STRUCTURAL DYNAMICS OF RETROVIRAL GENOMES

Jacob K. Grohman

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 Department of Biochemistry and Biophysics.

Chapel Hill
2012

Approved by:

Kevin M. Weeks, Ph.D.

Howard M. Fried, Ph.D.

Jason D. Lieb, Ph.D.

Nancy Allbritton, Ph.D.

Ron Swanstrom, Ph.D.

ABSTRACT

JACOB K. GROHMAN: Structural Dynamics of Retroviral Genomes
(Under the direction of Kevin M. Weeks)

Retroviral RNA genomes form myriad structures that are governed by critical interactions with either the nucleocapsid (NC) protein or the nucleocapsid domain of Gag polyprotein. I have pioneered two powerful technologies, based on SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension), to explore interactions between NC and the RNA genomes of the xenotropic murine leukemia virus related virus (XMRV) and the Moloney murine leukemia virus (MuLV). In the first approach, ultra-sensitive SHAPE, I developed a two-color capillary electrophoresis instrument with low attomole (10^{-18}) sensitivity. New analysis approaches and the high sensitivity of this instrument allowed me to obtain unprecedented single-nucleotide resolution structures from authentic RNA genomes of MuLV and XMRV. In the second approach, time-resolved SHAPE, I used a fast acting SHAPE reagent to reveal the structural biogenesis of NC binding to the MuLV retroviral RNA genome. The data generated using these SHAPE-based technologies allow me to propose detailed mechanisms for both the specific RNA binding and the chaperone activities of NC.

"We're surrounded. That simplifies the problem!"

-Chesty Puller, USMC

ACKNOWLEDGMENTS

I'd like to thank my girls, Perla and Madeline, with their love and support anything is possible.

Special thanks to my advisor and mentor, Kevin M. Weeks.

As always, I give thanks to God, country, and Corps.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

ºC	degree Celsius
μ_{ep}	electrophoretic velocity
μ_{eo}	electroosmotic velocity
µL	microliter
µM	micromolar
πr^2	capillary area
1M7	1-methyl-7-nitroisatoic anhydride
5-FAM	5 - Carboxyfluorescein
5'	The end of a linear polynucleotide strand at which the 5' hydroxyl group of the terminal nucleoside residue is normally phosphorylated.
6-JOE	6 - Carboxy - 4',5' - dichloro - 2',7' - dimethoxyfluorescein, succinimidyl ester
A	adenine
amol	attomole (10^{-18} moles)
BSA	bovine serum albumin
BzCN	benzoyl cyanide
C	input RNA concentration
cDNA	complementary DNA, single-stranded DNA formed from reverse transcription.
CE	capillary electrophoresis
DTT	dithriotreitol
E	electric field strength
EDTA	ethylenediaminetetraacetic acid
Ex virio RNA	Latin for experiments performed on gently extracted viral RNA
fmol	femtomole (10^{-15} moles)

G	guanosine
Gag	Gag polyprotein
HCl	hydrochloric acid
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HIV	Human Immunodeficiency Virus
HTLV	Human T-cell Leukemia Virus
In vitro	Latin for experiments performed in a test tube
In virio	Latin for experiments performed within authentic viruses
L	liter
Mg^{2+}	magnesium ion
$MgCl_2$	magnesium chloride
min	minute
mM	millimolar
mRNA	messenger RNA
MuLV	Moloney murine leukemia virus
NC	Nucleocapsid protein
NaCl	sodium chloride
NMR	nuclear magnetic resonance
nt	nucleotide
NTP	nucleotide triphosphate
PAL	Palindrome
PCR	polymerase chain reaction
pmols	picomoles (10^{-12} moles)
PPV	positive predictive value
Q	quantity injected
R, R^2	correlation coefficient

RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
s	second
SHAPE	Selective 2'-Hydroxyl Acylation analyzed by Primer Extension
SL	stem loop
T	thymidine
t_i	injection time
t_s	separation time, between cDNA injection and peak detection
TR-SHAPE	time-resolved selective 2'-hydroxyl acylation analyzed by primer extension
TBE	90 mM Tris-borate, 2 mM EDTA
TE	10 mM Tris (pH 7.5), 1 mM EDTA
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
US SHAPE	ultra-sensitive selective 2'-hydroxyl acylation analyzed by primer extension
U	uracil
v	volume
w	weight
XMRV	xenotropic murine leukemia virus-related virus
zmol	zeptomole (10^{-21} moles)

CHAPTER 1

INTRODUCTION

1.1 The role of retroviral RNA genomic structure in viral replication.

RNA structure plays an integral role throughout the replication cycle of all retroviruses (Figure 1.1),[1,2] including serious pathogens like Human Immunodeficiency Virus (HIV)[3] and Human T-cell Leukemia Virus (HTLV).[4,5] Aside from encoding viral proteins, the retroviral RNA genome must interact on a structural level with several cellular and viral components, including Gag and Nucleocapsid (NC) proteins.[6] Two distinct examples of genomic RNA structure auto-regulating its replication occur during (i) genomic RNA dimerization and (ii) packaging (Figure 1.1). (i) Retroviral genomes are composed of two identical copies of genomic RNA joined by base pairing and tertiary interactions in a noncoding region near the 5' ends of the two RNA genomes (Figure 1.2).[7-9] Once formed, the RNA dimer regulates reverse transcription, genomic recombination, and encapsidation of the genome into nascent viral particles (virions).[10,11] *In vitro* dimerization is greatly accelerated in the presence of the nucleocapsid (NC) protein,[12] which is copackaged with the viral RNA. The precise structural mechanism of dimer formation remains unknown. (ii) The genomic RNA destined for assembling new virions appears to be recognized structurally as a dimer.[13,14] Dimeric recognition ensures that exactly two copies of MuLV genomic RNA are packaged into every virion. The genomic RNA dimer is packaged into immature nascent virions[15]. The RNA is likely bound to the nucleocapsid domain of Gag in a

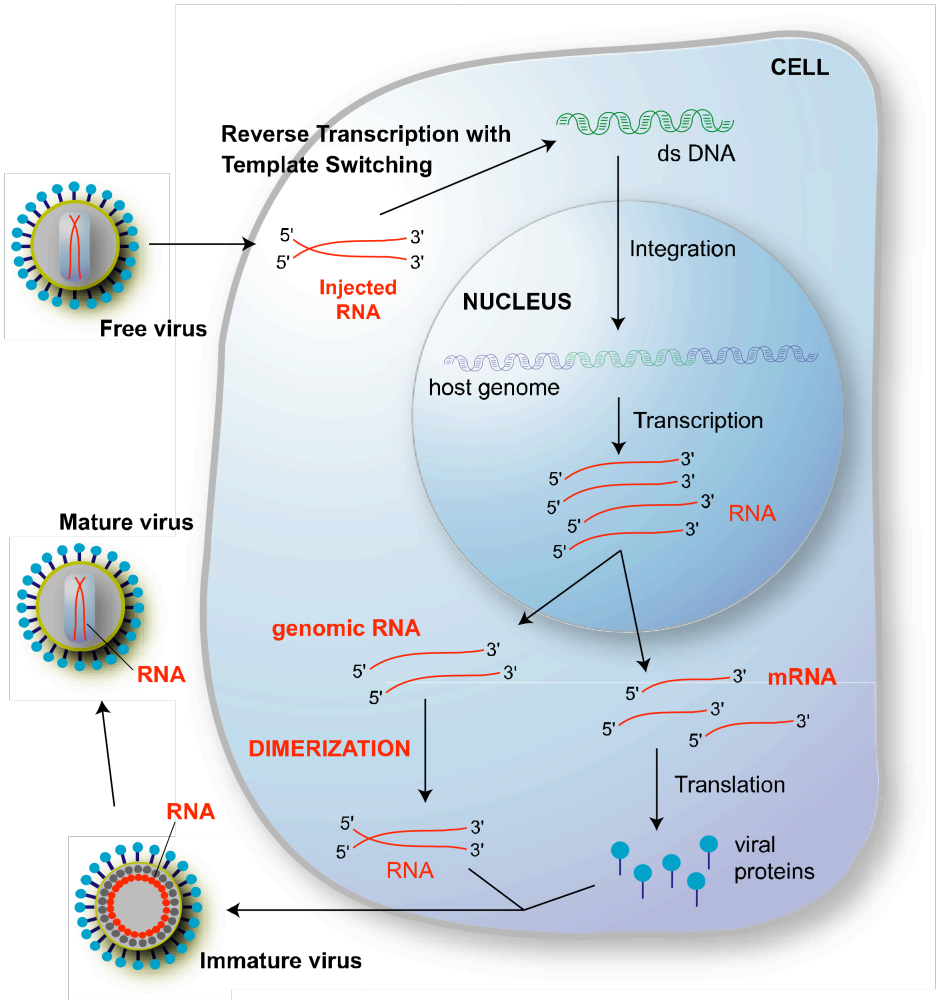


Figure 1.1 Overview of the retroviral replication cycle. Processes involving RNA conformational changes are emphasized in red. Key stages are labeled.

circular configuration within the immature virion (Figure 1.1). After budding from the host cell, Gag polyprotein undergoes viral protease-mediated cleavage.[16] The cleavage of Gag initiates dramatic structural changes of the virion, including changes in RNA structure and localization, to create the mature (infectious) particle (Figure 1.1). Since, only the mature virus is infective,[16] RNA structural changes have been proposed to be an important part of maturation. Genomic RNA structure of immature MuLV has not been characterized and only bulk measurements of global structure are currently available to provide information for distinguishing between the immature and mature MuLV particles. This limitation is based largely on the extraordinary difficulty of producing enough RNA to perform structural analysis.

An intimate understanding of how retroviral RNA effects its own fate during the replication cycle requires analysis of RNA structure both inside authentic viruses and during critical conformational changes. However, current technologies for understanding RNA structure are inadequate because they require amounts of RNA that are one thousand to one million-fold greater than that which is normally recoverable from virions and infected cells.

1.2 SHAPE Chemistry.

My lab has developed a quantitative chemistry to probe RNA structure known as Selective 2'-Hydroxyl Acylation analyzed by Primer Extension, or SHAPE.[17,18] SHAPE technology maps RNA structure in a simple two-step process, (i) RNA modification and (ii) primer extension. In the first step, exposure to electrophilic SHAPE reagents, like 1-methyl-7-nitroisatoic anhydride (1M7),[19] acylate the 2' hydroxyl of reactive RNA nucleotides creating 2'-O-adducts. Because every RNA nucleotide has a 2' hydroxyl, every nucleotide is

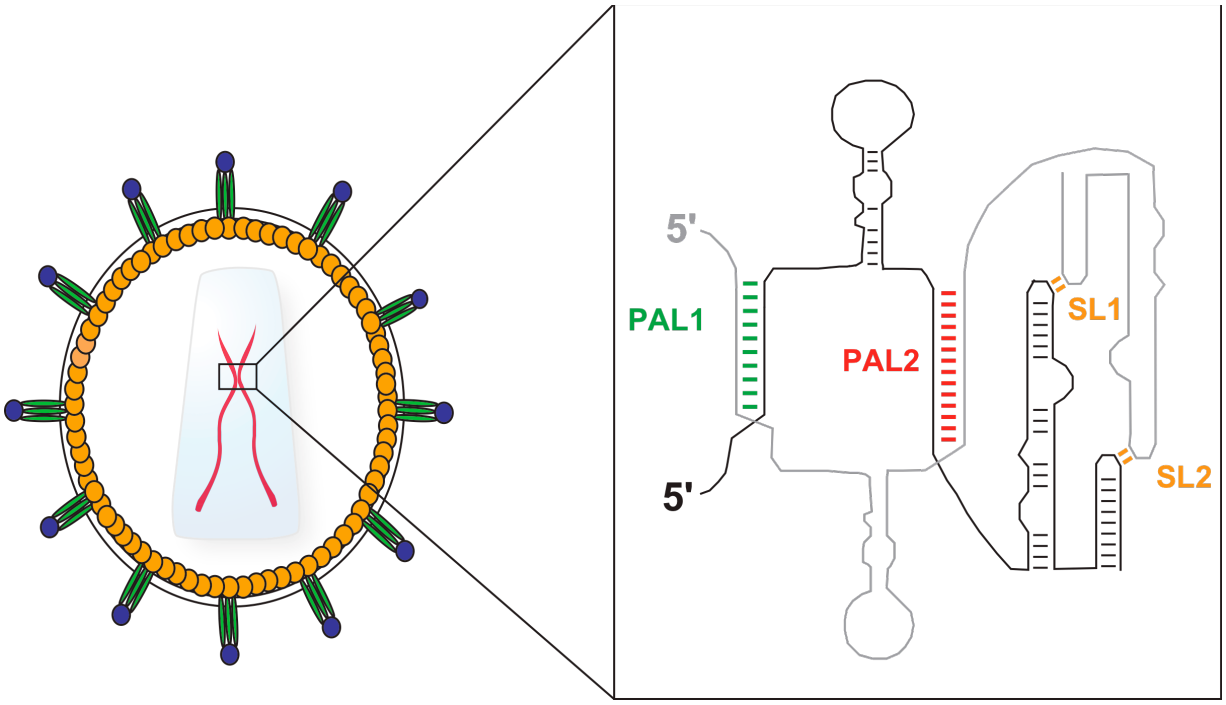


Figure 1.2 The gammaretroviral genomic dimer. Retroviruses package two identical copies of their genome into nascent infectious particles, or virions. Retroviral genomic RNA forms a dimeric structure by noncovalent interactions near their 5´ ends.

interrogated. The acylation reaction preferentially modifies flexible RNA nucleotides (Figure 1.3a). Flexible RNA nucleotides are most often located in single-stranded regions of RNA. The second step of SHAPE is primer extension of the modified RNA (Figure 1.3b). Fluorescently-labeled primers are used in a reverse transcription reaction to create a fluorescently-labeled structure-specific cDNA library[20]. The cDNA library is structure-specific because RNA adducts create stops in reverse transcription exactly one nucleotide prior to RNA modification. Therefore, the length of the cDNA corresponds to a position of modification on the RNA. A prevalent cDNA length therefore indicates a highly reactive RNA nucleotide located in a flexible, or single-stranded region of RNA. The labeled cDNA molecules are then separated in a single capillary of a DNA sequencer (Figure 1.3c, left panel). Each SHAPE experiment also includes a no reagent control and sequencing ladders to relate elution time to sequence. The resulting data, a plot of reactivities as a function of nucleotide position (Figure 1.3c, right panel), is used to propose and confirm RNA structures, locate protein binding sites, and monitor structural rearrangements.[21] I use the single nucleotide resolution SHAPE reactivities to create secondary structure "snapshots" of RNA (Figure 1.3d). This is done by inputting SHAPE reactivities as constraints into RNAstructure,[22] a folding algorithm that output secondary structures.

1.3 SHAPE requires a large amount of RNA.

SHAPE has been successful in analyzing many important retroviral RNA structures, Including the Moloney Murine Leukemia virus (MuLV),[23,24] but its uses have been limited *in virio* by the lack of the ability to detect small amounts of RNA. Like traditional probing experiments, SHAPE requires at least 3 pmols of RNA per experiment (Figure 1.4),

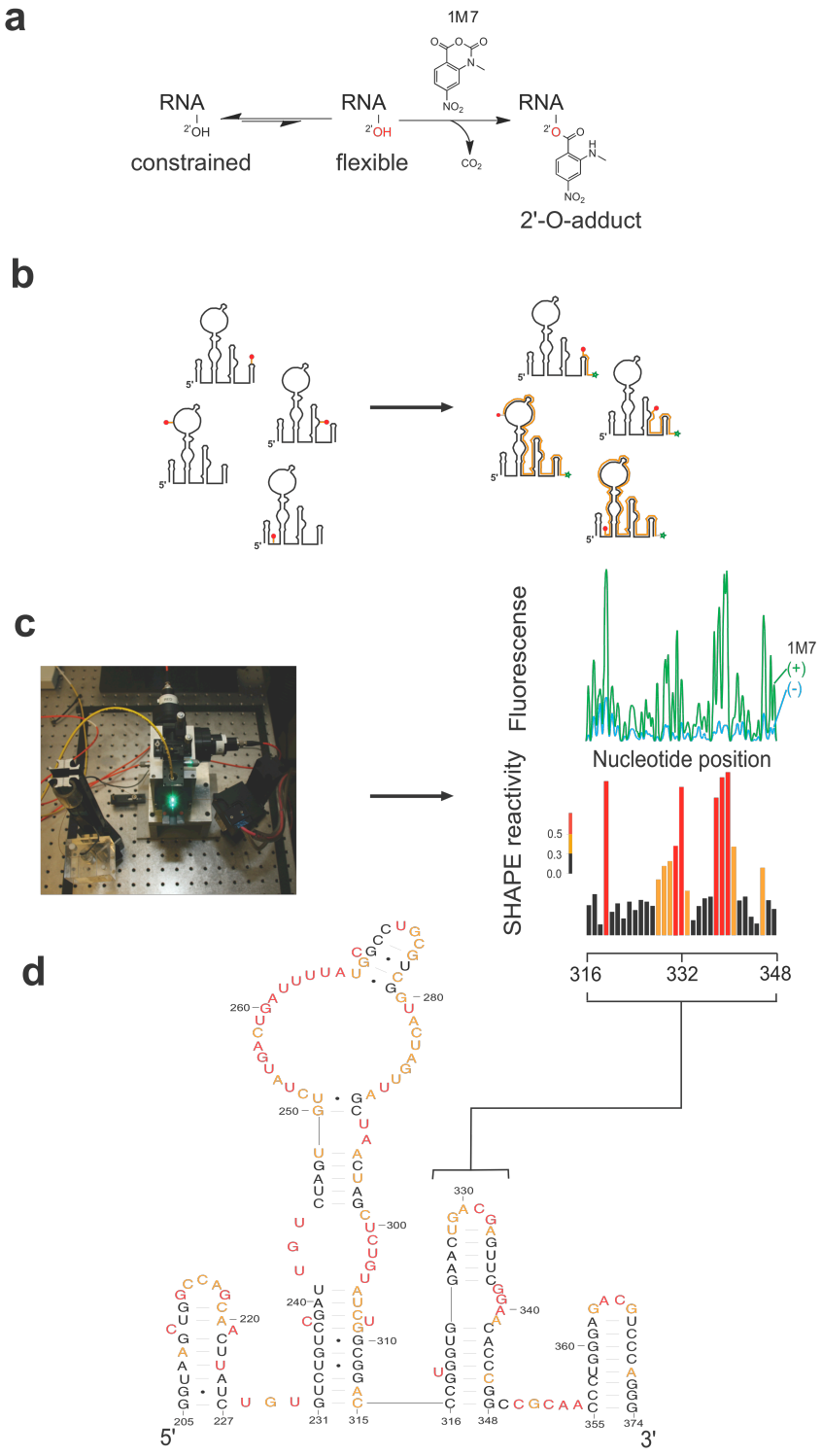


Figure 1.3 SHAPE chemistry.

which is 100,000 times more than what is easily isolatable. Therefore, the major hurdle is improving detection of the unique cDNAs to effect SHAPE on miniscule amounts of RNA.

1.4 Time-resolved SHAPE technology.

RNA structural transitions are an important aspect of cellular function and information transfer.[25] Dynamic regions within RNA molecules are often interaction sites for small molecules, proteins, and other RNAs.[26] Besides classical biochemical methods performed in conjunction with mutational data,[27] there are very few technologies adept at studying RNA conformational changes. With the advent of better structural technologies, including x-ray footprinting,[28] oligonucleotide hybridization, and time-resolved NMR studies,[29] however, the RNA structure field has experienced exponential growth. A new SHAPE-based technology, developed by my lab has also contributed to our ability to study RNA in motion.

Time-resolved SHAPE (TR-SHAPE), is a recently developed SHAPE-based technique for temporal structural and mechanistic studies of RNA.[30] In TR-SHAPE, a fast acting SHAPE reagent, benzoyl cyanide (BzCN), reacts by one of two pathways: 1) forms a 2'-O adduct with a flexible nucleotide or 2) rapidly self-inactivates by hydrolysis (0.25s half life).[31] TR-SHAPE can be used to take single second snapshots of RNA structure at single nucleotide resolution.

1.5 Research overview: development and application of two SHAPE-based technologies.

In this dissertation, I attempt to advance the field of RNA structural technologies with the development and application of two SHAPE-based technologies, ultrasensitive (US) and time-resolved (TR) SHAPE. In the second chapter, I create ultrasensitive SHAPE, which

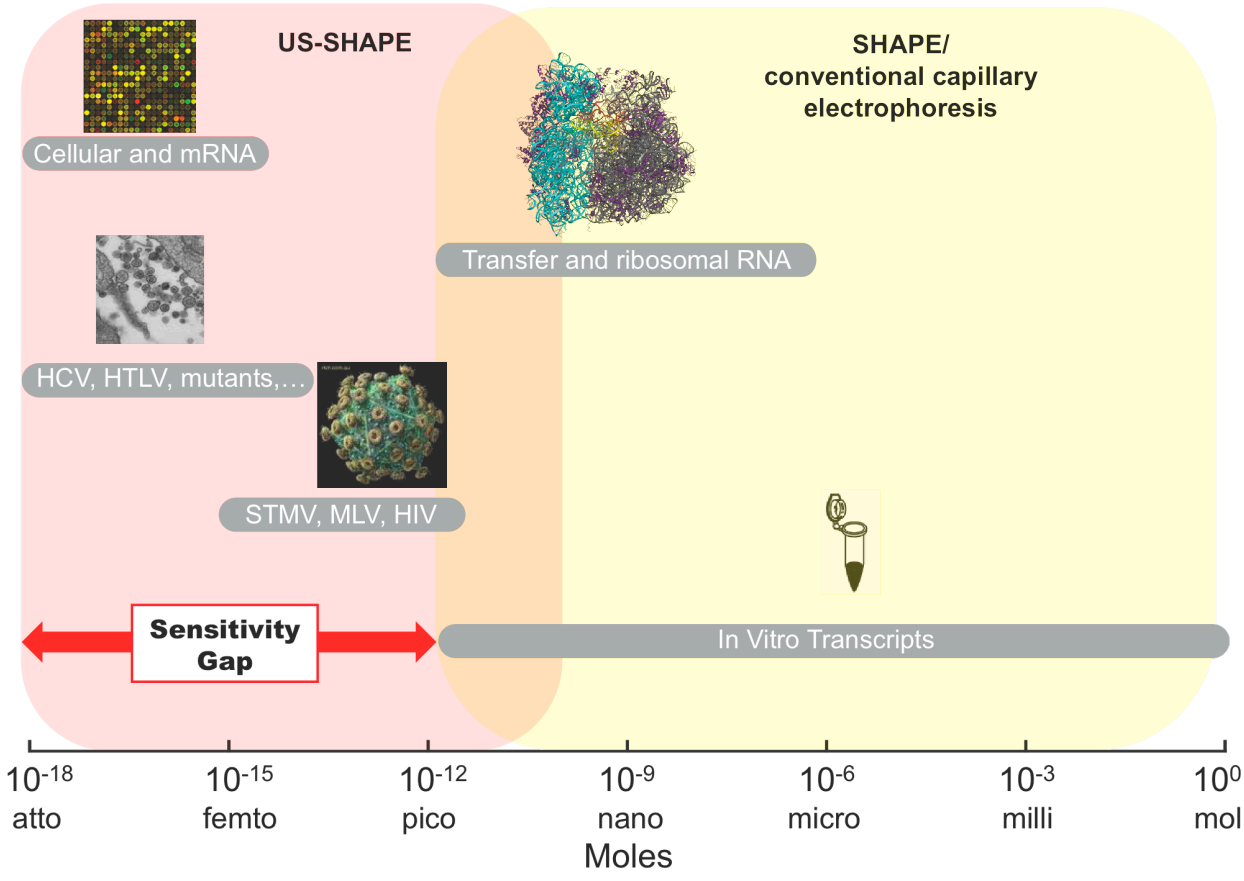


Figure 1.4 RNA abundance. Molar amounts of RNA obtained from different biological sources are shown in gray. Large colored ovals indicate the amount of RNA required for conventional and ultra-sensitive capillary electrophoresis technologies. The red arrow emphasizes the discrepancy between the amount of RNA obtainable and the amount required for structural analysis

has a sub-femtomole SHAPE sensitivity, more than four orders of magnitude more sensitive than traditional SHAPE. This technology allows me to create structural models of MuLV and XMRV within virions, as they interact on a structural level with Gag and NC proteins. In the third chapter, I apply US-SHAPE to create the first structural snapshot of MuLV genomic RNA packaged within nascent immature virions. In the fourth chapter, I advance a recent technology, time-resolved SHAPE, by creating an RNA "movie" of MuLV genomic RNA dimerization both with and without NC protein.

