For pH studies, samples at pH 5 and pH 6 were prepared by dialyzing wild-type AAV in citrate buffer (pH 5 or pH 6) supplemented with 137 mM NaCl. Samples at pH 7.2 were maintained in PBS. Error bars represent the standard deviation ($n = 3$). All AAV vectors were obtained from the UNC Vector Core, and

the size/nature of packaged genomes (ssDNA or scDNA) was blinded to the authors during these experiments.

Immunoblotting (dot blotting) of heat-treated AAV particles

Using a thermocycler, AAV packaging different transgenes was subjected to heat treatment. Wild-type AAV, empty AAV capsids, and AAV packaging single-stranded and self-complementary CMV-GFP genomes were heated to temperatures ranging from 40 to 73°C for 30 min prior to rapid cooling to 4°C (5×10^9 vg in 200 μ l per well). Each virus was heated in duplicate and applied to two separate dot blots. Samples were loaded onto two nitrocellulose membranes in a dot blot apparatus. Membranes were blocked using 5% dehydrated milk in 1 \times Tris-buffered saline (TBS)–0.1% Tween 20 (TBS-T) for 1 h. The monoclonal mouse antibody A20 recognizes a conformational epitope on the intact capsid, while the monoclonal mouse antibody A1 recognizes a linear epitope in the N terminus of the VP1 capsid protein³⁰. Membranes were soaked with primary antibodies (1:30 in 2% dehydrated milk in TBS-T) for 1 h. Horseradish peroxidase (HRP)-conjugated goat anti-mouse polyclonal antibody (1:5,000) was then bound to the primary antibody for 1 h, followed by 4 washes using TBS-T and detection of signal using the West Femto chemiluminescence detection kit (ThermoFisher Scientific).

Tungsten-shadowing electron microscopy.

AAV vectors packaging ssDNA and scDNA cassettes containing cytomegalovirus (CMV) promoter-driven green fluorescent protein (GFP) (3×10^9 vg) were diluted in 10 mM Tris-HCl (pH 7.5) to 1×10^{12} vg/ml. Dilutions were heated to 65°C for 5 min and then rapidly cooled on ice. Ten-microliter aliquots of these dilutions were then further diluted 1:3 in 10 mM Tris-HCl (pH 7.5), followed by 1:1 dilution with a buffer containing 4 mM spermidine, and mounted to charged-carbon-foil grids as published previously^{31,32}. Carbon grids were washed in water, dehydrated in a series of ethanol washes, air dried, and rotary shadow cast with tungsten. Samples were visualized on a Tecnai 12 transmission electron microscope at 40 kV. All microscopy images were captured using a Gatan Ultrascan 4000 charge-coupled device (CCD) camera and supporting software (Gatan Inc.).

DNA and capsid modeling

Following the report by Locker and Harvey³³, the AAV genome was coarsely modeled by an elastic bead-spring polymer chain model, where each bead represented six nucleotide pairs. The molecular mechanics potential function used in this model was of the form

$$E = \sum_i (E_{r_i} + E_{\theta_i} + E_{c_i}) + \sum_{i,j} E_{d_{i,j}} \quad (1)$$

The stretching (E_{r_i}) and (E_{θ_i}) bending terms were modeled as harmonic oscillators of the form

$$E_{\alpha_i} = k_{\alpha} (\alpha_i - \alpha_0)^2 \quad (2)$$

where the equilibrium bond distances ($\alpha = r$, r_i being the distance between monomers i and $i + 1$) and angles ($\alpha = \theta$, θ_i being the angle between monomers i , $i + 1$, and $i + 2$) and their spring constants were chosen to match the structure and properties of ssDNA and double-stranded DNA (dsDNA) under physiological conditions. Specifically, as described previously³³, we used statistical distributions of bond lengths in experimentally observed structures of DNA to derive stretching parameters of $k_r = 3.49$ kcal/mol/Å² and $r_0 = 20$ Å. Parameters for the bending term were chosen to match the persistence lengths and elastic properties of dsDNA and ssDNA observed experimentally, yielding values of $\theta_0 = 180^\circ$, $k_\theta = 14.6$ kcal/mol/rad² for stiff chains representing double-stranded DNA, and $k_\theta = 0.75$ kcal/mol/rad² for flexible chains representing single-stranded DNA. The exclusion term ($E_{d_{ij}}$) gives the chain volume and prevents self-crossing using a shifted, repulsive-only Lennard-Jones function (equation 3):

$$E_{d_{i,j}} = 4\epsilon \left[\left(\frac{\sigma}{d_{ij}} \right)^{12} - \left(\frac{\sigma}{d_{ij}} \right)^6 + \frac{1}{4} \right], d_{ij} < d_0 \quad (0 \text{ otherwise}) \quad (3)$$

where d_{ij} is the distance between monomers i and j , $\epsilon = 15$ kcal/mol, $\sigma = 22.27$ Å, and d_0 , the effective chain diameter, was chosen to be 25 Å, consistent with observed values of interaction distances in packed dsDNA viral genomes³⁴. No torsional restraints were included in the force field, and electrostatic and other long-range nonbonded interactions were also excluded in this simplified model. Chains of 350

monomers (representing 2,100 bp) were generated through a Monte Carlo random walk, with initial angles assigned from a Boltzmann distribution using the bending energy function for flexible chains given above and torsion angles randomly chosen from a uniform distribution. In this model, the icosahedral capsid was represented by a smooth sphere with a diameter of 216 Å ($R_c = 108$ Å), chosen to match the internal volume of the actual virus based on inspection of the X-ray crystal structure of the AAV-2 capsid (Protein Data Bank [PDB] no. 1LP3)³⁵. The capsid restraint term was modeled as in equation 4 as a purely repulsive semiharmonic potential:

$$E_{c_i} = k_c (R_i - R_c)^2, R_i > R_c \quad (0 \text{ otherwise}) \quad (4)$$

where R_i is the distance of monomer i from the center of the capsid and $k_c = 8.8$ kcal/mol/Å².

MD simulations

Molecular dynamics (MD) simulations were used to predict the organization and energies of encapsidated single-stranded and self-cDNA inside the AAV capsid. All simulations were carried out using the LAMMPS MD simulation package³⁶ with a time step of 500 fs and at a temperature of 300 K maintained by a Nosé-Hoover thermostat³⁷.

Models of virus particles encapsidating ssDNA genomes were created by placing randomly generated flexible chains representing ssDNA inside spherical semiharmonic potential boundaries of the form given in equation 4 and gradually reducing the confining radius, R_c , from an initial value of 2,000 Å to 254 Å in 1-Å decrements, equilibrating the DNA chain for 5 ns at each step. At a bounding radius of 254 Å, the DNA chain occupied 5% of the volume of the confining sphere, and it was allowed at this point to equilibrate over 50 ns of MD. The confining sphere was then shrunk further in 1-Å steps, following the same protocol of equilibrating for 5 ns after each 1-Å decrement and subsequently stopping to equilibrate for 50 ns at steps where the volume fraction of the DNA chain was a multiple of 5%. This was continued until the target R_c of 108 Å (containing 65% DNA by volume fraction) was reached.

Following confinement, the encapsidated ssDNA chains were converted step by step into dsDNA chains in order to simulate *in situ* base pairing of two adjacent ssDNA chains nucleated at the central ITR, as might occur in a self-complementary AAV vector. To accomplish this, the small spring constants of the bending terms in the flexible ssDNA chains ($k_{\theta} = 0.75$ kcal/mol/rad²) were converted, one angle at a time, to the larger value ($k_{\theta} = 14.6$ kcal/mol/rad²) characteristic of the stiffer dsDNA chains. The chains were allowed to equilibrate for 50 ns after each angle was stiffened, and the pressure (defined as the average total force exerted by the genome on the capsid walls divided by the surface area of the capsid) was calculated at intervals of 10% DNA base paired. Capsid pressure data were collected by using the radial positions of the monomers at each time step during the final 10 ns of the MD run to compute the average force exerted by the capsid restraining term (equation 4) and dividing by the surface area of the spherical capsid.

Psoralen/UV-A (PUVA) Crosslinking:

scAAV preparations were supplemented with 4'-aminomethyltrioxysalen to a final concentration of 0, 25, or 250µg/ml and incubated at 37 °C for 30 minutes then irradiated with a long-wave ultraviolet light at a distance of 7.6c m for 15minutes.

Incubation/SSB Binding:

scAAV samples were diluted to a concentration 3.3×10^{11} vg/ml with an estimated DNA concentration of 5 ng/ul in 10 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA (EMTE) and incubated at 90 °C for 20minutes in the presence or absence of 50ng/ul of a recombinant extremely thermostable single stranded binding protein (ET SSB, New England Biolabs) then gradually cooled to 0 °C. ETSSB was then added to an scAAV samples that had not been incubated with ET SSB. All reactions were fixed via incubation with 0.6% glutaraldehyde for 5 minutes at room temperature. The reactions were then passed through a CENTRI•SPIN 40 (Princeton Separations) size exclusion column equilibrated with EMTE using a modification of the manufacturer's protocol.

RESULTS

Thermally induced DNA release is dependent on genome length

AAV vectors have been previously shown to undergo structural transitions in response to limited heating²⁰⁻²². We adapted this approach to determine the impact of thermal stimuli on encapsidated DNA. Atomic force microscopy (AFM) images of scAAV vectors heated to 65°C demonstrated release of AAV genomes from intact capsids (Fig. 3.1A), similar to the case for other parvoviruses such as MVM¹⁸. Further, we utilized transmission electron microscopy (TEM) to investigate whether the heated capsids contained DNA (full) or not (empty). Capsids which are intact and contain an encapsidated genome appear as 25-nm opaque/white spheres (Fig. 3.1B, panel i). In contrast, capsids which have released their packaged DNA are permeable to uranyl acetate and therefore appear as 25-nm rings (Fig. 3.1B, panels ii to v). In addition, as shown in Fig. 3.1B (panels vi to x), several capsid fragments and other morphologies were observed upon heating.

Both ssDNA- and scDNA-packaging AAV capsids incubated at three different temperatures (37, 55, and 65°C) were then visualized by TEM to quantify the populations of full and empty particles (Fig. 3.2). At 55°C and 65°C, we observed an increase in the number of empty particles compared to that at 37°C. Three different ssAAV vectors with genomes ranging from 72% (3.4 kb) to 100% (4.7 kb) of the wild-type genome length were characterized (Fig. 3.2A). After counting the numbers of full and empty particles, we observed a trend wherein AAV capsid stability appears to increase as genome size is reduced (Fig. 3.3A). Notably, when heated to 55°C, wild-type AAV (4.7 kb) has only 40% intact capsids remaining, while the 3.4-kb ssAAV vector is unaffected, with nearly 100% intact capsids. Similarly, at 65°C, the wild-type AAV sample has ~16% intact capsids remaining, while the shorter, 3.4-kb ssAAV vector still has 75% intact capsids. This trend is also seen in case of scAAV vectors, where the smallest of the three tested genomes (4.1 kb) has 79% intact particles at 65°C, while the largest of the three (5.0 kb) has only 9% intact capsids remaining at 65°C (Fig. 3.2B and 3.3B).

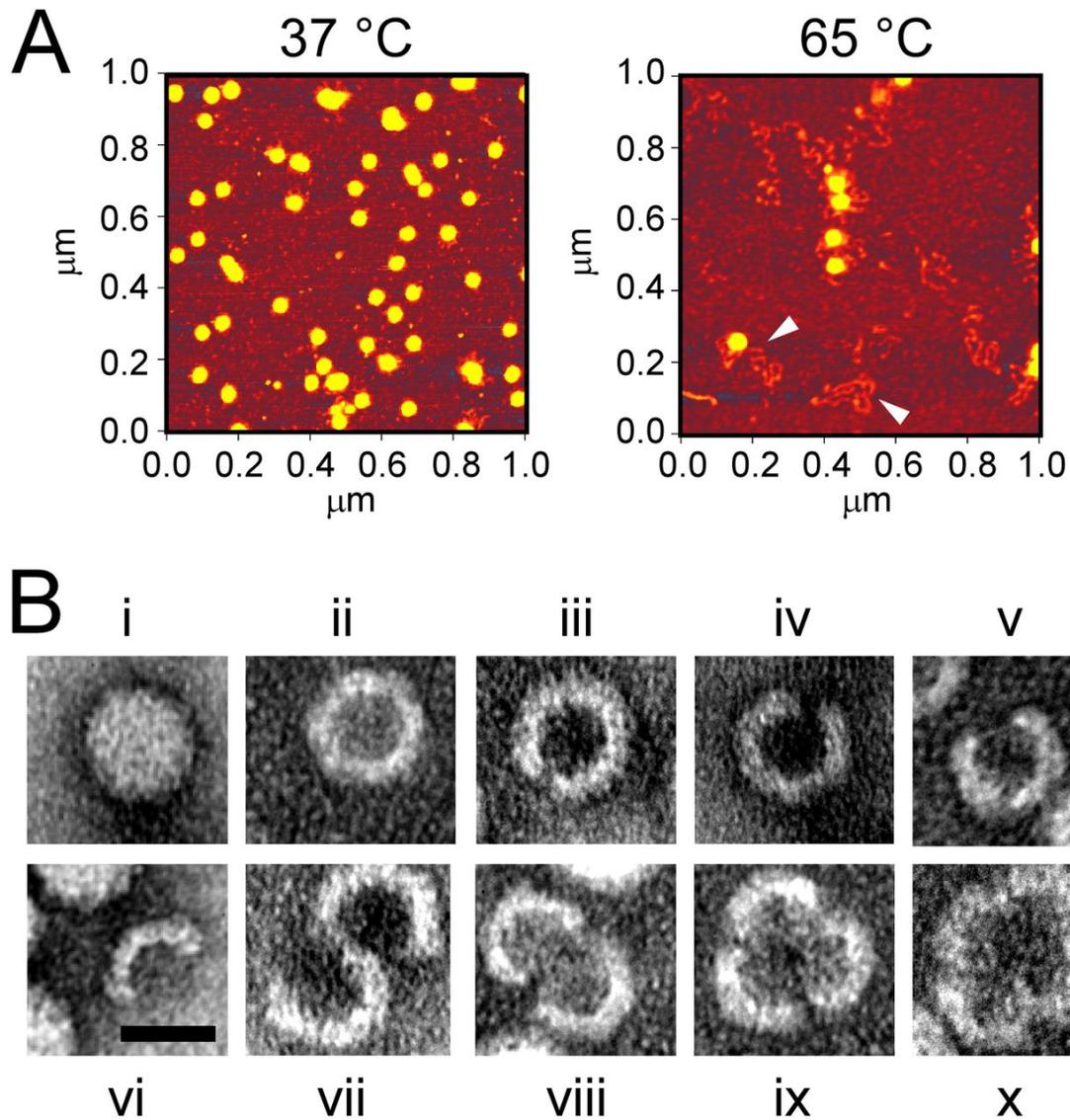


Figure 3.1 Ultrastructural characterization of AAV capsid uncoating. (A) Atomic force microscopy images of scAAV (sc-CMV-GFP) vectors heated to 37°C and 65°C. Viral particles are observed as 25-nm spheres. After heating to 65°C, viral genomes are detected both associated with viral particles and as free DNA (arrowheads). (B) Observed morphologies of heat treated AAV. Genome-containing AAV particles are impermeable to uranyl acetate (i), while empty AAV particles that have released DNA exhibit a ring-like architecture (ii). Genome release resulted in diverse morphologies characterized by small gaps in the capsid wall (iii to v) as well as capsid fragments forming multimeric aggregates (vi to x). The scale bar represents 20 nm.

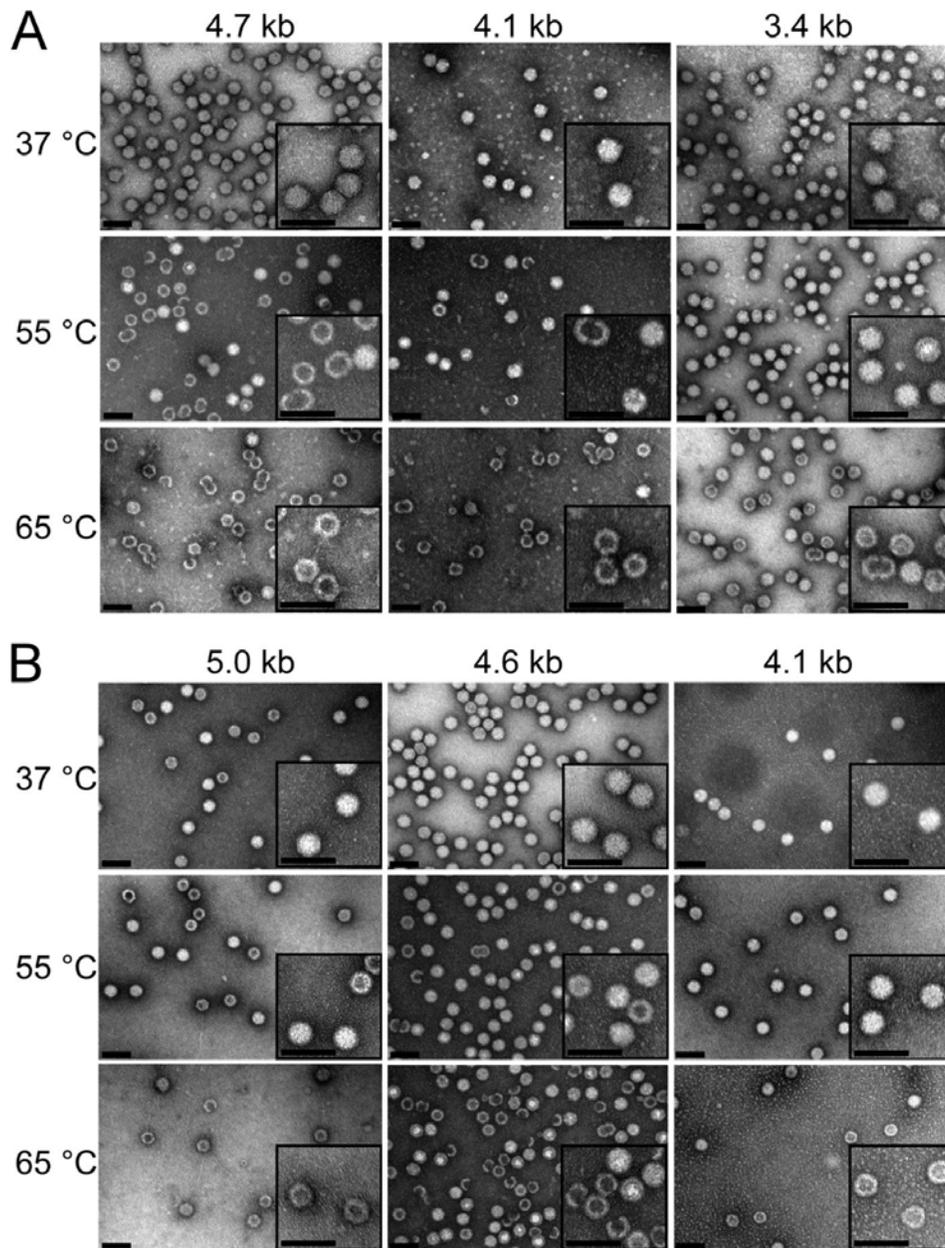


Figure 3.2 Effect of packaged genome length on AAV capsid uncoating. ssAAV (A) and scAAV (B) vectors were heated to different temperatures for 30 min prior to TEM imaging. Full AAV virions are viewed as 25-nm solid spheres, and empty virions are 25-nm donut-like structures. Differences in the relative amount of empty particles are apparent for different vector genome lengths. Genome lengths shown represent single-stranded DNA lengths (e.g., sc CMV-GFP is $2 \times 2,058$ bases + $3 \times$ ITRs = 4,551 bases). Images are representative of 10 to 20 captured images. Scale bars are 50 nm.

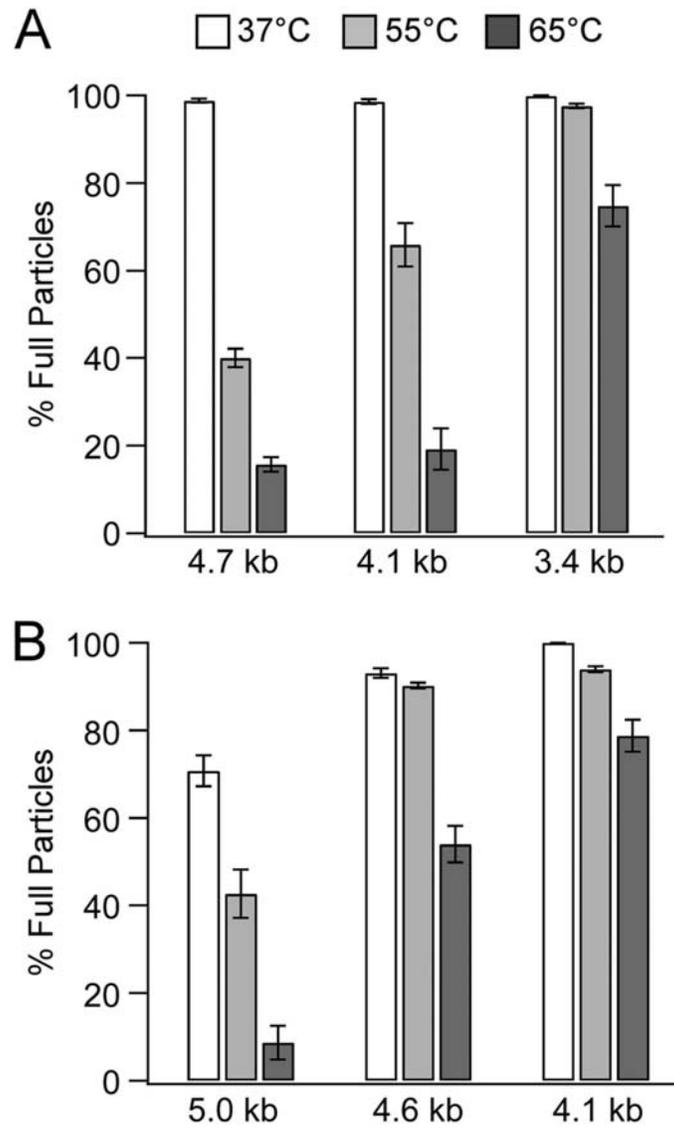


Figure 3.3 Quantitative analysis of TEM images. Thermally induced uncoating efficiency was assessed for intact-genome-containing (full) AAV particles packaging ssDNA (**A**) and scDNA (**B**) genomes of different length. Vectors were heated to 37, 55, or 65°C and visualized with TEM. Averages are taken from 10 to 20 images with particle counts for each sample ranging from 395 to 5,317 total particles. In all cases the average and standard deviation at each temperature was determined using the percentage of full particles from each image. Smaller capsid fragments (<50% of the shell) were excluded from counting. Genome lengths shown represent single-stranded DNA lengths. All values at 55°C and 65°C were determined to be significantly different from those at 37°C ($P < 0.001$ by a two-tailed Student t test).

scAAV vectors packaging a similar number of nucleotides are more thermostable than ssAAV vectors

To explore the phenomenon of AAV genome release further, we developed a more sensitive, fluorescence-based thermal melt assay. Briefly, while impermeable to large fluorophores in the native state, heated AAV capsids would release encapsidated DNA that is free to interact with SYBR gold. This dye is known to exhibit >1,000-fold enhancement in fluorescence upon binding ss/dsDNA, with an excitation maximum of ~495 nm and emission maximum of ~537 nm³⁸. To ensure that genome release is not rate limiting, the virus is held at each temperature for 5 min prior to reading the fluorescent signal. No large differences in fluorescence were observed between 5-min and 30-min heating periods (data not shown). At a critical temperature, we observed a sharp change in the fluorescence emission spectrum of AAV capsids packaging full-length genomes (Fig. 3.4A). When fluorescence intensities at the emission maxima were plotted as a function of temperature, we observed a sigmoidal profile with a characteristic thermal transition temperature (Fig. 3.4B). To further explore the applicability of this fluorimetric assay, we carried out similar studies evaluating thermally induced AAV genome release as a function of pH and osmotic pressure. As seen in Fig. 3.4C and D, AAV capsids appear to resist thermally induced genome release at lower pH and high external osmotic pressure. Next, we explored the phenomenon that self-complementary AAV (scAAV) vectors appeared to be more thermostable than ssAAV in earlier EM studies (Fig. 3.2). Representative plots of three different AAV vectors packaging ssDNA genomes of different lengths are shown in Fig. 3.5A. The transition temperature is defined as the temperature required for 50% of the AAV particles to release their preencapsidated genomes. As outlined above, since genome release is irreversible, this measurement can be used to obtain a pseudo- T_m (abbreviated as T_m) rather than a thermodynamically determined melting temperature. It is also important to note that although AAV virions are noted as full or empty in these low-resolution EM studies, it is possible that uncoated capsids and associated genomes exist in multiple states. Such a scenario is supported by previous studies demonstrating the existence of different packaged AAV genome states based on high-resolution cryo-EM studies³⁹.

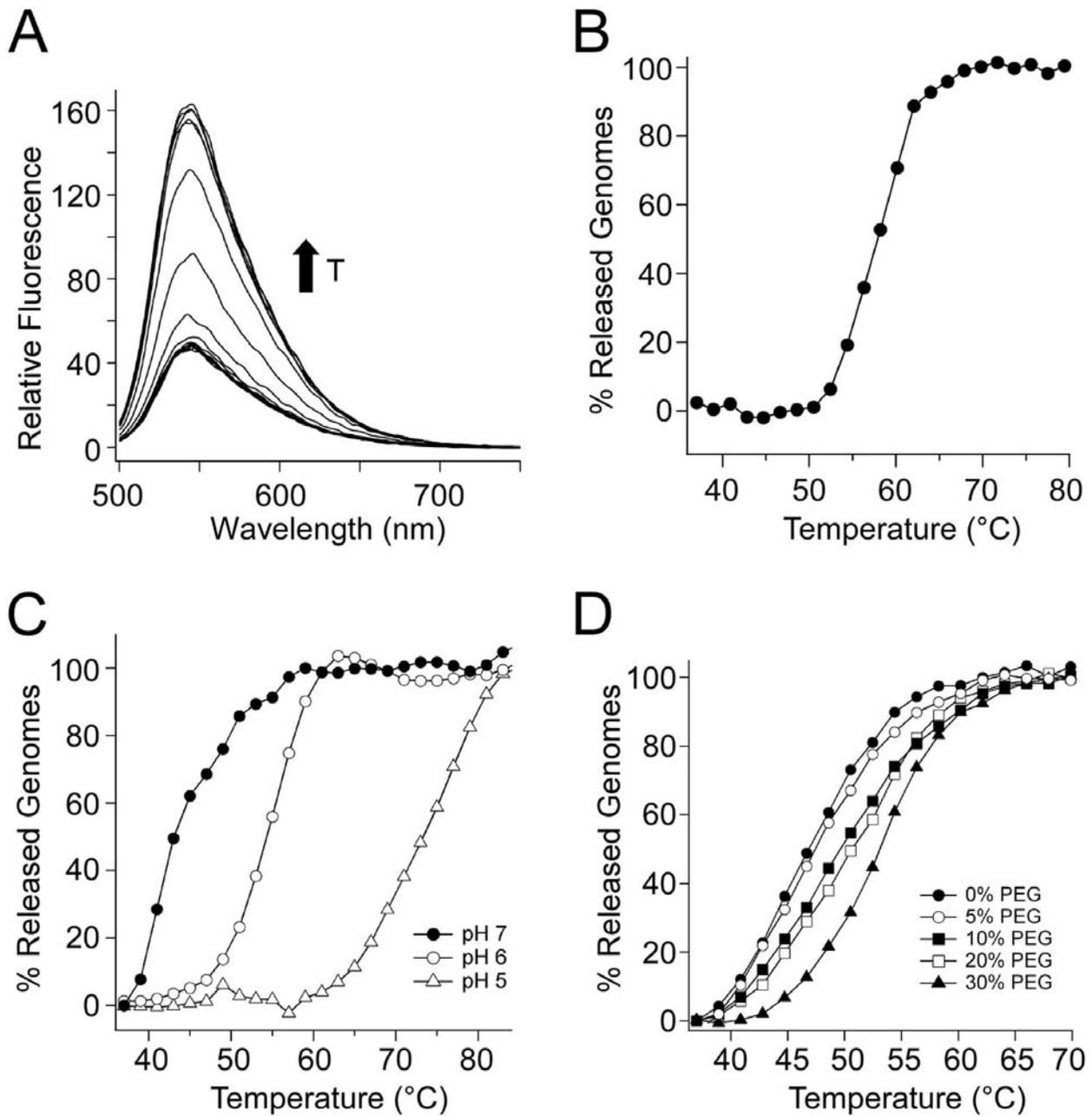


Figure 3.4 Fluorimetric detection of AAV genome release under different conditions. (A) In the presence of SYBR gold (25 μ M), AAV capsids were heated from 37°C to 95°C. At the point of genome release, an increase in fluorescence is observed at 550 nm (excitation, 495 nm), arising from the interaction between SYBR gold and the AAV genome. (B) When plotted and normalized to pretransition and posttransition baselines, a relatively sharp transition is observed in fluorescence at the emission wavelength. (C and D) Encapsidation stability is increased with decreasing pH (C) as well as increasing osmotic pressure (D). All experiments were repeated in triplicate.

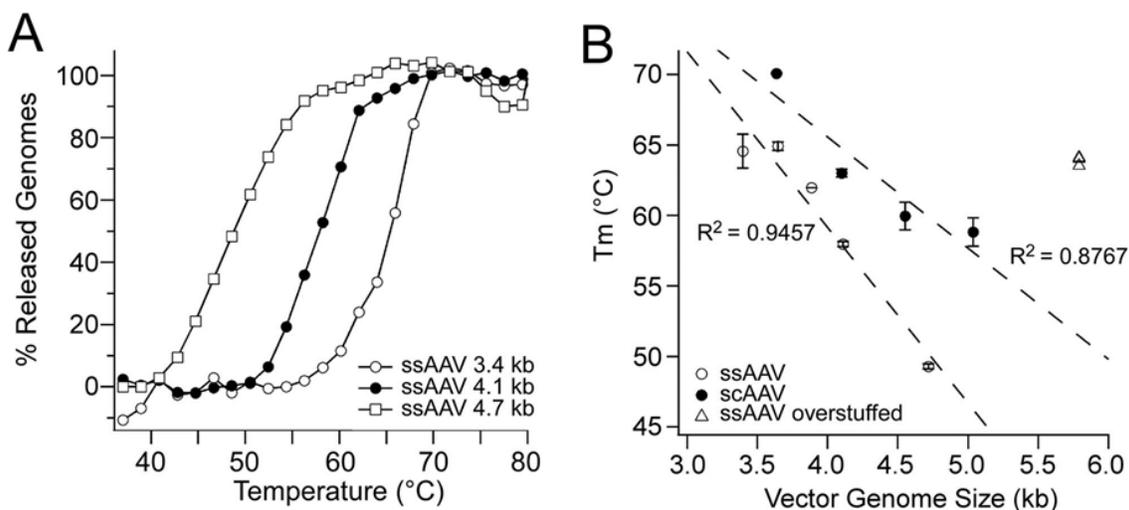


Figure 3.5 Fluorimetric analysis of AAV capsid uncoating as a function of genome length and self-complementarity. (A) Viral vectors were heated in 2°C increments and held for 5 min at each temperature prior to acquiring fluorescent signal. Three representative curves for different vector genome lengths are shown. Each vector was heated in triplicate along with a premelted control, which did not show a sharp transition. Curves were normalized to pre- and posttransition baselines to yield the percentage of released genomes. (B) Comparison of ssAAV, scAAV, and oversized ssAAV vectors shows an inverse linear correlation of T_m with vector genome size.

Thermal transitions for different ssAAV vectors acquired from the UNC vector core were then plotted as a function of genome length (Fig. 3.5B). A linear correlation is observed, where T_m is inversely proportional to genome length. These data correlate well with the TEM data shown above. A similar trend is seen with the panel of self-complementary AAV vectors (it should be noted that genome length is shown as total single-strand length in Fig. 3.5). We also evaluated the thermal profile of AAV vectors packaging oversized ss/scDNA genomes. As mentioned above, the packaging capacity of AAV vectors has been studied extensively. Although numerous reports have demonstrated successful transduction with vectors packaging oversized AAV genomes, the latter have been shown to be fragmented into subgenomic-length DNA^{12,14,15,17}. Therefore, it is not surprising that ssAAV vectors packaging a 5.8-kb genome exhibit a T_m similar to that of vectors packaging a smaller genome (3.4 kb). Surprisingly, scAAV vectors appear to be more thermally stable than ssAAV vectors in general. In addition, the thermal stability of scAAV vectors appears to be less sensitive to vector genome length than that of ssAAV vectors, as determined by the lower slope in the linear correlation (Fig. 3.5B).

Heat-induced exposure of VP1 N termini is not dependent on genome size or self-complementarity.

While the current studies are focused primarily on the effects of thermal stimuli on AAV genome release, heat-induced exposure of AAV VP1 N termini has been reported by several groups^{21,22}. Using a thermocycler, wild-type AAV, empty AAV capsids, and AAV packaging single-stranded or self-complementary CMV-GFP genomes were heated to temperatures ranging from 40 to 73°C. Samples were then subject to immunoblot analysis using the monoclonal mouse antibody A20, which recognizes a conformational epitope on the intact capsid, and the monoclonal mouse antibody A1, which recognizes a linear epitope in the N terminus of the VP1 capsid protein³⁰. As observed in Fig. 3.6, A20 staining is abrogated upon heating to temperatures of >60°C, indicating breakdown of intact capsids. Further, robust staining is observed at 60°C or higher, consistent with earlier studies by other groups. Interestingly, no significant difference was observed between samples regardless of full or empty particles or ssDNA or scDNA genomes. These results suggest that internal capsid pressure exerted by packaged AAV genomes might not play a direct role in externalization of VP1 N termini.

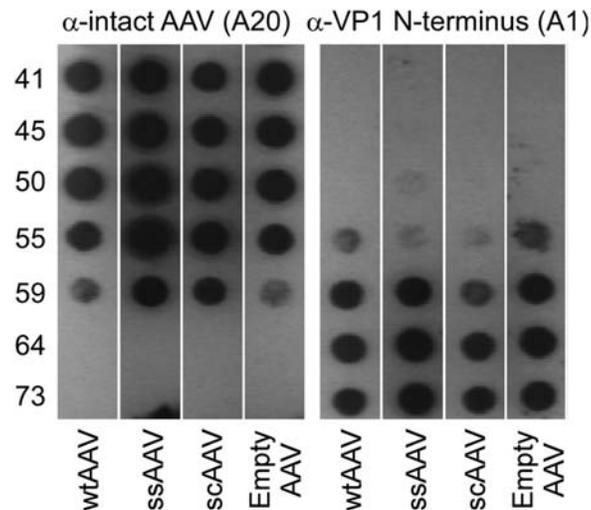


Figure 3.6 Exposure of the VP1 N termini is not dependent on genome length or self-complementarity. AAV capsids (5×10^9 vg/well) containing either the wild type (wtAAV), single-stranded CMV-GFP (ssAAV), self-complementary CMV-GFP (scAAV), or no genome at all were heated in a PCR mixture for 30 min. The monoclonal antibodies A20 and A1 were then used to probe the AAV capsid integrity as well as the state of exposure of the buried VP1 N-terminal domain required for infection. No significant difference was observed between genomes with different sizes or self-complementarity.

Tungsten-shadowing EM shows DNA secondary structure for dsDNA vectors

The intriguing difference in the thermal properties of ssAAV and scAAV vectors suggests that the nature of secondary structures formed within encapsidated viral DNA might affect capsid uncoating. To investigate this phenomenon further, we visualized AAV DNA after DNA release using a combination of transmission electron microscopy and tungsten shadowing (Fig. 3.7). After heating the ssDNA-packaging AAV vector to 65°C, the released viral genome was observed to collapse upon itself. This is likely due to kinetically driven base pairing within the ssDNA genome. In contrast, the scAAV genome clearly forms a dsDNA structure following release at 65°C. Furthermore, the mutated ITR located in the middle of the AAV genome is clearly observed. The thickness of the DNA in the images was consistent with the width of duplex DNA following coating with tungsten as measured in other studies³¹. Consistent with previous reports for other parvoviruses, the majority of viral DNA is observed associated with the capsid. These observations also suggest that a significant level of base pairing may exist in scAAV genomes prior to genome release.

Molecular dynamics provides insight into genome organization within the capsid

The persistence length of dsDNA is 50 nm (~147 bp), which corresponds to a minimum dsDNA ring size of approximately 16 nm in diameter⁴⁰. With a capsid shell that has an ~22-nm inner diameter, one would expect that packaging 2.3-kb dsDNA into the AAV capsid would exert a large amount of internal pressure within the capsid shell⁴¹. It is therefore counterintuitive that scAAV vectors are more thermally stable than ssAAV vectors. Additionally, the observation of dsDNA in electron micrographs after heating suggests that cDNA base pairing could occur within the capsid. We therefore used molecular dynamics simulations to gather insight into the nature of the DNA within the capsid.

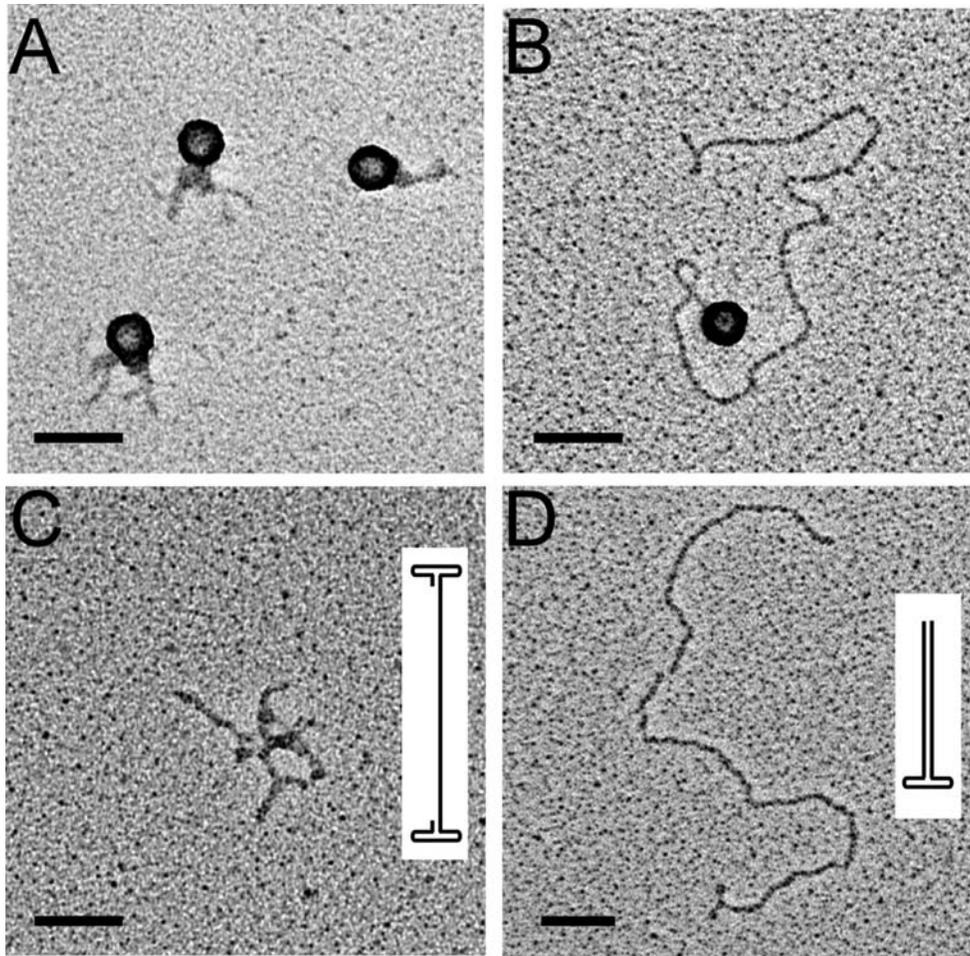


Figure 3.7 Tungsten-shadowing EM of released genomes from ssCMV-GFP (A and C) and scCMV-GFP (B and D) vectors. Viral particles were heated to 65°C and then imaged using tungsten-shadowing EM. Grids showed a combination of AAV virions shown as spheres, loose DNA, and virion-associated DNA. Scale bars represent 50 nm. Insets are cartoon representations of the predicted secondary structure.

In this model, the capsid is represented as a hollow sphere. Packaged DNA is represented as an elastic bead polymer chain of varying flexibility (high for ssDNA and low for dsDNA). Spheres containing ssDNA chains were reduced in diameter from 200 nm to 21.6 nm while performing molecular dynamics to sample conformational space and achieve thermal equilibrium before determining the capsid pressure, measured by the average total force exerted by the genome on the capsid walls divided by the surface area of the capsid. Encapsidated ssDNA chains were modified to have increasing degrees of double-stranded character by increasing the stiffness of the angle constraints between monomers, one at a time, simulating base pairing along an scAAV genome.

Counterintuitively, we find that as the angle stiffness is increased step by step along the confined chain, the pressure exerted by the chain on the capsid walls decreases until about 50% of the genome has been base paired, reaching a capsid pressure of 5.2 ± 0.1 atm (Fig. 3.8A). Further base pairing causes the capsid internal pressure to rise up to a maximum of 8.3 ± 0.7 atm. Visualizing the MD trajectory of the base pairing simulation provides insight into this phenomenon (Fig. 3.8B). Before base pairing begins, the ssDNA is largely disordered (blue licorice in Fig. 3.8B). The small bending energy penalty leaves the beads at the outer surface free to vibrate (due to thermal energy) against the capsid walls, creating pressure. As the chain is base paired by increasing the stiffness sequentially along the length of the chain, the stiffened segment attempts to reduce its bending energy by moving away from the capsid center, forming concentric rings near the surface of the containing sphere (red beads in Fig. 3.8B, panel ii). Unlike the flexible ssDNA chain, these dsDNA rings are not able to vibrate freely due to their increased stiffness, and thus they exert less pressure on the walls. As the degree of base pairing approaches 50%, the surface of the DNA sphere is almost completely covered by these stiff concentric rings, which stack to form a tight shell around the core of flexible ssDNA, damping their vibrations against the capsid walls (Fig. 3.8B, panel iii). As base pairing continues along the chain, a second inner layer of concentric rings begins to form, which pushes the outer layer further toward the capsid wall (Fig. 3.8B, panels iv and v). As the degree of base pairing increases, the force from this inner layer transmitted through the outer shell to the capsid increases, causing the capsid pressure to rise.

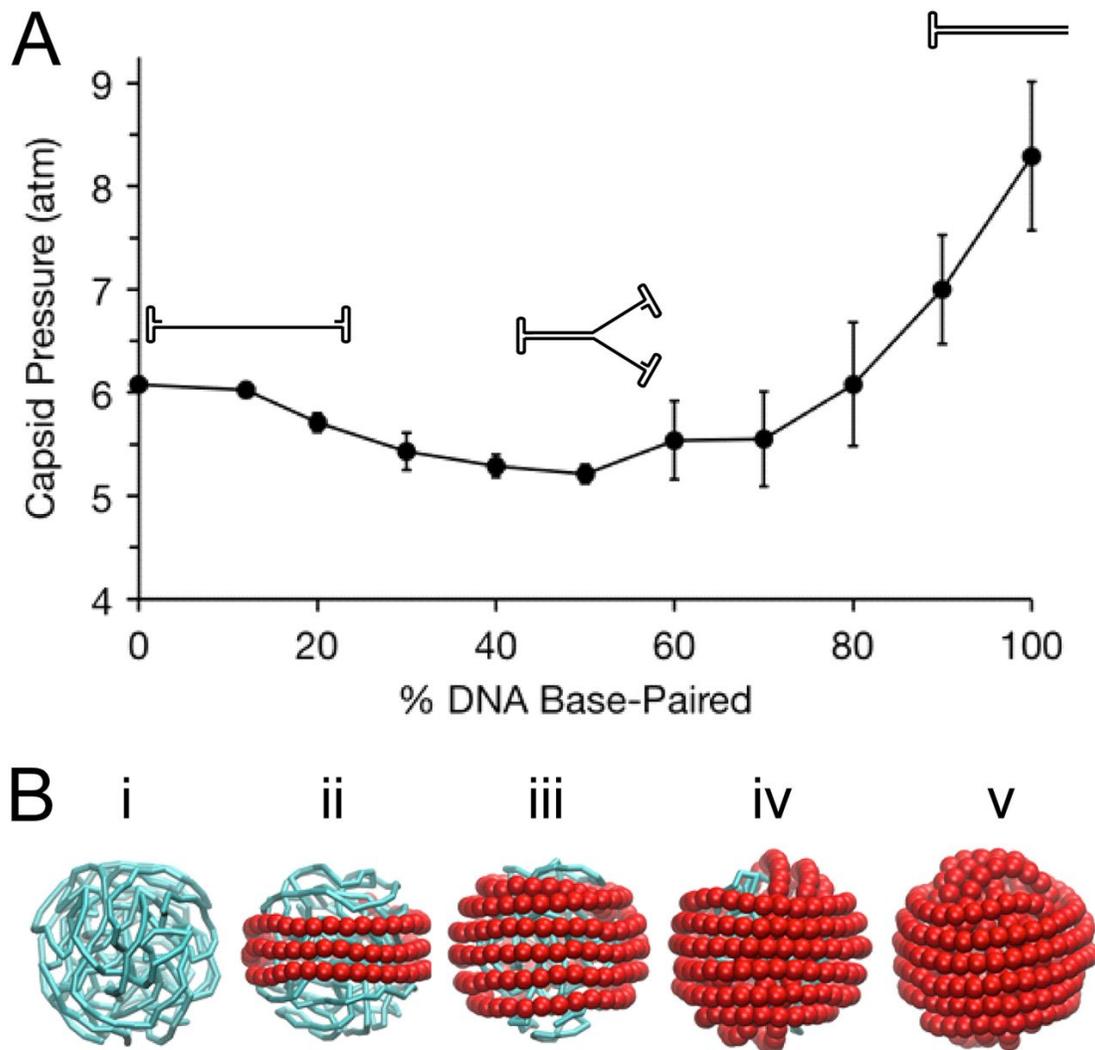


Figure 3.8 Computational modeling of internal capsid pressure and genome organization. (A) In molecular dynamics simulations, the stiffness of modeled ssDNA chains was increased step by step along the chain to simulate DNA base pairing in scAAV genomes. Pressure exerted on the capsid walls decreases until 50% of the chain has been converted, after which the capsid pressure increases. Error bars represent standard deviations ($n = 5$). Cartoon representations of predicted secondary structures of AAV genomes at different levels of base-pairing are shown above the curve. (B) Molecular dynamics simulation of progressive change in the organization of flexible ssDNA chains (cyan licorice) while being base paired into stiff dsDNA chains (red beads). Representative snapshots of chain organization within the model capsid at 0% (i), 25% (ii), 50% (iii), 75% (iv), and 100% (v) conversion of ssDNA to dsDNA are shown. dsDNA segments are seen to migrate to the periphery and adjacent to the inner capsid surface, where they may buffer the thermal motion of the flexible ssDNA chain and reduce the pressure exerted on the inner capsid walls during the initial stages of the conversion process. The development of an inner coil of dsDNA after 50% conversion may subsequently contribute to increasing capsid pressures.

We note that several simplifications have been made in constructing the model of the scAAV genome, most notably the omission of electrostatic interaction terms from the force field. Simulations of DNA packaging in bacteriophages in which electrostatics were included showed that DNA-DNA repulsions accounted for up to 51% of the total free energy cost of packaging, suggesting that these forces dominate when the genome is highly confined⁴². However, we also note that the genome conformations resulting from these simulations were topologically indistinguishable from those in similar systems where electrostatics was not accounted for⁴³. The electrostatic interaction energy, although a significant component of the energetics of packaging, depends primarily on the total DNA density and is largely agnostic to conformational details, which are dictated by the elastic bending terms in the force field. Since DNA bending is accurately represented in our model, we believe that, even in the absence of electrostatics, our simplified model correctly predicts the qualitative topology of the scAAV genome and its influence on the measured capsid pressure.

EM quantification of DNA release following thermal denaturation

Quantification of the adherence of AAV capsid protein to the grids used for our EM characterization revealed that scAAV adhered at consistently higher abundance than ssAAV (Figure 3.9A). This difference in capsid protein abundance may simply reflect the fact that AAV concentration was determined via a PCR based assay that quantified the presence of the viral genomic DNA. Regardless, the adherence of both scAAV and ssAAV was increased following mild thermal denaturation at 55 °C. However, denaturation at higher temperatures appeared to reduce capsid adherence. This may be due to disruption of the capsid structure, aggregation of the capsid proteins following denaturation or other unknown confounding factors. Thermal denaturation results in the release of viral genomes from both scAAV and ssAAV (Figure 3.9B). Denaturation at increasing temperatures results in the release of increasing amounts of both single- and double-stranded genomes, through the binding of single- stranded genomes did appear reduced when denatured at 65 °C versus 60°C. Following thermal denaturation a

large portion of the genomes released remained associated with capsid protein (Figure 3.9C), approaching 100% at the highest tested denaturation temperature.