1.5.1 Ultra-sensitive (US) SHAPE

Most single-nucleotide resolution RNA structure determination technologies cannot be used to analyze RNA from scarce biological samples, like viral genomes. To make quantitative RNA structure analysis applicable to a much wider array of RNA structure-function problems, I developed and applied ultra-sensitive selective 2'-hydroxyl acylation analyzed by primer extension (US-SHAPE) to structural analysis of authentic genomic RNA of the xenotropic murine leukemia virus-related virus (XMRV) (Figure 1.5).[32] For analysis of fluorescently labeled cDNAs generated in ultra-sensitivity SHAPE experiments, I developed a two-color capillary electrophoresis approach with zeptomole molecular detection limits and sub-femtomole sensitivity for complete SHAPE experiments involving hundreds of individual RNA structure measurements. Ultra-sensitivity SHAPE data correlated closely (R = 0.89) with data obtained by conventional capillary electrophoresis. Using US-SHAPE, I determined the dimeric structure of the XMRV packaging domain, examined dynamic interactions between packaging domain RNA and viral nucleocapsid protein inside virion particles, and identified the packaging signal for this virus (Figure 1.5). Despite extensive sequence differences between XMRV and the intensively studied Moloney murine leukemia virus, architectures of the regulatory domains are similar and reveal common principles of gamma retroviral RNA genome packaging.

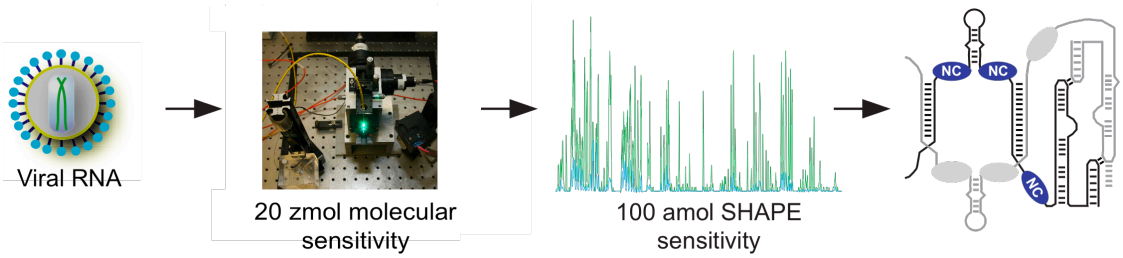


Figure 1.5 Ultra-sensitive SHAPE technology. Designed for the analysis of low-copy number authentic RNAs, the molecular and SHAPE sensitivities are shown. US-SHAPE has been used to determine the dimeric structure of the XMRV packaging and regulatory domain and located nucleocapsid (NC) binding sites within virions.

1.5.2 Secondary structure of MuLV genomic RNA within the immature virion.

Retroviruses package a structurally distinct dimeric RNA genome into noninfectious immature virions.[7,15] However, analysis of this structure is complicated by the relative scarcity of authentic retroviral RNA. In this chapter, I use ultra sensitive SHAPE (US-SHAPE) to obtain unprecedented structural information about the dimerization and packaging regulatory domain of the immature Moloney murine leukemia virus (MuLV) genomic RNA, both in virio and ex virio. I find a key structural difference within the palindromic sequence, PAL2, of immature viral RNA relative to mature viral RNA. In mature infectious MuLV particles, PAL2 forms an intermolecular duplex. However, in immature virions the PAL2 sequence is instead sequestered in an intramolecular stemloop. These US-SHAPE-derived structural models suggest that immature MuLV RNA is packaged into a kinetically-trapped intermediate state. The introduction of a MuLV mutant into authentic viruses in which the PAL2 domain is replaced with a string of adenosine residues alters the structural spacing in this region and displays a 10-fold decreased infectivity compared to native RNA.

1.5.3 A Guanosine-Centric Mechanism for RNA Chaperone Function.

Proteins with generalized, ATP-independent, RNA chaperone activity ubiquitously facilitate the RNA conformation changes and strand annealing steps required for essentially all the highly dynamic and structurally complex processes involving RNA.[33] The requirement for RNA chaperone activity is related, in part, to fundamental features of RNA structure. RNA structures are built up from only four fundamental building blocks such that many folds are compatible with a given sequence and misfolded structures are thermodynamically stable, creating long-lived intermediates.[34,35] However, a mechanistic understanding of the

widespread requirement for RNA chaperone function is unclear. The retroviral nucleocapsid (NC) proteins are paradigmatic RNA chaperones and are essential for many steps of retroviral replication including genome dimerization,[36] which enables specific packaging of two genomes into each virion. In this chapter, I use time-resolved selective hydroxyl acylation analyzed by primer extension (TR-SHAPE) to understand the effects of nucleocapsid on dimerization of the minimal dimerization motif of M-MuLV. TR-SHAPE allows second-scale resolution of conformational changes in RNA molecules. Over a 600-second timeframe, we find that the naked RNA undergoes four distinct time-dependent transitions over the course of dimerization (Figure 1.6a). Using model-free clustering, we discover that the entire data set, encompassing over 2700 single-nucleotide measurements, can be grouped into seven distinct time-dependent transitions that occur at four rates (Figure 1.6b). I find that dimerization of the Moloney murine leukemia virus (MuLV) genome proceeds via a complex mechanism characterized by multiple slow steps (Figure 1.6c). In the presence of the retroviral NC protein, the rate-limiting steps of dimerization collapse to a single, fast kinetic step. The initial interaction between NC and RNA primarily involves interactions with guanosine nucleotides, which forms the strongest, and most promiscuous, base pairs in any RNA. Substituting the more weakly pairing inosine for guanosine yields an RNA that folds rapidly, and in an NC-independent way, the same final structure as the native RNA. From this analysis, I propose a comprehensive dimerization model for gammaretroviruses, including a mode of action for the nucleocapsid chaperone protein.

1.6 The point: Biological impact of SHAPE-based structure technologies

Dimerization of genomic RNA is a critical step in nearly all retroviral infections and occurs against an enormous background of cellular RNAs. A mechanistic understanding of the dimerization process, with and without its copackaged nucleocapsid (NC) chaperone, is key to discovering the source of this specificity. Until now, a mechanistic understanding of

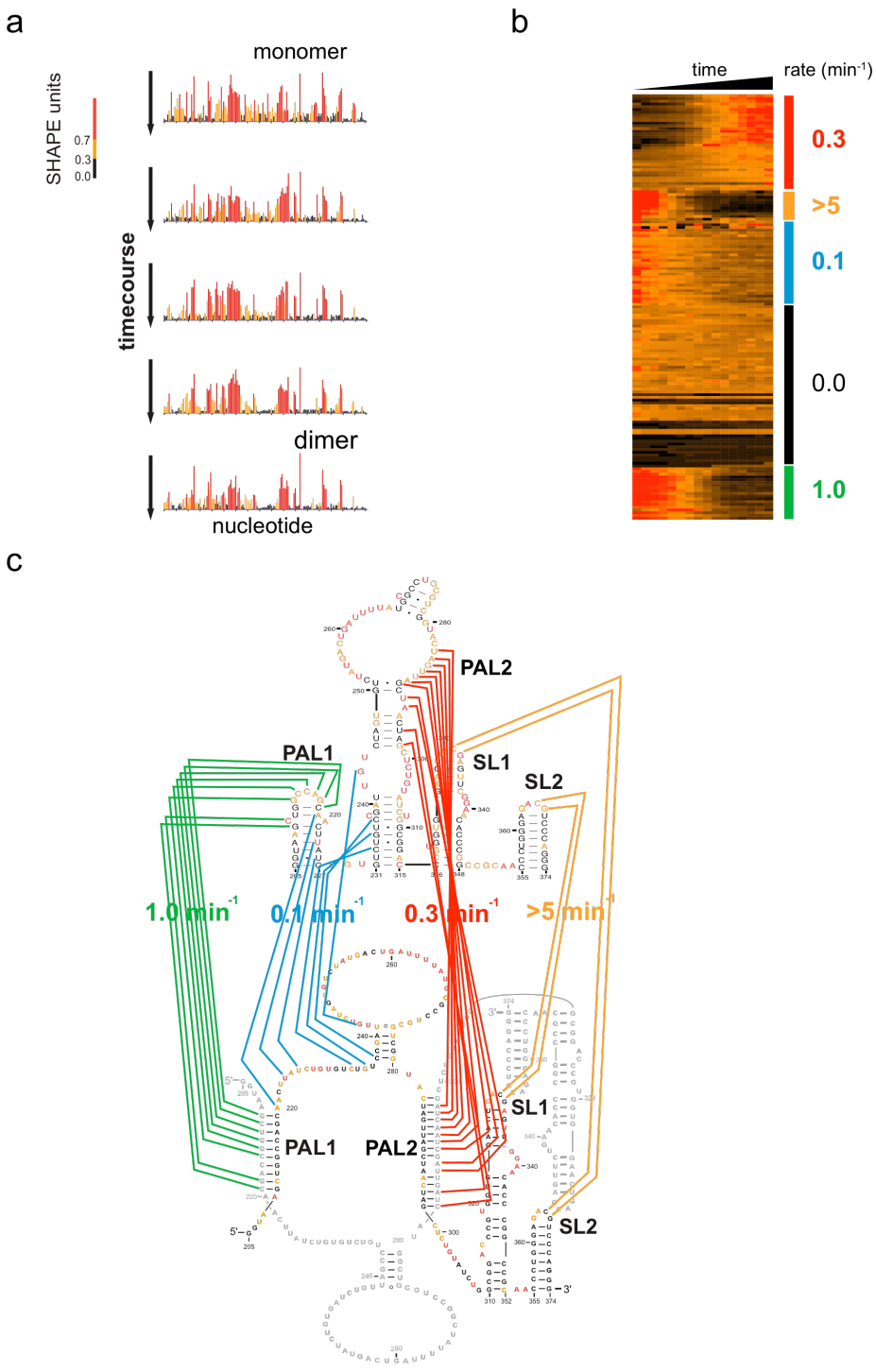


Figure 1.6 Time-resolved SHAPE technology. (a) Time-resolved reactivity profile of MuLV RNA dimerization. (b) K-means clustering of seven time-dependent behaviors. (c) MuLV genome dimerization proceeds via a complex mechanism characterized by multiple slow steps.

retroviral dimerization was intractable because RNA structural technologies were inept at studying RNA in motion. In chapter 4, I advance a SHAPE-based technology, time-resolved SHAPE (TR-SHAPE), to create the first ever "molecular movie" of RNA dimerization as it binds to NC. From this work, I discover a general mechanism for ATP independent RNA chaperone activity in which non-specific chaperones destabilize interactions between guanosine and other nucleotides to smooth out the RNA folding landscape. This mechanism appears to apply to diverse RNA chaperones and is related to fundamental features of RNA thermodynamics. It is likely that this generalized molecular mechanism of RNA chaperones will lead to novel approaches at preventing viral pathogenesis.

Specific genomic dimerization is the cornerstone for proper packaging into nascent retrovirions. Armed with a mechanistic, single-nucleotide resolution framework of how the genomic RNA dimer is formed, I next set out to understand how the dimeric structure interacts with its protein counterparts to form the simplest diploid organism, the infective retroviral virion. To fully understand the role of RNA conformational changes in retroviral biology, I needed to understand RNA structure and RNA-protein interactions within authentic virions.

Preliminary exploratory experiments have demonstrated that I can use SHAPE chemistry to map the equilibrium structure of mature MuLV RNA within authentic virions.[24,37] Strikingly, within the scanty 2% of the entire MuLV genome that was analyzed *in virio*, we detected specific differences in the structure of the mature genomic RNA inside viruses compared to the same protein-free extracted RNA. In chapters 2 and 3, I complement this work with the advent of ultra-sensitive SHAPE technology, designed for the analysis of scarce amounts of RNA extracted from authentic biological sources and within virions. Our results verify the packaging and regulatory signal RNA exists in a structurally distinct conformation in the immature state and I have created secondary structure models of the

RNA within this state. A mutant viral construct designed to mimic the immature state, with a structurally distinct dimerization domain element, was introduced into authentic viruses and found to exhibit a ten-fold deficiency in genomic replication. US-SHAPE-derived structural models and *in vivo* results suggest that immature MuLV RNA is packaged into a stable intermediate state. From these models, we have a more comprehensive understanding of how retroviruses package their genomic material and create infective progeny.

The future of this work will most likely involve the marriage of the two technologies, US-SHAPE and TR-SHAPE. The ability to develop single-nucleotide resolution structural models and conformational mechanisms within cells and infected virions will likely supplant analysis of artificial constructs and allow real time visualization of the replication of all RNA viruses.

1.7 REFERENCES

1. Coffin, J.M., Hughes, S.H., & Varmus, H.E. Retroviruses. (Cold Spring Harbor Press, Plainview, NY; 1997).

2. Roulston, A., Marcellus, R.C. & Branton, P.E. Viruses and apoptosis. *Annu Rev Microbiol* **53**, 577-628 (1999).

3. Watts, J.M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716 (2009).

4. Poiesz, B.J., Poiesz, M.J. & Choi, D. The human T-cell lymphoma/leukemia viruses. *Cancer Invest* **21**, 253-277 (2003).

5. Jarrett, R.F. Viruses and lymphoma/leukaemia. *J Pathol* **208**, 176-186 (2006).

6. D'Souza, V. & Summers, M.F. How retroviruses select their genomes. *Nat Rev Microbiol* **3**, 643-655 (2005).

7. Berkowitz, R., Fisher, J. & Goff, S.P. RNA packaging. *Curr Top Microbiol Immunol* **214**, 177-218 (1996).

8. Rein, A. Take two. *Nat Struct Mol Biol* **11**, 1034-1035 (2004).

9. Badorrek, C.S. & Weeks, K.M. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat Chem Biol* **1**, 104-111 (2005).

10. Mikkelsen, J.G., Lund, A.H., Duch, M. & Pedersen, F.S. Forced recombination of psi-modified murine leukaemia virus-based vectors with murine leukaemia-like and VL30 murine endogenous retroviruses. *J Gen Virol* **80 (Pt 11)**, 2957-2967 (1999).

11. Hibbert, C.S., Mirro, J. & Rein, A. mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J Virol* **78**, 10927-10938 (2004).

12. Bonnet-Mathoniere, B., Girard, P.M., Muriaux, D. & Paoletti, J. Nucleocapsid protein 10 activates dimerization of the RNA of Moloney murine leukaemia virus in vitro. *Eur J Biochem* **238**, 129-135 (1996).

13. Adam, M.A. & Miller, A.D. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J Virol* **62**, 3802-3806 (1988).

14. D'Souza, V. & Summers, M.F. Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **431**, 586-590 (2004).

15. Fu, W. & Rein, A. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* **67**, 5443-5449 (1993).

16. Yeager, M., Wilson-Kubalek, E.M., Weiner, S.G., Brown, P.O. & Rein, A. Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms. *Proc Natl Acad Sci U S A* **95**, 7299-7304 (1998).

17. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127**, 4223-4231 (2005).

18. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610-1616 (2006).

19. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129**, 4144-4145 (2007).

20. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. & Giddings, M.C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979-1990 (2008).

21. Wilkinson, K.A. et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**, e96 (2008).

22. Mathews, D.H. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101**, 7287-7292 (2004).

23. Badorrek, C.S., Gherghe, C.M. & Weeks, K.M. Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *Proc Natl Acad Sci U S A* **103**, 13640-13645 (2006).

24. Gherghe, C., Leonard, C.W., Gorelick, R.J. & Weeks, K.M. Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. *J Virol* **84**, 898-906 (2010).

25. Buck, J., Furtig, B., Noeske, J., Wohnert, J. & Schwalbe, H. Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc Natl Acad Sci U S A* **104**, 15699-15704 (2007).

26. Hall, K.B. RNA in motion. *Curr Opin Chem Biol* **12**, 612-618 (2008).

27. Treiber, D.K., Rook, M.S., Zarrinkar, P.P. & Williamson, J.R. Kinetic intermediates trapped by native interactions in RNA folding. *Science* **279**, 1943-1946 (1998).

28. Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R. & Brenowitz, M. Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J Mol Biol* **266**, 144-159 (1997).

29. Buck, J., Furtig, B., Noeske, J., Wohnert, J. & Schwalbe, H. Time-resolved NMR spectroscopy: ligand-induced refolding of riboswitches. *Methods Mol Biol* **540**, 161-171 (2009).

30. Mortimer, S.A. & Weeks, K.M. Time-resolved RNA SHAPE chemistry. *J Am Chem Soc* **130**, 16178-16180 (2008).

31. Mortimer, S.A. & Weeks, K.M. Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc* **4**, 1413-1421 (2009).

32. Grohman, J.K., Kottegoda, S., Gorelick, R.J., Allbritton, N.L. and Weeks, K.M. Attomole SHAPE reveals regulatory structures in the authentic XMRV RNA genome. *J Am Chem Soc* (2011).

33. Rocak, S. & Linder, P. DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol* **5**, 232-241 (2004).

34. Woodson, S.A. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol Life Sci* **57**, 796-808 (2000).

35. Johnson, T.H., Tijerina, P., Chadee, A.B., Herschlag, D. & Russell, R. Structural specificity conferred by a group I RNA peripheral element. *Proc Natl Acad Sci U S A* **102**, 10176-10181 (2005).

36. Cruceanu, M. et al. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res* **34**, 593-605 (2006).

37. Gherghe, C. et al. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci U S A* (2010).

CHAPTER 2

FEMTOMOLE SHAPE REVEALS REGULATORY STRUCTURES IN THE AUTHENTIC XMRV GENOME

2.1 The role of SHAPE in retroviral genomic RNA structure.

As is likely to be the case for essentially all coding and non-coding RNAs, higher-order secondary and tertiary structures in viral genomes play numerous roles in replication.[1-3] High-throughput selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) makes possible characterization of RNA structure at single-nucleotide resolution on the kilobase scale and has proven to be a powerful approach for understanding structure-function relationships in RNA.[4-7] In a SHAPE experiment, conformationally flexible (generally single-stranded) nucleotides react with 1-methyl-7-nitroisatoic anhydride (1M7),[8] or another electrophilic SHAPE reagent,[9,10] to create a covalent 2'-*O*-ribose adduct (Figure 2.1a). Sites of adduct formation are detected by primer extension using fluorescently labeled primers. SHAPE primer extension products can be separated on commercially available capillary electrophoresis sequencing instruments[11,12] resulting in an electropherogram in which peak intensities correlate with local flexibility at each individual nucleotide in the RNA (Figure 2.1b).[13] These data are processed to yield a reactivity profile for every RNA nucleotide (Figure 2.1c). SHAPE has been validated by analysis of RNAs with known structures and was used to discover structure-function relationships in an entire HIV-1 genome,[5,11] and in regulatory domains in the Moloney murine leukemia virus (MuLV).[14,15]

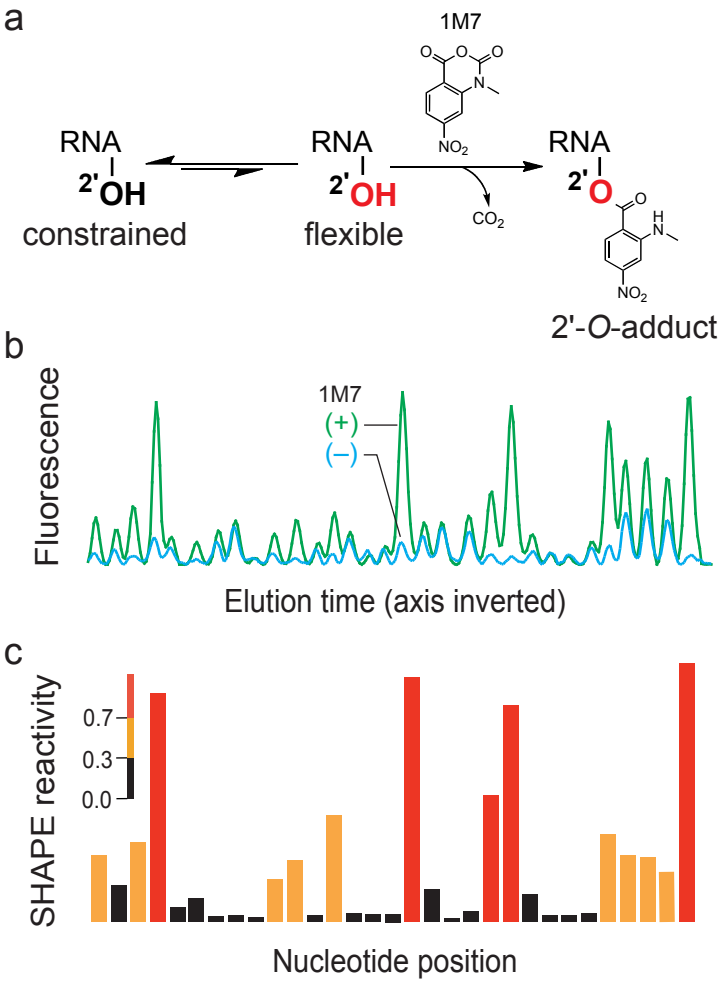


Figure 2.1 SHAPE overview**.** (a) Mechanism of RNA SHAPE chemistry. Hydroxyl-selective electrophiles modify the 2'-OH group of flexible RNA nucleotides (red) to create a 2'-*O*-adduct. (b) Subsequent detection of 2'-*O*-adducts by primer extension and separation by capillary electrophoresis yields an electropherogram of fluorescent peaks. Peak volumes correspond to SHAPE reactivity. (c) Integrated peak intensities after background subtraction.

2.2 The sensitivity of SHAPE.

The sensitivity of SHAPE is limited by the generally poor detection limits of conventional capillary electrophoresis instrumentation and requires ~3 pmol of RNA per experiment.[4] This limitation is significant: Tens of liters of infected cells in a highly optimized cell culture system were required to obtain sufficient genome copies (~100 pmol) for analysis of an entire HIV-1 genome.[5] High levels of background noise from light scattering, leakage of excitation light, and electronic noise reduce the ability of conventional instruments to detect fluorescently-labeled primer extension products.[16] When optimized for low-abundance species, capillary electrophoresis is, in principal, capable of detecting single molecules.[17]

To address the challenge of analyzing rare, biologically important RNAs, we have developed an ultra-sensitive RNA structure analysis platform that employs a two-color capillary electrophoresis instrument with zeptomole (10^{-21}) instrument detection limits and sub-femtomole sensitivity for complete SHAPE experiments comprised of hundreds of individual structure measurements. Here we describe application of this technology to analysis of a low-abundance gammaretrovirus, xenotropic murine leukemia virus related virus (XMRV).[18] Although XMRV is not likely to be a human pathogen,[19,20] as originally reported,[18,21] this virus is an excellent example of a recently identified infectious agent currently available in only low amounts. XMRV is evolutionarily related to the well-characterized MuLV,[22-25] suggesting that critical features of its structure and biology are conserved. Here, we used high-sensitivity SHAPE to analyze scarce XMRV RNA gently extracted from virions and inside viral particles. Despite extensive sequence differences between XMRV and the MuLV, the regulatory domain architectures are similar, revealing universal principles of retroviral RNA genome packaging.

2.3 Design and construction of an ultra-sensitive capillary electrophoresis instrument.

We sought to improve the detection limits of SHAPE experiments by designing and building a robust, yet highly sensitive, capillary electrophoresis instrument. A single-mode fiber optic cable with an integrated lens was used to focus the light from a dual color argon laser (488 and 514 nm) onto a capillary (Figure 2.2). The optical fiber enabled the laser to be placed at a distance and ensured robust alignment of the laser beam. A high-numerical aperture microscope objective oriented at a right angle to the laser beam collected light from the capillary and minimized collection of scattered excitation light. The microscope objective was mated to an optical cube (Figure 2.2) with a dichroic filter to reflect green wavelengths while transmitting yellow light. Collection lenses mounted into the cube focused the light onto a multimode optical fiber that acted as a spatial filter to remove scattered light and transfer fluoresced light to the detector. Once the optical path was aligned, the cube rigidly maintained the positions of the various components so that realignment was required infrequently. Filters specific for each fluorophore were placed between the optical cube and the detector photomultiplier tubes. The multimode fibers permitted the detectors to be placed at a distance from the optical cube facilitating alignment of the optical system and minimizing interference from high voltage capillary electrophoresis. The melding of these individual innovations yielded a robust and reliable optical system for the study of RNA structure.

2.4 Detection limits for free fluorophores and for a complex SHAPE library.

The limits of detection for the capillary electrophoresis instrument for the free 5-FAM and 6-JOE fluorophores used in our SHAPE experiments were determined by analysis of serially diluted free dyes. Dilutions were analyzed until a signal-to-noise ratio of 3:1 could no longer be maintained (Figure 2.3). For 5-FAM, the detection limit was 20

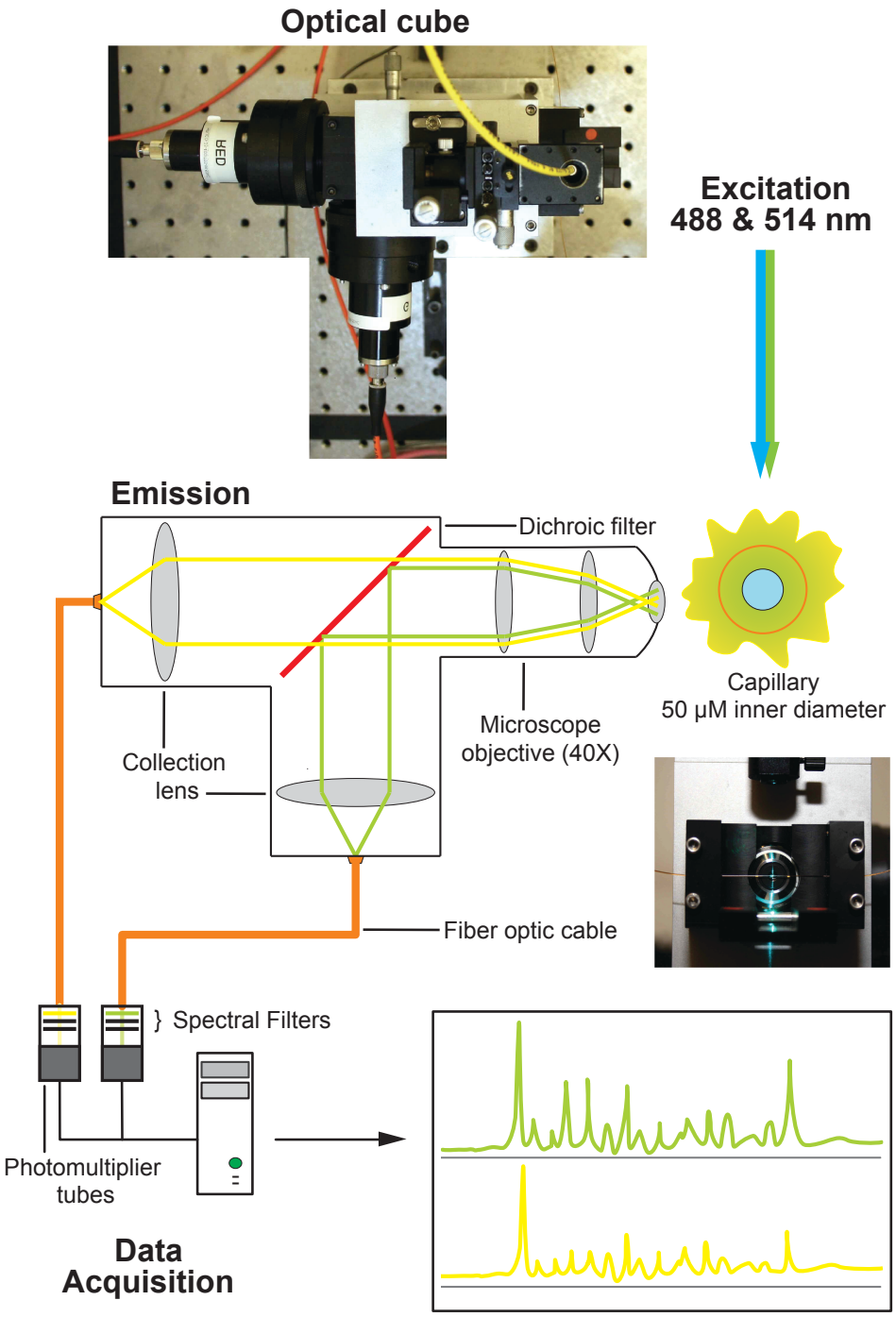


Figure 2.2 Robust, high sensitivity, capillary electrophoresis instrument design. Photograph and schematic of the high-sensitivity instrument. The optical cube design maximizes light detection and minimizes background noise within each of the three main capillary electrophoresis instrument components: excitation, emission and data acquisition.

zeptomoles (2×10^{-20} mol), or 10,000 molecules. For 6-JOE, the limit was ten-fold higher at 200 zmol (2×10^{-19} mol).

We then tested the ability of the system to separate, visualize and quantify hundreds of fluorescent peaks simultaneously as required in a SHAPE experiment. We performed dideoxy sequencing reactions[26] using an *in vitro* transcript corresponding to an RNA regulatory signal in the MuLV RNA.[25] For this validation experiment, 1 fmol of input RNA, or 1,000 times less than that required for a conventional instrument, was used. After primer extension, 59 ± 5% of the input RNA was converted to fluorescently labeled cDNA, (see Methods). Throughout this work, amounts are reported in terms of the input RNA. Triplicate measurements were highly reproducible with standard deviations less than 5% per nucleotide position (Figure 2.4a). We also compared data from a SHAPE reaction on the MuLV RNA obtained with the conventional instrument using 1 pmol RNA versus with the high-sensitivity instrument using 1 fmol of the same sample (Figure 2.4b). The two electropherograms are highly similar (R = 0.90) (Figure 2.4c).

We next used the *in vitro* MuLV transcript to determine the SHAPE limit of detection. We probed the structure of this RNA by SHAPE using 1M7,[8] and the resulting fluorescently labeled cDNA fragments were resolved by both conventional and high-sensitivity capillary electrophoresis. For a cDNA pool derived from 10 fmol of input RNA, we resolved peaks corresponding to all 170 nucleotides in the MuLV RNA using the high-sensitivity instrument (Figure 2.5a, left). Experiments corresponding to the (+) reagent and no-reagent control reactions were resolved in separate capillary electrophoresis runs, each also containing the same sequencing marker. These data were combined and processed as a single experiment (see Methods) to obtain a SHAPE reactivity profile (Figure 2.5b, left) and compared to data from a conventional instrument using 100 times more input RNA. The integrated intensities from the high sensitivity instrument correlated closely with data from a conventional instrument; the R-value is 0.95 (Figure 2.5c, left).

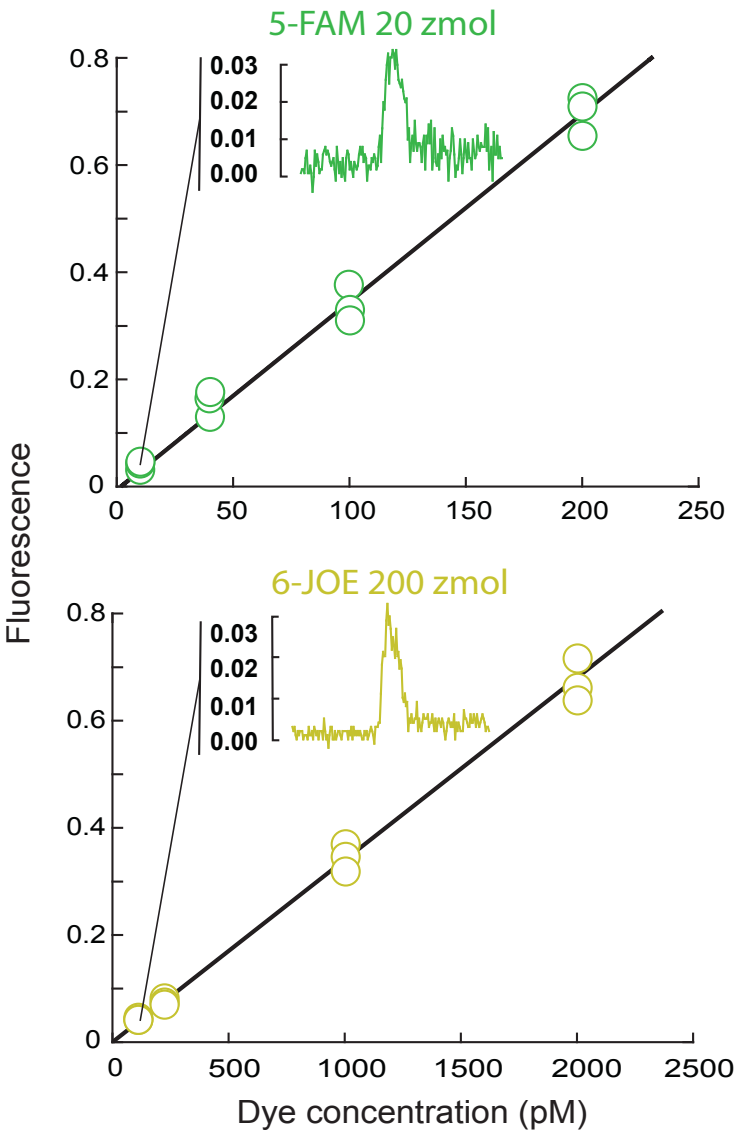


Figure 2.3 Instrument limit of detection. Electropherograms of free 5-FAM (green) and 6-JOE (yellow) fluorophores shown at their 3:1 signal-to-noise detection limits. Note that the x-axis scale is 10-fold larger in the lower panel.

Next, we sequentially diluted the SHAPE-produced cDNA library until a 3:1 signal-to-noise ratio was reached. This occurred at a dilution corresponding to 100 amol input RNA (Figure 2.5a and b, right panels). At this level of input RNA, the usable data eventually decays to a level below the instrument detection level due to drop-off[12,23] associated with the reverse transcription step (Figure 2.5a, right panel).

We estimated the amount of cDNA injected into the instrument at the 100 amol RNA dilution using electrokinetic injection calculations.[27] The quantity injected, Q (in moles), onto the capillary electrophoresis instrument is proportional to the sum of electrophoretic (μ_{ep}) and electroosmotic (μ_{eo}) velocities of the DNA in the polymer-filled capillary

$$Q = (\mu_{ep} + \mu_{eo})\, \pi r^2 E C t_i \qquad (1)$$

scaled by the capillary area (πr^2), the electric field strength (E), the input RNA concentration (C) and the injection time (t_i). The electro-osmotic mobility is approximately zero in this system (data not shown) and cDNA electrophoretic mobility is approximately equal to the length of the capillary divided by the separation time, t_s (the time between cDNA injection and fluorescent primer peak detection),[28] thus:

$$Q \approx (L/E t_s)\, \pi r^2 E C t_i \qquad (2)$$

For a solution containing 100 amol input RNA, approximately 1 amol of cDNA was physically injected into the instrument [L = 36 cm, E = 217 V/cm, t_s = 875 s (measured for the primer DNA, and thus a lower limit), C = 10 pM (100 amol / 10 µL) and t_i = 30 s]. Given that 20 zmol of cDNA are required to visualize a single peak (nucleotide) in the electropherogram, 3.4 amol (170 nucleotides × 20 zmol) represents the minimal sample quantity sufficient to report a SHAPE experiment for the MuLV packaging domain.

Thus, at 1 amol of cDNA injected onto the instrument, roughly one-third too few molecules are present to fully represent the entire SHAPE profile. Consistent with this

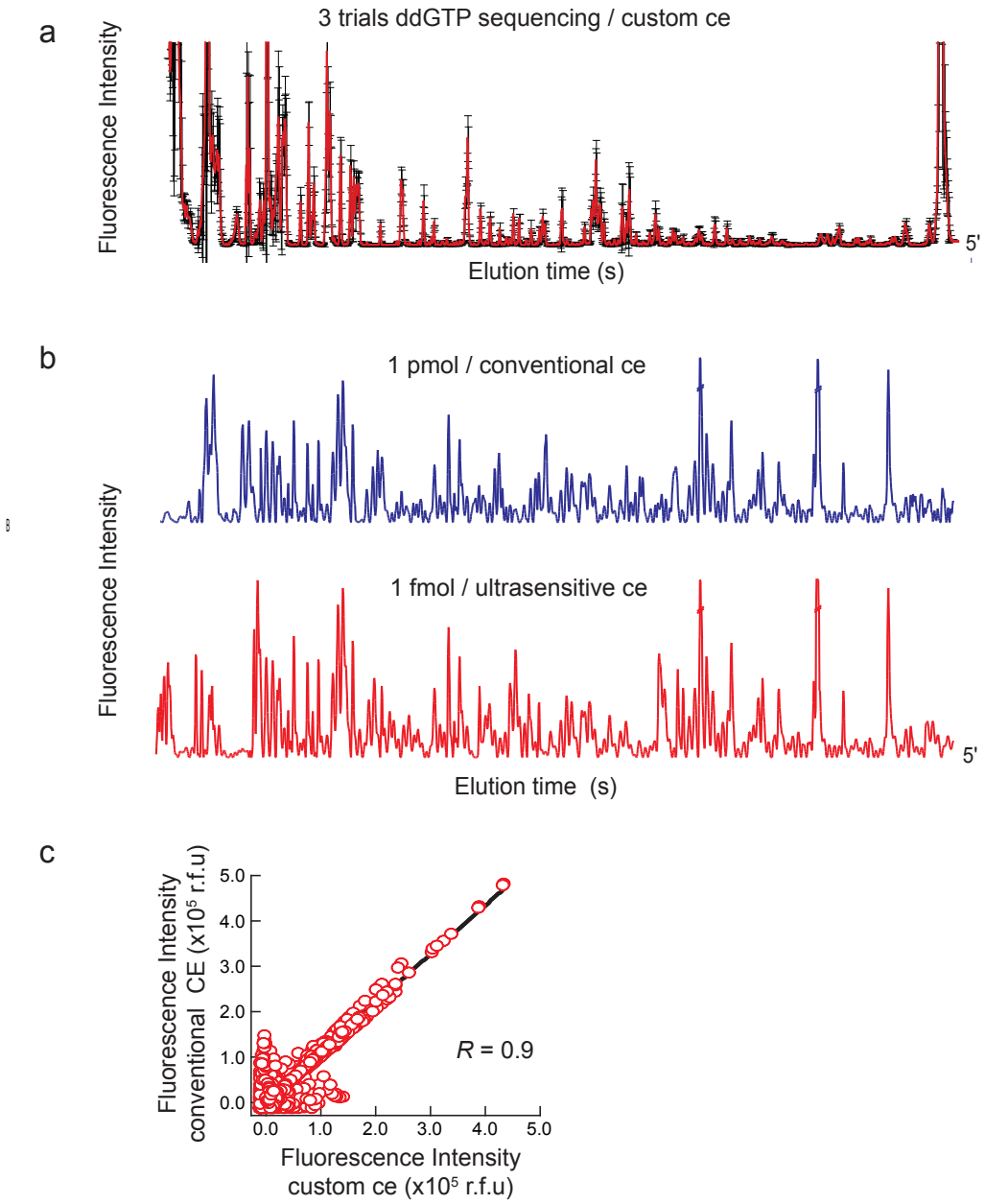


Figure 2.4 Precision and accuracy of high-sensitivity CE. (a) Mean and standard deviations for three dideoxy GTP (ddGTP) sequencing reactions. Error bars are given for every data point. Standard deviations were less than 5% of measured intensities throughout the trace. (b) Electropherograms for SHAPE reactions performed on the MuLV monomer using 5 mm 1M7 and detected using either conventional CE at 1 pmol RNA (blue) or high-sensitivity CE at 1 fmol input RNA (red). (c) Correlation of conventional and ultra-sensitive custom CE.

calculation for the 100 amol SHAPE limit of detection, ~60 nucleotides of SHAPE data in the electropherogram integrate to zero or near-zero values, reflecting signal decay that results from imperfect processivity of the reverse transcriptase enzyme used to generate the cDNA pool and a two-fold reduction in injection efficiency for the longest cDNAs (Figure 2.5a and b, right panels, gray bar). These positions also exhibit an accumulation of zero magnitude points along the x-axis in a correlation plot (Figure 2.5c, right). Strikingly, at this SHAPE limit of detection, the data are nevertheless still highly correlated with data obtained using conventional instrumentation at 1 pmol RNA (R = 0.89). Ultra-sensitive SHAPE thus yields accurate RNA structural information up to, and even beyond, the limit of fluorophore detection.

2.5 Accurate, nucleotide-resolution secondary structure prediction by ultra-sensitive SHAPE.

SHAPE reactivities are inversely correlated with the probability that a nucleotide is base paired, and it is therefore possible to use SHAPE information to predict RNA secondary structures accurately.[29,30] We used the nucleotide-resolution SHAPE data obtained at 10 fmol and 100 amol, equal to 100- and 10,000 fold lower concentrations than measurable using conventional instrumentation, to direct RNA secondary structure predictions for the MuLV packaging domain. This 170-nucleotide RNA corresponds to a regulatory domain at the 5' end of the MuLV RNA genome[31] that links two viral genome strands into a dimeric structure and comprises the "packaging signal" responsible for directing the full genome into nascent viral particles.[32] Point mutations in this domain can reduce packaging to non-specific levels.[25] The MuLV regulatory domain has been extensively characterized both structurally and biologically and forms a well-defined monomeric structure in the absence of magnesium ion.

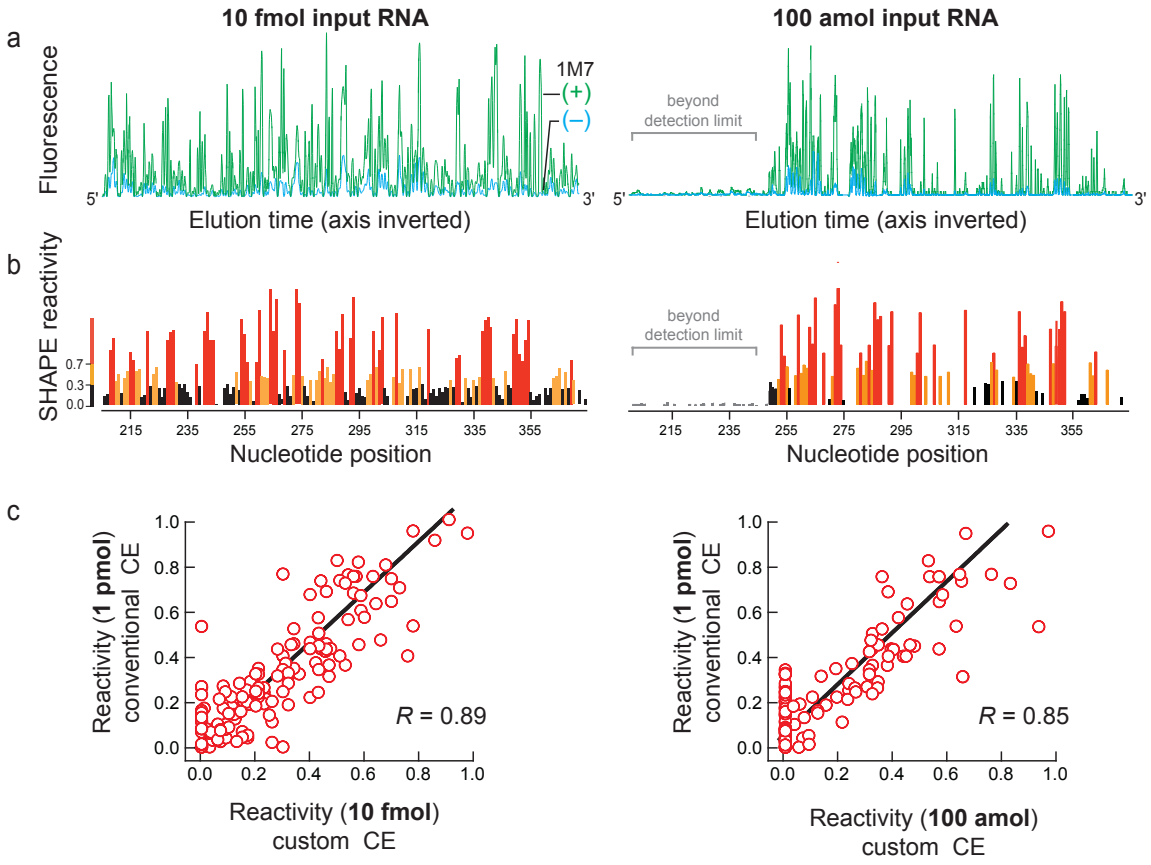


Figure 2.5 SHAPE limit of detection using MuLV. (a) Processed electropherograms and (b) quantitative SHAPE reactivities for MuLV RNA treated with 5 mM 1M7 are shown for cDNA pools corresponding to 10 fmol and 100 amol of input RNA. Regions labeled in gray in the 100 amol experiment represent data that fall below the statistical detection limit at the end of the electropherogram due to signal decay during primer extension.[11,12] (c) Correlation between SHAPE reactivities, measured at 10 fmol and 100 amol of input RNA, with an identical experiment resolved by conventional capillary electrophoresis instrumentation at 1 pmol of input RNA. The full datasets span a range from 0 to ~2 but only data to 1.0 are shown to emphasize the strong best-fit correlation for the most information-rich measurements. *R*-values for the full data range are higher, at 0.95 and 0.89, respectively.

The lowest free energy structures for the MuLV packaging domain were calculated using either thermodynamic nearest-neighbor constraints alone or thermodynamic nearest-neighbor constraints and SHAPE reactivity data included as pseudo-free energy constraints.[29] Structure predictions were evaluated by their sensitivity (the percent correctly predicted base pairs) and positive predictive value (PPV, the percentage of predicted pairs that are present in the accepted structure). In the absence of SHAPE data, the secondary structure of the MuLV RNA was predicted with sensitivity and PPV values of 36% and 35%, respectively. There were significant errors relative to the accepted structure,[14] as shown in Figure 2.6a. We then predicted the structure using SHAPE data as pseudo-free energy change constraints.[29,33] The resulting base pair sensitivities were 100 and 90%, respectively, and no incorrect base pairs were proposed (PPV = 100%) (Figure 2.6b and 2.6c). Consistent with the strong correlation observed between the ultra-sensitive SHAPE data and data obtained by using pmol RNA amounts, ultra-sensitive SHAPE yielded accurate secondary structure predictions with as few as 100 amol, or 10,000 molecules, of RNA.