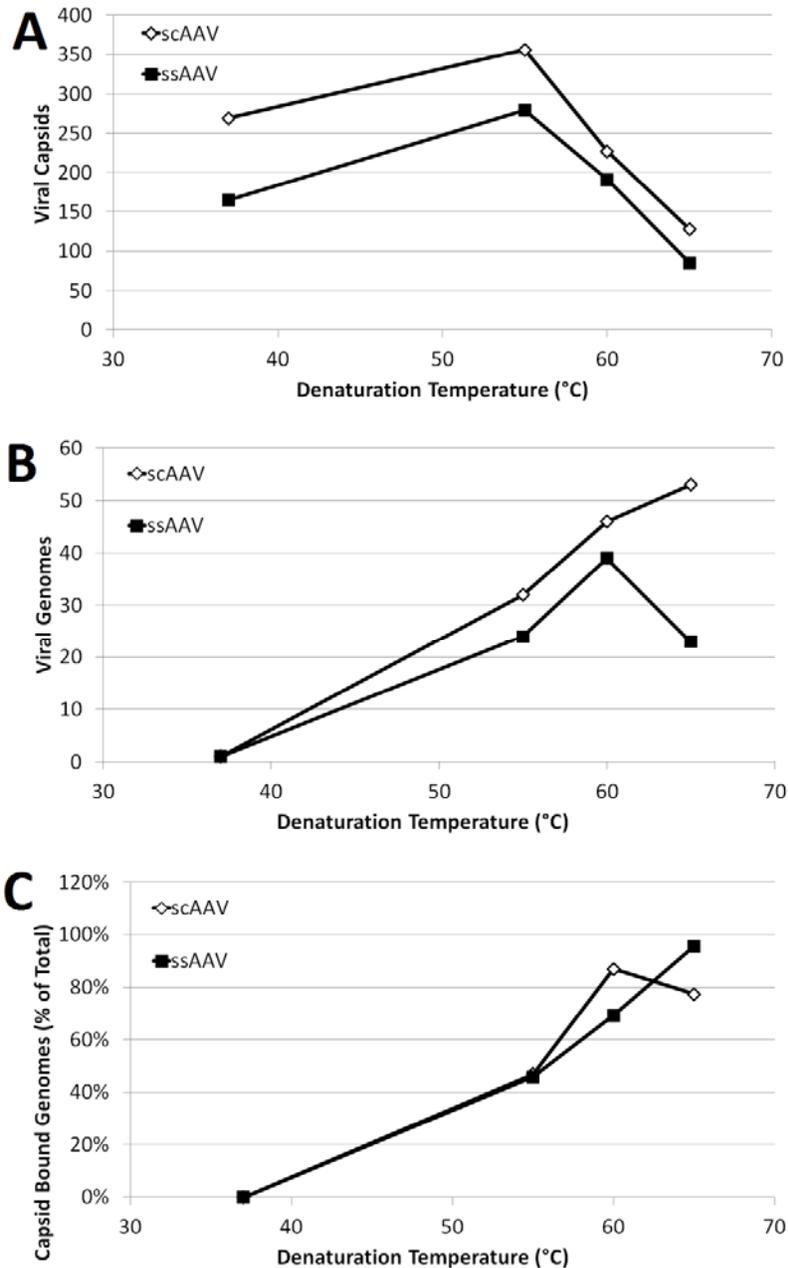


Figure 3.9: Characteristics of capsid adherence and genome release following thermal denaturation. (A) ssAAV and scAAV capsids adhere to EM grids in comparable numbers, and this adherence is dependent upon denaturation temperature. **(B)** scAAV and ssAAV genomes are released from their capsids by thermal denaturation. **(C)** scAAV and ssAAV genomes remain capsid-bound following release from the capsid in comparable abundance.

A careful examination dsDNA genomes following thermal denaturation revealed that they adopted a discreet set of structured (Figure 3.10A-E). The ends were either AAV Bound (Figure 3.10A) or free (Figure 3.10A), or they possessed a secondary (2') structure comparable to that expected at the inverted terminal repeats (Figure 3.10C), or a looped structure (Figure 3.10D) or a looped structure with a dsDNA tail (Figure 3.10E). Of genomes possessing two classifiable ends, the plurality of ends were of a free conformation, regardless of thermal denaturation, with a substantial minority of ends possessing the 2' structure (Figure 3.10F). Following denaturation increasing numbers of ends were observed to be AAV bound, with his minority overtaking the ends possessing the 2' structure. The proportion of genomes possessing particular end-structure combinations was affected by temperature. Prior to denaturation the plurality of genomes possessed a free end and an end with the expected 2' structure, with a sizable minority possessing two free ends (Figure 3.10G). Following thermal denaturation there was a profound increase in the abundance of AAV bound genomes possessing a free, and a concomitant slight increase in AAV bound genomes possessing an end with the expected 2' structure and a decrease in the abundance of other species. Analysis of the length of these genomes did not reveal any clear trends, whether when classified by end structure (Figure 3.10H) or end-structure combination (Figure 3.10I). However, these length-analysis data were possessed of considerable variability (not shown), which hinders meaningful comparisons.

Attempts at characterizing intra-capsid structure by EM

Incubating a solution of scAAV particles for 20 minutes at 90 °C then gradually cooling it to 0 °C produces a solution of proteins and annealed scAAV genomes that are amenable for examination using transmission electron microscopy (TEM). 4'-aminomethyltrioxysalen (psoralen) is an intercalating agent that can form a thymidine diadduct at TA sequences in double-stranded DNA upon exposure to long-wave ultraviolet light (UV-A), forming a stable interstrand crosslink. Treatment of intact scAAV particles with psoralen and UVA (PUVA) is expected to extensively crosslink the scAAV genome only if the genome is double stranded inside the capsid. Extremely-thermostable single stranded DNA binding protein (ET

SSB) is able to stably bind to single-stranded DNA at temperatures of up to 100°C, but has low affinity for double-stranded DNA. ET SSB is expected to fully coat the scAAV genome following PUVA treatment only if the genome is not extensively crosslinked. *Faux*-PUVA treated (0 µg/ml Psoralen) scAAV genomes appear as double-stranded DNA when incubated in the absence of ET SSB (Figure 3.11A) and ET SSB coated filaments when incubated with ET SSB (Figure 3.11B). scAAV genomes PUVA treated with either 25 µg/ml (Figure 3.11C) or 250 µg/ml (Figure 3.11D) psoralen show no appreciable difference from the *Faux*-PUVA treated scAAV genomes.

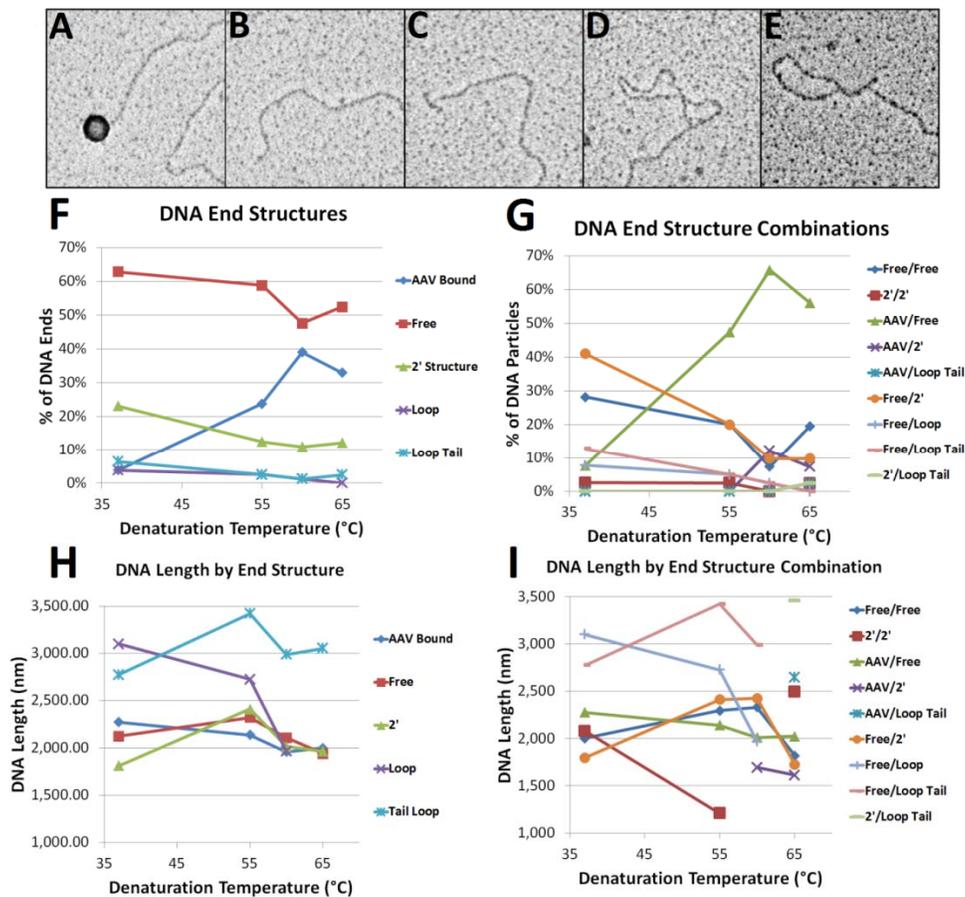


Figure 3.10: EM analysis of end structure and genome length. Genome end structures consist of AAV bound ends (A), free ends (B), ends with 2' structure consistent with an ITR (C), a loop structure (D) or a loop structure with a dsDNA fail (E). The relative abundance of these ends structures changes following thermal denaturation (F), and the relative abundance of genomes possessing particular combinations of these end structures also changes following thermal denaturation (G). The length of genomes possessing these end structures (H) or particular combinations of end structures (I) do not show clear differences following thermal denaturation.

The lack of extensive crosslinking after PUVA treatment may be due to a number of factors. The scAAV genome may be packaged as single-stranded DNA (ssDNA) in the capsid, the capsid may be impermeable to psoralen, or the PUVA treatment may have failed for technical reasons. That scAAV may be packaged as ssDNA is plausible; natural AAV genomes are single-stranded². However, psoralen potentiates UV inactivation of AAV, perhaps suggesting both that the genome is capable of limited base pairing in the capsid and that capsid is permeable to psoralen⁵. Future studies may test the success of PUVA treatment by crosslinking a sample of heat-denatured and annealed scAAV as a positive control. Ultimately the results of this preliminary examination do not contradict the model that scAAV genomes are packaged into AAV capsids as ssDNA, but its design lacks a control needed to generate results that could actually support that model.

DISCUSSION

Parvoviruses such as minute virus of mice (MVM) have been shown to release their ssDNA genomes in the 3'-to-5' direction, while leaving the empty virus shell intact¹⁸. Our studies indicate that AAV capsids respond in a similar manner to thermal stimuli. We observed different capsid fragments and morphologies that suggest that empty AAV shells might subsequently unravel from the site of genome release, presumably by loss of the viral proteins comprising the virion shell. Such capsid disassembly intermediates are not observed upon heating purified empty AAV particles (data not shown), which suggests that capsid disassembly might not precede genome release. Interestingly, heat-induced exposure of AAV VP1 N termini has also been reported by several groups^{21,22}. In the current study, both VP1 exposure and genome release were observed to occur upon heating capsids to approximately 60°C or higher (Fig. 3.4 and 6). Paradoxically, we observed that heat-induced exposure of VP1 N termini does not vary significantly between wild-type AAV, empty AAV virion shells, and representative ssAAV or scAAV vectors (Fig. 3.6). Based on these results, one possible scenario is that parvoviral VP1 exposure and genome release might be structurally distinct events with a certain degree of temporal overlap, as proposed earlier¹⁸. It is important to note that despite our experiments providing insight into biophysical

aspects of AAV capsid uncoating, their biological implications remain to be established. Nevertheless, the results underscore the importance of understanding capsid structural transitions that precede VP1 exposure and genome release as well as spatiotemporal aspects of such events in the AAV infectious pathway.

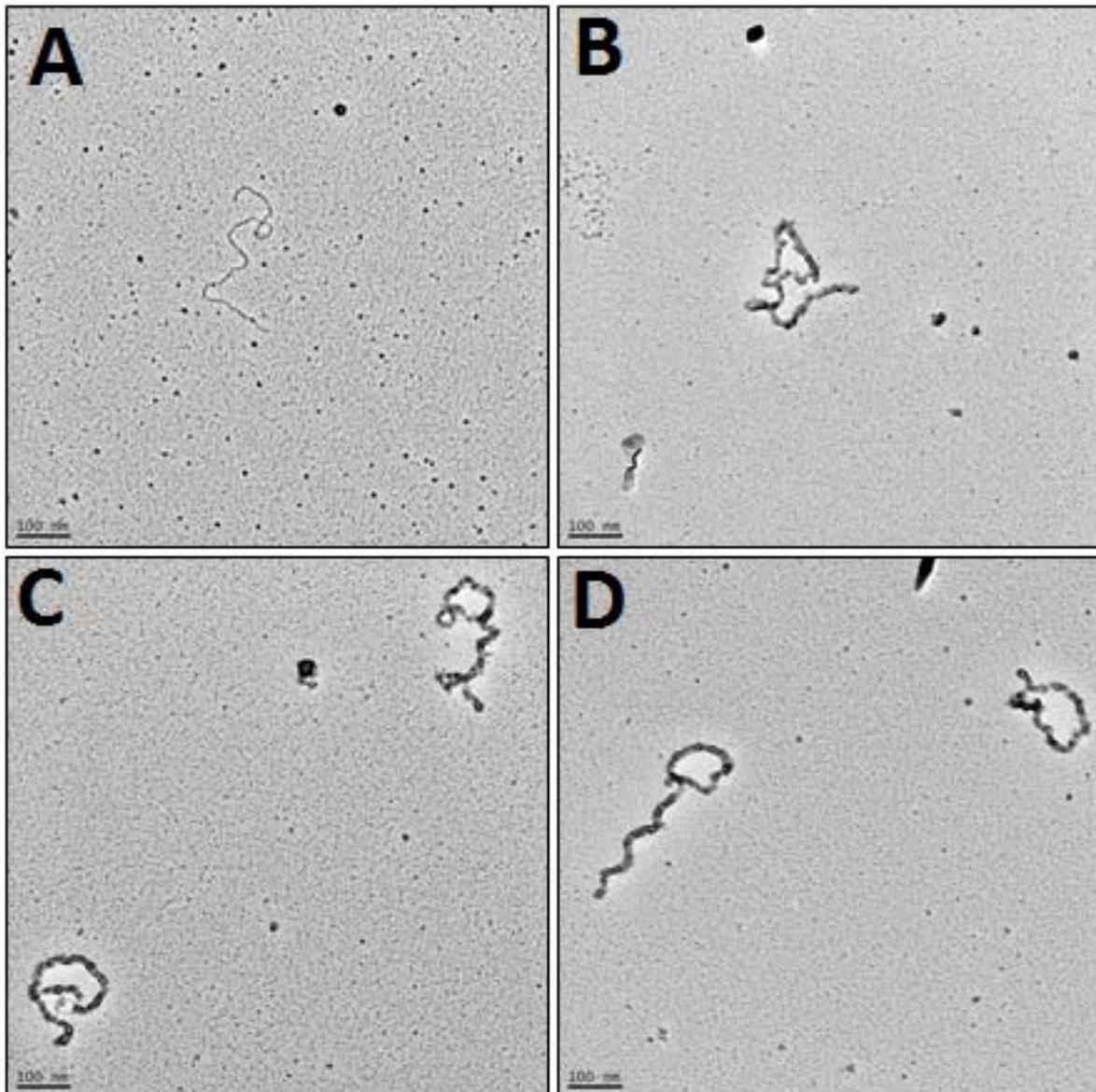


Figure 3.11: Psoralen fails to crosslinks scAAV genomes in intact capsids. (A) an scAAV genome. (B) an scAAV genome incubated in the presence of ET SSB shows ET SSB binding. Psoralen crosslinking with 25 $\mu\text{g/ml}$ (C) or 250 $\mu\text{g/ml}$ psoralen (D) do not affect ET SSB binding to the scAAV genome.

Using a combination of electron microscopy and fluorimetric assays, we observed a trend wherein subgenomic AAV particles require more thermal energy to release their genomes, regardless of whether they are packaging ssDNA or scDNA. In contrast as genome sizes approach wild-type AAV genome length (~4.7 kb), the capsids appear to acquire an optimal, metastable stable state that can efficiently uncoat. Packaged DNA genomes within bacteriophages can generate pressures of close to 50 atm^{41,44,45}, requiring only minimal stimuli for DNA ejection, but these pressures rapidly drop off as the length of phage DNA is decreased to 78% of wild type⁴¹. One might therefore expect that AAV vector genomes of different subgenomic size would greatly change the internal pressure of the capsid. Thus, it is likely that subgenomic AAV particles are defective in capsid uncoating and genome release within the host cell. Accordingly, incorporation of stuffer DNA and packaging of near-wild-type genomes into AAV capsids might improve infectivity. Further, we also observed that acidic pH and external osmotic pressures exerted by ~30% PEG 8000 solutions inhibited AAV capsid genome release. One possible explanation for the stabilizing effect of low pH on AAV capsids might be increased interactions between packaged DNA and positively charged histidine side chains within the capsid interior. Additionally, we surmise that the observed increase in capsid stability and genome retention at acidic pH may be beneficial from a physiological standpoint. Enhanced capsid stability within endosomal/lysosomal compartments during intracellular trafficking could help protect viral DNA from degradation by nucleases prior to release into the nucleus. The osmotic pressure effect is corroborated by earlier studies with bacteriophage capsids, wherein genome ejection was observed to be inhibited by pressures comparable to that of the cytoplasm of bacteria⁴⁶. Another interesting observation was that AAV vectors packaging genomes larger than wild type exhibit stability equivalent to that of subgenomic particles. This result can be explained by earlier studies demonstrating that packaged genomes for oversized AAV vectors are truncated^{12,14,15,17}.

Lastly, the observation that scAAV vectors are more thermally stable than ssAAV vectors was unexpected. Since dsDNA is inherently stiffer than ssDNA, one would expect packaging a fully base-paired dsDNA genome into an AAV capsid to be energetically unfavorable. Our results from molecular modeling studies support the notion that internal pressures associated with capsids packaging dsDNA

genomes could indeed be higher than those associated with ssDNA-packaging vectors. However, closer analysis of the modeling data revealed that a hybrid AAV genome comprised of ssDNA and dsDNA (50% base pairing) would exert an internal capsid pressure that is lower than that with 100% ssDNA alone. These theoretical predictions are supported indirectly by tungsten-shadowing EM studies, wherein we observed predominantly dsDNA character in released scAAV genomes. In contrast, the ssAAV genome was observed to form a disordered structure consistent with folding of DNA into a local thermodynamic minimum. Therefore, it is tempting to speculate that partially base-paired scAAV genomes might nucleate rapid annealing upon release to form a complete dsDNA structure in the host cell. Although these results suggest that scAAV vectors might be more resistant to uncoating than ssAAV vectors, this apparent defect is likely offset by bypassing second-strand synthesis, which results in rapid onset of transgene expression. In summary, these studies provide a path forward toward characterization of AAV capsid uncoating and genome release in a more relevant physiological setting. When integrated with intracellular trafficking studies and high-resolution structural data, it is possible that the results described here might provide a more complete picture of AAV infection and possibly guide the design of optimal AAV vector genomes for gene therapy applications.

REFERENCES

- [1] Bowles DE, Rabinowitz JE, Samulski RJ. 2006. *The genus Dependovirus*, p 15–24 In Kerr JR, Cotmore SF, Bloom ME, Linden RM, Parrish CR, editors. (ed), Parvoviruses. Hodder Arnold Publishers, London, United Kingdom
- [2] Samulski RJ, Srivastava A, Berns KI, Muzyczka N. 1983. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* 33:135–143
- [3] Xiao X, Xiao W, Li J, Samulski RJ. 1997. A novel 165-base-pair terminal repeat sequence is the sole cis requirement for the adeno-associated virus life cycle. *J. Virol.* 71:941–948
- [4] Agbandje-McKenna M, Kleinschmidt J. 2011. AAV capsid structure and cell interactions. *Methods Mol. Biol.* 807:47–92
- [5] Mingozzi F, High KA. 2011. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nat. Rev. Genet.* 12:341–355
- [6] Ding W, Zhang L, Yan Z, Engelhardt JF. 2005. Intracellular trafficking of adeno-associated viral vectors. *Gene Ther.* 12:873–880
- [7] Nonnenmacher M, Weber T. 2012. Intracellular transport of recombinant adeno-associated virus vectors. *Gene Ther.* 19:649–658
- [8] McCarty DM. 2008. Self-complementary AAV vectors; advances and applications. *Mol. Ther.* 16:1648–1656
- [9] Hauck B, Zhao W, High K, Xiao W. 2004. Intracellular viral processing, not single-stranded DNA accumulation, is crucial for recombinant adeno-associated virus transduction. *J. Virol.* 78:13678–13686
- [10] Ferrari FK, Samulski T, Shenk T, Samulski RJ. 1996. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol.* 70:3227–3234
- [11] Fisher KJ, Gao GP, Weitzman MD, DeMatteo R, Burda JF, Wilson JM. 1996. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. *J. Virol.* 70:520–532
- [12] Dong B, Nakai H, Xiao W. 2010. Characterization of genome integrity for oversized recombinant AAV vector. *Mol. Ther.* 18:87–92
- [13] Grieger JC, Samulski RJ. 2005. Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. *J. Virol.* 79:9933–9944
- [14] Hirsch ML, Agbandje-McKenna M, Samulski RJ. 2010. Little vector, big gene transduction: fragmented genome reassembly of adeno-associated virus. *Mol. Ther.* 18:6–8
- [15] Lai Y, Yue Y, Duan D. 2010. Evidence for the failure of adeno-associated virus serotype 5 to package a viral genome \geq 8.2 kb. *Mol. Ther.* 18:75–79

- [16] Wu J, Zhao W, Zhong L, Han Z, Li B, Ma W, Weigel-Kelley KA, Warrington KH, Srivastava A. 2007. Self-complementary recombinant adeno-associated viral vectors: packaging capacity and the role of rep proteins in vector purity. *Hum. Gene Ther.* 18:171–182
- [17] Wu Z, Yang H, Colosi P. 2010. Effect of genome size on AAV vector packaging. *Mol. Ther.* 18:80–86
- [18] Cotmore SF, Hafenstein S, Tattersall P. 2010. Depletion of virion-associated divalent cations induces parvovirus minute virus of mice to eject its genome in a 3'-to-5' direction from an otherwise intact viral particle. *J. Virol.* 84:1945–1956
- [19] Carrasco C, Carreira A, Schaap IA, Serena PA, Gomez-Herrero J, Mateu MG, de Pablo PJ. 2006. DNA-mediated anisotropic mechanical reinforcement of a virus. *Proc. Natl. Acad. Sci. U. S. A.* 103:13706–13711
- [20] Horowitz ED, Finn MG, Asokan A. 2012. Tyrosine cross-linking reveals interfacial dynamics in adeno-associated viral capsids during infection. *ACS Chem. Biol.* 7:1059–1066
- [21] Johnson JS, Li C, DiPrimio N, Weinberg MS, McCown TJ, Samulski RJ. 2010. Mutagenesis of adeno-associated virus type 2 capsid protein VP1 uncovers new roles for basic amino acids in trafficking and cell-specific transduction. *J. Virol.* 84:8888–8902
- [22] Kronenberg S, Bottcher B, von der Lieth CW, Bleker S, Kleinschmidt JA. 2005. A conformational change in the adeno-associated virus type 2 capsid leads to the exposure of hidden VP1 N termini. *J. Virol.* 79:5296–5303
- [23] Grieger JC, Choi VW, Samulski RJ. 2006. Production and characterization of adeno-associated viral vectors. *Nat. Protoc.* 1:1412–1428
- [24] Samulski RJ, Chang LS, Shenk T. 1989. Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *J. Virol.* 63:3822–3828
- [25] Pulicherla N, Shen S, Yadav S, Debbink K, Govindasamy L, Agbandje-McKenna M, Asokan A. 2011. Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. *Mol. Ther.* 19:1070–1078
- [26] DiPrimio N, Asokan A, Govindasamy L, Agbandje-McKenna M, Samulski RJ. 2008. Surface loop dynamics in adeno-associated virus capsid assembly. *J. Virol.* 82:5178–5189
- [27] Gradinaru V, Zhang F, Ramakrishnan C, Mattis J, Prakash R, Diester I, Goshen I, Thompson KR, Deisseroth K. 2010. Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* 141:154–165
- [28] Gray SJ, Foti SB, Schwartz JW, Bachaboina L, Taylor-Blake B, Coleman J, Ehlers MD, Zylka MJ, McCown TJ, Samulski RJ. 2011. Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. *Hum. Gene Ther.* 22:1143–1153
- [29] Torikai K, Ito M, Jordan LE, Mayor HD. 1970. Properties of light particles produced during growth of type 4 adeno-associated satellite virus. *J. Virol.* 6:363–369