2.6 Ultra-sensitive SHAPE analysis of authentic XMRV.

We then analyzed a 184-nucleotide region near the 5' end of the XMRV RNA genome that is sequentially homologous to the MuLV dimerization and packaging regulatory domain. SHAPE reactivity measurements were obtained using 50 fmol aliquots of XMRV genomic RNA that had been gently extracted from virions purified from cell culture,[34] using the 22RV1 virus-producing cell line.[35] Integrated SHAPE reactivities (Figure 2.7a), taken together with comparison to prior structural analysis of MuLV, were used to develop a model for the structure of this regulatory domain. SHAPE probing data of RNA isolated from virions indicated clearly that the XMRV genome was present as a dimer in mature viral particles. The overall architecture is similar to that of MuLV and includes four key dimer interaction elements. Two palindromic sequences, termed PAL1

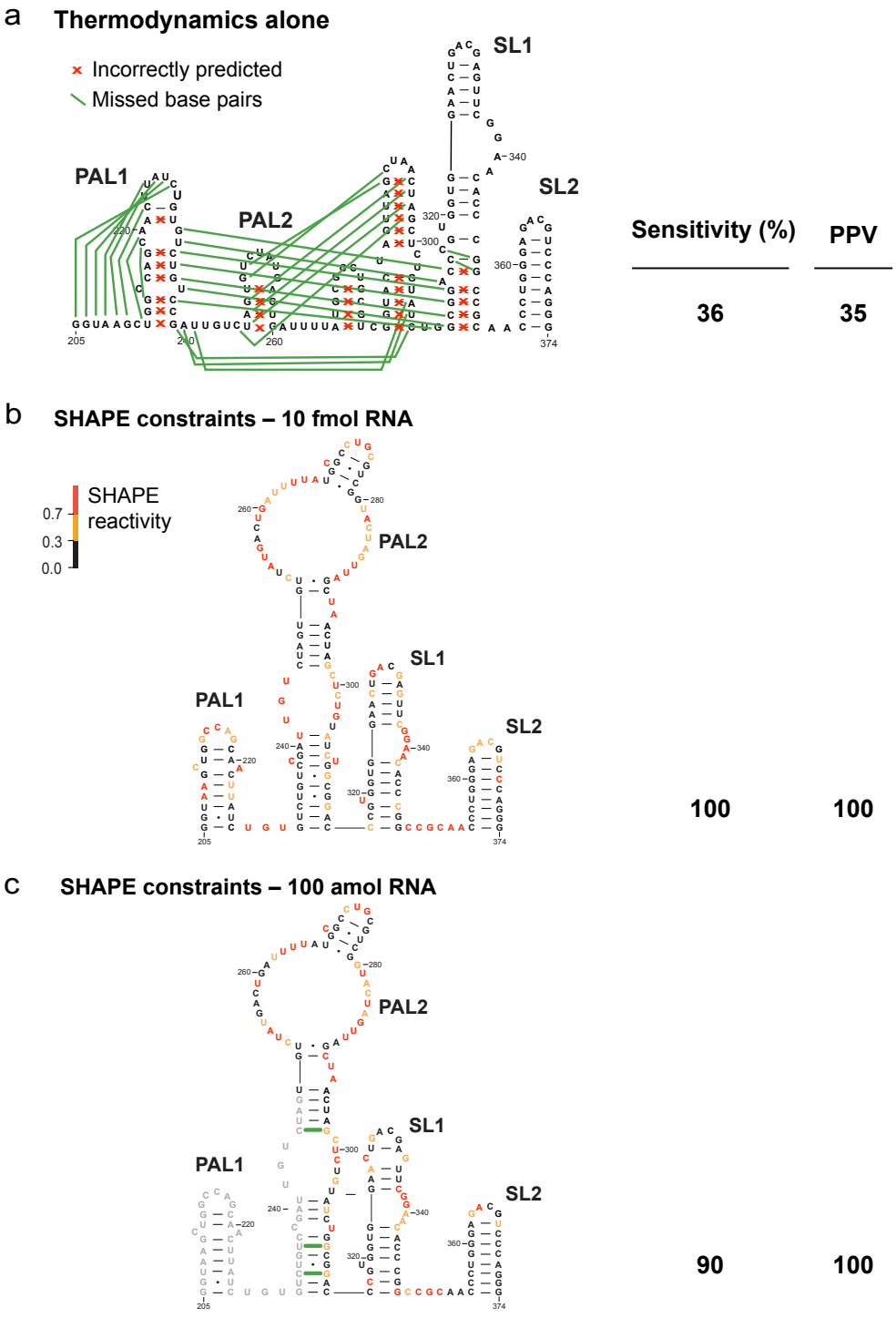


Figure 2.6 SHAPE-directed folding of the MuLV genomic RNA monomer. (a) Secondary structure of the MuLV monomer in the packaging domain as predicted using thermodynamic constraints alone. Missed base pairs and incorrectly predicted base pairs relative to the accepted structure[14] are denoted by green lines and red x's, respectively. (b and c) MuLV monomer structures predicted using SHAPE constraints to direct RNA secondary structure predictions. SHAPE data were obtained at (b) 10 fmol input RNA and (c) 100 amol input RNA.

and PAL2, form intermolecular duplexes in the dimeric structure, as evidenced by their uniformly low SHAPE reactivities (Figure 2.7b). In contrast, when these structures are forced into a monomeric conformation, the lowest free energy models contain several structures with reactivities that are incompatible with the SHAPE data (Figure 2.8). The apical loops in SL1 and SL2 form intermolecular loop-loop kissing interactions, as these nucleotides are unreactive towards SHAPE and by analogy with previous work on MuLV (Figure 2.6b).[23]

We next analyzed the structure of the XMRV genome inside viral particles, again using fmol quantities of RNA. Intact viral particles were treated with the 1M7 SHAPE reagent, the modified RNA was extracted from the virions and primer extension products were quantified by ultra-sensitive SHAPE. Because SHAPE data are highly reproducible, the net effect of different environments – including a complex *in virio* environment – can be readily characterized by subtracting one SHAPE profile from another to yield a difference plot (Figure 2.7c). The difference analysis revealed three major regions that were less reactive, or protected, in virions than in the gently extracted genomic RNA and one segment that was more reactive in virions (in blue, Figure 2.7c). These experiments, performed on fmol amounts of a structurally uncharacterized authentic retrovirus, provided strong clues as to the mechanism of genome packaging in XMRV.

2.7 Discussion

2.7.1 Authentic RNAs are difficult to study.

Most biologically important RNAs are present at very low levels. As a result, studies on these RNAs have been limited to indirect functional studies or direct experimental studies using *in vitro* transcribed RNA as models. Recent work blending SHAPE with next-generation sequencing has lowered detection limits to the tenth-picomole (10^{-13}) level.[7] Analysis of scarce RNAs might be further optimized by RT-PCR approaches but care must be taken so that processing steps do not reduce the

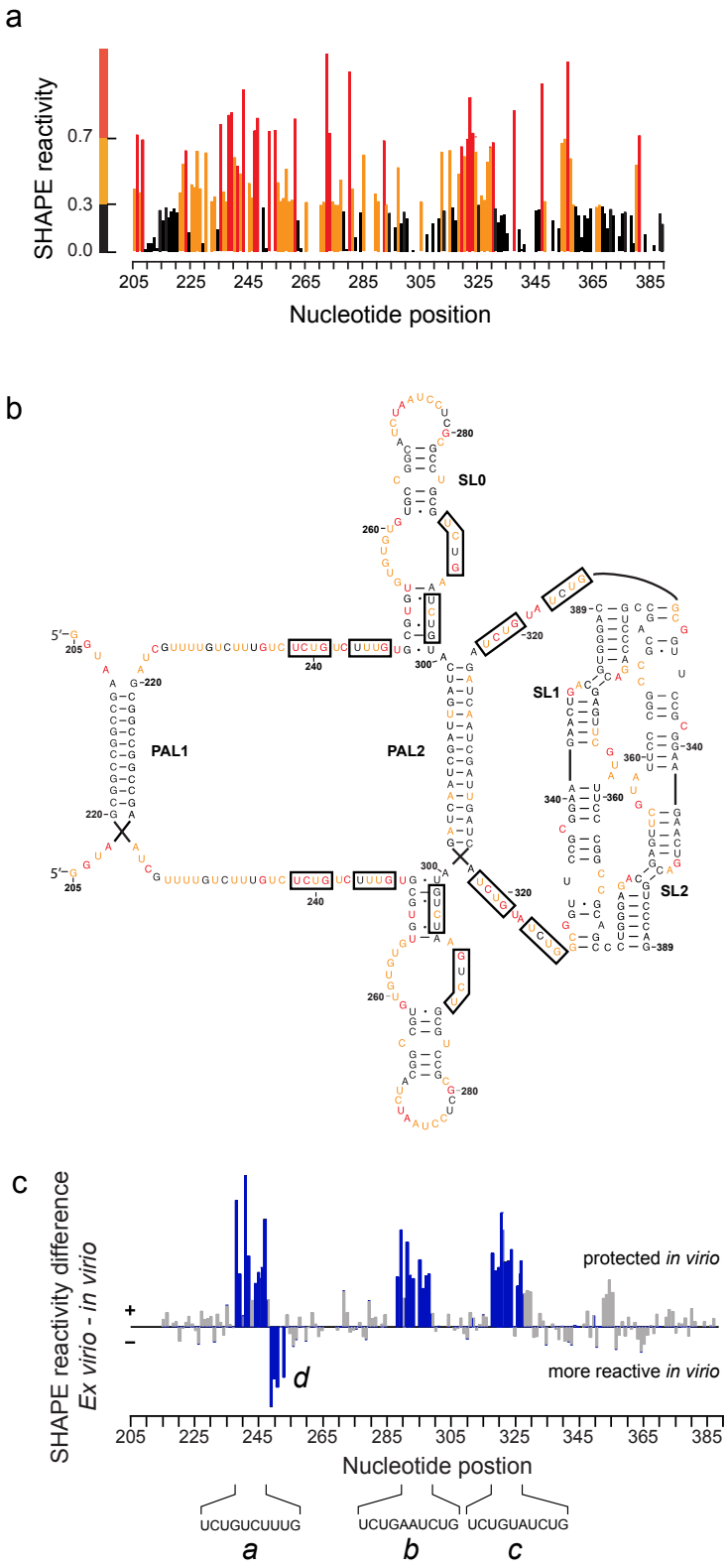


Figure 2.7 Structure analysis of authentic XMRV genomic RNA. (a) Integrated SHAPE data. (b) Secondary structure model for the XMRV RNA genome. (c) SHAPE reactivity differences between the RNA gently extracted from virions versus the genomic RNA modified inside virion particles.

quantitative accuracy of the resulting structural information.[36] In this study, we increased the molecular sensitivity of the SHAPE nucleotide-resolution RNA structure probing technology to sub-femtomole (10^{-16}) levels by implementing multiple innovations to achieve highly sensitive capillary electrophoresis detection of the fluorescently labeled cDNA primer extension products.

2.7.2 Biological implications of US-SHAPE.

Using ultra-sensitive SHAPE, we examined the structure of the dimerization and packaging regulatory domain of XMRV genomic RNA gently extracted from virions and also within virions. The ultra-sensitive SHAPE data obtained on authentic viral RNA strongly support a model in which XMRV is present as a dimer in mature virions (compare Figures 2.7b and Figure 2.8). Genomic dimerization is mediated by four key intermolecular interactions at PAL1, PAL2, SL1 and SL2 sequences (Figure 2.7b). Remarkably, even though there are extensive sequence differences and nucleotide insertions in XMRV relative to MuLV (Figure 2.9), the structures of their respective packaging domains are conserved. It is likely that a similar dimeric architecture is evolutionarily conserved among all gammaretroviruses.

The dimerization and packaging domain in the MuLV virus has been intensively studied and packaging of the RNA genome involves, in part, a binding interaction between the nucleocapsid (NC) domain of Gag and a specific sequence of nucleotides in the RNA.[37] The MuLV NC protein binds tandem repeats of the sequence UCUG. The strongest interactions occur at the first U and final G of the motif and this single-stranded element must be flanked by helical regions for highest-affinity binding.[25] There are two such elements in the MuLV RNA genome and these elements constitute the RNA packaging signal for this virus.[25]

In contrast, inspection of the SHAPE reactivity differences in the XMRV packaging domain for the *ex virio* and *in virio* RNA revealed three strong sets of

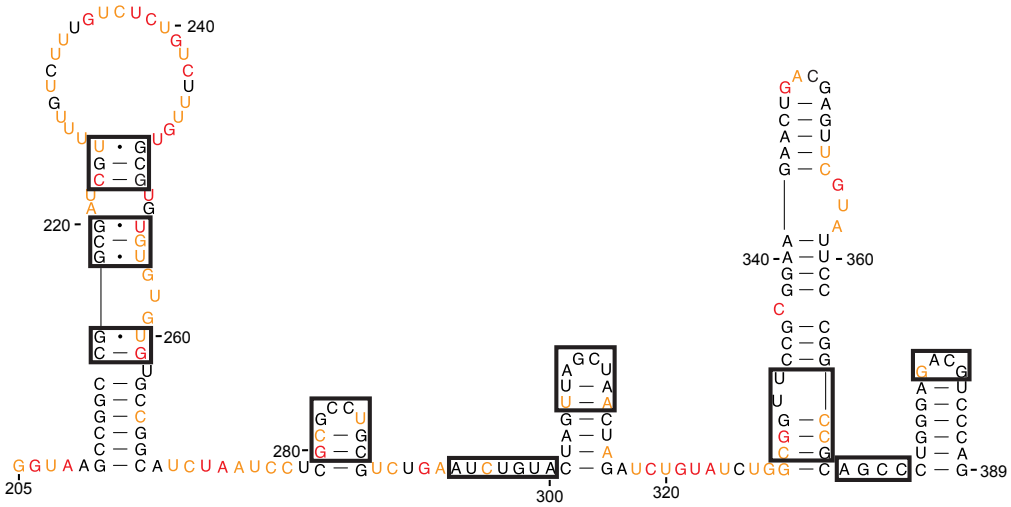


Figure 2.8 XMRV SHAPE data forced into a monomeric conformation. Regions of disagreement with the SHAPE data are boxed.

protections. By analogy with previous work on MuLV[25], virion-specific protection from SHAPE reactivity is likely due to protein binding. Virion-specific increases in reactivity are most directly attributed to RNA conformational changes. Two of these three motifs are unique to XMRV as compared to MuLV. The first unique motif (Figure 2.7c, motif *a*) has the sequence UCUGUCUUUG and is positioned between the structured elements PAL1 and SL0 (see boxed nucleotides 238-247, Figure 2.7b). The second motif, UCUGAAUCUG, has no analog in MuLV (Figure 2.7c, motif *b*) and is partially occluded by the SL0 helix in the deproteinized RNA structural model (Figure 2.7b, nucleotides 289-298). Difference plot analysis suggested that the base of helix SL0 is disrupted upon NC binding and has a different conformation inside the virion. (Figure 2.7c, region *d*). This model is supported by *in virio* differences within SL0 compared with the *ex virio* state (Figure 2.10, double-headed arrow). It is also consistent with the known nucleic acid chaperone properties of NC.[38] The third motif in (Figure 2.7c, motif *c*), UCUGUAUCUG, is identical in both position and sequence to that in MuLV (Figure 2.7b, nucleotides 318-327). These three flexible regions in XMRV, consisting of some form of the sequence UCUGRRURUG (Figure 2.10, see regions *a*, *b* and *c*), are flanked by structured regions and interact most strongly with NC at the first U and last G nucleotides within these motifs (capitalized in Figure 2.10, lower panel).

In conclusion, ultra-sensitive SHAPE has made possible structural analysis of a rare, recently discovered, authentic genomic viral RNA. SHAPE characterization of the XMRV RNA within virion particles revealed functional determinants of a structure critical for packaging. XMRV appears to use a genome packaging strategy that is similar to that employed by MuLV, but has distinct and more complex features (Figure 2.10). We anticipate that by making analysis of femtomole quantities of RNA possible, high sensitivity SHAPE technologies will accelerate functional and structural characterization

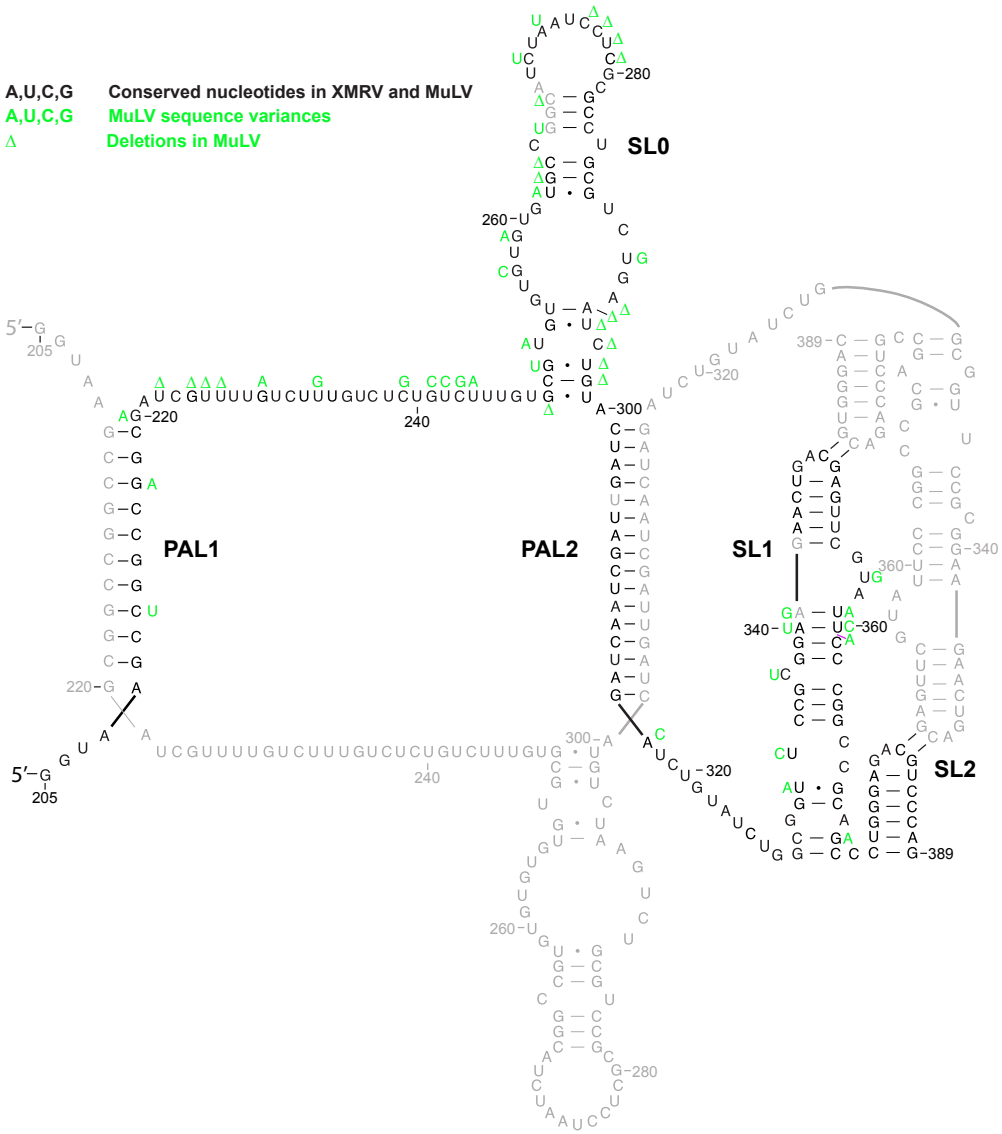


Figure 2.9 Comparison of XMRV and MuLV packaging domain sequences. Sequence and secondary structure model of the XMRV packaging domain. The second strand is shown in gray for clarity. Nucleotides that differ in MuLV compared to XMRV are in green; deletions relative to MuLV are shown as green triangles. Although there are numerous insertions and deletions in XMRV compared to MuLV, the architecture of the packaging domains are the same.

of many low abundance RNAs, including viral genomes, rare biological mutants, and, with further refinement, complex mixtures like messenger RNAs and clinical samples.

2.8 Experimental Procedures.

2.8.1 Instrument design and construction.

The high-sensitivity capillary electrophoresis laser-induced fluorescence instrument is of custom design and construction (see Figure 2.11).[39] The detector utilizes a right-angle excitation design (Figure 2.2). A single-mode fiber-optic cable with an integrated lens (3.5-cm focal length, Oz optics) was used to focus the beam from a multiline argon laser (488 and 514 nm, JDS Uniphase) into the lumen of a capillary. Light from the capillary was collected using a microscope objective (40x, 0.75 na, Plan Fluor, Nikon). The laser and capillary were positioned with respect to each other using an XYZ stage (Newport). A dichroic filter (FF550-Di01-25X36, Semrock) was used to divide the collected light into the 5-FAM and 6-Joe wavelengths. Collection lenses (two collimators for single-mode fiber/ HPUCO-23-400/700-M-50AC-30OD, Oz Optics) were used to direct the light into multimode optical cables (200/230 μm multimode IRVIS fiber/ QMMJ-33-IRVIS-200/230-3-1, Oz Optics). The entry-way into the fiber optic cable also acted as a pin hole to spatially filter the light and remove residual scattered excitation light. Bandpass filters (525 and 565 nm; FF01-525/20-25, FF01-565/24-25, Semrock) appropriate for each fluorophore were used to further remove excitation light prior to entry into the detector. The light was detected, amplified and converted to current using a photomultiplier tube (PMT, H7732, Hamamatsu) mounted in a custom housing with shutters and filter slots. Current from the PMT was amplified and converted to voltage using a preamplifier (Advanced Research Instruments). The signal was then digitized using an analog to digital board (National Instruments) and the

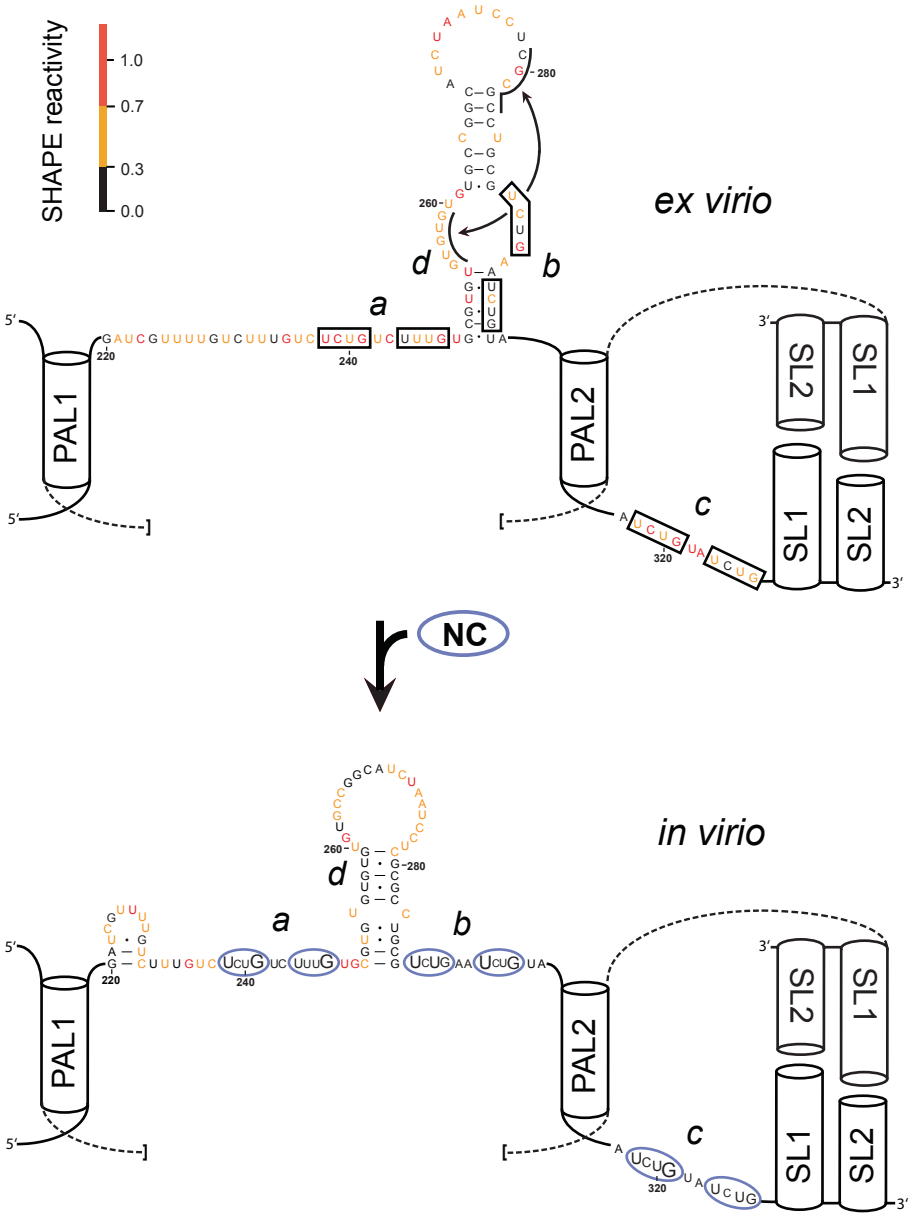


Figure 2.10 Model for XMRV retroviral genome packaging. Secondary structure model for the NC domain of Gag bound to the XMRV packaging signal RNA. Regions with the largest changes in SHAPE reactivity for the RNA analyzed inside virions, relative to the ex virio RNA, are identified (*a*-*d*, bottom panel) with the same labels used in Figure 2.7c. Conserved helical elements that stabilize the dimer state and do not differ in conformation when *in virio* and *ex virio* data are compared are shown schematically as cylinders; these helices flank single-stranded regions bound by NC and are components of the protein binding site.[25]

instrument was controlled via a custom Labview program designed in conjunction with the UNC Chemistry department electronics shop (available upon request). Prefabricated 310 Genetic Analyzer capillaries 47 cm in length with a 50-μm inner diameter (Applied Biosystems) were constructed to have an optical window 36 cm from the inlet. Two nonconductive columns were mounted on either side of the optical unit to accommodate the capillary inlet, outlet, and custom-built capillary washing system. The capillary electrophoresis system was powered by a high voltage power supply (Spellman). A full set of mechanical drawings is available upon request.