- [30] Wobus CE, Hugle-Dorr B, Girod A, Petersen G, Hallek M, Kleinschmidt JA. 2000. Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection. *J. Virol.* 74:9281–9293
- [31] Compton SA, Ozgur S, Griffith JD. 2010. Ring-shaped Rad51 paralog protein complexes bind Holliday junctions and replication forks as visualized by electron microscopy. *J. Biol. Chem.* 285:13349–13356
- [32] Griffith JD, Lee S, Wang YH. 1997. Visualizing nucleic acids and their complexes using electron microscopy. *Curr. Opin. Struct. Biol.* 7:362–366
- [33] Locker CR, Harvey SC. 2006. A model for viral genome packing. *Multiscale Modeling Simulation* 5:1264–1279
- [34] Cerritelli ME, Cheng N, Rosenberg AH, McPherson CE, Booy FP, Steven AC. 1997. Encapsidated conformation of bacteriophage T7 DNA. *Cell* 91:271–280
- [35] Xie Q, Bu W, Bhatia S, Hare J, Somasundaram T, Azzi A, Chapman MS. 2002. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* 99:10405–10410
- [36] Plimpton S. 1995. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 117:1–19
- [37] Nosé S. 1984. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52:255–268
- [38] Tuma RS, Beaudet MP, Jin X, Jones LJ, Cheung CY, Yue S, Singer VL. 1999. Characterization of SYBR gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Anal. Biochem.* 268:278–288
- [39] Gerlach B, Kleinschmidt JA, Bottcher B. 2011. Conformational changes in adeno-associated virus type 1 induced by genome packaging. *J. Mol. Biol.* 409:427–438
- [40] Theodorakopoulos N, Peyrard M. 2012. Base pair openings and temperature dependence of DNA flexibility. *Phys. Rev. Lett.* 108:078104.
- [41] Ivanovska I, Wuite G, Jonsson B, Evilevitch A. 2007. Internal DNA pressure modifies stability of WT phage. *Proc. Natl. Acad. Sci. U. S. A.* 104:9603–9608
- [42] Petrov AS, Harvey SC. 2008. Packaging double-helical DNA into viral capsids: structures, forces, and energetics. *Biophys. J.* 95:497–502
- [43] Locker CR, Fuller SD, Harvey SC. 2007. DNA organization and thermodynamics during viral packing. *Biophys. J.* 93:2861–2869
- [44] Purohit PK, Inamdar MM, Grayson PD, Squires TM, Kondev J, Phillips R. 2005. Forces during bacteriophage DNA packaging and ejection. *Biophys. J.* 88:851–866

- [45] Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C. 2001. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* 413:748–752
- [46] Evilevitch A, Lavelle L, Knobler CM, Raspaud E, Gelbart WM. 2003. Osmotic pressure inhibition of DNA ejection from phage. *Proc. Natl. Acad. Sci. U. S. A.* 100:9292–9295
- [47] Horowitz, E. D., Rahman, K. S., Bower, B. D., Dismuke, D. J., Falvo, M. R., Griffith, J. D., Harvey, S. C., and Asokan, A. (2013) Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release, *J Virol* 87, 2994-3002.

Chapter 4: A Guanosine-Centric Mechanism for RNA Chaperone Function¹

Introduction

Outside the cellular environment or in the absence of chaperone proteins, most RNAs fold via complex pathways involving multiple, long-lived intermediates. RNA chaperone proteins with non- or semispecific RNA binding activities accelerate adoption of the thermodynamically most stable RNA structure by lowering the energetic barriers between RNA states and by facilitating rearrangement of misfolded states¹⁻⁴. Retroviruses package two RNA genomes in each virus particle⁵. These genomes dimerize near their 5' ends, and dimerization is catalyzed by an RNA chaperone, nucleocapsid (NC), which is derived from the retroviral Gag protein that coassembles with the viral RNA to generate replication-competent virus^{2,6,7}. By following the dimerization of a region of the Moloney murine leukemia virus (MuLV) genomic RNA at single-nucleotide resolution, we uncovered a simple mechanism for how a retroviral nucleocapsid chaperone protein functions.

MATERIALS AND METHODS

Retroviral RNA transcripts

The MuLV RNA fragment analyzed here is 331 nts and spans the dimerization and packaging regulatory domain (~170 nts). The RNA includes 5' and 3' flanking sequences of 46 and 115 nucleotides that facilitate primer extension analysis of SHAPE adducts. This RNA has a structure similar to the same region in authentic genomic RNA isolated from virions^{11,12}. The transcript was synthesized and purified

¹ The following chapter describes work done in collaboration with Dr. Jacob Grohman, Mr. Colin Lickwar, Dr. Jason Lieb, Dr. Brent Znosko and Dr. Kevin Weeks. I significantly contributed the purification of the hnRNP A1 used in the experiments reported below³³.

as described¹². For the inosine-substituted RNA, transcription was performed in the presence of inosine triphosphate instead of guanosine triphosphate.

NC and UP1 proteins

The MuLV retroviral nucleocapsid protein was purified as described²⁴. The UP1 open reading frame was amplified by PCR from a vector containing the hnRNP A1 cDNA (GenBank NM002136) and cloned into the Bam HI and Not I sites of pET28a (Invitrogen) to yield an Nterminal (His)₆-tagged version. UP1 was expressed in BL21(DE3)pLysS *E. coli* cells (Invitrogen) in 2 L LB using standard approaches. Cells were recovered by centrifugation and resuspended in 100 ml of 10 mM Tris-HCl (pH 7.4), 5 mM β -mercaptoethanol, 200 mM NaCl (supplemented with EDTA-free protease inhibitors; Roche) and lysed by incubation with 1 mg/ml egg-white lysozyme (20 min at 4 °C) and 5 min of sonication (Branson Sonifier 450, 75% duty cycle, tip setting of 6, at 0 °C). This lysate was separated via centrifugation (Beckman SW28 rotor at 131,000 $\times g$, 1 hr at 4 °C), and the (His)₆-UP1 protein was affinity purified using 1.0 ml Ni-NTA agarose beads (Qiagen; equilibrated in resuspension buffer overnight at 4 °C). Following three 50 ml washes [in 10 mM Tris-HCl (pH 7.4), 500 mM NaCl, 5 mM β -mercaptoethanol, 5% (v/v) glycerol, 25 mM imidazole], UP1 was eluted from the beads with 10 ml of 10 mM Tris-HCl, 150 mM NaCl, 250 mM imidazole (pH 7.5). The eluate was dialyzed against 4 L 20 mM HEPES (pH 7.4), 0.2 mM EDTA, 10% (v/v) glycerol, 1 mM DTT, 0.1 M KCl overnight at 4 °C using a 10-kDa cutoff dialysis cassette (Thermo Scientific). The protein was quantified by measurement of 280 nm absorbance; aliquoted; flash frozen in liquid nitrogen; and stored at -80°C until use.

Time-resolved SHAPE

MuLV domain RNAs (120 and 40 pmol for the 600 nM and 200 nM reactions, respectively) were renatured in RNase-free water by heating at 95 °C for 3 min in a total volume of 150 μ L and cooled on ice for 3 min. The resulting monomers were then equilibrated at 37 °C for 3 min by adding 40 μ L of a 5 \times folding buffer, omitting magnesium ion [50 mM HEPES (pH 7.5), 200 mM potassium acetate (pH 7.5)].

A pre-reaction (0 sec) time point was taken before magnesium ion addition to capture the initial monomer state. Time-resolved SHAPE experiments were then initiated by the addition of MgCl₂ to 5 mM (using 10 μL at 100 mM) at 37 °C, and the RNA structure was interrogated over a time course spanning 7 sec to 1 hr. For each time point, SHAPE modification was performed by adding 9 μL (5.4 and 1.8 pmol for the 600 and 200 nM reactions respectively) of the evolving MuLV RNA reaction to 1 μL BzCN (200 mM in DMSO) at 25 °C; immediate vigorous pipetting ensured mixing. Reaction with BzCN is complete within ~1 sec (13). No-reaction controls used neat DMSO. RNA was recovered by ethanol precipitation [2.5 vol ethanol, 1 μL glycogen (20 mg/mL), incubation at -20 °C for 60 min, and centrifugation at 20,000 ×g]. Pellets were resuspended in 6 μL 1/2× TE [5 mM Tris, 0.5 mM EDTA (pH 8.0)] buffer. The dimerization reactions in the presence of a saturating amount of NC or UP1 protein (80 μM in 1 μL) were performed identically except the total initial volume of RNA in water was 149 μL and the proteins were added as a mixture with magnesium (11 μL total volume). For reactions containing NC or UP1, after SHAPE modification, bound protein was removed by digestion with proteinase K [60 μg (Invitrogen); 10 min, 37 °C] followed by extraction with phenol:chloroform:isoamyl alcohol (25:24:1) prior to ethanol precipitation.

Optimal concentration of NC to facilitate dimerization

The equilibrium dissociation constant for NC binding to short sequences containing guanosine is 100-400 nM¹⁷, and thus most of the NC protein is likely bound under the conditions used in our experiments. We determined the concentration of NC that yielded the largest acceleration of the MuLV dimerization reaction empirically. Observed dimerization rates increased as the NC concentration was increased from 2 to 8 μM and did not increase further as the concentration was increased to 16 μM. An 8 μM NC concentration was therefore chosen as the standard condition for these experiments.

Primer extension and data processing

MuLV primers (5'-GGUGC ACCAA AGAGU CCAAA AGC-3', LNA nucleotides are underlined, 5'-end labeled with 5-FAM or 6-JOE) were complementary to the 3' end of the MuLV dimerization domain (nucleotides 422 to 445)²⁵. Primers (1 μ L; 10 pmol) were annealed to MuLV domain RNA (6 μ L; 5.4 pmol in 1/2 \times TE buffer) by heating at 65 $^{\circ}$ C and 45 $^{\circ}$ C for 5 min each and then were snap-cooled on ice. Reverse transcription buffer [3 μ L; 200 mM Tris (pH 8.0), 250 mM KCl, 10 mM MgCl₂, 2 mM each dNTP, 20 mM DTT] was added at 0 $^{\circ}$ C, and primer extension was performed with Superscript III (Invitrogen) reverse transcriptase (0.5 μ L; 100 U) at 45 $^{\circ}$ C for 1 min, followed by incubation at 52 $^{\circ}$ C and 65 $^{\circ}$ C for 10 min each. Reactions were quenched by addition of 3 M sodium acetate (pH 5.2) and cooling to 4 $^{\circ}$ C. A sequencing marker was generated by adding 0.5 μ L ddGTP (10 mM) to the primer extension reaction mixture using unmodified RNA. The BzCN and DMSO reaction mixes were each combined with equal amounts of ddGTP-terminated sequencing ladders, precipitated with ethanol, and resuspended in deionized formamide (10 μ L). cDNA fragments were resolved on an Applied Biosystems 3130 capillary electrophoresis (CE) instrument. Time-resolved SHAPE data were processed using *QuShape*²⁶ (available at chem.unc.edu/rna/qushape). The initial raw sequence trace was corrected for dye variation and signal decay. Peak intensities were scaled, aligned, integrated, and normalized using a suite of optimized statistical algorithms. After analysis of the reference trace, subsequent time points were automatically analyzed by alignment to the saved project file as a reference trace²⁶. Each CE separation contained a reaction performed in the presence or absence of BzCN (labeled with 5-FAM) and a sequencing reaction (labeled with 6-JOE) performed using ddGTP.

RNA structure modeling

To develop secondary structure models for intermediate states, SHAPE reactivity information was used to impose a pseudo-free energy change constraint in conjunction with nearest neighbor thermodynamic parameters in the secondary structure modeling program *RNAstructure*^{19,27}. To develop approximate secondary structure models for the inosine substituted RNA (Fig. S3), we replaced

guanosine nearest-neighbor parameters with their corresponding inosine parameters in a modified version of *RNAstructure*; loop and bulge parameters used the existing values for guanosine. A-U, G-C, G-U, and I-U nearest neighbor parameters were reported previously^{18,22,28}. I-C nearest neighbor parameters were calculated as described¹⁸. Average nearest neighbor free energy change contributions ($\Delta G^{\circ 37}$) and standard deviations (Fig. 4A) were calculated from all possible nearest neighbor combinations containing the base pair of interest (G-C, I-C, or A-U) adjacent to a Watson-Crick pair.

***k*-means clustering**

SHAPE profiles were sorted by individual nucleotide kinetic behaviors by *k*-means clustering using *Cluster 3.0*¹⁶. We employed the following kinetics-specific features: (i) under the *kmeans* option, we organized the kinetic data for each nucleotide by rows (termed "genes" in *Cluster*); (ii) under the Adjust Data option, we used Normalize Genes; and (iii) we used the absolute correlation (uncentered) similarity metric to organize each row (or nucleotide-specific kinetic dataset). Clusters were visualized using *TreeView*¹⁶ [M. Eisen, *Cluster and TreeView online manual* (1998)]. To determine the optimal number of clusters, *k*-means clustering was initially performed starting from $k = 2$ through 10 (9 trials total). For each k value, clustering was reiterated both 100 and 1000 times, the optimal solutions were found twice, and the clustering solution with the smallest sum of within-cluster distances was accepted. The optimal cluster number for each experiment was chosen independently of the initial k value. For example, for the RNA-alone reactions, seven clusters were apparent in the optimal solution, which occurred at $k = 7$. Seven was also the optimal number of clusters when higher k values (including $k = 8, 9,$ and 10) were initially input. Forced division of the data by k values >7 resulted in overcategorizing nucleotides with slightly varying reactivities (net reactivity changes of less than 0.2 SHAPE units) as having a distinctive kinetic behavior. For the NC, UPI facilitated, and inosine variant RNAs, the same procedure for finding the optimal number of clusters consistently yielded an optimal solution at $k = 3$. Further increasing the k value in the initial search did not result in formation of any apparent new clusters. For most clusters, observed rates for individual nucleotides were the same, within error. In the case of

cluster 1, the kinetic profiles were characterized by two fast rates; this cluster was divided into two groups, 1a and 1b. For each grouping, net rates are reported as the mean \pm standard deviation over all nucleotides in the cluster.

Dimerization followed by non-denaturing electrophoretic mobility shift

The native MuLV domain RNA, internally labeled with ^{32}P (~ 0.5 nM), was mixed with unlabeled RNA (600 nM in 15 μL of water). Reactions were treated exactly as per the time-resolved SHAPE experiments, outlined above. Briefly, samples were heated to 95 $^{\circ}\text{C}$ to denature, rapidly cooled on ice, and equilibrated with 5 μL of a 5 \times folding buffer without magnesium ion [50 mM HEPES (pH 7.5), 200 mM potassium acetate (pH 7.5), at 37 $^{\circ}\text{C}$]. The reaction was initiated by the addition of MgCl_2 to 5 mM (5 μL , 25 mM) at 37 $^{\circ}\text{C}$, for a final volume of 25 μL . Time point aliquots (3 μL) were mixed with 5 μL of 30% (v/v) glycerol (containing marker dyes), loaded directly onto a running non-denaturing gel (5% polyacrylamide; 29:1 acrylamide/bisacrylamide, in TBE; 0.4 mm \times 28.5 cm \times 23 cm gel), and resolved by electrophoresis at 4 $^{\circ}\text{C}$ for 2 h at 20 W. Observed rates are reported as the mean \pm standard deviation from three independent experiments.

RESULTS AND DISCUSSION

We studied an RNA construct spanning the 170-nucleotide (nt) MuLV dimerization region⁸⁻¹⁰ and including 5' and 3' flanking sequences of 46 and 115 nucleotides, respectively. This RNA dimerizes under physiological-like conditions in vitro and has a structure similar to that of genomic RNA isolated from virions^{11,12}. Point mutations in this region of the MuLV genome eliminate its selective packaging into virions¹⁰. We followed dimerization at single-nucleotide resolution using time-resolved, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE)^{13,14}. A fast-acting reagent, benzoyl cyanide (BzCN), that either reacts to form a 2'-*O*-adduct at conformationally flexible nucleotides or undergoes rapid self-inactivation by hydrolysis (with a 0.25-s half-life), was used¹⁴. Each time point, obtained over reactions spanning tens of minutes, thus yields a structural snapshot of ~ 1 s duration.

SHAPE profiles for the initial monomer and final dimer forms agree well with accepted structures for the MuLV dimerization domain (Fig. 4.S1 and text 4.S1). Five key regions underwent large-scale structural changes during dimerization (Fig. 4.1). The loops of hairpins SL1 and SL2 (positions 329 to 332 and 363 to 366, respectively) were reactive in the monomer and became unreactive during dimerization (within 7 s), consistent with formation of a stable intermolecular loop-loop kissing interaction¹⁵. Two palindromic sequences, PAL1 (positions 210 to 219) and PAL2 (positions 283 to 298), were initially reactive but became unreactive because of intermolecular duplex formation in the dimer. Conversely, two regions that form the “anchoring helix” (positions 231 to 251 and 290 to 315) in the monomer became more reactive upon dimer formation (Fig. 4.1 and fig. 4.S1).

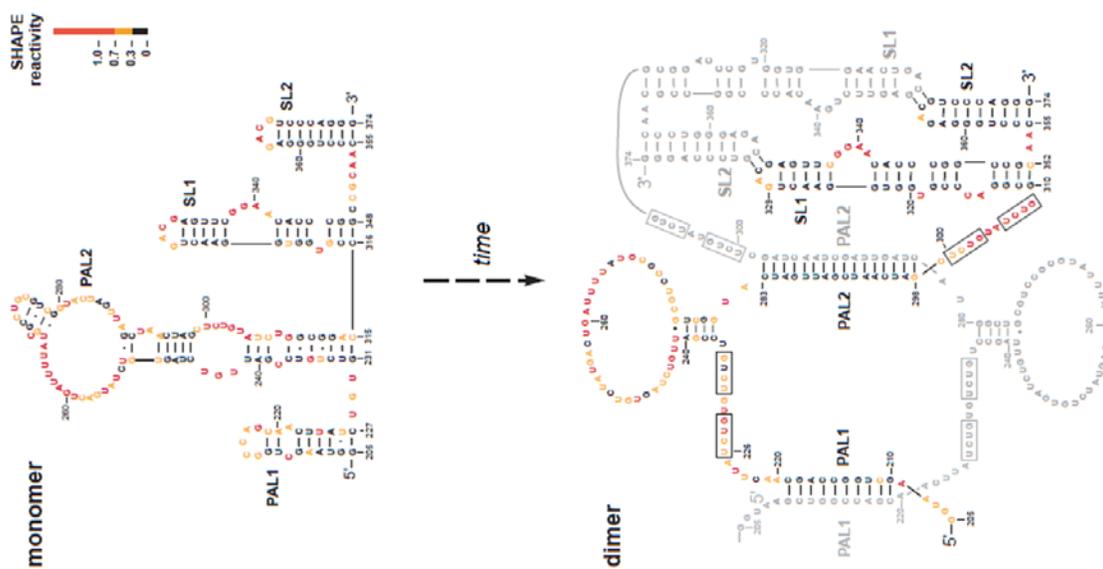


Figure 4.S1: Secondary structures of monomer and dimer states. Nucleotides are colored by SHAPE reactivity. RNA motifs that undergo the largest structural changes during dimerization are labeled. For clarity, only one strand of the dimer is colored.

Text 4.S1: SHAPE profiles for the initial monomer and final dimer forms agree well with accepted structures for the MuLV dimerization domain, including those obtained using authentic genomic RNA^{10,12,25} (Fig. 4.S1). Nucleotides with high SHAPE reactivities are conformationally dynamic^{29,30} and generally occur in single-stranded regions of the RNA secondary structure (Fig. 4.S1, red nucleotides). Conversely, nucleotides with low SHAPE reactivities (Fig. 4.S1, black) tend to be base paired. Each time point is effectively a structural snapshot of ~1 second duration. These 1-second snapshots are sufficient to monitor many features of RNA dynamics; however, there are also likely to be additional features of the reaction that occur more rapidly than measured here.

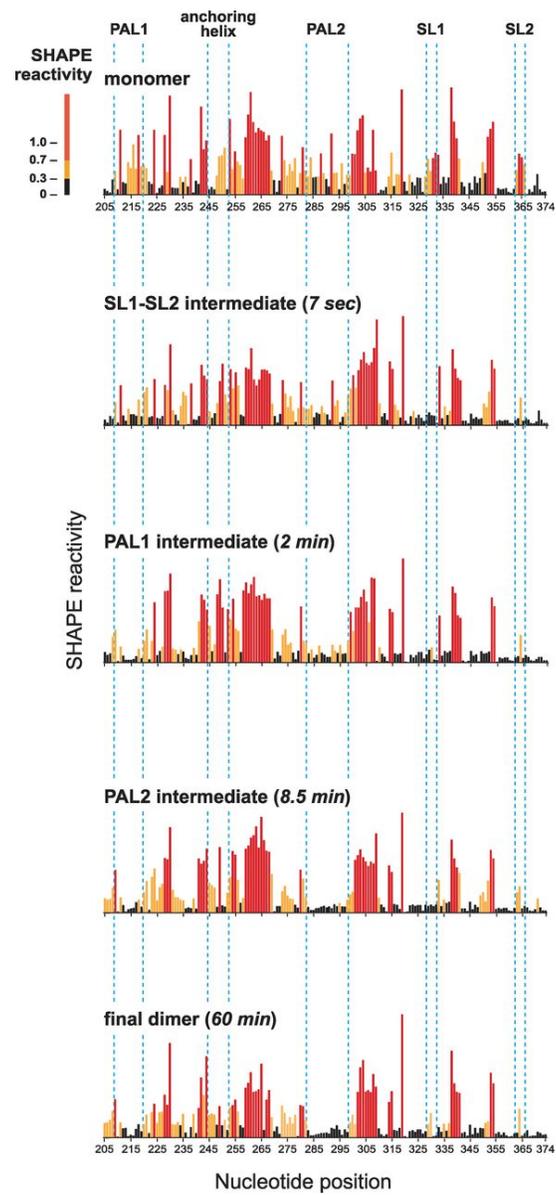


Fig. 4.1 Time-resolved SHAPE analysis of MuLV RNA dimerization. SHAPE reactivities are shown for monomer (no magnesium), dimer (60 min), and representative time points in which specific structural intermediates predominate. Key structural interactions that change during dimerization are highlighted within sets of dashed blue lines.

We obtained SHAPE data for every nucleotide within the 170-nt MuLV domain in 16 1-s snapshots yielding more than 2700 structural data points. We grouped nucleotides with similar kinetic behaviors by *k*-means clustering¹⁶. In the presence of 5 mM Mg²⁺ and without a protein chaperone, there were seven distinct kinetic behaviors involving four net rates (Fig. 4.2A). Rates were identical, within error, over a three-fold change in RNA concentration (fig. 4.S2), indicating that most conformational changes reflect pseudo-unimolecular transitions between two interacting RNAs. The fastest rate of ≥ 5 min⁻¹ (Fig. 4.2A; cluster 1a, in orange on structures at bottom) occurred at nucleotides at the apexes of SL1 and SL2, suggesting formation of a complex between two RNAs before the first time point. PAL1 nucleotides became less reactive at a net rate of 1.6 ± 0.4 min⁻¹ (Fig. 4.2A, cluster 1b, green on structures). The anchoring helix and PAL2 nucleotides demonstrated opposing kinetic behaviors (rates of 0.30 ± 0.03 min⁻¹) (Fig. 4.2A, clusters 2 and 3, in red), suggestive of a single process involving both structures. Positions in a large, flexible domain (positions 251 to 282) showed slower kinetic behavior with a net rate of 0.11 ± 0.02 min⁻¹ (Fig. 4.2A, cluster 4, in black on structures). Finally, nucleotides in clusters 5 and 6 showed biphasic kinetic behavior in which the SHAPE reactivity first increased and then decreased over time, or vice versa, with rates of 1.6 and 0.1 min⁻¹. Time-resolved SHAPE analysis of the MuLV domain thus reveals that dimerization is complex, slow, and characterized by multiple structurally distinct transitions and intermediates.