2.8.2 Instrument limit of detection.

Experiments to determine instrumental detection limits were performed using free 5-FAM and 6-JOE fluorophores. Samples, diluted from a 10 μM stock, were hydrodynamically injected into the system by elevating the capillary at the sample inlet by 10 cm for 12 s to permit sample introduction via siphoning. The volume of sample injected at 0.17 nL/sec[27] into the 50-μm capillary using gravity injection was thus approximately 2 nL. Experiments were carried out in triplicate (Figure 2.4). The capillary was filled with 100 mM borate (pH 8) and samples were separated at 6 kV for 10 min. Samples were diluted until an integrated fluorescence signal-to-noise ratio of 3:1 was no longer maintained. The signal-to-noise ratio was calculated by dividing the integrated fluorescence by standard deviation of background signal from the first 100 s after injection.

2.8.3 Retroviral RNA transcripts.

The *in vitro* MuLV RNA construct is 331 nucleotides and spans the dimerization and packaging regulatory domain (~170 nts) including 5' and 3' flanking sequences of 46

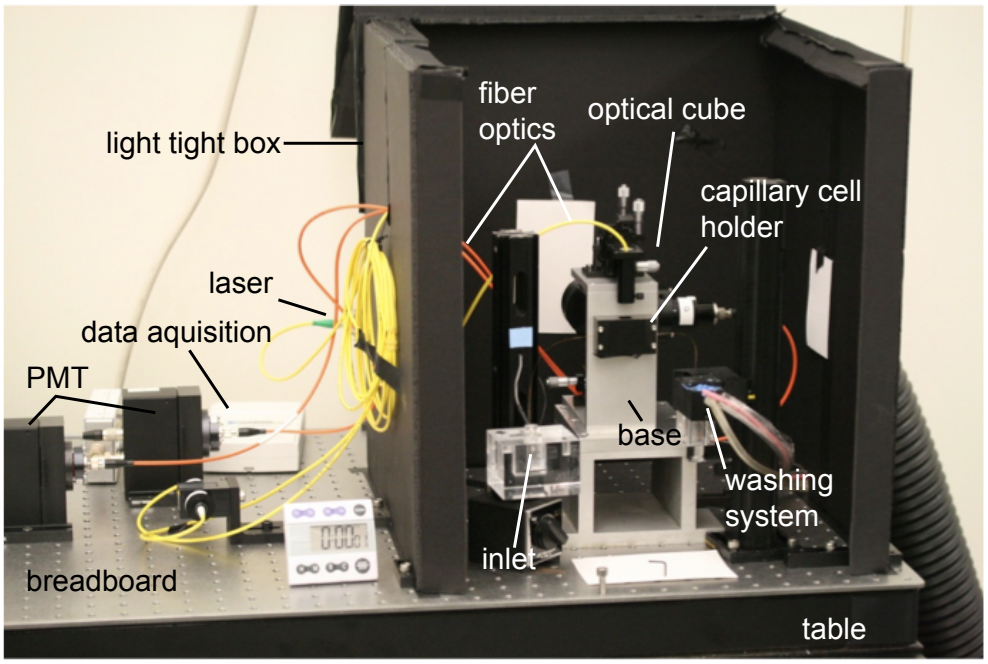


Figure 2.11 Ultra-sensitivity CE and key instrument components. Image of the instrument with major components labeled.

and 115 nucleotides, respectively.[14,17,24] The transcript was generated and purified as described.[25]

2.8.4 Isolation of XMRV.

XMRV was obtained from the constitutively producing XMRV cell line, 22Rv1 (ATCC),[35] cultured in RPMI media with 10% (vol/vol) fetal bovine serum. Culture fluids were collected and filtered, first through 1-μm then 0.45-μm filters, before virus was isolated by continuous-flow ultracentrifugation.[34] Virus was subtilisin-treated as described[40] and collected by centrifugation through a 3.6-mL 20% (wt/vol) sucrose (in PBS without Ca^{2+} or Mg^{2+}) cushion at 390,000 × *g* for 2 h at 4 °C. The supernatant was removed, the viral pellet was overlaid with ice-cold PBS (without Ca^{2+} or Mg^{2+}) and the sample was centrifuged for an additional 1 hr at 4 °C. The supernatant was aspirated and the pellets were resuspended in 600 μL HNS buffer [50 mM HEPES (pH 8.0), 200 mM NaCl, 0.1 mM EDTA, 10% (v/v)] for further manipulations, described below.

2.8.5 SHAPE analysis.

XMRV RNA, gently extracted from XMRV virions as described previously,[15] was incubated in 1× folding buffer [50 mM HEPES (pH 7.5), 200 mM potassium acetate (pH 7.5), and 5 mM $MgCl_2$] at 37 °C for 20 min prior to modification. The resuspended virus in HNS buffer (300 μL) was pipetted into 33.3 μL 50 mM 1M7 (1-methyl-7-nitroisatoic anhydride) in DMSO (plus reagent) or neat DMSO (minus reagent), mixed by pipetting and incubated at 37 °C for 15 min to modify RNA. Virions were collected by ultracentrifugation as described.[40] Virion resuspensions were preincubated at 37 °C for 5 min and 300 μL of virus solution was pipetted into 33 μL solutions of either 1M7 (50 mM) or DMSO at 37 °C. Solutions were mixed by pipetting and incubated for 15 min at 37 °C. Samples turn from faint yellow to dark orange upon reaction with 1M7. Samples were resuspended in 100 μL of 1× lysis buffer [50 mM Hepes (pH 7.5), 200 mM NaCl, 3 mM $MgCl_2$] and lysed with 1%

(wt/vol) SDS and 20 mg/mL proteinase K at 22 °C for 30 min. The lysate was extracted four times with phenol/chloroform/isoamyl alcohol (25:24:1), followed by four extractions with pure chloroform. Samples were then precipitated with ethanol and stored in 70% ethanol at -20 °C until primer extension.

In vitro generated MuLV RNA monomer transcripts (2 pmol) were renatured by heating at 95 °C for 3 min in 12 µL water, cooled on ice for 3 min and equilibrated at 37 °C for 3 min in a folding buffer without magnesium [50 mM HEPES (pH 7.5), 200 mM potassium acetate (pH 7.5)]. SHAPE modification was performed by treating 9 µL MuLV RNA with 1 µL 1M7 (50 mM in DMSO), followed by incubation at 37 °C for 2 min. RNA was recovered using a precipitation approach optimized for the quantitative recovery of low abundance RNAs [2.5 vol ethanol, 1 vol isopropanol, 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.

2.8.6 Primer extension.

The protocol was described in detail previously.[4,14] DNA primers incorporated locked nucleic acid nucleotides (LNA, underlined) for use on low-copy RNA to increase binding affinity and stringency.[41-43] The efficiency of primer extension using LNA-containing primers was calculated to be 59 ± 5% based on the fraction full-length extension product relative to the sum of this value and one-half the unextended primer peak (to correct for a 2:1 primer to RNA molar ratio during primer extension), using an electropherogram from a no-reagent control reaction. MuLV primers (5'-G<u>G</u>U<u>G</u>C <u>A</u>C<u>C</u>AA AGAGU CCAAA AGC-3', 5'-end labeled with 5-FAM or 6-JOE) annealed 3' of the MuLV dimerization domain (nucleotides 422 to 445) XMRV primers (5'-G<u>A</u>A<u>G</u>T <u>C</u>T<u>C</u>TG TCTCT CGTCC CCC-3') annealed to nucleotides 443 to 468 in the genome sequence. Primers (0.5 µL; 2 fmol) were annealed to XMRV RNA (6 µL; 1 fmol in 1/2× TE buffer) by heating

at 65 °C and 45 °C for 5 min and then snap-cooled on ice. Reverse transcription buffer [3 µl; 200 mM Tris (pH 8.0), 250 mM KCl, 10 mM $MgCl_2$, 2 mM each dNTP, 20 mM DTT] was added at 0 °C, and primer extension was performed with Superscript III (Invitrogen) reverse transcriptase (0.5 µl; 100 U) at 45 °C for 1 min, followed by incubation at 52 °C and 65 °C, for 10 min each. MuLV transcript RNA was subjected to primer extension exactly as described above except that primer and RNA concentrations were 2 and 1 pmol, respectively. Reactions were quenched by cooling to 4 °C and addition of 3 M sodium acetate (pH 5.2). A sequencing marker was generated by adding 0.5 µL dideoxy-GTP (10 mM) to the primer extension reaction mixture using unmodified RNA. The 1M7 and DMSO reaction mixes were each combined with equal amounts of dideoxy-GTP-terminated sequencing ladders, precipitated with ethanol, and resuspended in deionized formamide (20 µl).

2.8.7 Capillary electrophoresis.

For conventional capillary electrophoresis, cDNA fragments were resolved on an Applied Biosystems 3130 instrument. This instrument is fully automated for use with high abundance cDNA samples of 1 pmol or greater. The samples were loaded onto microtiter plates and injected onto the instrument using standard instrument protocols. MuLV and XMRV cDNA fragments, ranging in concentration from 100 amol to 50 fmol of input RNA, were resolved by custom high-sensitivity capillary electrophoresis. For ultra-sensitive SHAPE, samples of 1 to 10 µL were electrokinetically injected (10.2 kV for 30 s) onto a 47 cm (50-µm inner diameter) uncoated capillary containing Pop7 (Applied Biosystems) as the internal polymer. The detection window was placed at 36 cm. Voltage was applied using positive polarity (10.2 kV for 60 min) and immersing the capillary ends in a 1× running buffer (Applied Biosystems). The PMT voltages were 0.9 kV and the laser power was 2 mW. Excitation was achieved using an argon laser at 488 and 514 nm wavelengths focused within the capillary at the detection window. 5-FAM was used to detect cDNAs

corresponding to SHAPE structure probing and the lower sensitivity dye, 6-JOE, was used for sequencing. A single high-sensitivity SHAPE experiment was thus resolved in two separate capillary electrophoresis runs, one each for the plus- and no-reagent reactions. Sensitivity for sequencing is not a critical issue because sequencing can be performed with an *in vitro* transcript or with DNA rather than a scarce biological sample.

2.8.8 Data processing and structure prediction.

Raw sequence traces were corrected for dye variation and signal decay; peak intensities were integrated using ShapeFinder.[12] Traces were analyzed by combining data from two separate capillary runs. Each capillary electrophoresis separation contained (*i*) a reaction performed in the presence or absence of 1M7 (detected using a FAM-labeled primer) and (*ii*) a sequencing reaction (labeled with JOE) performed using dideoxy GTP. Data from the two runs were aligned using the identical sequencing lanes and then analyzed as described.[12] SHAPE reactivities were normalized by dividing by the average intensity of the 10% most reactive nucleotides, after excluding outliers found by box plot analysis.[29] SHAPE reactivity information was used to impose a pseudo-free energy change constraint in conjunction with nearest neighbor thermodynamic parameters in the prediction program RNAstructure.[29,30,33]

2.8.9 Acknowledgments.

We are indebted to Freddy Pinero in the UNC Chemistry Instrument shop for mechanical drawings and construction of custom instrument pieces, Alex Villa for creation of the Labview software, Jenny Mauger for helpful discussions on Labview programming, and Julian W. Bess, Jr., (AIDS and Cancer Virus Program, Biological Products Core) for providing and purifying XMRV stocks for this work.

2.9 REFERENCES.

1. Coffin, J.M., Hughes, S.H., & Varmus, H.E. Retroviruses. (Cold Spring Harbor Press, Plainview, NY; 1997).

2. D'Souza, V. & Summers, M.F. How retroviruses select their genomes. *Nat Rev Microbiol* **3**, 643-655 (2005).

3. Roberts, L. & Holcik, M. RNA structure: new messages in translation, replication and disease. Workshop on the role of RNA structures in the translation of viral and cellular RNAs. *EMBO Rep* **10**, 449-453 (2009).

4. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610-1616 (2006).

5. Watts, J.M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716 (2009).

6. Mauger, K.M.W.a.D.M. Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.* **44** (2011).

7. Lucks, J.B. et al. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci U S A* **108**, 11063-11068 (2011).

8. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129**, 4144-4145 (2007).

9. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127**, 4223-4231 (2005).

10. Mortimer, S.A. & Weeks, K.M. Time-resolved RNA SHAPE chemistry. *J Am Chem Soc* **130**, 16178-16180 (2008).

11. Wilkinson, K.A. et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**, e96 (2008).

12. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. & Giddings, M.C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979-1990 (2008).

13. Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G. & Weeks, K.M. Strong correlation between SHAPE chemistry and the generalized NMR

order parameter (S2) in RNA. *J Am Chem Soc* **130**, 12244-12245 (2008).

14. Gherghe, C. & Weeks, K.M. The SL1-SL2 (stem-loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer. *J Biol Chem* **281**, 37952-37961 (2006).

15. Gherghe, C., Leonard, C.W., Gorelick, R.J. & Weeks, K.M. Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. *J Virol* **84**, 898-906 (2010).

16. MacTaylor, C.E. & Ewing, A.G. Critical review of recent developments in fluorescence detection for capillary electrophoresis. *Electrophoresis* **18**, 2279-2290 (1997).

17. Borland, L.M., Kottegoda, S., Phillips, K.S. & Allbritton, N.L. Chemical analysis of single cells. *Annu Rev Anal Chem (Palo Alto Calif)* **1**, 191-227 (2008).

18. Urisman, A. et al. Identification of a novel Gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *PLoS Pathog* **2**, e25 (2006).

19. Weiss, R.A. A cautionary tale of virus and disease. *BMC Biol* **8**, 124 (2010).

20. Paprotka, T. et al. Recombinant origin of the retrovirus XMRV. *Science* **333**, 97-101 (2011).

21. Lombardi, V.C. et al. Detection of an infectious retrovirus, XMRV, in blood cells of patients with chronic fatigue syndrome. *Science* **326**, 585-589 (2009).

22. Ly, H. & Parslow, T.G. Bipartite signal for genomic RNA dimerization in Moloney murine leukemia virus. *J Virol* **76**, 3135-3144 (2002).

23. Badorrek, C.S. & Weeks, K.M. Architecture of a gamma retroviral genomic RNA dimer. *Biochemistry* **45**, 12664-12672 (2006).

24. Badorrek, C.S., Gherghe, C.M. & Weeks, K.M. Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *Proc Natl Acad Sci U S A* **103**, 13640-13645 (2006).

25. Gherghe, C. et al. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci U S A* (2010).

26. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology* **24**, 104-108 (1992).

27. Weinberger, R. Practical Capillary Electrophoresis, Vol. 2. (Academic Press, San Diego; 2000).

28. Ghosal, S. Electrokinetic flow and dispersion in capillary electrophoresis. *Annu Rev Fluid Mech* **38**, 309-338 (2006).

29. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A* **106**, 97-102 (2009).

30. Low, J.T. & Weeks, K.M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150-158 (2010).

31. Badorrek, C.S. & Weeks, K.M. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat Chem Biol* **1**, 104-111 (2005).

32. Adam, M.A. & Miller, A.D. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J Virol* **62**, 3802-3806 (1988).

33. Mathews, D.H. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101**, 7287-7292 (2004).

34. Chertova, E. et al. Envelope glycoprotein incorporation, not shedding of surface envelope glycoprotein (gp120/SU), Is the primary determinant of SU content of purified human immunodeficiency virus type 1 and simian immunodeficiency virus. *J Virol* **76**, 5315-5325 (2002).

35. Sramkoski, R.M. et al. A new human prostate carcinoma cell line, 22Rv1. *In Vitro Cell Dev Biol Anim* **35**, 403-409 (1999).

36. Weeks, K.M. RNA structure probing dash seq. *Proc Natl Acad Sci U S A* **108**, 10933-10934 (2011).

37. D'Souza, V. & Summers, M.F. Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* **431**, 586-590 (2004).

38. Levin, J.G., Guo, J., Rouzina, I. & Musier-Forsyth, K. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* **80**, 217-286 (2005).

39. Sims, C.E. et al. Laser-micropipet combination for single-cell analysis. *Anal Chem* **70**, 4570-4577 (1998).

40. Ott, D.E. et al. Analysis and localization of cyclophilin A found in the virions of human immunodeficiency virus type 1 MN strain. *AIDS Res Hum Retroviruses* **11**, 1003-1006 (1995).

41. Latorra, D., Arar, K. & Hurley, J.M. Design considerations and effects of LNA in PCR primers. *Mol Cell Probes* **17**, 253-259 (2003).

42. Levin, J.D., Fiala, D., Samala, M.F., Kahn, J.D. & Peterson, R.J. Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers. *Nucleic Acids Res* **34**, e142 (2006).

43. Fratczak, A., Kierzek, R. & Kierzek, E. LNA-modified primers drastically improve hybridization to target RNA and reverse transcription. *Biochemistry* **48**, 514-516 (2009).

CHAPTER 3

IMMATURE MULV GENOMIC RNA PACKAGED AS A STABLE INTERMEDIATE STRUCTURE

3.1 Retroviral genomic RNA is packaged in an immature state.

During viral replication, a structurally distinct genomic RNA dimer is packaged into immature nascent virions.[1,2] As only mature virus has been found to be infective, RNA structural changes have been proposed to be an important part of maturation[3,4] and subsequent pathogenesis. The RNA structure of immature MuLV has not been characterized and only bulk measurements of global structure[4] are currently available to provide information to help distinguish between immature and mature MuLV particles. Likewise, there is no information available to determine whether RNA structures detected *in vitro* are in the immature or mature dimeric state. Presumably, in light of the short life-span of the immature state, current RNA structures of MuLV are of the mature dimer. Furthermore, the packaged state of MuLV genomic RNA has not been analyzed. This limitation is based largely on the extraordinary difficulty of producing enough RNA to perform structural analysis using traditional techniques. Considering the tight link between viral structure and function, the inability to analyze these structural states presents a considerable barrier to fully understanding viral pathogenesis. Here I use new technology on MuLV RNA to try to address this formerly intractable problem.

3.2 Ultra-sensitive SHAPE for highly sensitive structural analysis.

Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) allows characterization of RNA structure at single nucleotide resolution.[5-8] In a traditional SHAPE experiment, target RNA is first equilibrated with a physiologically-relevant folding solution[9]

(e.g., 50 mM HEPES, 200 mM KOAc, 5 mM $MgCl_2$). When added to RNA, 1-methyl-nitroisatoic anhydride (1M7),[10] or other SHAPE reagents,[9,11] react with flexible (single-stranded) nucleotides to form a bulky RNA adduct. Adduct locations are then mapped as stops in a subsequent primer extension reaction using fluorescently labeled primers.[12] Structure-specific cDNAs are separated by capillary electrophoresis and the resulting electropherogram is processed to yield a reactivity profile detailing the likelihood of base pairing for every RNA nucleotide (Figure 3.1a). SHAPE has been successful in analyzing many important retroviral RNA structures, including the entire HIV-1 genome[6] and the mature Moloney Murine Leukemia virus (MuLV) dimerization and regulatory domain.[13] However, structural analysis of authentic viral RNA has remained problematic as purification from infected cells and virions typically yields only attomoles to femtomoles (10^{-18} to 10^{-15}) of material, and capillary electrophoresis-based structural technologies, like SHAPE, require picomoles of RNA.[5]

I have improved the detection limits of capillary electrophoresis (CE) based-techniques by four orders of magnitude directly by designing and building a custom CE instrument with sub-femtomole SHAPE sensitivity.[14] I now apply this technology, ultra-sensitive SHAPE (US-SHAPE), to analysis of the secondary structure of the dimerization and packaging domain of the Moloney murine leukemia virus (MuLV), using samples assayed with SHAPE reagents both after extraction from and while within immature retrovirus particles. I find that this important regulatory domain has a unique structure while it remains an immature, non-infectious particle. From this high-resolution structural information, we created a series of mutants in authentic viruses designed to mimic the immature state and test our structural model. We discovered a mutant which is at least 10-fold deficient in replicative fitness. I then devised a structural model of the packaged immature genomic RNA that is functionally relevant as determined by mutation studies.

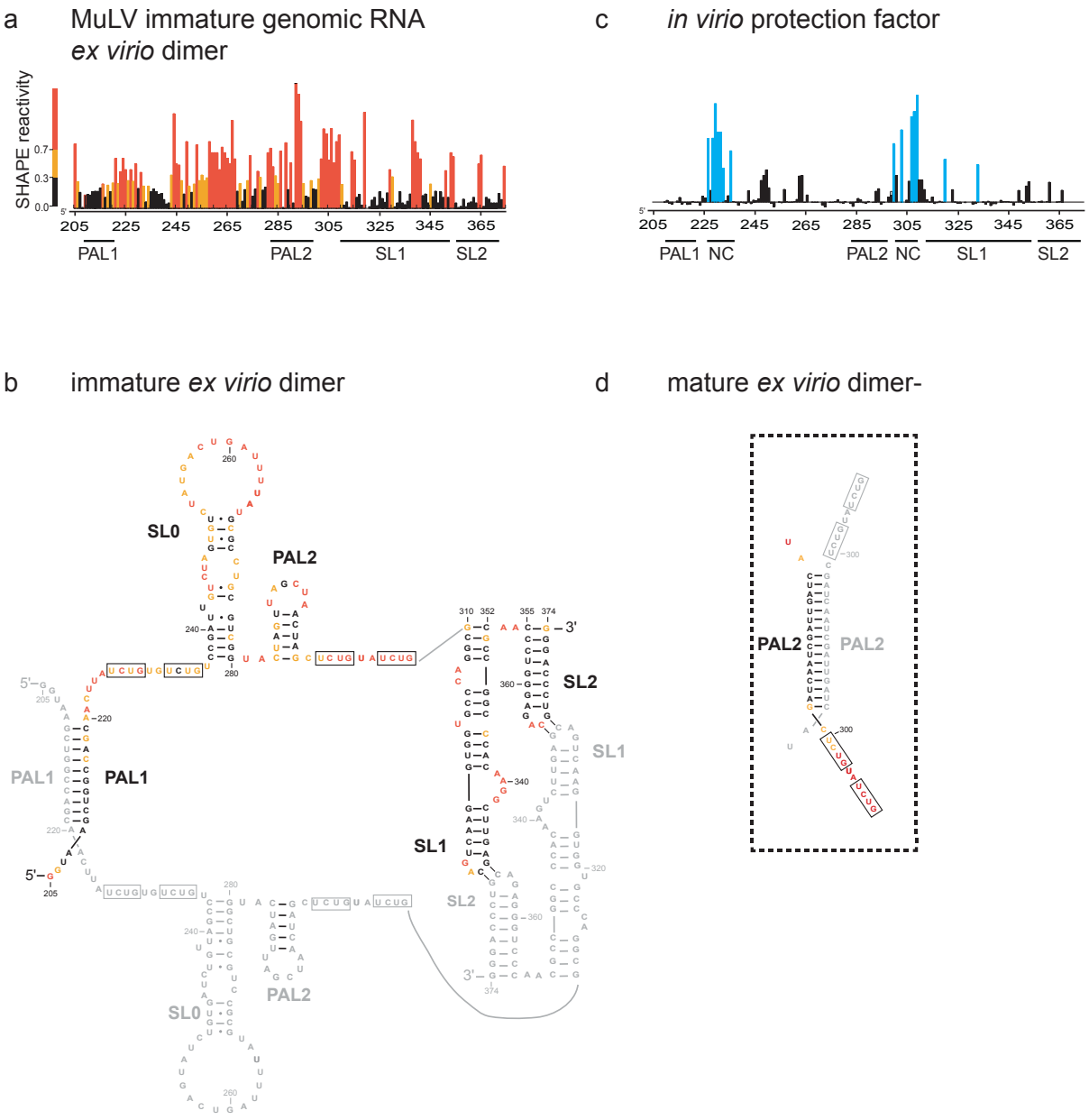


Figure 3.1 Structure analysis of authentic MuLV immature genomic RNA as a dimer. (a) Integrated SHAPE data. (b) Secondary structure model for the immature MuLV RNA genome. (c) Protection factor calculated by normalized SHAPE reactivity differences between the RNA gently extracted from virions versus the genomic RNA modified inside virion particles. (d) Secondary structure model for the mature MuLV palindromic 2 (PAL2) sequence in an intermolecular duplex.

3.3 US-SHAPE analysis of *ex virio* immature MuLV authentic RNA.

Immature MuLV RNA used here is created by transfection of a well-known protease mutant that produces an inactive protease that prevents Gag polyprotein cleavage, and thus viral maturation (see methods).[1,3,15,16] In the laboratory, the result of a large-scale, specialized preparation of transfected cells is only a few hundred femtomoles because protease deficient virus is only capable of a single round of replication. I utilized US-SHAPE, developed precisely for the structural analysis of low-copy number authentic RNAs, to analyze the MuLV dimerization and packaging regulatory domain.[17,18] This region, spanning 170 nucleotides, was selected because it is the most prominent interaction site between the two genomic strands. SHAPE reactivity measurements were obtained using 50 fmol aliquots of MuLV genomic RNA that had been gently extracted from virions purified from cell culture. Processed SHAPE reactivities (Figure 3.1a) were used to develop a model for the structure of this regulatory domain and compared to previously determined structural models of this domain in the mature state.[13,19] SHAPE probing data of RNA isolated from immature virions display a similar overall architecture to that of mature MuLV RNA[13] with one major exception; the structural model of the dimerization and packaging regulatory domain contains only three key dimerization elements (PAL1, SL1. and SL2) whereas the mature dimer state contains four such interaction elements (PAL1, PAL2, SL1, and SL2). The apical loops in SL1 and SL2 form intermolecular loop-loop kissing interactions in both the immature and mature state, as these nucleotides are unreactive towards SHAPE and are analogous with previous observations in MuLV (Figure 3.1b, SL1 and SL2).[20] A palindromic sequence, termed PAL1, forms a 10 base-pair intermolecular duplex in both dimeric structures, as evidenced by their uniformly low SHAPE reactivities (Figure 3.1b, PAL1). A second palindromic sequence, PAL2, forms an intramolecular stemloop in RNA extracted from immature virus (Figure 3.1b, nucleotides 283 through 298), while this sequence forms an

intermolecular duplex in the mature *ex virio* dimer extracted from mature particles (Figure 3.1d).

3.4 US-SHAPE analysis of *in virio* immature MuLV RNA.

I next analyzed the structure of the MuLV genomic dimer within immature viruses, again using femtomolar quantities of RNA as described.[13,14] Intact viral particles were treated with 1M7 SHAPE reagent, the modified RNA was extracted, and the primer extension products were quantified. Highly reproducible SHAPE profiles from the *ex* and *in virio* RNAs were used to create differential plots to reveal the net effect of complexation in the *in virio* environment. For example, protein binding sites, indicated by a decrease in SHAPE reactivity in the *in virio* modified RNA relative to the *ex virio* modified RNA, are readily detected using a protection factor plot. Protection factor analysis was done by subtracting the *in virio* from the *ex virio* SHAPE profile and dividing the difference by the *in virio* profile. This analysis revealed two major regions (blue in Figure 3.1c) that are protected within the virus particle. The protections (Figure 3.1c, labeled NC) are consistent with known nucleocapsid (NC) binding sites for MuLV (Figure 3.1b, boxed).[19] Taken together, I conclude that the genomic RNA dimer is bound to the NC domain of Gag protein within immature virions, but the immature dimer, containing an unformed PAL2 duplex, is structurally distinct from the mature state.

3.5 Immature MuLV RNA resembles a late kinetic step in genomic dimerization.

In recent work from our lab, I used time-resolved SHAPE to develop a single nucleotide resolution mechanistic model for dimerization of an *in vitro* MuLV RNA genome.[21] The culmination of this work was a structural “movie” of genomic RNA dimerization in the presence and absence of nucleocapsid (NC) protein. Here, I extend our analysis to consider the immature MuLV structural state by comparing the SHAPE reactivity profiles of immature MuLV RNA and profiles taken during the dimerization pathway. Strikingly, I found that the immature RNA structure closely resembles a late kinetic step in the dimerization of the

genome (Figure 3.2a), with a linear correlation coefficient of 0.84 (Figure 3.2b). If I correlate the immature structure to each of the 16 consecutive structural "snapshots" taken during dimerization I find that the structure correlates to a single step late in the monomer to dimer transition (Figure 3.2c). I conclude that the immature state is an intermediate in the dimerization of MuLV genomic RNA.

3.6 Mutation and replicative fitness of authentic immature genomic RNA.

I next set out to validate our secondary structural model of the dimerization domain by creating a series of mutations intended to mimic the secondary structure of MuLV RNA in the immature state. Mutants were designed with two structural goals: first we incorporated one or more mutations within the PAL2 sequence that intended to affect the formation of the PAL2 intermolecular duplex and second, I devised mutations to allow or disrupt nucleocapsid binding (Figure 3.3a).

The resulting MuLV PAL2 variants were transfected into live viruses and allowed to replicate in a sarcoma positive leukemia negative (S+L-) focus assay (see methods).[22] The focus inducing units are counted five days after transfection and their resulting replication efficiencies, normalized to the native sequence, are determined (Figure 3.3b). Mutations were made that should strengthen, completely delete, or completely disrupt the PAL2 intermolecular duplex seen in the immature state. Increasing the G-C base-pair content of the PAL2 stemloop was used to strengthen the PAL2 stemloop and these constructs resulted in a viable virus still capable of replication with similar efficiency to the native sequence (70 +/- 40%) (Figure 3.3b, mutated). Surprisingly, the deletion variant (Figure 3.3b, deleted), with a completely excised PAL2 region (nucleotides 283 - 298), also replicated with an efficiency (70 +/- 20%) similar to the native sequence (Figure 3.3b, native). In contrast, one mutant construct wherein the PAL2 sequence was replaced with a string of A's (Figure 3.3a and b, all to A's) was reproducibly found to be replication deficient (10 +/- 10%) as compared to the native sequence. In summary, we found that while neither

strengthening nor complete deletion of PAL2 was capable of preventing the infectivity of the MuLV retrovirus complete disruption of secondary structure of PAL2 produced a reproducible 10-fold replication deficiency. This is likely because it forced altered spacing between NC binding motifs; these date provide clues to the function of the RNA in the immature state.

3.7 Discussion

3.7.1 US-SHAPE for authentic immature RNA structure.

Retroviral RNA genomes form myriad structures that are governed by critical interactions with either the nucleocapsid protein or the nucleocapsid domain of Gag polyprotein.[15,23,24] I have made use of a powerful new method based on SHAPE[9] technology to explore interactions between NC and the immature RNA genome of the Moloney murine leukemia virus (MuLV). In this approach, ultrasensitive SHAPE (US-SHAPE), I used a two-color capillary electrophoresis instrument with femtomole (10^{-15}) sensitivity for SHAPE experiments on authentic viral RNA.[14] The high sensitivity of this instrument has allowed me to characterize authentic immature RNA genomes within MuLV virions at single-nucleotide resolution.

Before now only two pieces of empirical evidence were available to distinguish between the immature and mature states of the MuLV virus. First, cryo-electron microscopy studies of the immature MuLV virion show concentric circles of uncleaved Gag polyprotein subunits including matrix, capsid, and nucleocapsid proteins.[1] Second, genomic RNA extracted from immature virions has a lower thermal stability than their mature counterparts,[3] suggesting a significant difference in secondary structure. Our results verify the packaging and regulatory signal RNA exists in a structurally distinct conformation in the immature state and I have created secondary structure models of the RNA within the immature state (Figure 3.1b).

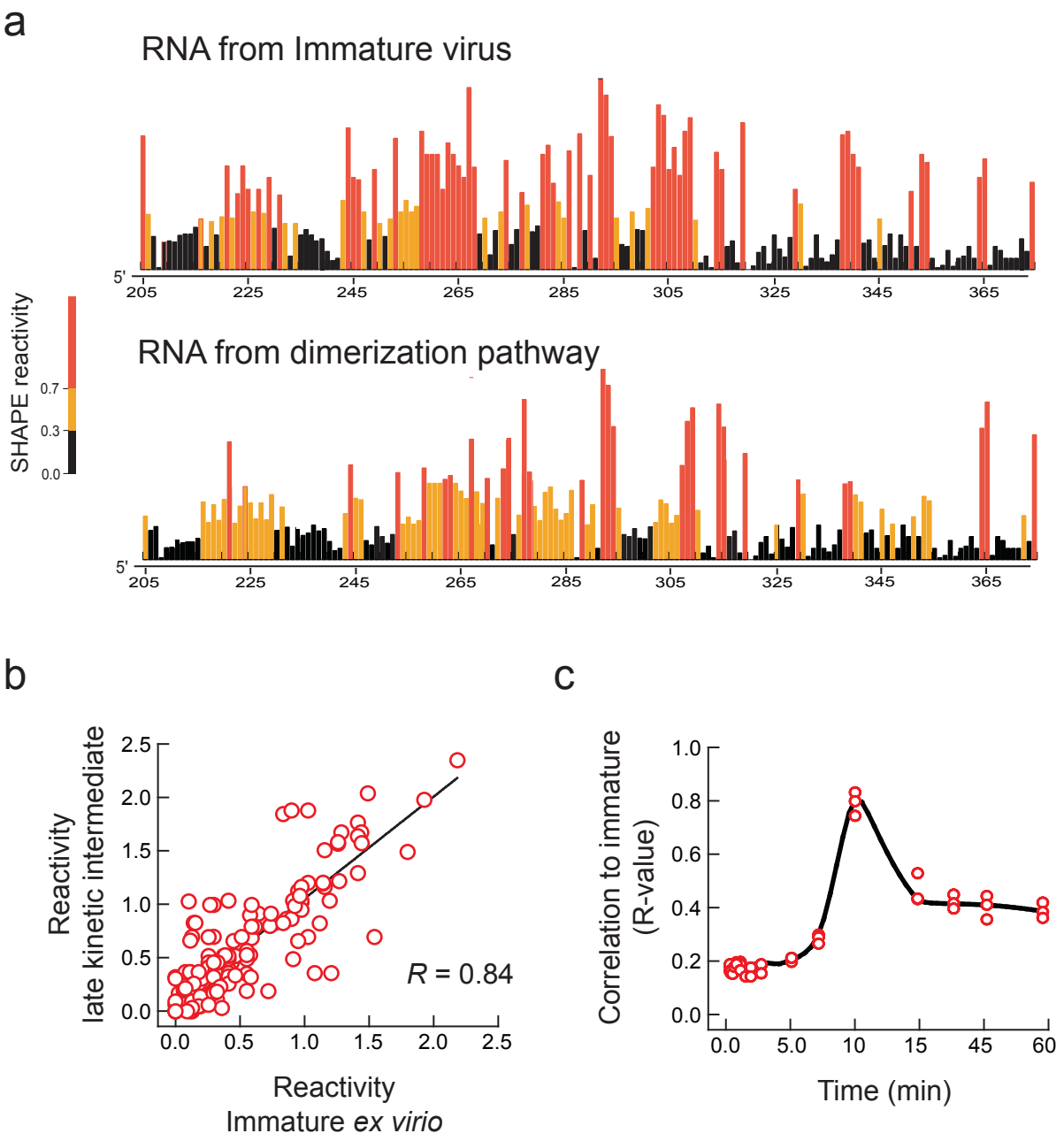


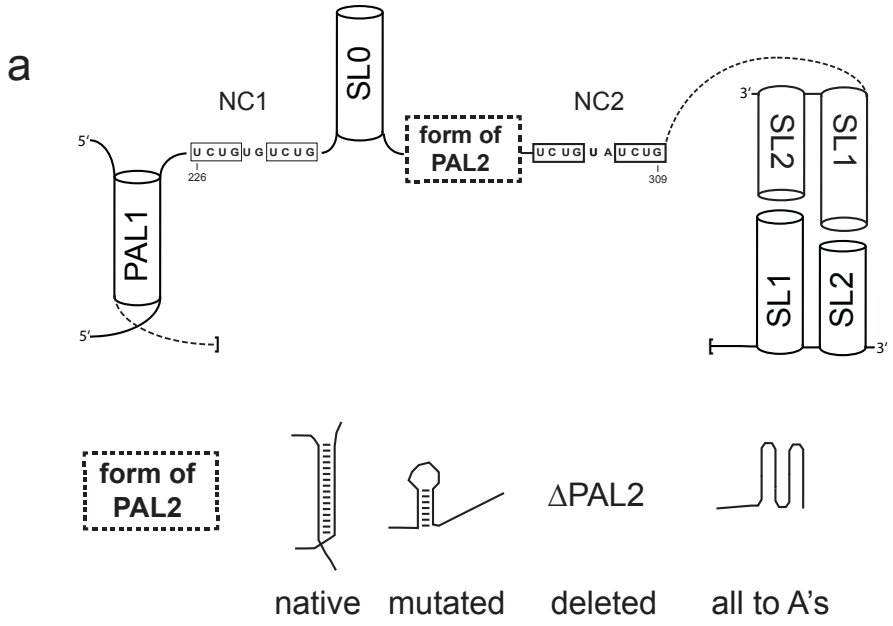
Figure 3.2 Comparison of immature MuLV RNA SHAPE reactivities with SHAPE reactivities of the MuLV dimerization pathway. (a) (Top) Integrated SHAPE data of authentic MuLV immature genomic RNA as a dimer. (Bottom) Integrated SHAPE data of a late timepoint in the dimerization of MuLV. (b) Correlation of SHAPE data in (a). (c) Correlation of immature RNA SHAPE reactivity to every timepoint in the dimerization of MuLV.

3.7.2 Immature RNA is bound to NC domain of Gag polyprotein.

Our *in virio* results confirm that the RNA is bound to the NC domain of Gag polyprotein while within the immature virion (Figure 3.1c). This binding interaction occurs primarily between the NC domain of Gag and tandem UCUG repeats within the dimerization and packaging regulatory domain. One approach to increasing binding affinity is to eliminate off-target binding events, and it appears that the viral RNA uses this approach by flanking each single-stranded binding domain with helical regions, thus sequestering these sites in base-pairing interactions (Figure 3.1b, boxed).[19] This is particularly interesting as RNA-protein interactions have often been found to be non-specific and electrostatic-based, hence specificity in this case is derived from the RNA directly.

3.7.3 Mimicking the immature state.

We next attempted to capture the immature state by creating various viral constructs with mutated PAL2 regions and analyzing their respective effects within authentic virions. We created three MuLV variants (Figure 3.3a) and compared their replication efficiency to the native MuLV sequence in an *in vivo* replication assay (Figure 3.3b). The first mutant contained alternative base-pairings in the PAL2 sequence intended to both strengthen this intramolecular duplex and mimic the immature form of the virus. This mutation was expected to be deficient in replication, however it replicated with near native efficiency (Figure 3.3b, mutated). Contrary to several previously published results we found that deletion of the entire PAL2 sequence had little effect on viral replication, with near native efficiency being observed (Figure 3.3b, deleted). Only disruption of the PAL2 duplex via a mutation that changed every residue in the duplex to an adenosine residue caused the virus to exhibit a slight deficiency in replication (~10-fold) of its genome (Figure 3.3b, all to A's), although this mutant was still capable of multiple rounds of replication.



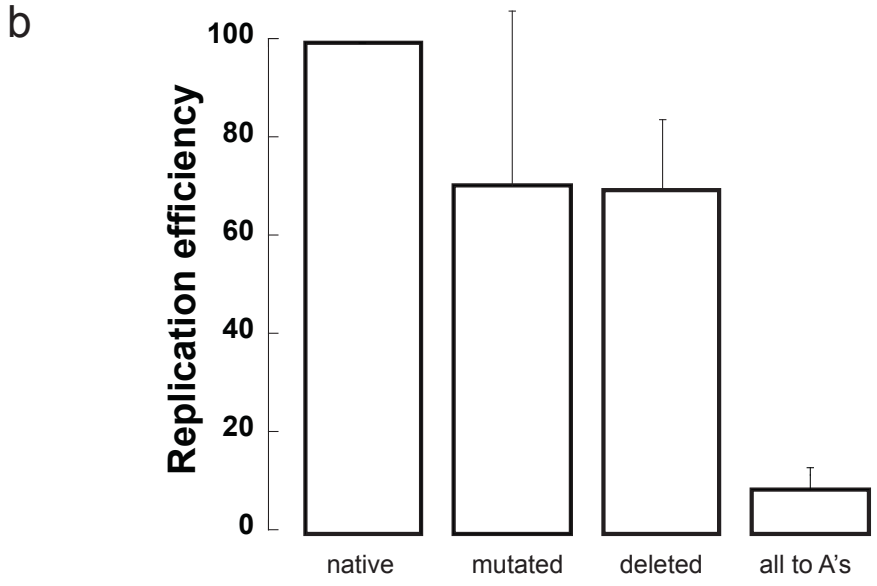


Figure 3.3 Mutation and replication efficiency of authentic immature genomic RNA. (a) Structural model for MuLV genomic RNA. Structured regions are shown as cylinders. Known NC binding sites are boxed. Four mutations are shown with different conformations of PAL2 (dashed box). (b) Replication efficiencies of each of the four PAL2 mutants are shown with error bars as calculated from an S+L- focus assay.[22]

3.7.4 Defining a precise structural spacing for NC binding to immature RNA.

These unexpected results suggested to us that factors other than specific sequence identity are the primary determinants of immature viral dimerization. We hypothesize that a precise 'structural spacing' of the NC binding sites is required for these RNA-protein interactions. Natively, this structural spacing includes two tandem unstructured UCUG motifs flanked by highly structured regions. When the structural spacing is fulfilled by flanking structured regions, as in the native, mutated, and deleted variants (Figure 3.3a), NC is capable of binding and chaperoning the RNA to its mature infectious state (Figure 3.3b). However, when the structural spacing between tandem UCUG motifs is changed sufficiently, as it was when the entire sequence was change to A's (Figure 3.3a, all to A's), the packaging prerequisite is not fulfilled, nucleocapsid does not bind, and thus viral replication is deficient (Figure 3.3b).

3.7.5 Biological implications of the immature state.

We have conducted a fairly wide range of genomic RNA mutational studies which fail to completely disrupt the virus's ability to replicate, resulting in a general hypothesis that the structure of the RNA within the immature virion is not critical to maturation of viral particles. In contrast, our current work indicates that a precise structural spacing is critical for NC binding and maturation of the retrovirus. In addition, our data suggests that the immature RNA secondary structure is a legitimate intermediate in the viral replication pathway as it closely resembles a late kinetic step in the dimerization of the MuLV genome. We propose a functional model where the immature RNA is packaged as a "kinetically trapped" RNA state containing an intramolecular helix that, when bound to the nucleocapsid domain of the polymeric Gag protein, is physically constrained by the Matrix protein and embedded in the viral membrane (Figure 3.4a). Once Gag polyprotein is cleaved into its three constituent proteins the viral genome is released and free NC protein chaperones the genomic RNA into its mature infectious form (Figure 3.4b). Both NC domain and free NC binding to genomic

RNA is controlled by a precise structural spacing of two tandem unstructured UCUG motifs flanked by structural elements. Without this precise structural configuration, NC does not properly recognize and chaperone the RNA to its final infectious state.

3.8 Experimental Procedures.

3.8.1 Creation of Immature MuLV RNA.

Immature MuLV genomic RNA was created using a well-characterized mutant in which the active site of the viral protease contains an aspartate to leucine substition (D32L).[1] When the mutated plasmid is transfected into NIH 3T3 mouse cells, an inactive form of the viral protease is produced (PR-) which is incapable of cleaving Gag polyprotein. The abrogation of Gag prevents maturation and the virion contains immature MuLV genomic RNA. After transfection of pMuLV-D32L, infected cells produce virus. MuLV genomic RNA is packaged into nascent immature virions along with PR-. Because PR- cannot cleave Gag, the virion remains in the immature and non-infectious state. Because immature virus is not infectious, only one round of immature virions is produced per transfection. The viral titer is therefore in attomoles (~10,000 molecules) and a large-scale prep is required to produce femtomole amounts of immature genomic RNA. Immature virions are pelleted by centrifugation at 25,000 rpm at 4 °C and resuspended in PBS. The immature RNA is then modified within the immature virions by addition of 5 mM 1M7. For both *ex virio* RNA or after *in virio* modification, the virions are lysed by incubating at 37 °C for 30 minutes in a buffer containing Tris pH 7.4, EDTA, SDS. NaCl, proteinase K. Cell lysate is cleared by centrifugation and the viral RNA is purified by phenol-chloroform extraction and ethanol precipitation. Immature MuLV RNA is then analyzed by US-SHAPE.[14]

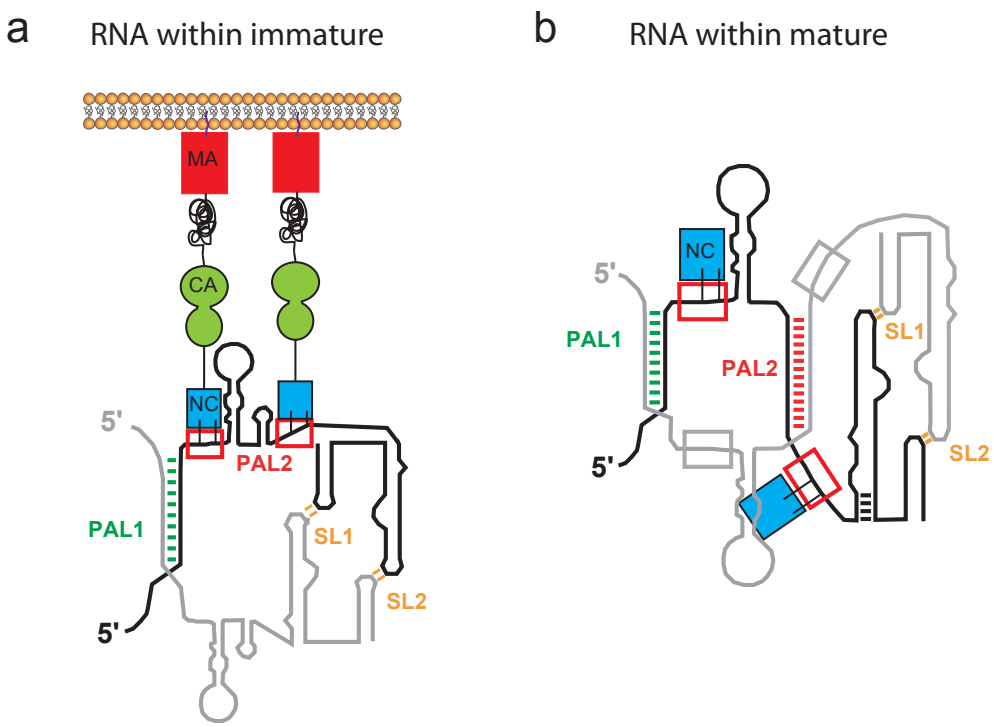


Figure 3.4 Model for MuLV retroviral genome packaging. Secondary structure model for the (a) NC domain of Gag polyprotein bound to the MuLV packaging signal RNA within the immature virion and (b) NC protein bound to the MuLV packaging signal RNA with the mature virion.

3.8.2 US-SHAPE analysis of immature RNA.

MuLV RNA, gently extracted from immature MuLV PR- virions as described previously,[13] was incubated in 1x folding buffer [50 mM HEPES (pH 7.5), 200 mM potassium acetate (pH 7.5), and 5 mM MgCl2] at 37 °C for 20 min prior to modification. The resuspended virus in HNS buffer (300 µL) was pipetted into 33.3 µL 50 mM 1M7 (1-methyl-7-nitroisatoic anhydride) in DMSO (plus reagent) or neat DMSO (minus reagent), mixed by pipetting and incubated at 37 °C for 5 min to modify RNA. Virions were collected by ultracentrifugation as described.[25] Virion resuspensions were preincubated at 37 °C for 5 min and 300 µL of virus solution was pipetted into 33 µL solutions of either 1M7 (50 mM) or DMSO at 37 °C. Solutions were mixed by pipetting and incubated for 15 min at 37 °C. Samples turn from faint yellow to dark orange upon reaction with 1M7. Samples were resuspended in 100 µL of 1x lysis buffer [50 mM Hepes (pH 7.5), 200 mM NaCl, 3 mM $MgCl_2$] and lysed with 1% (wt/vol) SDS and 20 mg/mL proteinase K at 22 °C for 30 min. The lysate was extracted four times with phenol/chloroform/isoamyl alcohol (25:24:1), followed by four extractions with pure chloroform. Samples were then precipitated with ethanol and stored in 70% ethanol at -20 °C until primer extension.