We next performed an analogous set of experiments initiating dimerization by simultaneous addition of magnesium ion and the MuLV NC protein. With the addition of chaperone, clustering of the SHAPE data revealed that the NC protein collapsed dimerization into a single kinetic process that occurred at a net rate of 1.6 ± 0.4 min⁻¹ (Fig. 4.2B). There was no evidence of the slow and multirate processes that characterized the RNA-only reaction. Initial binding interactions between the NC chaperone and RNA monomer were readily detected in a difference analysis in which the SHAPE profile, immediately after NC binding, was subtracted from that of the reactivity profile of the free RNA (Fig. 4.3A). Of the 29 nucleotides with the largest changes in SHAPE reactivity, 19 (or 66%) are guanosine residues (Fig. 4.3, A and B, and text 4.S2), consistent with studies showing that NC contains a cleft that

binds guanosine¹⁷. Sites of protection (positive peaks) likely correspond to sites of stable binding by NC during the 1-s window of the time-resolved SHAPE experiment; the smaller number of guanosine residues with higher reactivity in the presence of NC (negative peaks) likely reflect either a rapid binding and release or NC-induced conformational changes.

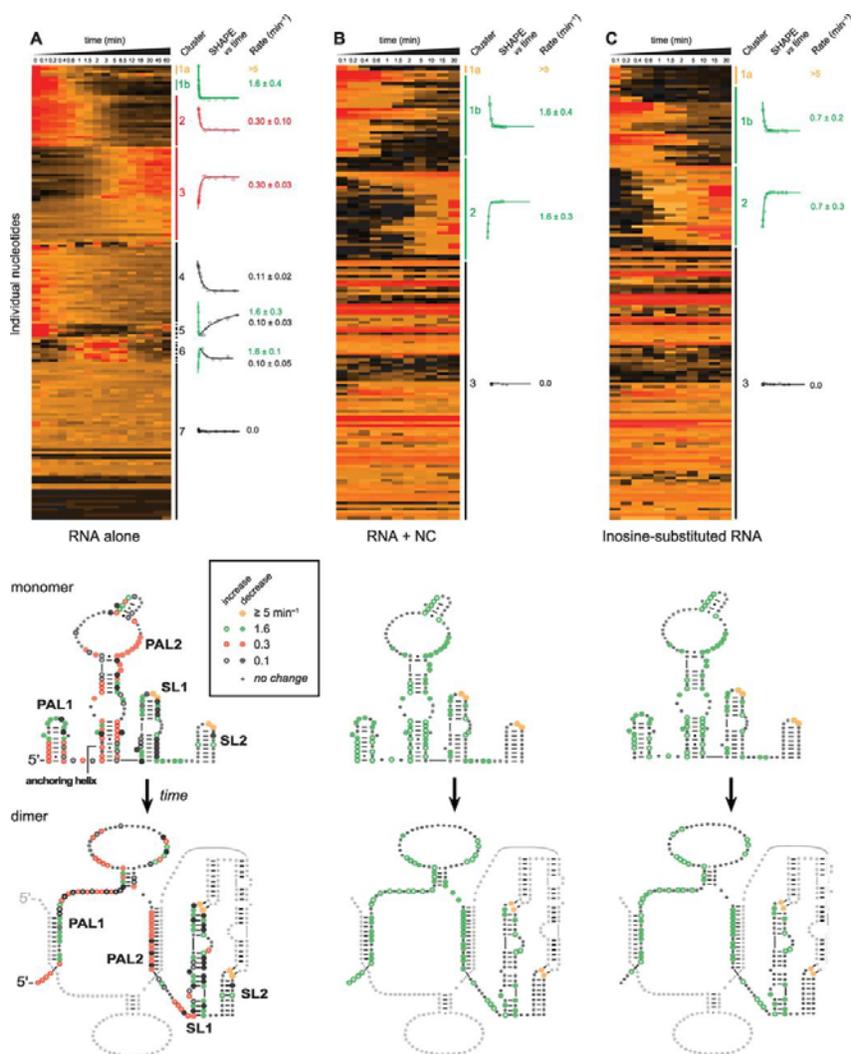


Fig. 4.2 Model-free clustering of nucleotide-resolution kinetic profiles for dimerization. SHAPE reactivities at 600 nM MuLV RNA (A) in the absence of and (B) in the presence of 8 μ M NC, and (C) for an RNA containing inosine in place of guanosine. Each data point is shown on a scale (black to red) corresponding to its SHAPE reactivity (see Fig. 4.1). The y axis shows every nucleotide (170 positions) in the MuLV dimerization domain RNA in an order determined by *k*-means clustering rather than linear sequence. Major kinetic clusters are labeled, and representative kinetic profiles and observed net rates are shown for each cluster. Rates are reported as the mean for all nucleotides in each cluster \pm the standard deviation. Positions of nucleotides in each cluster are shown in structural cartoons below each kinetic profile, colored by rate: orange > green > red > black. For clarity, only one strand of the dimer is colored.

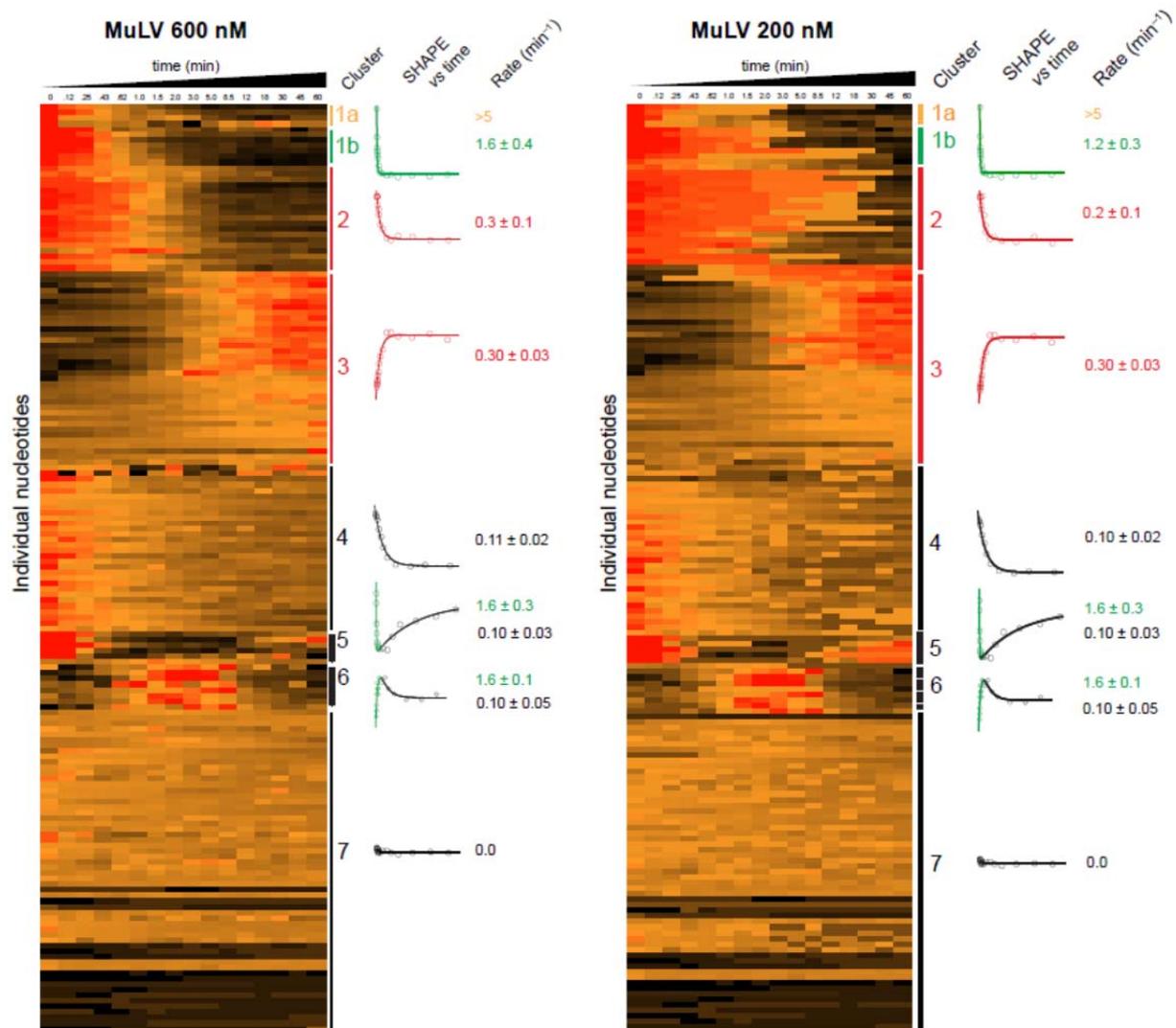


Figure 4.S2: Time-resolved SHAPE of the native sequence MuLV RNA dimerization at 600 nM and 200 nM. Representative kinetic profiles and net rates are shown. Observed net rates were identical, within error, over a 3-fold RNA concentration range. Data shown in the 600 nM panel are the same as data shown in Fig. 4.2A. Nucleotides are shown in the order determined by *k*-means clustering rather than the 5' to 3' nucleotide sequence order.

The preference of NC to interact at guanosine residues prompted us to consider whether NC exerts its RNA chaperone activity by destabilizing interactions between guanosine and other nucleotides. We explored the dimerization reaction using an RNA in which all guanosine residues were replaced by inosine, in essence removing a single amine group from each guanosine position. Inosine-cytosine pairs are iso-structural with guanosine-cytosine pairs, but are ~ 1 kcal/mol less stable (Fig. 4.4A); inosine also pairs more weakly with uridine than guanosine¹⁸. The guanosine-to-inosine substitution will thus reduce both the strength and the promiscuity of alternative base pairs during the RNA folding reaction. The inosine-substituted RNA formed essentially the same final dimer structure as the guanosine-containing MuLV domain as indicated by SHAPE-directed modeling¹⁹ (fig. S3A), and individual nucleotide SHAPE reactivities for the inosine and native MuLV domain dimers are strongly correlated ($R^2 = 0.88$) (Fig. 4.4B). Although the overall secondary structures for inosine and native RNAs in the monomer states are similar (fig. 4.S3, B and C), SHAPE reactivities correlate poorly ($R^2 = 0.26$) (Fig. 4.4C). However, adding NC to the guanosine-containing monomer converts this RNA to a structure that has a SHAPE profile highly similar to that of the inosine RNA monomer ($R^2 = 0.87$) (Fig. 4.4D). The inosine-substituted RNA is thus a good model both for the NC-destabilized native RNA in the monomer state and for the final dimer.

Text 4.S2: In the dimerization domain, 50 of 170 nts (29%) of residues are guanosine, such that the observed preferential interactions with 66% and 52% of guanosine by NC and UP1 (Fig. 3), respectively, are greater than expected by chance; p-values (exact binomial test) are 0.00005 and 0.02, respectively. In addition, for both proteins, each of the seven most strongly protected nucleotides (positive amplitudes in Figs. 4.3A and 4.3D) are guanosine.

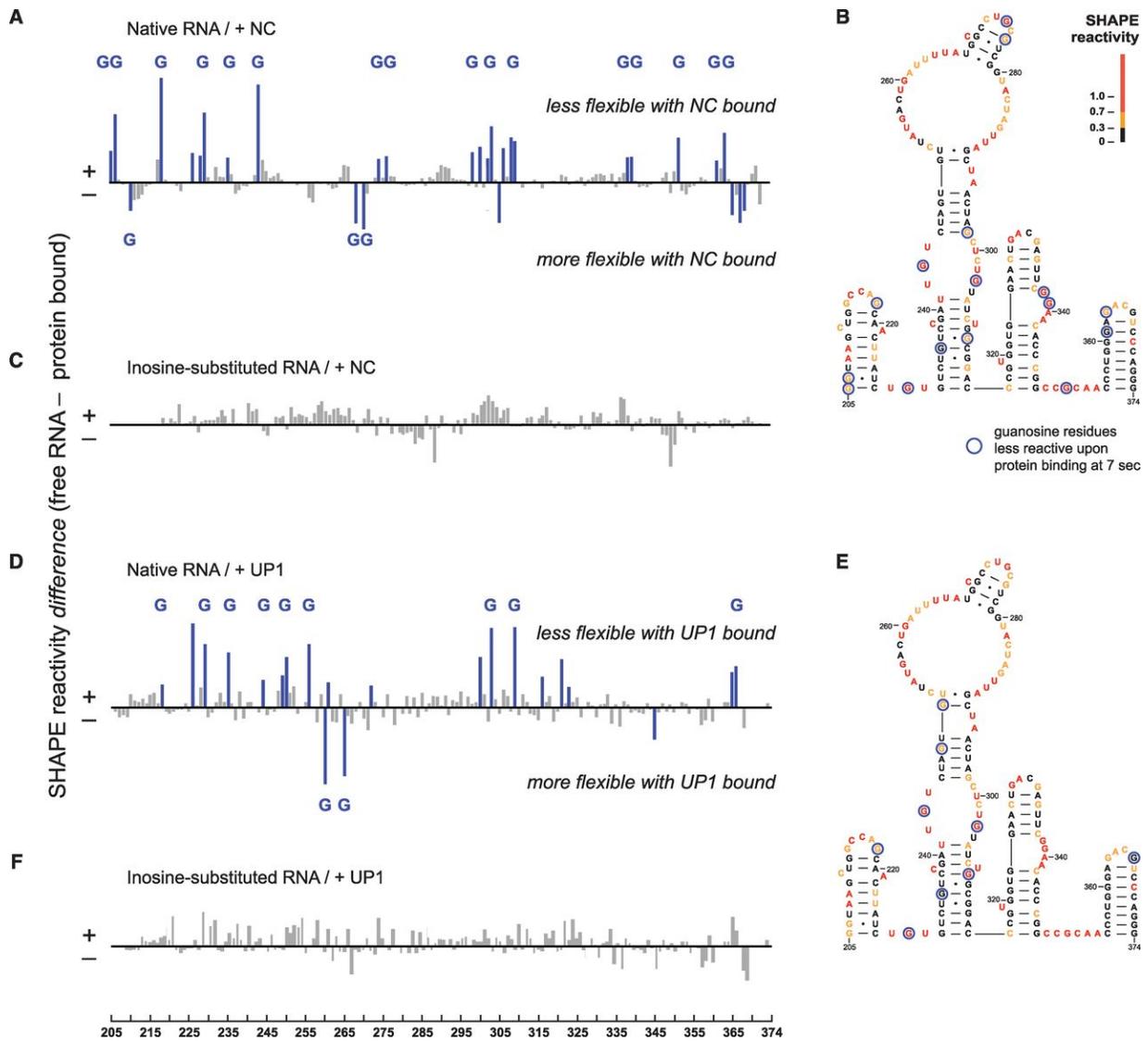


Fig. 4.3 Initial interactions between NC and UP1 with the MuLV monomer. (A and D) SHAPE difference plots illustrating the effect of (A) NC or (D) UP1 binding to the native dimerization domain 7 s after protein addition. Sites of strongest interaction, corresponding to SHAPE differences greater than 20%, are highlighted in blue; those that occur at guanosine residues are labeled with a G. (B and E) Superposition of strongest initial interaction sites for (B) NC and (E) UP1 on a MuLV dimerization domain secondary structure model. Structures are colored by SHAPE reactivity before protein binding. (C and F) SHAPE difference plots illustrating the lack of an effect of (C) NC or (F) UP1 binding to the inosine-substituted RNA.

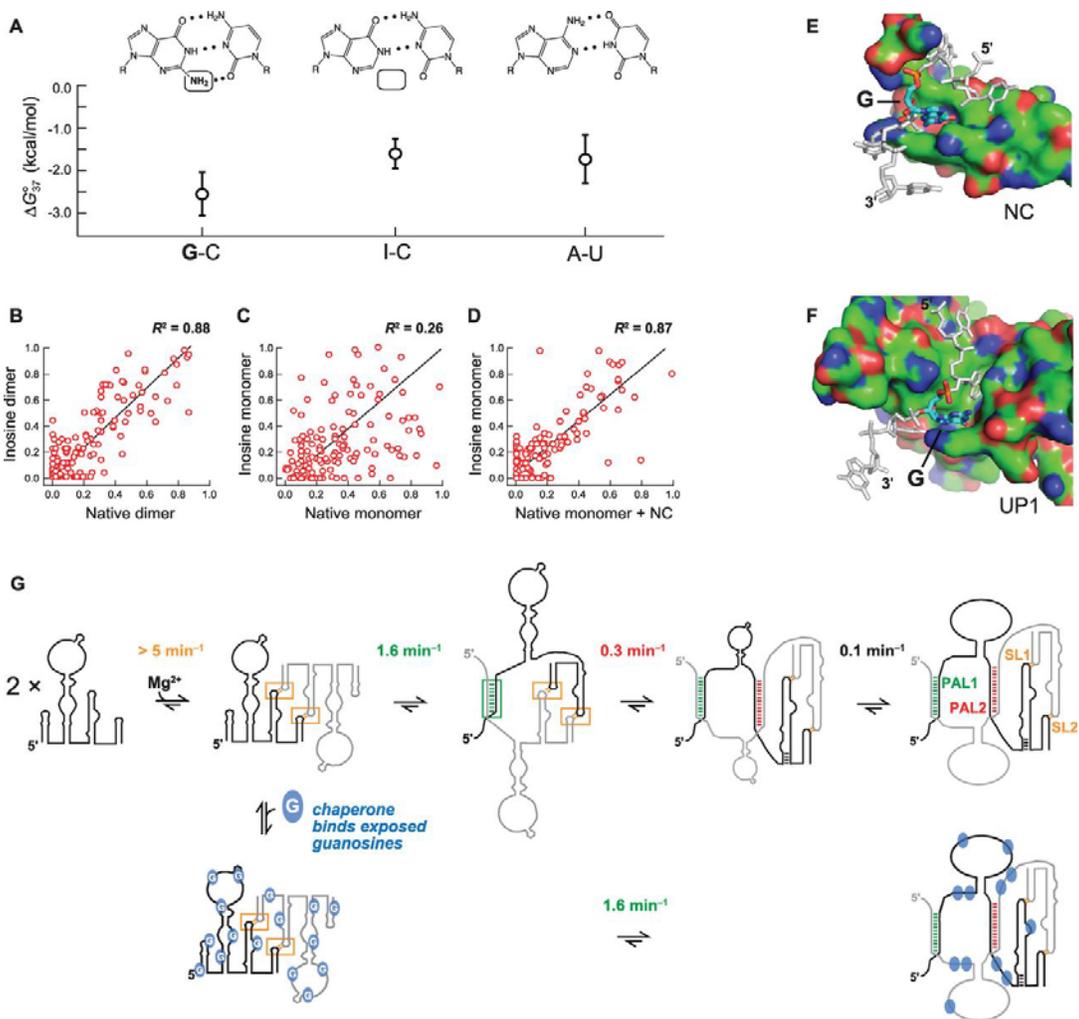
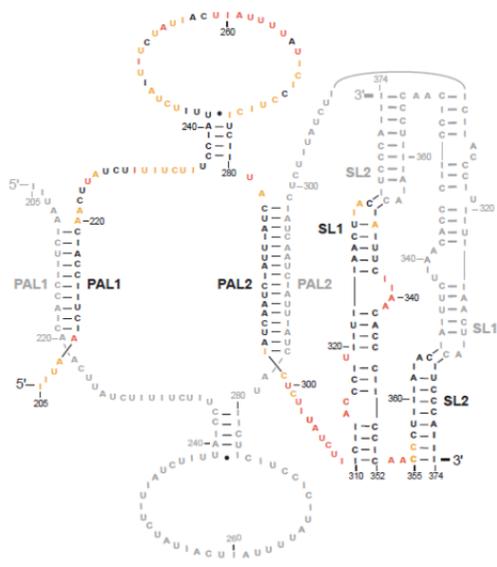


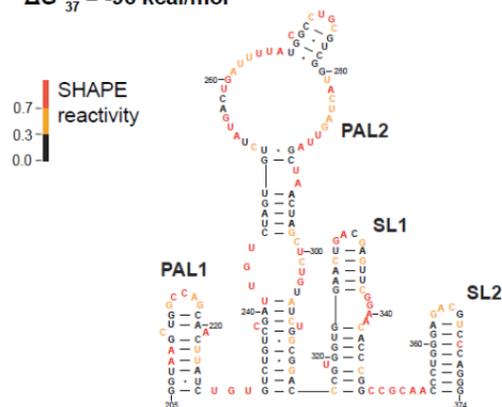
Fig. 4.4 Role of guanosine in RNA structure and mechanism of chaperone-mediated RNA folding. (A) Average stabilities (ΔG_{37}°) of nearest neighbor base pair combinations involving one G-C, I-C, or A-U pair and one Watson-Crick pair in 1 M NaCl^{18,22}, with representative structures. Standard deviations for the nearest-neighbors combinations are shown with lines. (B to D) Correlations between SHAPE reactivity profiles of (B) native and inosine-substituted dimers (obtained at 30 min dimerization time points), (C) native and inosine-substituted monomers (obtained just before addition of Mg^{2+}), and (D) the native RNA after a 7-s interaction with NC versus the protein-free inosine-substituted monomer (both in the presence of Mg^{2+}). The inosine dimer and native dimer in the presence of NC (not shown) also show a strong correlation ($R^2 = 0.89$) reflecting that NC binds at relatively few sites in the native dimer (fig. 4.S6). (E and F) Structures of (E) NC¹⁷ and (F) UP1²³ chaperones, emphasizing that both have a guanosine-binding pocket and that flanking nucleotides interact in an extended conformation. NC and UP1 bind guanosine in distinct ways involving anti and syn nucleotide conformations, respectively. (G) RNA-only (top) and chaperone-catalyzed (bottom) MuLV genome assembly mechanisms. Net rates are reported for each step. The overall reaction proceeds sequentially as indicated by (i) the change in reaction order (from second to first, yielding a large increase in effective RNA concentration) upon formation of the initial SL1-SL2 kissing interaction in the first step and (ii) the observation of biphasic profiles (Fig. 4.2) that include both the 1.6 and 0.1 min^{-1} processes. Evidence for a specific order of the 0.3 min^{-1} process is less strong, and this step may occur in parallel with the 0.1 min^{-1} conformational change.

A Inosine MuLV dimer



B Native MuLV monomer

$\Delta G^{\circ}_{37} = -96$ kcal/mol



C Inosine-substituted MuLV monomer

$\Delta G^{\circ}_{37} \approx -76$ kcal/mol

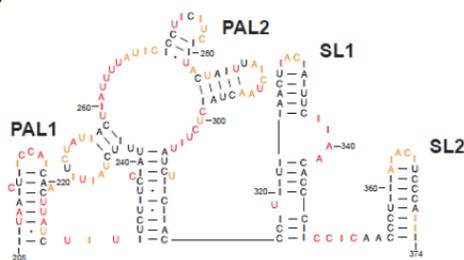


Figure 4.S3: Native and inosine-substituted RNA secondary structures. (A) SHAPE-directed secondary structure model¹⁹ for the inosine-substituted dimer. Nucleotides are colored by SHAPE reactivity; only one strand of the dimer is colored. Comparison of the SHAPE-directed secondary structures and free energies (ΔG°_{37}) for the (B) native (guanosine-containing) and (C) inosine-substituted monomers. Free energies (ΔG°_{37}) were calculated using either the standard parameters in *RNAstructure*²⁷ for the native RNA or a modified *RNAstructure* program incorporating inosine nearest-neighbor parameters.

Time-resolved SHAPE analysis of dimerization of the inosine-substituted RNA in the absence of NC revealed a single, fast kinetic step involving similar nucleotides as NC-mediated dimerization of the native sequence RNA (compare Fig. 4.2B and 2C). The dimerization rate of the inosine RNA was accelerated by a factor of 7 relative to that of the free native RNA. The NC protein does not affect the structure of the inosine-substituted RNA (Fig. 4.3C). A nondenaturing gel-based analysis confirmed that addition of NC protein had no effect on the rate of formation of the final dimer state for the inosine-substituted RNA (fig. S4). Replacement of guanosine with inosine thus both abrogates most of the need for the RNA chaperone activity of the NC protein and converts the RNA into a form that folds via a simple and direct pathway (compare Fig. 4.2A and 2C).