The primer extension protocol was described in detail previously.[5,20] DNA primers incorporated locked nucleic acid nucleotides (LNA, underlined) for use on low-copy RNA to increase binding affinity and stringency.[26-28] MuLV primers (5'-G<u>G</u>U<u>G</u>C <u>A</u>C<u>C</u>AA AGAGU CCAAA AGC-3', 5'-end labeled with 5-FAM or 6-JOE) annealed 3' of the MuLV dimerization domain (nucleotides 422 to 445) Primers (0.5 µL; 2 fmol) were annealed to MuLV RNA (6 µL; 1 fmol in 1/2x TE buffer) by heating at 65 °C and 45 °C for 5 min and then snap-cooled on ice. Reverse transcription buffer [3 µl; 200 mM Tris (pH 8.0), 250 mM KCl, 10 mM $MgCl_2$, 2 mM each dNTP, 20 mM DTT] was added at 0 °C, and primer extension was performed with Superscript III (Invitrogen) reverse transcriptase (0.5 µl; 100 U) at 45 °C for 1 min, followed by incubation at 52 °C and 65 °C, for 10 min each. Reactions were cooled to 4 °C

and quenched by addition of 3 M sodium acetate (pH 5.2). A sequencing marker was generated by adding 0.5 µL dideoxy-GTP (10 mM) to the primer extension reaction mixture in a reverse transcription reaction with unmodified RNA. The 1M7 and DMSO reaction mixes were each combined with equal amounts of dideoxy-GTP-terminated sequencing ladders, precipitated with ethanol, and resuspended in deionized formamide (20 µl).

3.8.3 Capillary electrophoresis.

MuLV cDNA fragments, extended from 50 fmol of input RNA, were resolved by custom high-sensitivity capillary electrophoresis as described.[14] For ultra-sensitive SHAPE, 10 µL samples, in deionized formamide, were electrokinetically injected (10.2 kV for 30 s) onto a 47 cm (50-µm inner diameter) uncoated capillary containing Pop7 (Applied Biosystems) as the internal polymer. The detection window was placed at 36 cm. Voltage was applied using positive polarity (10.2 kV for 60 min) and immersing the capillary ends in a 1x running buffer (Applied Biosystems). The PMT voltages were 0.9 kV and the laser power was 2 mW. Excitation was achieved using an argon laser at 488 and 514 nm wavelengths focused within the capillary at the detection window. 5-FAM was used to detect cDNAs corresponding to SHAPE structure probing and the lower sensitivity dye, 6-JOE, was used for sequencing.

3.8.4 Data processing and structure prediction.

Raw sequence traces were corrected for dye variation and signal decay; peak intensities were integrated using ShapeFinder.[29] Traces were analyzed by combining data from two separate capillary runs. Each capillary electrophoresis separation contained (*i*) a reaction performed in the presence or absence of 1M7 (detected using a FAM-labeled primer) and (*ii*) a sequencing reaction (labeled with JOE) performed using dideoxy GTP. Data from the two runs were aligned using the identical sequencing lanes and then analyzed as described[29]. SHAPE reactivities were normalized by dividing by the average intensity of the 10% most reactive nucleotides, after excluding outliers found by box plot analysis.[30]

SHAPE reactivity information was used to impose a pseudo-free energy change constraint in conjunction with nearest neighbor thermodynamic parameters in the prediction program RNAstructure.[30-32]

3.8.5 S+/L- viral replication assay.

The assay was performed as described.[22] When cultured, sarcoma positive leukemia negative (S+/L-) cells cannot release foci of murine sarcoma virus (MSV) unless they are superinfected with MLV. After superinfection, focus forming units form in direct proportion to the concentration used for superinfection. This assay is thus used as a rapid and sensitive technique for the replicative fitness of viral mutants. S+L- cells are plated in 60 mM plastic petri dishes using 10^5 cells per dish. The next day, the S+L- cells are infected with either native or mutated MuLV. The foci are counted five days after infection using microscopy. The foci appear as lytic areas in the form of grapelike clusters of round cells. Foci were normalized to native viral sequence and reported as a percentage with errors calculated by triplicate measurements.

3.9 REFERENCES

1. Fu, W. & Rein, A. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* **67**, 5443-5449 (1993).

2. Berkowitz, R., Fisher, J. & Goff, S.P. RNA packaging. *Curr Top Microbiol Immunol* **214**, 177-218 (1996).

3. Hibbert, C.S., Mirro, J. & Rein, A. mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J Virol* **78**, 10927-10938 (2004).

4. Yeager, M., Wilson-Kubalek, E.M., Weiner, S.G., Brown, P.O. & Rein, A. Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms. *Proc Natl Acad Sci U S A* **95**, 7299-7304 (1998).

5. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610-1616 (2006).

6. Watts, J.M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716 (2009).

7. Mauger, K.M.W.a.D.M. Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.* **44** (2011).

8. Lucks, J.B. et al. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci U S A* **108**, 11063-11068 (2011).

9. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127**, 4223-4231 (2005).

10. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129**, 4144-4145 (2007).

11. Mortimer, S.A. & Weeks, K.M. Time-resolved RNA SHAPE chemistry. *J Am Chem Soc* **130**, 16178-16180 (2008).

12. Wilkinson, K.A. et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**, e96 (2008).

13. Gherghe, C., Leonard, C.W., Gorelick, R.J. & Weeks, K.M. Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. *J Virol* **84**, 898-906 (2010).

14. Grohman, J.K., Kottegoda, S., Gorelick, R.J., Allbritton, N.L. and Weeks, K.M. Attomole SHAPE reveals regulatory structures in the authentic XMRV RNA genome. *J Am Chem Soc* (2011).

15. Coffin, J.M., Hughes, S.H., & Varmus, H.E. Retroviruses. (Cold Spring Harbor Press, Plainview, NY; 1997).

16. Campbell, S., Oshima, M., Mirro, J., Nagashima, K. & Rein, A. Reversal by dithiothreitol treatment of the block in murine leukemia virus maturation induced by disulfide cross-linking. *J Virol* **76**, 10050-10055 (2002).

17. Evans, M.J., Bacharach, E. & Goff, S.P. RNA sequences in the Moloney murine leukemia virus genome bound by the Gag precursor protein in the yeast three-hybrid system. *J Virol* **78**, 7677-7684 (2004).

18. Badorrek, C.S., Gherghe, C.M. & Weeks, K.M. Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *Proc Natl Acad Sci U S A* **103**, 13640-13645 (2006).

19. Gherghe, C. et al. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci U S A* (2010).

20. Gherghe, C. & Weeks, K.M. The SL1-SL2 (stem-loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer. *J Biol Chem* **281**, 37952-37961 (2006).

21. Grohman, J.K., Gorelick, R.J., Lickwar, C., Lieb, J., and Weeks, K.M. Enhanced Dimerization of a Retroviral Genome via Nucleocapsid-RNA Interaction. (manuscript in preparation).

22. Bassin, R.H., Tuttle, N. & Fischinger, P.J. Rapid cell culture assay technic for murine leukaemia viruses. *Nature* **229**, 564-566 (1971).

23. Paillart, J.C., Marquet, R., Skripkin, E., Ehresmann, C. & Ehresmann, B. Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* **78**, 639-653 (1996).

24. D'Souza, V. & Summers, M.F. How retroviruses select their genomes. *Nat Rev Microbiol* **3**, 643-655 (2005).

25. Ott, D.E. et al. Analysis and localization of cyclophilin A found in the virions of human immunodeficiency virus type 1 MN strain. *AIDS Res Hum Retroviruses* **11**, 1003-1006 (1995).

26. Latorra, D., Arar, K. & Hurley, J.M. Design considerations and effects of LNA in PCR primers. *Mol Cell Probes* **17**, 253-259 (2003).

27. Levin, J.D., Fiala, D., Samala, M.F., Kahn, J.D. & Peterson, R.J. Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers. *Nucleic Acids Res* **34**, e142 (2006).

28. Fratczak, A., Kierzek, R. & Kierzek, E. LNA-modified primers drastically improve hybridization to target RNA and reverse transcription. *Biochemistry* **48**, 514-516 (2009).

29. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. & Giddings, M.C. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* **14**, 1979-1990 (2008).

30. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A* **106**, 97-102 (2009).

31. Low, J.T. & Weeks, K.M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150-158 (2010).

32. Mathews, D.H. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101**, 7287-7292 (2004).

CHAPTER 4

A GUANOSINE-CENTRIC MECHANISM FOR RNA CHAPERONE FUNCTION

4.1 RNA chaperones.

RNA molecules function as the central information conduit in biology.[1,2] This information is encoded both in the primary sequence and in higher order structures involving both base pairing and tertiary interactions. Higher order structure is built up from only four fundamental building blocks, and both biologically functional and non-functional structures are highly stable. On their own, many RNAs fold via apparently complex pathways involving multiple, long-lived intermediates in folding landscapes that are characterized as rugged. Proteins with RNA chaperone activity dramatically simplify these complex pathways by lowering the barriers between RNA states and by facilitating rearrangement of misfolded states. An important class of RNA chaperones is those that bind non-specifically and in an ATP-Independent way to RNA. Non-specific RNA chaperone proteins accelerate strand annealing and exchange reactions, destabilize misfolded states, and facilitate interconversions between RNA states. However, the molecular mechanisms for these reactions are poorly understood. In this work, I uncover a remarkably simple mechanism by which retroviral NC proteins, an important class of RNA chaperone, facilitates an RNA genome conformational isomerization essential for viral replication.

4.2 The role of RNA chaperones in retroviral biology.

As is the case with many RNAs, the RNA genomes of many viruses contain dynamic higher order structures that are essential for the function of these large RNAs.[3] In one

example, retroviruses including HIV-1 and the Moloney murine leukemia virus (MuLV) obligatorily package two RNA genomes in each virus particle. Packaging is governed, in part, by physically linking the two RNA strands via a non-covalent dimer structure involving nucleotides near the 5' end of the genome.[4-6] Like many processes involving RNA conformational changes, dimerization is catalyzed by an RNA chaperone, the nucleocapsid domain of the retroviral Gag protein,[7] which is copackaged with the viral RNA.[8] The isolated nucleocapsid (NC) protein, which functions independently in other viral replication steps, facilitates dimerization in vitro[7] and mutations in NC abrogate viral replication *in vivo*.[9] The RNA chaperone activities of the NC proteins are well-established;[10,11] however, the molecular basis for chaperone function remains unresolved.

Genome dimerization in MuLV is mediated by base pairing and tertiary interactions at a 170 nucleotide RNA domain located in a conserved noncoding region near the 5' end of the genome.[12-14] The simplified transcript of the MuLV dimerization domain used here functionally recapitulates the genomic dimerization element as it dimerizes under physiological-like conditions *in vitro*,[15-17] overlaps with sequences sufficient to mediate packaging *in vivo*,[18] and has a structure similar to that of authentic genomic RNA isolated from virions.[19] Point mutations in this domain eliminate selective packaging.[20]

4.3 Initial characterization of genomic RNA dimerization.

Conformational changes in large RNAs can be characterized in a general way by resolving distinct states by non-denaturing gel electrophoresis (Figure 4.1a-d). In the presence of magnesium ion and upon incubation at 37 °C, the RNA monomer was converted to a dimer at an observed rate (k_{obs}) of 0.15 min^{-1} (Figure 4.1a). An intermediate (I) is also visible (Figure 4.1a, I) In the presence of a saturating concentration of NC, equal to ~1 protein per six RNA nucleotides[21], dimerization is accelerated ten-fold (k_{obs}= 1.6 min^{-1}) and the intermediate is no longer visible (Figure 4.1 e). Thus, as assessed globally, NC

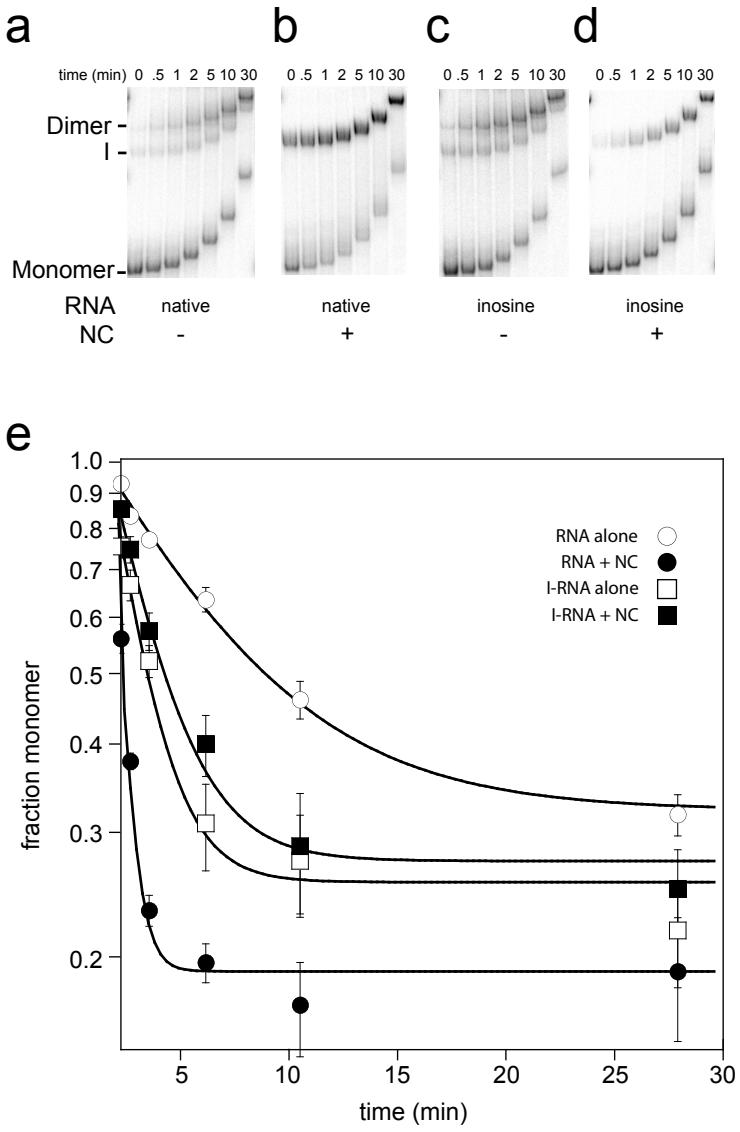


Figure 4.1 Visualization of MuLV dimerization by non-denaturing electrophoresis. (a) Representative gels are shown for wild type (wt), inosine variant (ITP), wt + 8 μM NC, and ITP + 8 μM NC at 600 nM MuLV RNA (b). 1st order exponential curves are shown on a logarithmic axis describing each of the rates of dimerization by gel shift.

accelerates dimerization and reduced the relative population of at least one intermediate state (Figure 4.1a and b).

4.4 Complex, time-dependent changes in the dimerization domain.

I developed a nucleotide-resolution model for assembly of the MuLV dimerization domain using time-resolved SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension).[22] Time-resolved SHAPE employs 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 (0.25s half life).[23] The short in-solution lifetime of the reagent means that time-resolved SHAPE yields one second, single nucleotide resolution snapshots of RNA structure.[24]

I initially established models for the structures[19,25] of the initial monomer state, obtained in the absence of magnesium ion, and the final dimer state reflecting the structure of the RNA after a 30 min incubation (Figure 4.2a). SHAPE reactivities and the secondary structure models of the monomer and dimer correlate well. For example, nucleotides with high SHAPE reactivities are conformationally dynamic,[26] and tend to occur in single-stranded regions of the RNA in both the monomer and dimer states (Figure 4.2a, red nucleotides). Conversely, nucleotides with low SHAPE reactivities occur preferentially in base-paired structures (Figure 4.2a, in black). These models agree well with prior accepted structures for the MuLV dimerization domain,[17] including that obtained using authentic genomic RNA.[19]

Dimerization was initiated by addition of magnesium ion. One second snapshots over the course of dimerization were obtained by mixing a portion of the evolving reaction with an aliquot of the reagent, BzCN, resulting in a reactivity profile spanning the entire RNA at each time point (Figure 4.2b). Five sets of key interactions undergo large-scale structural changes during dimerization (Figure 4.2b, dashed blue regions). The two stem-loops, SL1 and SL2 (positions 329-332 and 363-366), are unstructured in the monomer and form an

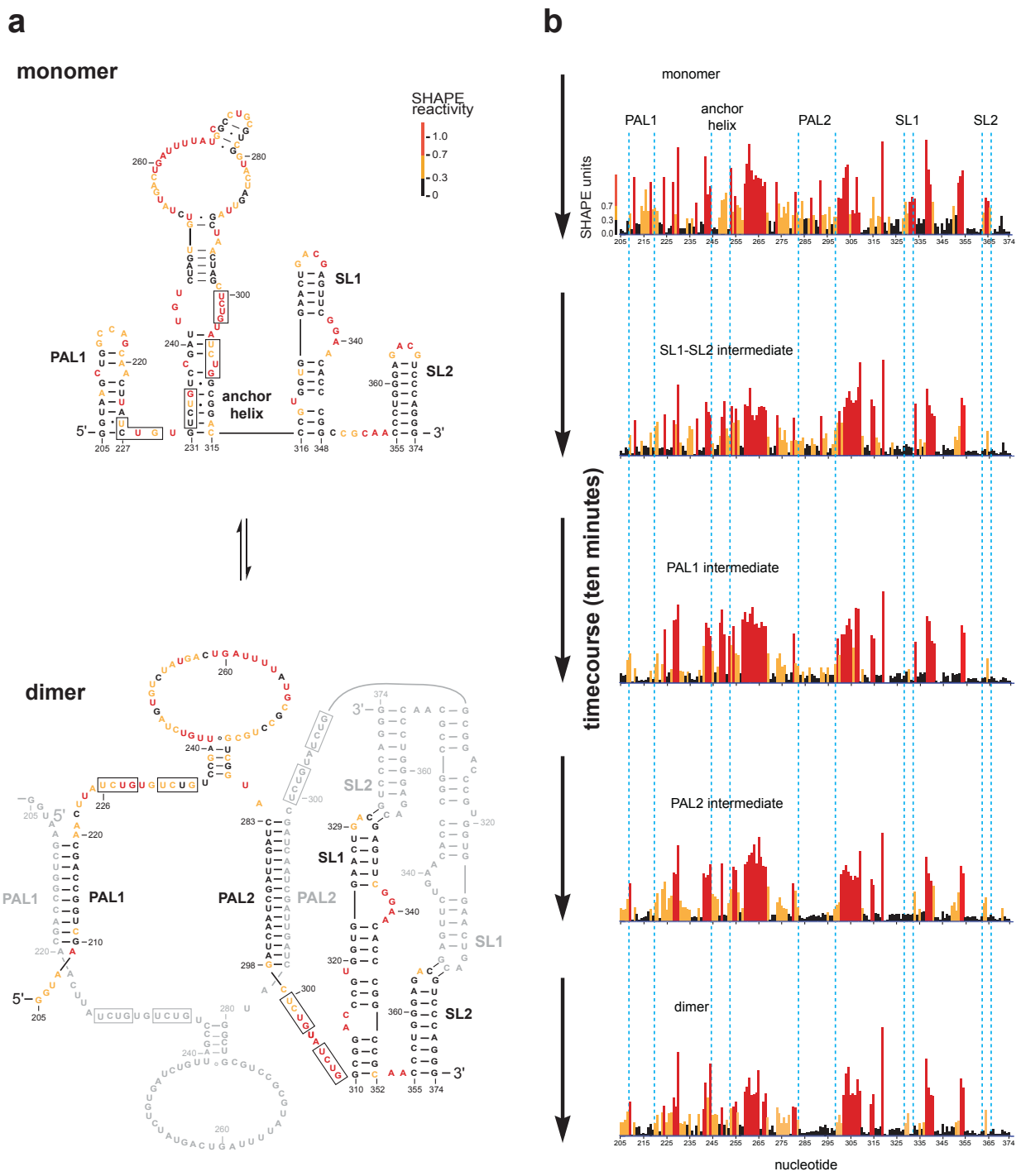


Figure 4.2 Time-resolved SHAPE analysis of MuLV dimerization. (a) Secondary structures of starting monomer and final dimer states. Nucleotides are colored by SHAPE reactivity. Key structural features are labeled, and boxes highlight known NC binding motifs. (b) Time course of MuLV dimerization. SHAPE reactivities are shown for monomer, dimer and representative intermediate time points. Key structural interactions are denoted with dashed blue lines.

intermolecular loop-loop kissing interaction in the dimer (Figure 4.2a, SL1 and SL2). A palindromic sequence, PAL1 (positions 210-219), is initially located within a reactive stem-loop in the monomer, but forms a stable, unreactive, intermolecular duplex in the dimer (Figure 4.2a, PAL1). A second palindromic sequence, PAL2 (positions 283-298), initially located in a large flexible region in the monomer also forms a stable intermolecular duplex in the dimer (Figure 4.2a, PAL2). Conversely a stable stem, termed the anchoring helix (positions 231-251 and 290-315), transitions from a constrained to a more flexible state (Figure 4.2a, intermediates).

I obtained reactivity data for every nucleotide within the 170 nt domain over sixteen one-second dimerization snapshots to yield over 2,700 structural data points. I grouped nucleotides exhibiting similar kinetic behaviors by k-means clustering,[27] illustrated using the same coloring scheme used to indicate SHAPE reactivities on the secondary structures of the monomer and dimer states (Figure 4.2a). The dimerization profile is complex and involves seven kinetic behaviors, or clusters, involving four rates (Figure 4.3a and Table 4.1). Although formally a bimolecular, or second-order process, all measureable rates are identical, within error, over a three-fold change in RNA concentration (Figure 4.4). The fastest measured process involved protection of nucleotides at the apex of SL1 and SL2 and was complete by the first time point and therefore occurs faster than ~5 min^{-1} (Figure 4.3a, cluster 1a). These observations are consistent with a model in which the first dimerization step involves rapid, nearly diffusion controlled, intramolecular interactions involving the SL1-SL2 stem loops followed by a pseudo-unimolecular structural changes to form the mature dimer.