The unwinding domain of the heterogeneous nuclear ribonucleoprotein (hnRNP) A1 protein (UP1) contains an arginine-rich RNA recognition motif and has potent RNA chaperone activity^{1,20}. UP1 has no structural similarity with NC except that both proteins contain clefts that bind guanosine (Fig. 4.4, E and F). As in the presence of NC, UP1-mediated dimerization of the MuLV domain proceeded in a single, fast kinetic step (fig. 4.S5) accelerated by a factor of ~ 20 ($k_{\text{obs}} > 2 \text{ min}^{-1}$) relative to the RNA alone. Of the initial interaction sites (at ~ 7 s) for UP1 on the native monomer RNA, 52% were guanosine residues (Fig. 4.3D and text 4.S2). The set of guanosines contacted most strongly by UP1 included some but not all of the guanosines contacted by NC (compare Fig. 4.3B and 3E). UP1 had no effect on the dimerization rate of the inosine-substituted RNA (Fig. 4.3F and fig. 4.S4). UP1 is not known to play a role in structure rearrangements for the MuLV RNA genome, yet is a potent facilitator of RNA dimerization of the MuLV domain and does so by a mechanism similar to the cognate NC chaperone.

Our data support a model of MuLV genomic dimerization in which two MuLV monomers initially associate rapidly via loop-loop interactions; subsequent steps for RNA-only folding are complex, involve multiple intermediates, and proceed slowly (Fig. 4.4G). In the presence of the chaperone, RNA dimerization was accelerated by a factor of more than 10 and appeared to occur in a single kinetic step, indicating that the chaperone function accelerated multiple classes of slow RNA conformational changes (Fig. 4.4G). Our data indicate that RNA chaperones NC and UP1 both act by binding to exposed

guanosine residues in RNA, thereby destabilizing stronger base pairings and creating a simplified folding pathway (text 4.S3). The two proteins contact distinct, partially overlapping sets of guanosine residues in their initial interactions with RNA; thus, many possible guanosine-binding activities may support RNA chaperone function. The NC and UP1 chaperones also bind to the final native sequence dimer in patterns that are distinct from their initial interactions with the monomer state (fig. 4.S6). These data suggest that chaperone binding does not discriminate between folded and misfolded RNA states per se but that guanosine nucleotides are ultimately arranged in the final structure in such a way that chaperone binding (or inosine substitution) does not overly destabilize the final RNA structure. In this way, a guanosine-centric mechanism for RNA chaperone function is analogous to the mechanism of some chaperones that facilitate protein folding that destabilizes interactions involving hydrophobic amino acid residues²¹. In these cases, both RNA and protein chaperones simply interact with residues especially prone to forming stable intermediate and non-native states.

Text 4.S3: The NC and UP1 chaperone exhibited all of the activities attributed to non-specific RNA-binding chaperone proteins – including RNA annealing, strand exchange, destabilization of intermediates, facilitating complex structural rearrangements, and the ability to act on large RNA regions simultaneously^{1,3,4} (and see Figs. 4.2 and 4.S6), and were largely mediated by interactions between NC or UP1 and guanosine nucleotides. The guanosine-centric mechanism outlined here imposes very few requirements on a potential RNA chaperone, primarily that it contains a binding pocket for guanosine (see Figs. 4.4E and 4.4F), and therefore is likely to apply to a wide variety of chaperone proteins. There are other mechanisms by which RNA chaperones can function, however. These include via ATP-dependent mechanisms³¹, as macromolecular complexes³², and as non-specific RNA binding proteins that interact locally with RNA via mechanisms that are independent of guanosine nucleotides.

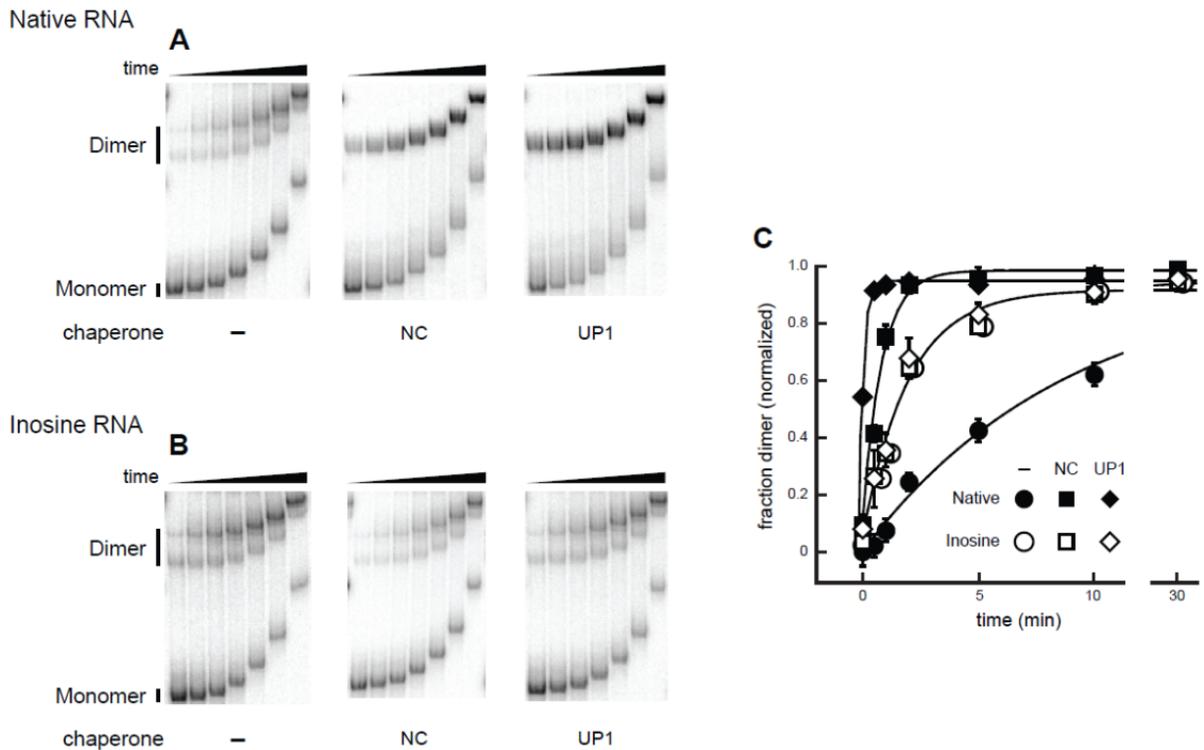


Figure 4.S4: Visualization of MuLV genomic RNA dimerization by non-denaturing electrophoresis. Dimerization time courses are shown for 600 nM (A) native and (B) inosinesubstituted RNAs with and without 8 μ M NC or UP1. (C) Fraction of RNA dimer as a function of time. Points are fit assuming a first order process. Observed net rates for the native, native + NC, and native + UP1 experiments were 0.12 ± 0.05 , 1.4 ± 0.3 , and $>2 \text{ min}^{-1}$ respectively; and for the inosine RNA, inosine RNA + NC, and inosine RNA + UP1 were 0.8 ± 0.1 , 0.7 ± 0.1 , and $0.8 \pm 0.2 \text{ min}^{-1}$ respectively. These rates correspond well to a subset of the rate process detected by nucleotide-resolution time-resolved SHAPE (Figs. 4.2, 4.S2, and 4.S5).

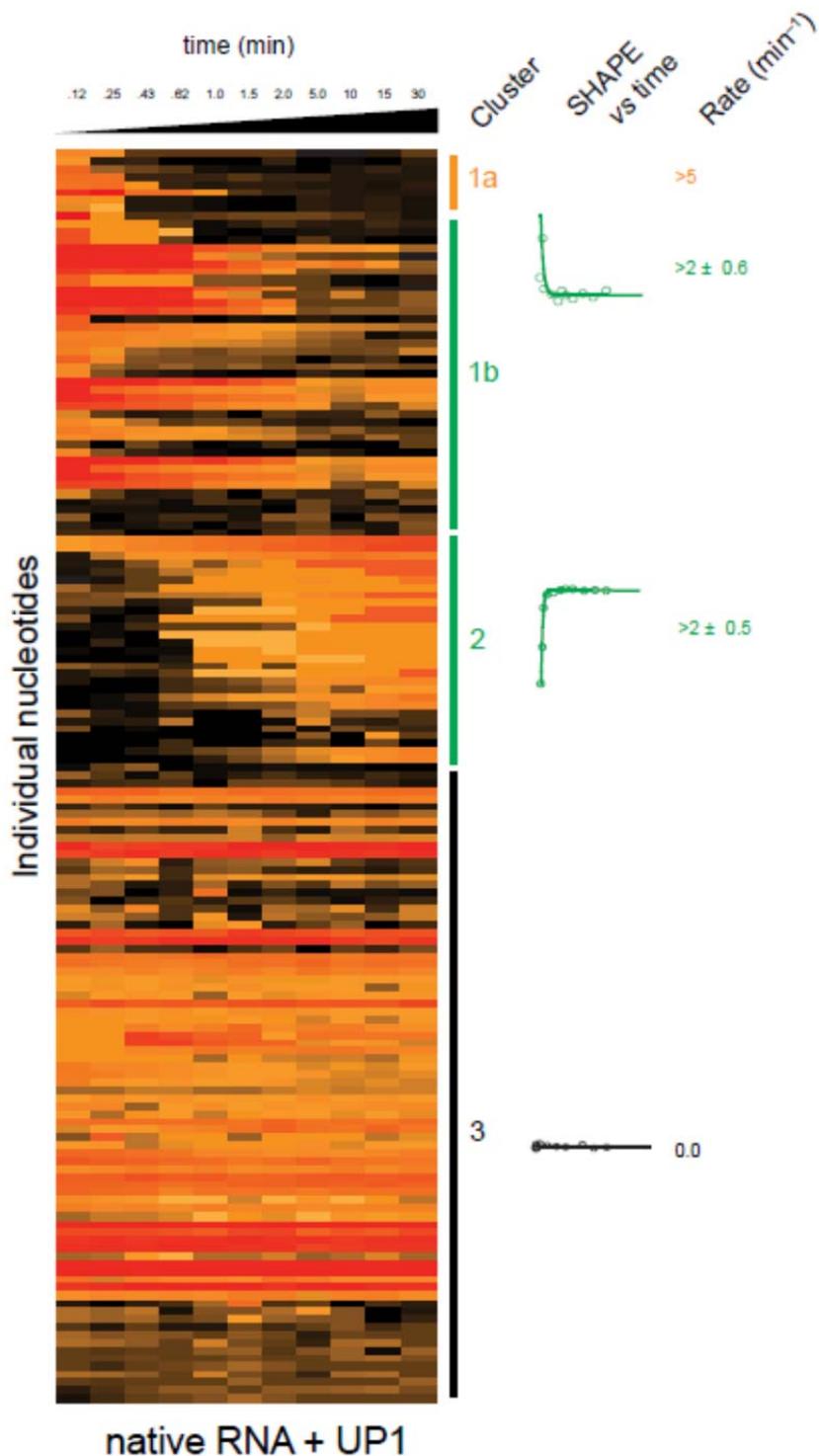


Figure 4.S5: Time-resolved SHAPE of UP1-mediated RNA dimerization of the native MuLV RNA. Each data point is shown on a scale (black to red) corresponding to its SHAPE reactivity. Major kinetic clusters are labeled and representative kinetic profiles and observed net rates are shown for each cluster. The effect of UP1-facilitation is readily seen by comparing this figure to Fig. 4.2A. Nucleotides are shown in the order determined by *k*-means clustering.

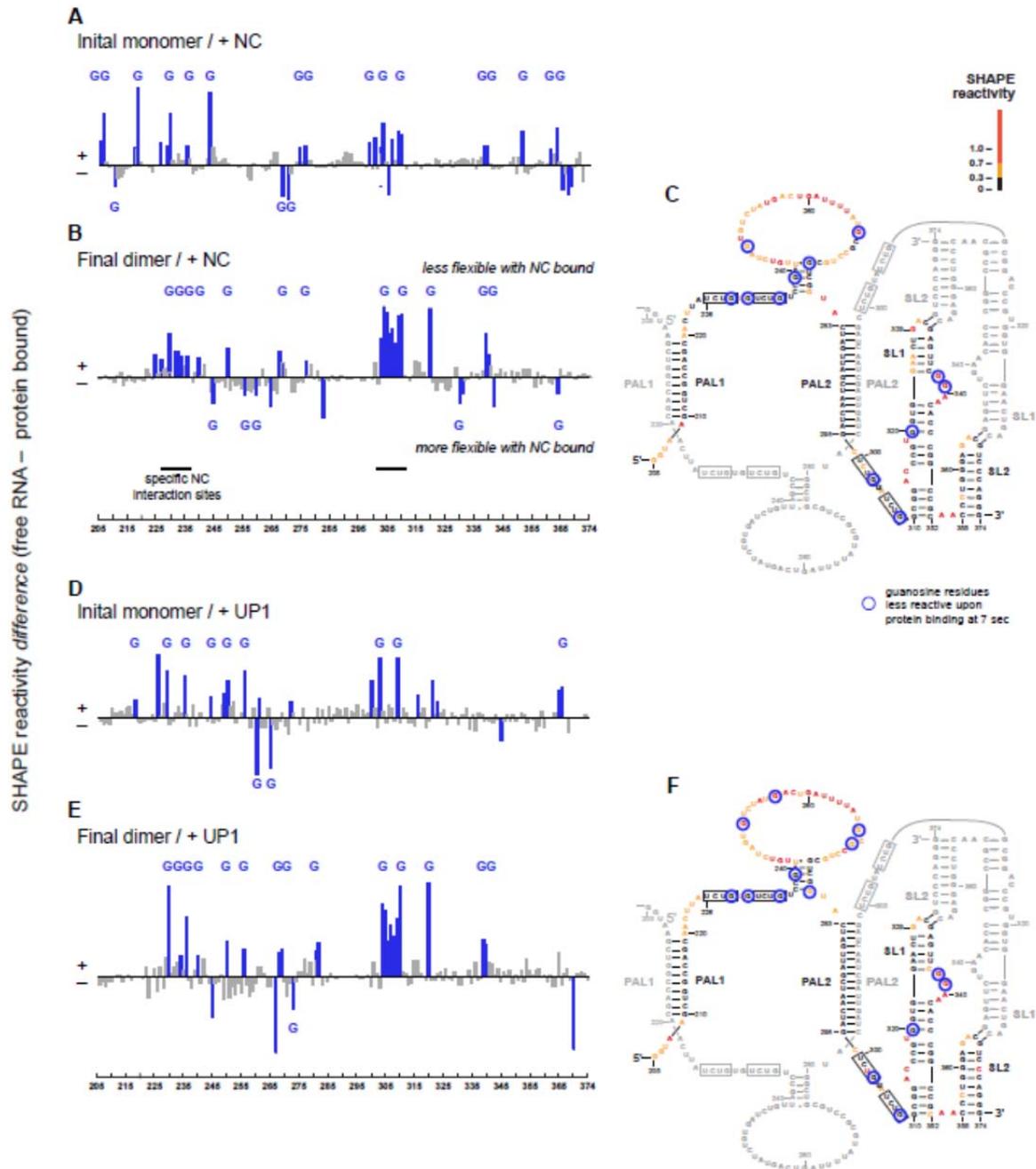


Figure 4.S6: Interactions between NC and UP1 with the MuLV monomer and dimer states.

SHAPE difference plots illustrating the effect of (A, B) NC or (D, E) UP1 binding to the initial monomer and final dimer states. Site of strongest interaction, corresponding to SHAPE differences greater than 20%, are highlighted blue; those that occur at guanosine residues are labeled with a G. Black bars in panel B denote specific NC binding sites that mediate RNA genome packaging¹⁰. Superposition of strongest interaction sites for (C) NC and (F) UP1 on a model for the MuLV domain secondary structure in the final dimer state. Data shown in panels A and D are also shown in Fig. 4.3.

REFERENCES

- [1] D. Herschlag, RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 270, 20871 (1995).
2. A. Rein, L. E. Henderson, J. G. Levin, Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.* 23, 297 (1998).
- [3] L. Rajkowitsch *et al.*, RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* 4, 118 (2007).
- [4] S. A. Woodson, Taming free energy landscapes with RNA chaperones. *RNA Biol.* 7, 677 (2010).
- [5] J. M. Coffin, S. H. Hughes, H. E. Varmus, *Retroviruses* (Cold Spring Harbor Press, Plainview, NY, 1997).
- [6] A. Rein, Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol.* 7, 700 (2010).
- [7] J. A. Thomas, R. J. Gorelick, Nucleocapsid protein function in early infection processes. *Virus Res.* 134, 39 (2008).
- [8] M. A. Adam, A. D. Miller, Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J. Virol.* 62, 3802 (1988).
- [9] C. S. Hibbert, J. Mirro, A. Rein, mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J. Virol.* 78, 10927 (2004).
- [10] C. Gherghe *et al.*, Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19248 (2010).
- [11] C. S. Badorrek, K. M. Weeks, Architecture of a gamma retroviral genomic RNA dimer. *Biochemistry* 45, 12664 (2006).
- [12] C. Gherghe, C. W. Leonard, R. J. Gorelick, K. M. Weeks, Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. *J. Virol.* 84, 898 (2010).
- [13] S. A. Mortimer, K. M. Weeks, Time-resolved RNA SHAPE chemistry. *J. Am. Chem. Soc.* 130, 16178 (2008).
- [14] S. A. Mortimer, K. M. Weeks, Time-resolved RNA SHAPE chemistry: Quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* 4, 1413 (2009).
- [15] P. T. Li, C. Bustamante, I. Tinoco, Jr., Unusual mechanical stability of a minimal RNA kissing complex. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15847 (2006).
- [16] M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Cluster analysis and display of genomewide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14863 (1998).
- [17] A. Dey, D. York, A. Smalls-Mantey, M. F. Summers, Composition and sequence-dependent binding of RNA to the nucleocapsid protein of Moloney murine leukemia virus. *Biochemistry* 44, 3735 (2005).

- [18] D. J. Wright, J. L. Rice, D. M. Yanker, B. M. Znosko, Nearest neighbor parameters for inosine x uridine pairs in RNA duplexes. *Biochemistry* 46, 4625 (2007).
- [19] K. E. Deigan, T. W. Li, D. H. Mathews, K. M. Weeks, Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* 106, 97 (2009).
- [20] D. S. Portman, G. Dreyfuss, RNA annealing activities in HeLa nuclei. *EMBO J.* 13, 213(1994).
- [21] F. U. Hartl, M. Hayer-Hartl, Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* 16, 574 (2009).
- [22] T. Xia *et al.*, Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37, 14719 (1998).
- [23] J. Ding *et al.*, Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* 13, 1102 (1999).
- [24] K. M. Stewart-Maynard *et al.*, Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J. Virol.* 82, 10129 (2008).
- [25] J. K. Grohman, S. Kottegoda, R. J. Gorelick, N. L. Allbritton, K. M. Weeks, Femtomole SHAPE reveals regulatory structures in the authentic XMRV RNA genome. *J. Am. Chem. Soc.* 133, 20326 (2011).
- [26] F. Karabiber, J. L. McGinnis, O. V. Favorov, K. M. Weeks, QuShape: Rapid, accurate, and best practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* 19, 63 (2013).
- [27] J. S. Reuter, D. H. Mathews, RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129 (2010).
- [28] J. L. Chen *et al.*, Testing the nearest neighbor model for canonical RNA base pairs: Revision of GU parameters. *Biochemistry* 51, 3508 (2012).
- [29] E. J. Merino, K. A. Wilkinson, J. L. Coughlan, K. M. Weeks, RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* 127, 4223 (2005).
- [30] C. M. Gherghe, Z. Shajani, K. A. Wilkinson, G. Varani, K. M. Weeks, Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J. Am. Chem. Soc.* 130, 12244 (2008).
- [31] P. Linder, E. Jankowsky, From unwinding to clamping: The DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* 12, 505 (2011).
- [32] X. Qu *et al.*, The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* 475, 118 (2011).
- [33] Grohman, J. K., Gorelick, R. J., Lickwar, C. R., Lieb, J. D., Bower, B. D., Znosko, B. M., and Weeks, K. M. (2013) A guanosine-centric mechanism for RNA chaperone function, *Science* 340, 190-195.

Chapter 5: An Investigation of Heterogenous Ribonucleoprotein A1 and Unwinding Protein 1 DNA Binding Characteristics

INTRODUCTION:

Heterogenous ribonucleoprotein complex (hnRNP) A1 is one member of a family of hnRNPs that associates with heterogenous nuclear RNA (hnRNA), also known as precursor RNA (pre-RNA). hnRNPs facilitate the excision of introns from hnRNA and the assembly of mature messenger RNAs (mRNAs) via process known as splicing¹. However, hnRNP A1 has also been found to mediate a wide variety of other cellular and viral processes. Significantly, hnRNP A1 has been found to bind specifically to telomeric DNA² and to facilitate telomere maintenance³, possibly by promoting telomerase activity⁴. A proteolytic fragment of hnRNP A1 called Unwinding Protein 1 (UP1), shares some of these properties. At the time our investigation of hnRNP A1 was begun (Early 2010), the role of hnRNP A1 in telomere biology remained only poorly understood. We hoped to characterize the binding properties of hnRNP A1 on telomeric DNAs. This initial characterization would have provided a foundation for later work, including determining relative binding affinity of POT1 and hnRNP A1 to telomeric ssDNA overhangs. However, the binding of hnRNP A1 to a variety of both telomeric and non-telomeric ssDNAs was not amenable to examination via electron microscopy. Moreover, several groups published papers regarding the role on hnRNP A1 in telomere biology after this project was begun^{5, 6}, or reported that they were aggressively investigating this topic at the 2011 Cold Spring Harbor meeting on Telomeres and Telomerase. For these and other reasons we abandoned our investigation of hnRNPA1 in July of 2011. Future studies with G-rich telomere-derived RNA (TERRA), however, remain of potential interest. The cloning and generation of a simple hexahistidine tagged hnRNP A1 expression vector, done in this work, has provided a useful tool for future work at UNC and elsewhere.

MATERIALS AND METHODS:

DNAs and Templates

pRST5, pRST5Nick and pGLGAP were prepared as previously described^{8, 9}. To generate the strand-displaced templates pGL GAP and pRST5Nick was nicked using Nb.BbvCI (New England Biolabs) and pRST5 was nicked with Nb.BsmBI (New England Biolabs), and the resulting 3' hydroxyl was used to prime an elongation reaction using DNA Polymerase I Exo(-) (New England Biolabs) containing dATP, dTTP and dGTP. to generate a model telomere, pRST5 was cut with BsmBI (New England Biolabs) and exonucleolytically digested with T7 Exonuclease (New England Biolabs) to generate a G-rich telomeric and a quasi-random 3' tail on either end of the linear dsDNA.

Proteins:

A quantity of purified hnRNP A1 was received as a generous gift from Dr. Adrian Krainer, which was used in some of our initial experiments. Most data is derived from hnRNP A1 expressed in and purified from *E. coli*. Previous work had employed calf tissue and a more-complicated purification procedure. A plasmid containing the hnRNP A1 cDNA (NM_002136) was purchased from a commercial vendor (Open Biosystems: Cat#IHS1380-97433308). This cDNA or a fragment of this cDNA coding for UP1 were ligated into a HIS-Tagging protein expression vector, pET28a (Invitrogen), via conventional cloning. Proper assembly of the final vectors was confirmed via bidirectional sequencing (MWG Operon). The assembled vectors were individually transformed into chemically competent BL21(DE3)PlysS cells via heat shock and grown on LB-Agar plates supplemented with 10 µg/ml kanamycin sulfate (Kan). Individual colonies were used to serially inoculate LB+Kan cultures until 2x500 ml cultures reached an optical density (OD) at 595 nm of 0.6. Protein expression was then induced in these cultures via addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. Cultures were induced at 37 °C for 2 h in a shaking incubator. Cells were then recovered by

centrifugation and resuspended in lysis buffer (50mM Tris-HCl at pH 7.4, 8mM β -MeOH, 300mM NaCl) supplemented with 1x protease inhibitors (EDTA Free Complete [®] Tablets) and stored at -80 °C. Cells were thawed and lysed via addition of 1 mg/ml egg-white lysozyme and sonication and the crude lysate separated via ultracentrifugation in an SW-27 rotor at 27,000 RPM for 2 h. The supernatant was collected and passed through a 20 ml Fast Flow DEAE-Sepharose (Amersham Pharmacia) equilibrated in lysis buffer and was collected in the flowthrough fraction. This flow through was then applied to 2 ml of NiNTA agarose (Qiagen) equilibrated in lysis buffer supplemented with 50 mM imidazole. The protein was then eluted following thorough washing of the beads with 50mM Imidazole and a stepwise gradient to 500mM imidazole. The obtained protein was then dialyzed overnight against storage buffer (20% glycerol, 300mM NaCl, 5mM β ME, 20mM Tris-HCl pH 7.4), aliquoted, flash frozen over liquid nitrogen, and stored at -80 °C. Proteins were quantified via Bradford (Biorad) calibrated against a bovine serum albumin control.

Electron Microscopy:

Carbon was evaporated onto freshly cleaved mica under vacuum in a modified DV-502A (Denton), and this carbon foil was then floated onto a surface of double-distilled water. 400 mesh EM grids (Ted Pella Inc.), were then briefly dipped in 5% polyurethane diluted in toluene and then partially dried via blotting on filter paper. These grids were then placed face down atop the floating carbon film and allowed to adhere overnight at room temperature. The grids were then recovered using a tensioned film of plastic wrap, transferred to filter paper and allowed to dry at room temperature. Hand-made size exclusion columns for use in cleaning samples prior to examination by electron microscopy were manufactured via addition of approximately 2 ml of A5M Agarose Beads (Agarose Beads Technology) to a disposable 2.4 ml transfer pipette (Fisher 13-711-5AM) stoppered with glass wool. These columns were then equilibrated with 10x column volumes of EM-grade TE (EMTE; 10 mM Tris-HCL at pH 7.5, 0.1 mM EDTA) and clamped at the tip until use. As necessary samples were fixed prior to cleaning via addition of 0.3 % to 0.6 % (v/v) glutaraldehyde and incubation at either room temperature or 0 °C for 5

min. Samples were passed over A5M columns via addition of a 50 μ l sample to the center of the top of the bead bed of an equilibrated and flowing A5M column, followed by dropwise addition of EMTE to the column. Purified samples were recovered in fractions obtained immediately after an approximately 650 μ l void volume. In preparation for examination of samples by electron microscopy, carbon foil coated copper grids were subjected to glow discharge in a modified DV-502A (Denton) to render the carbon hydrophilic. Purified samples and samples not requiring purification were then countercharged via addition of 1x spermidine buffer (2.5 mM spermidine, 75 mM NaCl, 50 mM KCl, 10 mM Tris-HCl at pH 7.5, 1 mM MgCl₂, 0.25 mM CaCl₂). Countercharged samples were then immediately applied to charged carbon grids and incubated at RT for 3 min, then briefly washed in distilled water, then incubated in a separate distilled water bath for 3 min. The samples were then dehydrated in successive 25%, 50%, 75%, and 95% (v/v) ethanol baths for 5 min per bath. Samples were then air dried and tungsten shadow cast in a modified DV-502a (Denton) equipped with a rotating stage. Shadowed grids were then stored under vacuum at room temperature until examination using either a Tecnai-12 (FEI) or a CM-12 (Phillips) transmission electron microscope.

Dynamic Light Scattering

The samples were injected into a Superdex 75 column followed by two 5 ml Sephadex G50 columns (GE Healthcare). The column outlet was connected to a Dawn EOS multiangle, static, light scattering detector (Wyatt Technology Corporation, Santa Barbara, CA). Average molar mass measurements were determined from aligned elution profiles within ASTRA for Windows software (version 5.3.4.16; Wyatt), using the Debye plot. Detectors 3 to 8 were used.

RESULTS:

hnRNP A1 & UP1 bind to ssDNA

We sought to validate the DNA binding activities of our purified hnRNP A1 via a combination of electrophoretic mobility shift assays (EMSAs) and electron microscopy (EM). hnRNP A1 binding to M13

ssDNA was observed to be non-cooperative. As successively higher concentrations of hnRNP A1 were incubated with M13 ssDNA the ssDNA was shifted into complexes of progressively lower electrophoretic mobility, consistent with stable hnRNP A1 binding (Figure 5.1). The detected bands were not derived from contaminating nucleic acids, as a reaction containing high concentrations of hnRNP A1 but no M13 yielded no detectable band (Figure 5.1, lane 2). The observed shifts in electrophoretic mobility were protein mediated, as subsequent incubation of a reaction containing 300 µg/ml hnRNP A1 with SDS and proteinase K increased abolished the observed mobility shift (Figure 5.1, lane 9). UP1 exhibited broadly comparable binding activities (data not shown).

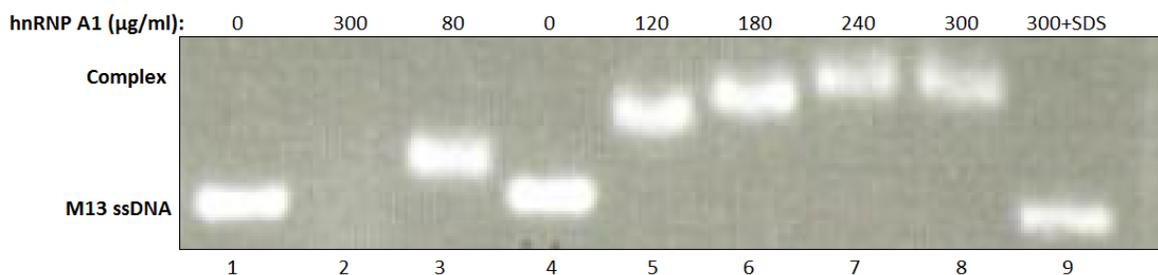


Figure 5.1 Electrophoretic mobility shift assay demonstrating hnRNP A1 ssDNA binding. hnRNP A1 induces progressively severe retardation of M13 electrophoretic mobility with increasing protein concentration (lanes 2-8). The observed bands consist of M13 ssDNA, as the band is not observed when M13 is omitted even in the presence of high concentrations of hnRNP A1 (lane 2). The observed shift in mobility is protein-mediated, as the shift can be abolished upon incubation with SDS and proteinase K (lane 9).

EM examination revealed that hnRNP A1 bound to M13 as large, heterogenous, and occasionally oligomeric complexes (Figure 5.2). While hnRNP A1 binding to M13 did occasionally generate an open structure reminiscent of that formed when *E.coli* single-stranded DNA binding proteins binds to M13 (e.g. Figure 5.2C), most particles were condensed and not amenable to examination by electron microscopy. In solution, hnRNP A1 appeared as a large quasi-spherical particle, with an apparent volume greater than anticipated for its hypothetical molecular weight of 38 kDa (Figure 5.2,B).

It has previously been reported the hnRNP A1 exhibits binding cooperatively¹⁰, but that this binding cooperativity is lost upon freezing or prolonged storage¹¹. The reason for this change in binding

properties has not been described in available literature, but may involve protein aggregation or denaturation. To investigate possible aggregation or oligomerization of hnRNP A1, we submitted a quantity of hnRNP A1 to Dr. Mehmet Kesimer for analysis via dynamic light scattering (DLS). The obtained DLS data was of low quality, possibly stemming from low protein concentration and inefficient chromatographic separation of the protein species (Figure 5.3) However, the data was still interpretable and suggested that in solution our hnRNP A1 existed predominately as monomeric protein with an estimated mass of 46 ± 3 kDa (Peak 1). While this is greater than the expected mass of 38 kDa, this number is affected by the low quality of the data. hnRNP A1 was also found to exist as an approximately 640 kDa complex (Peak 2), possibly due to the presence of some residual contaminating nucleic acids in the purified proteins or DNA independent hnRNP A1 oligomerization. The presence of contaminating nucleic acids would be expected to nucleate binding of hnRNP A1 and permit the formation of larger complexes. This second peak was calculated to constitute approximately 21% of the total protein in solution, but as this peak overlaps the shoulder of Peak 1, such quantification is likely overestimated. Finally, DLS also reported the existence of a peak of very high mass, in excess of several megadaltons (Peak 3). As with peak 3, this may be due to the presence of contaminating nucleic acids or it may represent a very high mass contaminant not removed by the hnRNP A1 purification procedure. The calculated masses and abundances of these proteins must be taken with caution, as they are substantially affected by analysis. Subsequent re-analysis suggested Peak 1 had a mass of 37 kDa, much closer to the expected 38 kDa mass (Dr. Kesimer, personal communications). Re-analysis also reduced the expected abundance of Peak 3 protein to approximately 20 nanograms. Regardless, these data suggest that the purified hnRNP A1 did not suffer from high amounts of aggregation during purification or storage.

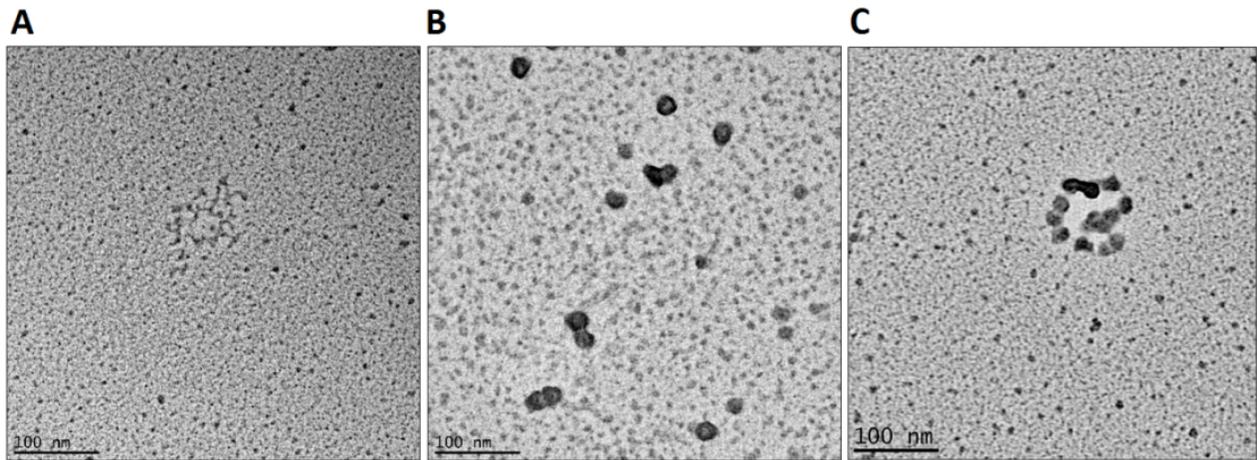


Figure 5.2: hnRNP A1 binds to M13 ssDNA. (A) M13ssDNA. **(B)** hnRNP A1 not bound to M13. **(C)** hnRNP A1 bound to M13 ssDNA.

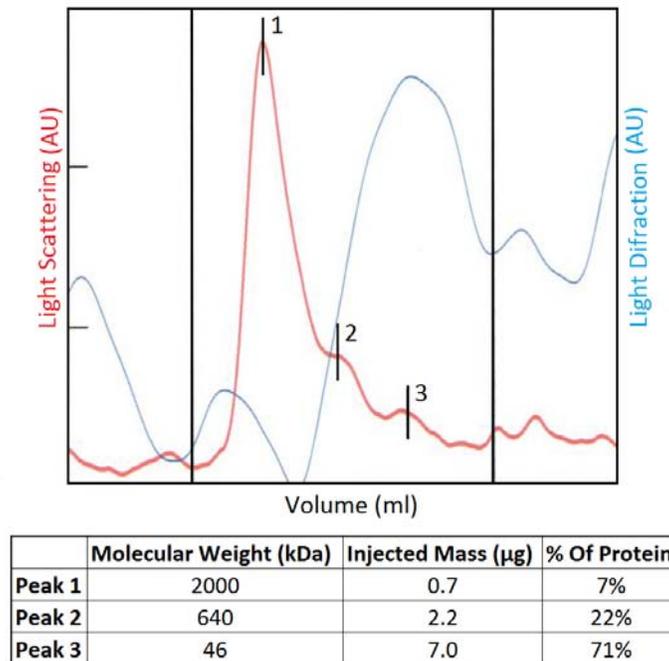


Figure 5.3 Dynamic light scattering reveals mostly homogenous hnRNP A1. hnRNP A1 was observed to exist as three populations of particles. Peak 1 possessed an apparent molecular weight of in excess of 2 megadaltons, but accounted for less than 10% of the total protein in solution. Peak 2 possessed an apparent mass of 640 kilodaltons and accounted for 22% of the total protein. Peak 3 possessed an apparent mass of 46 kilodaltons and accounted for about 71% of total protein.

hnRNP A1 binds preferentially to telomeric ssDNA

We next sought to investigate the binding characteristics of hnRNP A1 and UP1 on templates containing telomeric and non-telomeric ssDNA and dsDNAs. hnRNP A1 binding was observed to bind to strand-displaced telomeric DNA (Figure 5.4B) as a zone of higher electron density not observed in the absence of hnRNP A1 (Figure 5.4D). In contrast, hnRNP A1 was not observed to bind significantly to strand-displaced non-telomeric ssDNA (Figure 5.4A). However, a small zone of high electron density was observed on strand-displaced non-telomeric ssDNA in the absence of hnRNP A1 (Figure 5.4C), that was comparable to that observed in the presence of hnRNP A1.

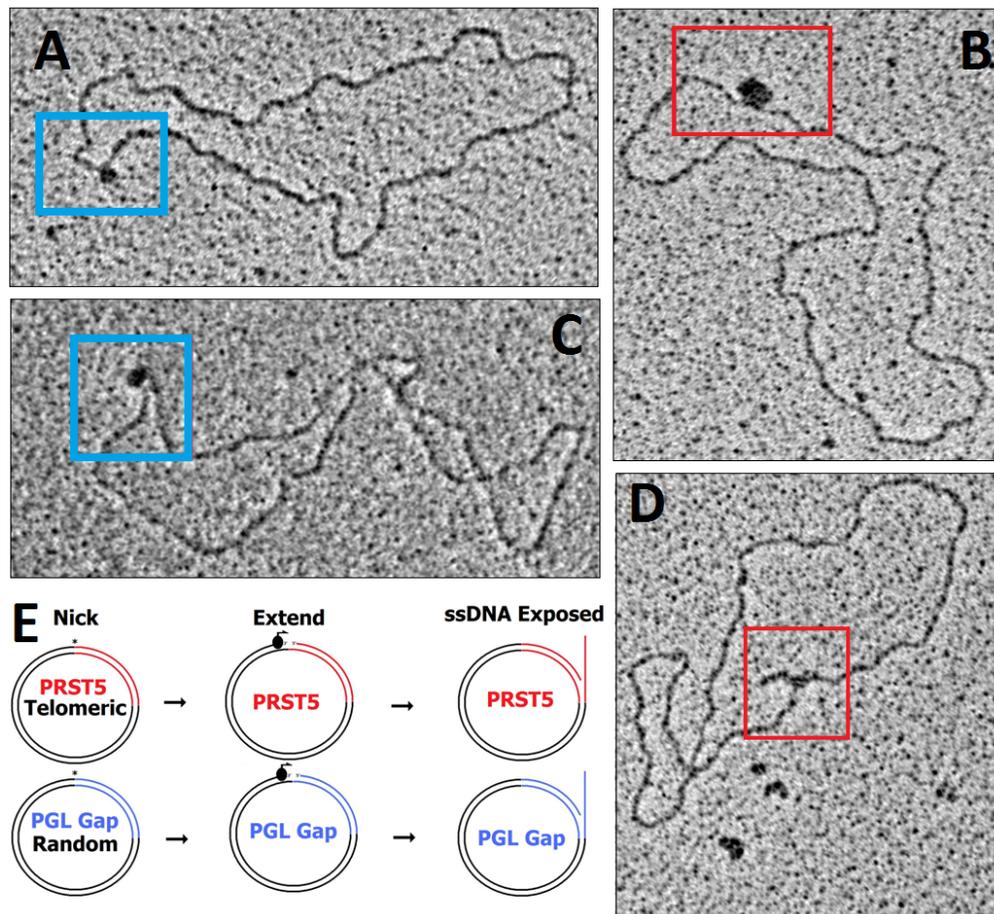


Figure 5.4 hnRNP A1 binds to strand displaced telomeric but not non-telomeric DNA. (A) hnRNP A1 does not bind to strand displaced non-telomeric ssDNA. (B) hnRNP A1 does not bind to strand displaced non-telomeric ssDNA. (C) hnRNP A1 does not bind to strand displaced non-telomeric ssDNA. (D) hnRNP A1 does not bind to strand displaced non-telomeric ssDNA. (E) Diagram of strand displaced templates; templates are nicked with Nb. Bbvcl then strand displaced with Klenow Exo(-) with dATP, dTTP and dGTP. Strand displacement stops at a defined nucleotide leaving ssDNA tails of defined length.

To investigate hnRNP A1's apparent binding preference for telomeric ssDNA, we tested its binding to a linear dsDNA possessing telomeric 3' ssDNA tails of telomeric or non-telomeric sequence on either end (Figure 5.5C). At a fixed concentration, hnRNP A1 was observed to bind to one end of the template more frequently than would be expected if hnRNP A1 binding to the two ssDNA tails was non-specific (Figure 5.5A). Likewise, hnRNP A1 bound to both ends less frequently than would be expected if hnRNP A1 binding was non-specific. These data suggest that hnRNP A1 binds preferentially to one end of the template, through this experiment did not explicitly identify which.

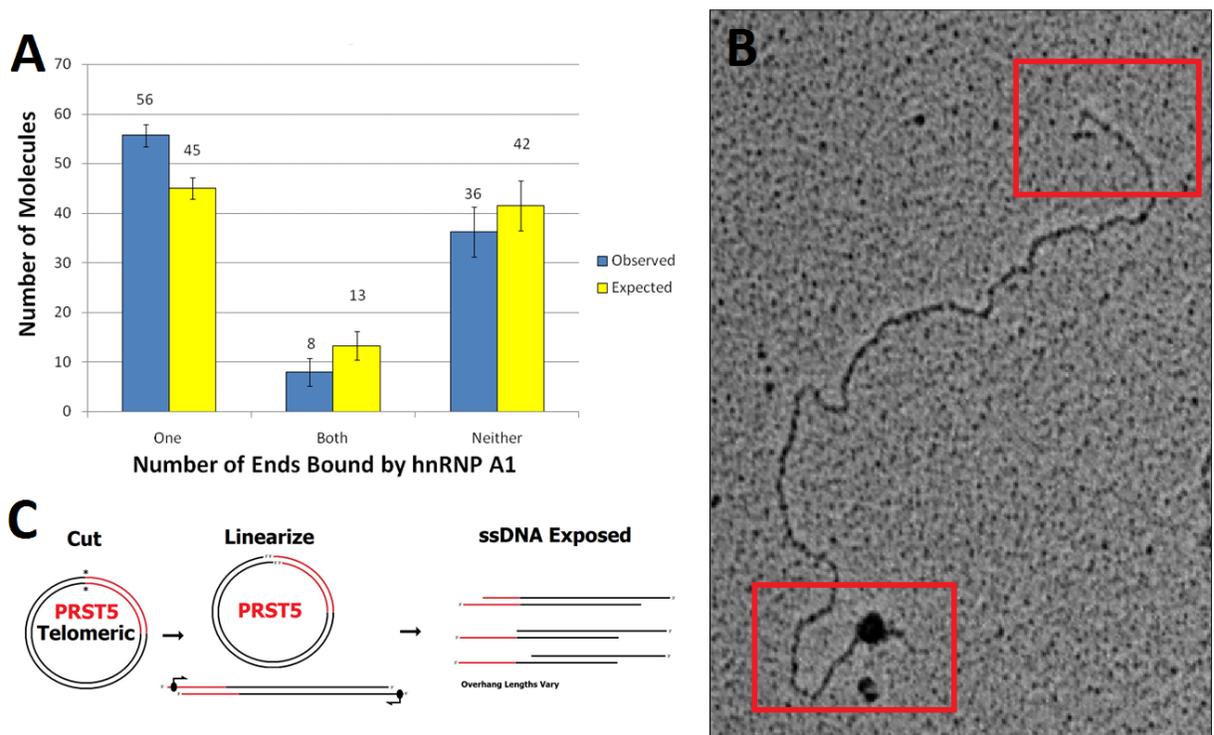


Figure 5.5 hnRNP A1 binds preferentially to one end of a template with telomeric and non-telomeric ssDNA on opposite sides of a linear template. (A) Abundance of particles with hnRNP A1 bound to one, both or neither compared with those expected if binding was random. (B) particle with an example of hnRNP A1 binding on one end (bottom) and not on the other end (top). (C) Diagram of template; plasmid it cut with BsmBI then exonucleolytically digested with Exonuclease 3 to expose ssDNA.

In a subsequent experiment we characterized the binding of hnRNP A1 and SSB to similar templates possessing only a 3' telomeric ssDNA tail (Figure 5.6). On these templates, the tails could be readily detected by the binding of *E. coli* single-stranded DNA binding protein (SSB) (Figure 5.6A),

which stretched the ssDNA to form a filamentous structure. In contrast, hnRNP A1 binding resulted in the formation of the ball-like complexes previously observed (Figure 5.6B). These structures were distinctly different from those observed on the template in the absence of any added protein (Figure 5.6C).

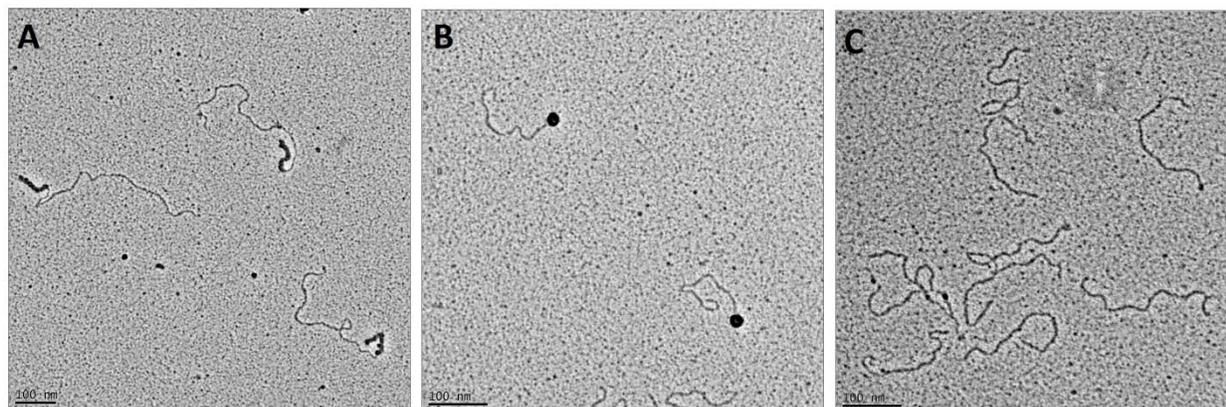


Figure 5.6 Comparative binding of *E. coli* and hnRNP A1 to exposed telomeric ssDNA. (A) *E. coli* SSB bound to exposed telomeric ssDNA. (B) hnRNP A1 bound to exposed telomeric ssDNA. (C) Templates with exposed telomeric ssDNA.

We attempted to optimize buffer conditions to generate more specific hnRNP A1 binding by varying salt, glycerol and dithiothreitol (DTT) concentration (Figure 5.7). Increasing the sodium chloride concentration in a reaction buffer containing 25 mM HEPES and 0 mM DTT (Figure 5.7A) to 0.5 mM (Figure 5.7B), 2.5 mM (Figure 5.7C) or 5 mM (Figure 5.7D) seemed to result in greater hnRNP A1 binding but not the formation of any interesting binding structures. Likewise increasing the glycerol concentration to 1% (Figure 5.7E), 5% (Figure 5.7F) or 10% (Figure 5.7G) did not appear to significantly affect hnRNP A1 binding. Finally, similarly increasing the sodium chloride concentration to 100 mM (Figure 5.7H), 200 mM (Figure 5.7I) or 300 mM (Figure 5.7E) suggested that inclusion of 200 mM NaCl may improve hnRNP A1 binding, though this trend was not clear.

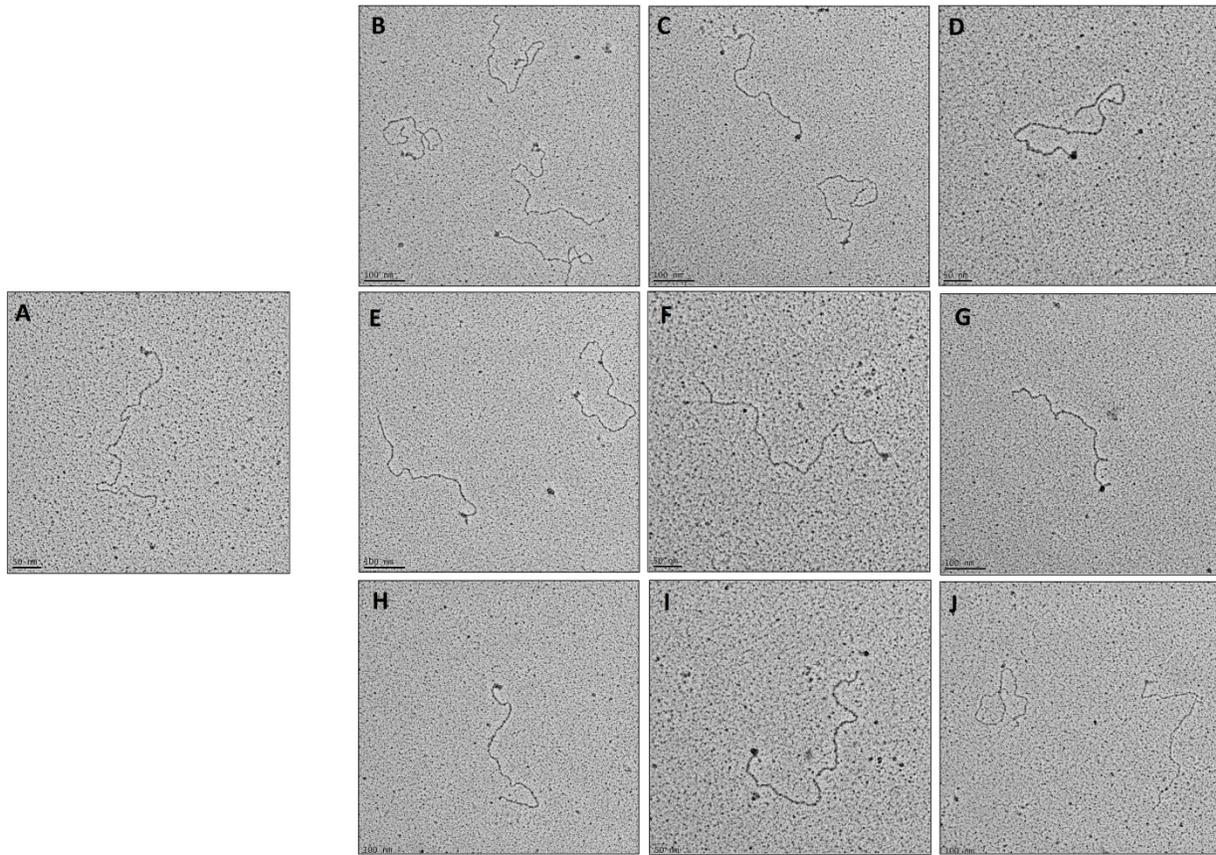


Figure 5.7: Optimization of hnRNP A1 binding conditions. (A) binding in 25 mM HEPES. (B) +0.5 mM DTT (C) +2.5 mM DTT. (D) +5 mM DTT. (E) +1% Glycerol. (F) +5% Glycerol. (G) +10% Glycerol. (H) +100 mM NaCl. (I) +200 mM NaCl. (J) +300 mM NaCl.

We likewise attempted to optimize buffer conditions to generate specific UP1 binding by varying salt, glycerol and dithiothreitol (DTT) concentration (Figure 5.7). Increasing the sodium chloride concentration in a reaction buffer containing 25 mM HEPES and 0 mM DTT (Figure 5.8A) to 2.5 mM (Figure 5.8B) or 5 mM (Figure 5.8C) did not appear to affect UP1 binding. Likewise increasing the glycerol concentration to 1% (Figure 5.8D), 5% (Figure 5.8E) or 10% (figure 5.8F) did not appear to significantly affect UP1 binding. Finally, as with hnRNP A1 increasing the sodium chloride concentration to 100 mM (Figure 5.8G), 200 mM (Figure 5.8H) or 300mM (figure 5.8I) suggested that inclusion of 200 mM NaCl may improve UP1 binding, though this trend was not clear.

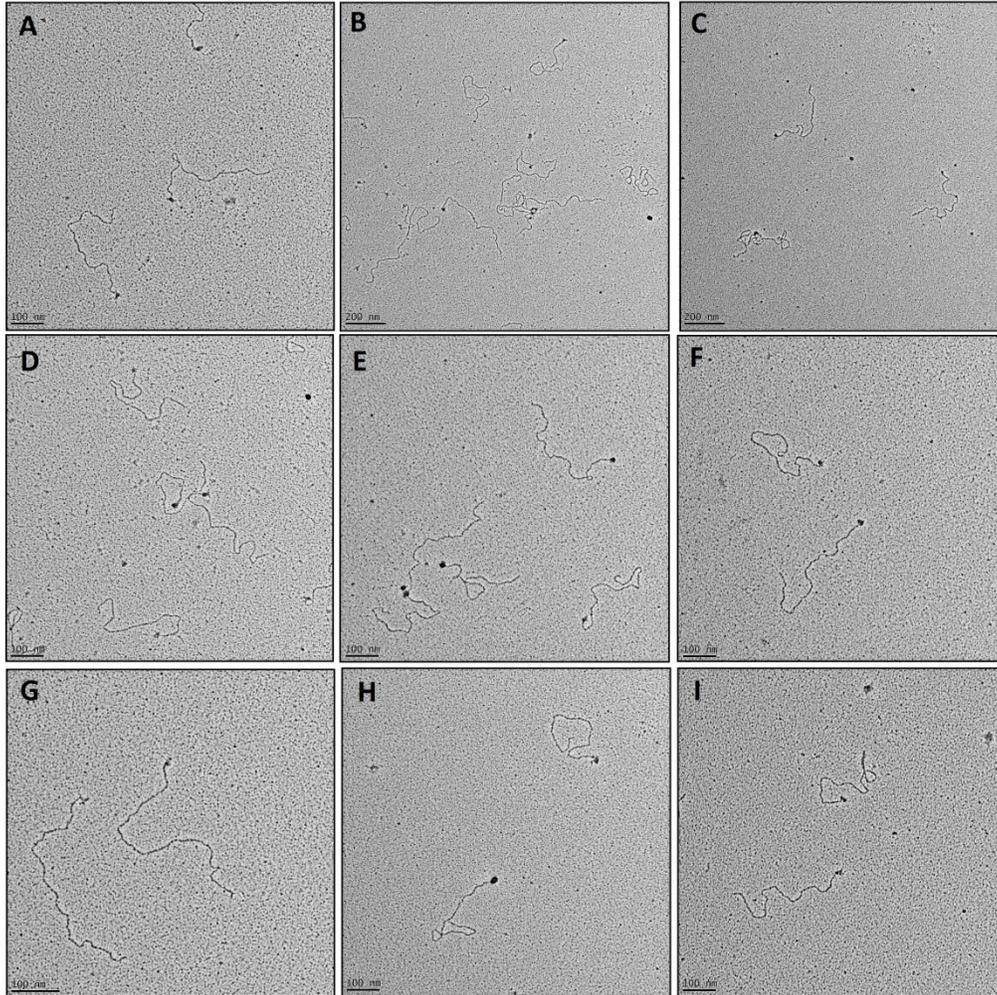


Figure 5.8: Optimization of UP1 binding conditions. (A) binding in 25 mM HEPES. **(B)** +2.5 mM DTT **(C)** +5 mM DTT. **(D)** +1% Glycerol. **(E)** +5% Glycerol. **(F)** +10% Glycerol. **(G)** +100 mM NaCl. **(H)** +200 mM NaCl. **(I)** +300 mM NaCl.

Possible deleterious effects of N-terminal tagging

Dr. Adrian Krainer reported that the binding and catalytic activities of hnRNP A1 may be disrupted by N-terminal tagging (personal communication). hnRNP A1 and UP1 binding are mediated primarily by two RNA recognition motifs (RRMs) in the N-terminal region of the protein^{12, 13}. While the amino terminus of hnRNP A1 and UP1 are solvent exposed and separated from the RRM domains spatially, the addition of a 3.5 kDa hexahistidine tagged linker may disrupt either the binding of these domains or affect the oligomerization of these proteins. Additionally, hnRNP A1 is post-translationally

modified *in vivo*, via processes that are not present in *E. coli*, but which are active in eukaryotic expression systems such as insect cells (reviewed in¹). We did not observe qualitative differences in binding of n-terminally hexahistidine tagged *E. coli* purified hnRNP A1 or untagged baculovirus purified hnRNP A1 obtained from Dr. Krainer, though our data was not of sufficient quality to discount such a possibility (data not shown).

DISCUSSION

Our attempt at characterizing the DNA binding characteristics of hnRNP A1 and UP1 *in vitro* using electron microscopy did not yield data of sufficient quality for publication. Despite numerous attempts at optimizing buffer and reaction conditions, to promote specific binding of these proteins into discernable structures, the proteins always bound as large complexes. These complexes may represent physiologically relevant oligomerization, but they are also consistent with undesirable aggregation. hnRNP A1 has been reported to bind cooperatively on both RNA¹⁰ and DNA¹¹, but the solution structure of full-length hnRNP A21 has not been solved and oligomerization domains for hnRNP A1 have not been identified or characterized¹⁴. As such, it remains unclear whether hnRNP A1 does indeed multimerize in solution, or upon binding and likewise whether this multimerization is required for hnRNP A1's binding activities.

While it is possible that further optimization may have permitted us to generate higher quality data, this project was terminated in large part due to competition from other laboratories. In March 2011 a collaborative effort involving the laboratories of Dr. Sandy Change, Dr. Jan Karlseder, Dr. Zhou Songyang reported that hnRNP A1 and telomere derived RNA (TERRA) may facilitate removal of RPA from telomeric DNA to permit POT1 binding⁶. As our preliminary results did not generate any data that was compelling we chose to put the project on hold.

Sometime later, we received word from our collaborator, Dr. Adrian Krainer, that our N-terminal tagging of hnRNP A1 and UP1 may have adversely affected their activities. Dr. Krainer's research has long focused on *E. coli* purified but untagged hnRNP A1¹⁶. While it is plausible that our N-terminal

tagging strategy did in fact adversely affect hnRNP A1's binding activities, our purified UP1 proved to be active as an RNA chaperone in a separate project⁷. Whether this is because the binding activities of hnRNP A1 and its proteolytic fragment, UP1, are different or whether the chaperone activities and binding activities are independent of each other is unclear. While we could have re-purified hnRNP A1 and UP1 using techniques similar to Dr. Krainer's, it remained unclear whether we could actually use EM to answer any scientifically interesting questions.

REFERENCES

- [1] Jean-Philippe, J., Paz, S., and Caputi, M. (2013) hnRNP A1: the Swiss army knife of gene expression, *Int J Mol Sci* 14, 18999-19024.
- [2] Dallaire, F., Dupuis, S., Fiset, S., and Chabot, B. (2000) Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro, *J Biol Chem* 275, 14509-14516.
- [3] LaBranche, H., Dupuis, S., Ben-David, Y., Bani, M. R., Wellinger, R. J., and Chabot, B. (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase, *Nat Genet* 19, 199-202.
- [4] Zhang, Q. S., Manche, L., Xu, R. M., and Krainer, A. R. (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity, *RNA* 12, 1116-1128.
- [5] Choi, Y. H., Lim, J. K., Jeong, M. W., and Kim, K. T. (2012) HnRNP A1 phosphorylated by VRK1 stimulates telomerase and its binding to telomeric DNA sequence, *Nucleic Acids Res* 40, 8499-8518.
- [6] Flynn, R. L., Centore, R. C., O'Sullivan, R. J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J., and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA, *Nature* 471, 532-536.
- [7] Grohman, J. K., Gorelick, R. J., Lickwar, C. R., Lieb, J. D., Bower, B. D., Znosko, B. M., and Weeks, K. M. (2013) A guanosine-centric mechanism for RNA chaperone function, *Science* 340, 190-195.
- [8] Stansel, R. M., de Lange, T., and Griffith, J. D. (2001) T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang, *EMBO J* 20, 5532-5540.
- [9] Fouche, N., Ozgur, S., Roy, D., and Griffith, J. D. (2006) Replication fork regression in repetitive DNAs, *Nucleic Acids Res* 34, 6044-6050.
- [10] Okunola, H. L., and Krainer, A. R. (2009) Cooperative-binding and splicing-repressive properties of hnRNP A1, *Mol Cell Biol* 29, 5620-5631.
- [11] Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., and Wilson, S. H. (1988) Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in *Escherichia coli* and cooperative binding to single-stranded nucleic acids, *J Biol Chem* 263, 1063-1071.
- [12] Joka, L., Dong, A. P., Mayeda, A., Krainer, A. R., and Xu, R. M. (1997) Crystallization and preliminary X-ray diffraction studies of UP1, the two-RRM domain of hnRNP A1, *Acta Crystallogr D Biol Crystallogr* 53, 615-618.
- [13] Xu, R. M., Joka, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs, *Structure* 5, 559-570.
- [14] Mayeda, A., Munroe, S. H., Caceres, J. F., and Krainer, A. R. (1994) Function of Conserved Domains of HnRNP A1 and Other HnRNP a/B Proteins, *Embo Journal* 13, 5483-5495.

- [15] Redon, S., Zemp, I., and Lingner, J. (2013) A three-state model for the regulation of telomerase by TERRA and hnRNPA1, *Nucleic Acids Res* 41, 9117-9128.
- [16] Mayeda, A., and Krainer, A. R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2, *Cell* 68, 365-375.

Chapter 6: Conclusions and Final Thoughts

Presented in this thesis are data from selected work from my tenure as a graduate student in the laboratory of Dr. Jack Griffith, at the University of North Carolina at Chapel Hill. The overriding goal of my dissertation was to gain an understanding of functional interactions between TRF2 and Rad51 recombinase. I chose to focus on the *in vitro* interaction of these proteins in a displacement loop (D-loop) assay. The activity of Rad51 has been characterized using *in vitro* D-loop and strand displacement assays, and the activities of TRF2 had previously been characterized using a similar D-loop assay by the laboratory of Dr. Eric Gilson and in cell extracts by Dr. Jan Karlseder. Prior to undertaking this characterization a student, Ravindra Amunugama, of our collaborator, Dr. Rick Fishel, reported modifications to established D-loop protocols that increased the efficiency of those reactions. This made it plausible to examine the functional interaction of TRF2 and Rad51 in a D-loop assay modified from previous TRF2 and Rad51 characterizations. Using this assay we sought to determine *in vitro* whether TRF2 and Rad51 exhibit functional cooperation, as suggested by earlier work from the laboratory of Dr. Karlseder, or whether TRF2 and Rad51 may exhibit functional antagonism, which might be expected given their previously characterized *in vitro* activities. Dr. Gilson had reported that TRF2 promotes telomeric D-loop formation by promoting positive supercoiling within telomeric dsDNA, while Rad51 has previously been reported to promote D-loop formation optimally on negatively supercoiled DNA.

Homologous recombination/repair in telomere maintenance & protection

The results of our characterization of the functional interaction of TRF1, TRF2 and Rad51 include several novel findings that offer insight into how HR may promote telomere maintenance and protection, and how TRF2 may promote non-telomeric DSB repair. These results agree with some

previous findings and contradict others, and suggest that the activities of these proteins *in vivo* must be modulated by factors absent from our and other *in vitro* characterizations.

Reviewers of our paper raised the criticism that our results appeared to contradict a report from Dr. Karlseder's lab that Rad51 and TRF2 functionally cooperate to promote telomeric D-loop formation in nuclear extracts¹. However, it is difficult to compare the results of these assays for methodological reasons. Most importantly, nuclear extracts contain a host of proteins that may influence the activities of TRF2 and Rad51 that were absent from our characterization. This notion is supported by a recent report that TRF2 inhibits HR mediated processes at humanized yeast telomeres². Determining the reason or reasons for these discrepancies will require a more-detailed understanding of the mechanisms of the functional interactions between TRF2 and proteins in the HR pathway.

Our results revealed that TRF1, TRF2 and Rad51 exhibit functional interactions. However, we were unable to identify specific domains that mediate these interactions. We showed that TRF2 inhibited Rad51 mediated telomeric D-loop formation in a Myb domain dependent manner. However, we also observed that TRF1, which possesses a homologous Myb domain to TRF2, actually promoted Rad51-mediated telomeric D-loop formation. In contrast, we observed that the basic domain was dispensable for TRF2's inhibition of Rad51 mediated telomeric D-loop formation. Taken together these data suggest that TRF2's Myb domain is necessary but not sufficient for inhibiting Rad51-mediated telomeric D-loop formation. It appears likely that the Myb domain is instead required for the recruitment of some portion of the linked or dimerization domain that itself inhibits Rad51-mediated telomeric D-loop formation. Work from Dr. Eric Gilson's lab suggests that TRF2's dimerization domain mediates TRF2's ability to promote positive supercoiling within telomeric dsDNA³. As Rad51 promotes D-loop formation most efficiently when acting on negatively supercoiled templates, we speculate that TRF2's ability to promote positive supercoiling may also be required for its inhibition of Rad51-mediated telomeric D-loop formation. However, the dimerization domain is unamenable to mutation and loss of dimerization also disrupts Myb domain binding. We tested the ability of a TRF2 mutant allegedly hypomorphic for supercoiling induction but competent for dimerization, which was described by Dr. Gilson at the 2013 CSH meeting

on Telomeres and Telomerase, to inhibit Rad51-mediated telomeric D-loop formation. However, this mutant inhibited Rad51-mediated telomeric D-loop formation in a manner comparable to TRF2. Ultimately, in the future identification of the domain that mediates TRF2's inhibition of Rad51-mediated telomeric D-loop formation may require cloning and expressing additional TRF1 and TRF2 dimerization and linker domain mutant proteins, possible including domain swap proteins.

The fluorescent D-loop assay we have used in this assay was adapted from previous characterization of both Rad51⁴ and TRF2³, with advice from Dr. Amunugama. We chose to modify the previously described techniques mostly due to the tedious nature of using conventional radiolabeled D-loop assays. Such assays use ³²P labeled oligonucleotides, which must be of very high specific activity to generate satisfactory results. As such the labeled oligonucleotides must be used almost immediately after labeling and must be frequently re-made. Additionally, the agarose gels used for such assays must be dried prior to imaging, and we found that the drying process tended to promote diffusion of the oligo throughout the gel and that the oligo was poorly retained within the gel throughout the drying process. Additionally, imaging the gels required a several hour or overnight exposure. We circumvented all of these technical issues by using commercially purified Cy3 labeled oligo, which did not decay and which could be rapidly imaged in an undried gel. This innovation allowed us to greatly increase the throughput of our experiments. This assay could be modified to conduct characterizations of functional interactions between Rad51 and BRCA2 or the Rad51 paralogs, topics which are of considerable interest.

Intra-capsid AAV genome organization is unamenable to EM characterization

During my collaboration with Dr. Horowitz I attempted to characterize the intra-capsid organization of self-complimentary AAV genomes using electron microscopy using several techniques. These characterizations did not yield usable results, but were complicated by lack of a good positive control. The most promising of our protocols involved using psoralen and ultraviolet A light (PUVA) in an attempt to crosslink portions the self-complimentary AAV genome in an intact viral capsid. After PUVA treatment we then thermally denatured the virions at high temperatures in the presence of an

extremely thermostable single-stranded DNA binding protein (ETSSB), which would immediately bind to any available ssDNA in the scAAV genomes. In a negative control which was not UV irradiated and where we used no SSB we slowly cooled the DNA to permit proper base pairing after denaturation. The negative control was observed to consist of fully double-stranded DNA, as expected. A positive control that was not UV irradiated but was exposed to ETSSB was observed to consist of fully ET SSB coated DNA, as expected. Experimental samples that were UV irradiated following treatment with varying concentrations of psoralen and exposed to ETSSB were indistinguishable from the positive control. These data suggested that the PUVA treatment failed to crosslink the scAAV genomes inside the capsid proteins. However, we failed to run a control to demonstrate the PUVA treatment could crosslink a non-encapsidated scAAV genome. To do this we should have thermally denatured and renatured scAAV genomes, as in our negative control, then subsequently subjected them to PUVA treatment, denaturation and SSB. This control would be expected to show evidence of PUVA crosslinking, which would manifest as incomplete ETSSB binding along the scAAV genome.

Ultimately, our collaborators decided not to repeat this experiment with our recommended controls, likely because data from a discrete molecular dynamics (DMD) model suggested that the scAAV genomes were partially double stranded when packaged in the viral capsid. However, there was evidence in the literature suggesting the PUVA approach had merit¹². Psoralen has been reported to potentiate UV inactivation of AAV, suggesting that the AAV genome is capable of at-least limited base pairing inside the capsid and that the capsid is indeed psoralen permeable¹². As the DMD model used in the publication possessed numerous simplifications¹³, it may prove worthwhile to validate the finding using the PUVA technique. This experiment would require collaboration with the Laboratories of Dr. Aravind Asokan or another AAV researcher as the quantity of AAV required for the PUVA procedure is non-trivial. However, such an undertaking is of defined and limited scope and short duration. It would require little if anything in the way of new reagents and would only require optimization of the PUVA procedure and validation of the appropriate positive and negative controls.

REFERENCES

- [1] Verdun, R. E., and Karlseder, J. (2006) The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres, *Cell* 127, 709-720.
- [2] Saint-Leger, A., Koelblen, M., Civitelli, L., Bah, A., Djerbi, N., Giraud-Panis, M. J., Londono-Vallejo, A., Ascenzioni, F., and Gilson, E. (2014) The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres, *Cell Cycle* 13.
- [3] Amiard, S., Doudeau, M., Pinte, S., Poulet, A., Lenain, C., Faivre-Moskalenko, C., Angelov, D., Hug, N., Vindigni, A., Bouvet, P., Paoletti, J., Gilson, E., and Giraud-Panis, M. J. (2007) A topological mechanism for TRF2-enhanced strand invasion, *Nat Struct Mol Biol* 14, 147-154.
- [4] Amunugama, R., He, Y., Willcox, S., Forties, R. A., Shim, K. S., Bundschuh, R., Luo, Y., Griffith, J., and Fishel, R. (2012) RAD51 protein ATP cap regulates nucleoprotein filament stability, *J Biol Chem* 287, 8724-8736.
- [5] Tan, J. H., Wooley, J. C., and LeStourgeon, W. M. (2000) Nuclear matrix-like filaments and fibrogranular complexes form through the rearrangement of specific nuclear ribonucleoproteins, *Mol Biol Cell* 11, 1547-1554.
- [6] Flynn, R. L., Centore, R. C., O'Sullivan, R. J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J., and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA, *Nature* 471, 532-536.
- [7] Huang, P. R., Hung, S. C., and Wang, T. C. (2010) Telomeric DNA-binding activities of heterogeneous nuclear ribonucleoprotein A3 in vitro and in vivo, *Biochim Biophys Acta* 1803, 1164-1174.
- [8] Kruger, A. C., Raarup, M. K., Nielsen, M. M., Kristensen, M., Besenbacher, F., Kjems, J., and Birkedal, V. (2010) Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level, *Eur Biophys J* 39, 1343-1350.
- [9] Ting, N. S., Pohorelic, B., Yu, Y., Lees-Miller, S. P., and Beattie, T. L. (2009) The human telomerase RNA component, hTR, activates the DNA-dependent protein kinase to phosphorylate heterogeneous nuclear ribonucleoprotein A1, *Nucleic Acids Res* 37, 6105-6115.
- [10] Le, P. N., Maranon, D. G., Altina, N. H., Battaglia, C. L., and Bailey, S. M. (2013) TERRA, hnRNP A1, and DNA-PKcs Interactions at Human Telomeres, *Front Oncol* 3, 91.
- [11] Redon, S., Zemp, I., and Lingner, J. (2013) A three-state model for the regulation of telomerase by TERRA and hnRNPA1, *Nucleic Acids Res* 41, 9117-9128.
- [12] Callahan, S. M., Wonganan, P., Obenauer-Kutner, L. J., Sutjipto, S., Dekker, J. D., and Croyle, M. A. (2008) Controlled inactivation of recombinant viruses with vitamin B2, *J Virol Methods* 148, 132-145.
- [13] Horowitz, E. D., Rahman, K. S., Bower, B. D., Dismuke, D. J., Falvo, M. R., Griffith, J. D., Harvey, S. C., and Asokan, A. (2013) Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release, *J Virol* 87, 2994-3002.