Following the initial, rapid reduction in SHAPE reactivities involving the SL1-SL2 step loops, nucleotides in PAL1 also undergo rapid protection at an observed average rate of 1.6 ± 0.4 min^{-1} (Figure 4.3a, cluster 1b). Nucleotides in clusters 2 and 3 undergo a conformational change at the rate of 0.30 ± 0.03 min^{-1}, but with two opposing behaviors, one

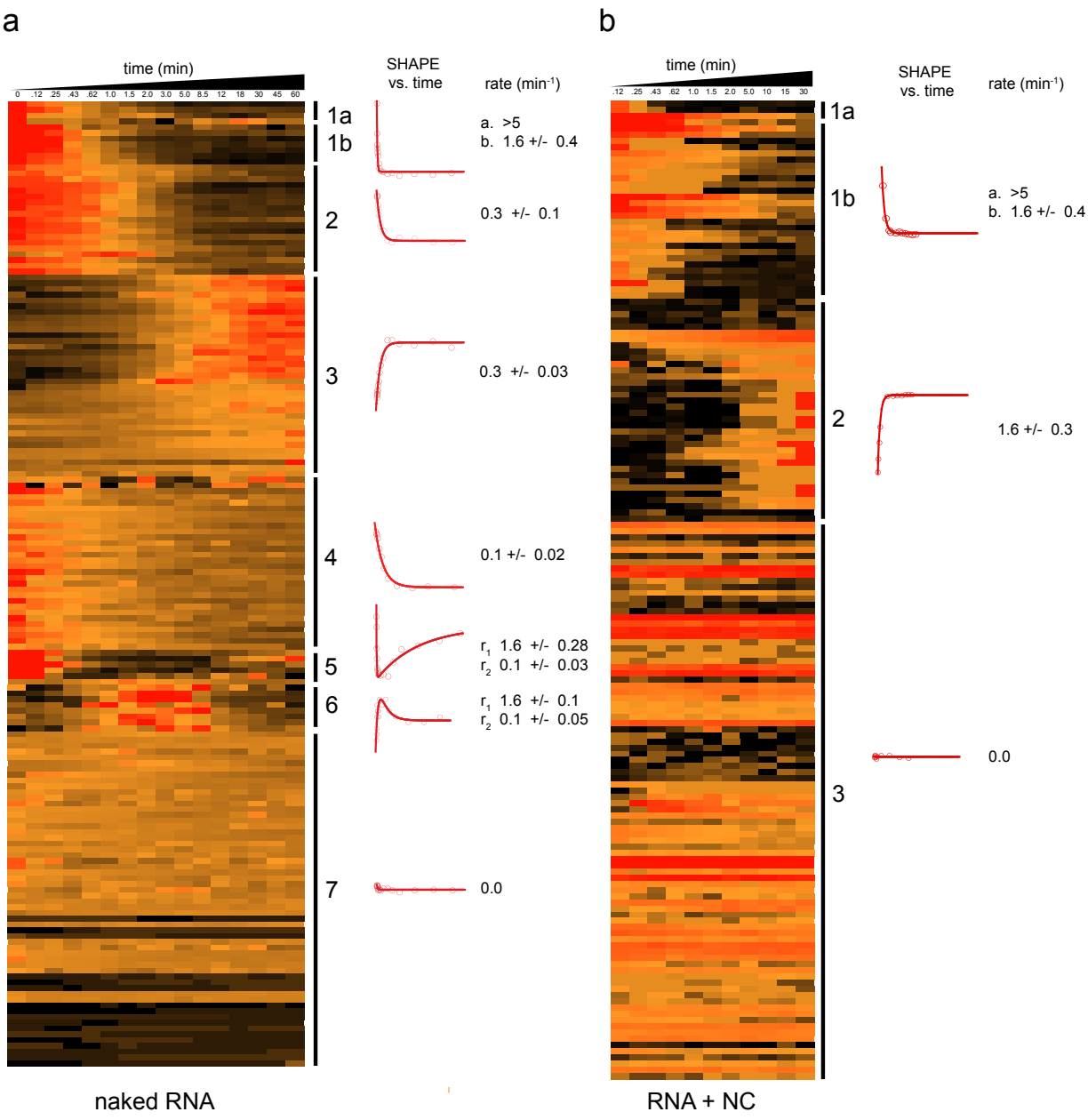


Figure 4.3 Model-free clustering of nucleotide-resolution profiles for dimerization. (a) MuLV RNA (600 nM) alone and (b) as facilitated by the retroviral NC protein (8 μM). Each data point is shown on a scale (black to red) corresponding closely to SHAPE reactivities. A representative kinetic curve and rate is shown for each cluster.

decreasing and the other increasing in net local nucleotide flexibility (Figure 4.3a). These clusters include nucleotides in the anchoring helix and in PAL2, suggestive of a single process including both structures. Cluster 4, including nucleotides in the large flexible domain (nts 251 - 282), shows a statistically significant slower kinetic behavior with a rate constant of 0.11 ± 0.02 min^{-1}. Finally, two clusters, 5 and 6, showed complex kinetic behavior in which the SHAPE reactivity either first increased and then decreased over time, or vice versa. These clusters reflect conformational changes occurring at the 1.6 and 0.1 min^{-1} rates, consistent with nucleotides whose structures are influenced by a combination of the processes characterized by single rate constants (Figure 4.3a).

In sum, time-resolved nucleotide-resolution analysis of MuLV dimerization emphasizes that this process is complex, slow, and characterized by multiple kinetically distinct transitions and intermediates.

4.5 Nucleocapsid-mediated MuLV dimerization.

We next performed an analogous set of experiments except that dimerization was initiated by addition of magnesium ion and the MuLV NC protein, a well-established and efficient RNA chaperone.[4] The resulting SHAPE data were subjected to clustering and, strikingly, revealed that the NC protein collapsed dimerization into a single kinetic process occurring at 1.6 ± 0.4 min^{-1}. Nucleotides involved in this process both decrease (cluster 1) and increase (cluster 2) in SHAPE reactivity; roughly half of the nucleotides show no change in shape activity during dimerization (Figure 4.3b). Dimerization thus occurs in a single, fast transition in the presence of NC, with a complete apparent absence of the slow and multi-rate processes that characterized the RNA-only reaction.

To understand the profound contribution that NC makes to streamlining the dimerization reaction, we monitored NC binding immediately after combining the protein and RNA, corresponding to a ~2 sec time point. These experiments were performed in the absence of magnesium ion to reduce the rate of dimerization. The nucleotide-resolution

Table 4.1: Nucleotides ordered by model free clustering

Cluster #	600 nM RNA nucleotides	600 nM RNA + 8 μM NC nucleotides
1a	331, 332, 365, 366	332, 365, 366
1b	209-211, 214-216, 229, 231, 232, 237	209-216, 238, 248-251, 265, 275, 283, 284, 285, 287-289, 292, 294, 295, 297, 328, 349, 350-352
2	238, 240, 272, 284-290, 292, 296, 298, 312, 329, 349, 368	205-208, 217-219, 222-225, 228, 229, 232, 233, 234, 239, 240, 241, 243, 244, 245, 268, 270, 272, 302, 305, 314, 315
3	205-208, 223, 225, 226, 227, 233, 235, 239, 241, 244-247, 295, 305-309	The rest
4	217-220, 222, 224, 234, 243, 254, 257, 265, 268, 270, 280, 283, 291, 293, 300, 314, 315, 317, 318, 322, 325-326, 328, 330, 335, 341, 323, 348-352, 369	
5	251, 271, 313, 320-321	
6	237, 252, 258, 279, 299, 301, 311, 344	
7	The rest	

effects of NC binding are readily detected in a difference analysis in which the SHAPE profile after NC binding is subtracted from that of the free RNA. Strikingly, of the 29 nucleotides showing the largest changes in SHAPE reactivity, 19 (or 66%) correspond to guanosine residues (G labels, Figure 4.5a). Preferential binding to guanosine is consistent with studies showing that the MuLV NC protein contains a cleft that binds selectively to this nucleotide.[28] Sites of protection (positive peaks, Figure 4.5a) likely correspond to stable binding during the 1 sec window of time-resolved SHAPE experiment; whereas, the smaller number of sites showing increased flexibility at guanosine residues in the presence of NC (negative peaks) likely reflect a rapid binding and release process.[29]

4.6 A guanosine-to-inosine RNA variant accelerates native-like dimerization.

One of the most physically profound features of nucleic acid structure are the special characteristics of base pairs involving guanosine. First, guanosine-cytosine base pairs are, on average, ~ 1 kcal/mol more stable than A-U pairs which corresponds to a factor of greater than 10 in equilibrium constant. Second, guanosine pairing is promiscuous and forms stable pairs with both cytosine and uridine.[30,31] Both of these factors has the potential to increase the complexity of an RNA folding pathways by stabilizing non-native structures and by increasing the number of alternative base paired states, respectively.

The strong preference of NC to interact at guanosine residues (Figure 4.5) coupled with these distinctive features of guanosine pairing prompted us to consider whether NC exerts its RNA chaperone activity by stabilizing interactions between guanosine and the other nucleotides. We therefore explored the dimerization reaction using an RNA in which all guanosine residues were replaced by inosine. Inosine-cytosine pairs are iso-structural, but roughly 1 kcal/mol weaker than guanosine-cytosine pairs and inosine does not pair stably with uridine (Figure 4.6a).

The rate of dimerization of the inosine RNA was accelerated by 6-fold relative to that of the native RNA (Figure 4.1a,c). Addition of NC protein had no effect on the rate of

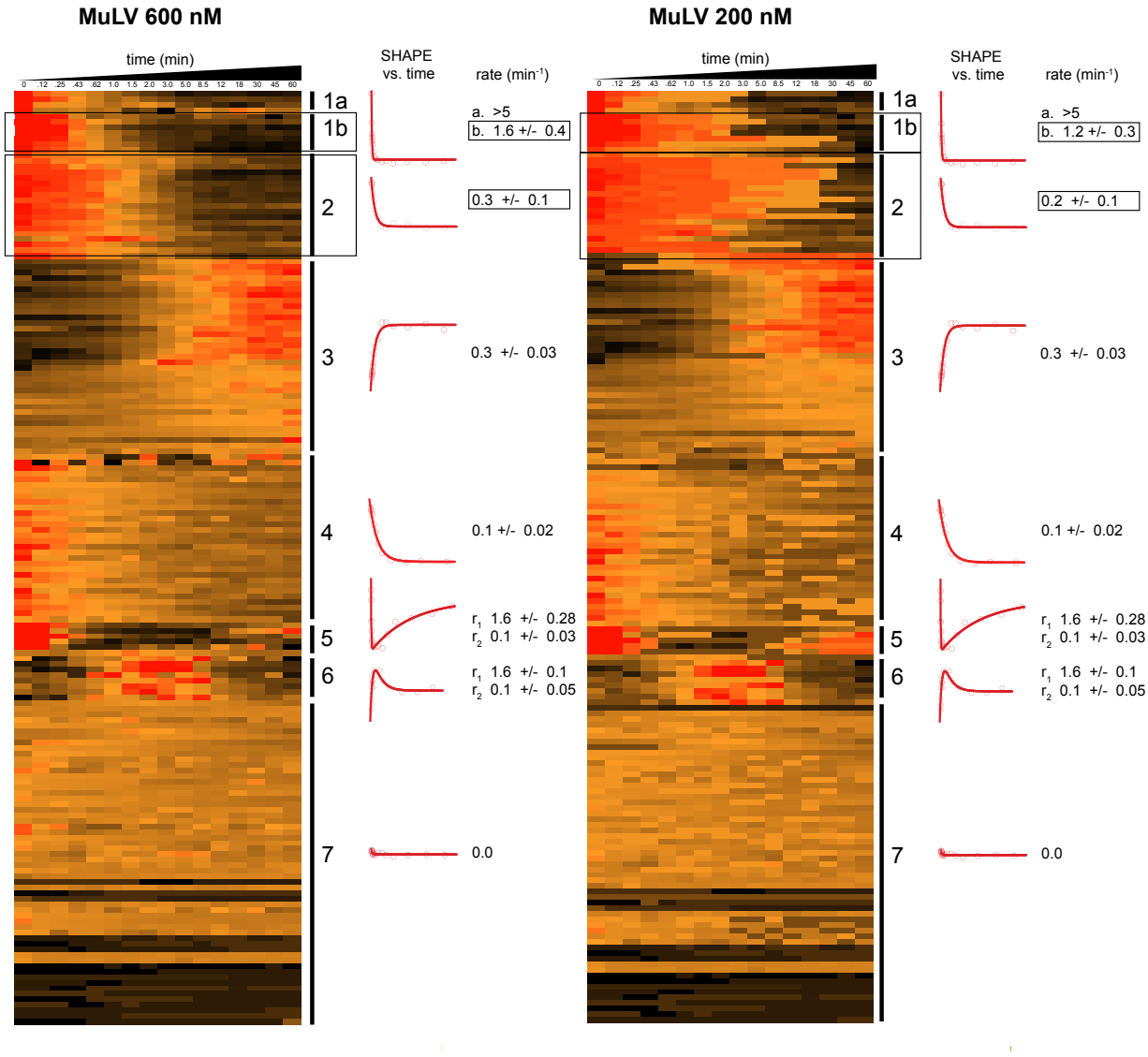


Figure 4.4 Time-resolved SHAPE of MuLV RNA dimerization at 600 and 200 nM (labeled as in Figure 4.3). Model-free clustering analysis of both 600 and 200 nM MuLV RNA were compared with little to no concentration dependence. Clusters displaying slight rate differences (1b and 2) are boxed.

formation of the final state, although NC did resolve the apparent build-up of the PAL1 intermediate (Figure 4.1d).

We confirmed that the inosine RNA forms the same final dimer as the native RNA by probing the structure of the final inosine dimer by SHAPE. The inosine RNA forms essentially the same dimer structure and, moreover, the absolute nucleotide-resolution SHAPE measurements for the inosine and native dimers correlate strongly (R^2 = 0.88) (Figure 4.6b). In strong contrast, the SHAPE reactivities of the inosine and native RNAs in the monomer state correlate poorly (R^2 = 0.26) (Figure 4.6c), even though the two monomers fold into highly similar secondary structures. The inosine RNA monomer is destabilized by ~20 kcal/mol relative to the native sequence dimer. Strikingly however, addition of NC to the native sequence RNA yields a SHAPE profile that is highly similar to that of the inosine RNA (R^2 = 0.87; Figure 4.6d).

These data support the remarkable conclusion that a significant component of the RNA chaperone function of the MuLV NC protein can be explained by interactions that destabilize interactions between guanosine nucleotides and other positions in the RNA, that this function can be mimicked by simply substituting guanosine for the more weakly interacting inosine analogue, and that replacement of a single functional group in RNA (Figure 4.5a) can abrogate most of the RNA chaperone active of the NC protein.

4.7 Discussion.

4.7.1 The behavior of RNA in biology.

Two fundamental features dominate the behavior of RNA in biology. First, full function requires that RNAs fold back on themselves to form information-rich structures that mediate biological interactions and carry out essential catalytic and recognition functions. Second, in perhaps the majority of cases, RNA molecules on their own fold incompletely, slowly, and by populating non-native intermediate states. These latter features are remedied by essential interactions with RNA chaperones. Substantial progress has been made in

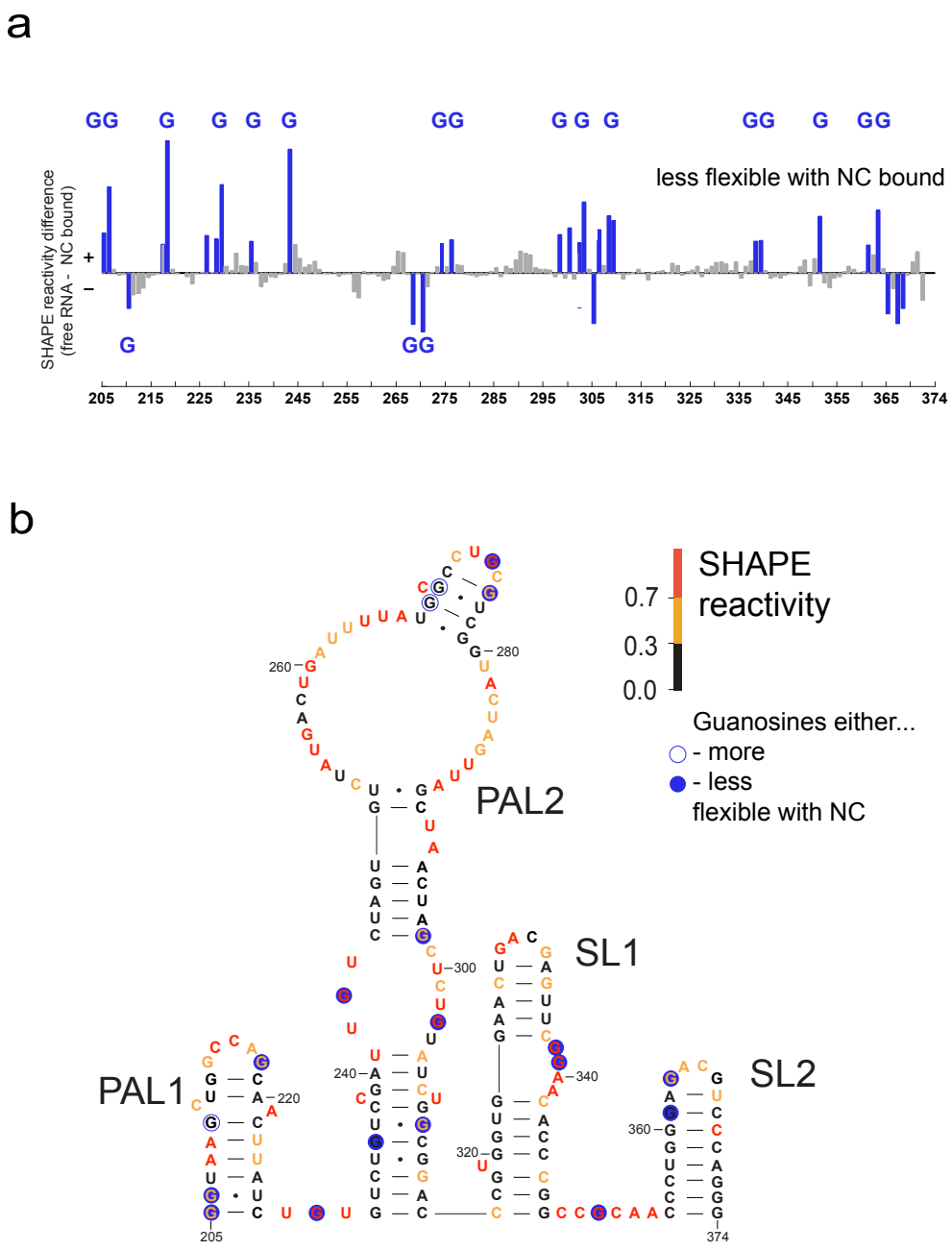


Figure 4.5 Initial interactions between NC and the MuLV monomer. (a) A SHAPE difference plot illustrating the effect of NC binding after protein addition (0 ºC). Positive differences reflect nucleotides less flexible in the presence of NC. Differences greater than 20% are highlighted blue; of these, guanosine residues are labeled with a G.

understanding the mechanisms of ATP-dependent chaperone proteins[32,33] and significant work has also defined the essential activities important for the function of the non-specific ATP-independent class of RNA chaperones.[34,35] However, the molecular mechanisms by which generalized RNA chaperones facilitate RNA folding is not well understood. All RNA chaperone proteins ultimately function by facilitating structural re-equilibration reactions that allow an RNA to escape from stable, kinetically trapped states.

4.7.2 Time-resolved SHAPE model of MuLV genomic dimerization.

Only recently have RNA structural technologies made it possible to visualize how RNAs fold at nucleotide resolution in real time.[36-38] Most RNA structural studies to date therefore have emphasized analyses of the initial and final RNA states and the relative populations of these states. Here, we have employed time-resolved SHAPE technology, coupled with model-free clustering, to create a full kinetic analysis of dimerization of a retroviral genome at nucleotide resolution in single-second snapshots. We use this data to develop a complete assembly model for retroviral dimerization and to decipher a mechanism for the interaction of retroviral genomic RNA with its nucleocapsid chaperone.

Time-resolved SHAPE supports a model of MuLV genomic dimerization in which two MuLV monomers are rapidly brought together via loop-loop interactions at the 3' end of the dimerization motif and via a second loop-loop interaction at their 5' ends. The RNA-only model is complex: Seven distinct kinetic behaviors are observed and can be incorporated into a single unified model of MuLV RNA dimerization (Figure 4.7a). Upon magnesium addition, the SL1-SL2 stem loop-loop kissing interaction forms at the near-diffusion controlled limit (10^9 $M^{-1}min^{-1}$) and a second fast loop-loop interaction occurs that yields the PAL1 intermolecular duplex (1.6 min^{-1}). A large-scale structural rearrangement, consistent with the observed mobility shift resolved by non-denaturing electrophoresis, creates the PAL2 duplex accompanied by melting of the anchoring helix (0.3 min^{-1}). These fast events

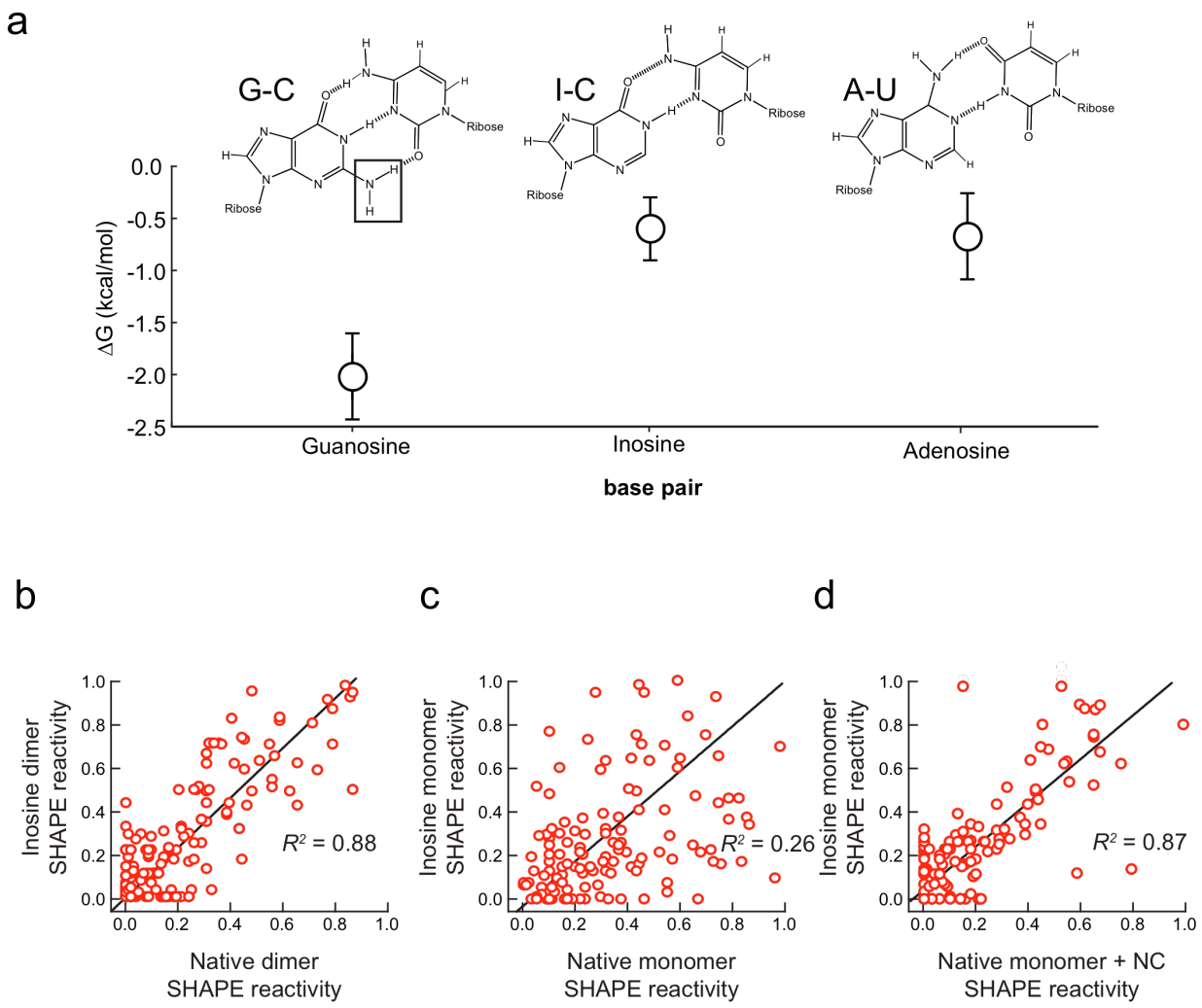


Figure 4.6 Effect of inosine on RNA structure and on global conformation of the MuLV dimerization domain. (a) Watson Crick structure and free energy for G-C, I-C, and A-U base pairs. Correlations between (b) the inosine and native dimer, (c) the inosine and native monomer, and (d) the inosine monomer and native monomer 2s after NC addition.

then promote slower (0.1 min^{-1}) small-scale interactions that create the final dimer. This complex mechanism is transformed -- and greatly simplified -- in the presence of its biological chaperone for this process, the retroviral NC protein. RNA dimerization is accelerated by greater than ten-fold: The multistep dimerization process now occurs in a single kinetic step.

4.7.3 A Guanosine-Centric Mechanism for RNA Chaperone Function.

Previous work on MuLV dimerization has suggested that NC binds the retroviral RNA at guanosine residues.[19,20] The critical clue in understanding the molecular mechanism of the chaperone function of NC was the observation that the initial interaction between NC and the MuLV RNA occurred predominately at guanosine residues, consistent with work showing NC contains a specific binding pocket for this nucleotide[28] and substitution of guanosine residues within the NC-binding motif abrogates binding and packaging of this RNA in vivo.[20] Single second snapshots taken after addition of NC to the MuLV monomer, show distinct changes in the reactivity profile concomitant with NC binding at flexible guanosine residues both within and outside of known NC binding motifs (Figure 4.5). An RNA variant, in which guanosine residues are substitute by inosine, dimerizes more efficiently than wt and is no longer affected by NC. From this, we infer a molecular mechanism for nucleocapsid-mediated dimerization in which NC binds to guanosine residues throughout the retroviral RNA dimerization domain and destabilizes interactions between this nucleotide and the rest of the RNA. NC binding both destabilizes slow-resolving structures rich in guanosine and allows a great number of structures to form, and thereby facilitate more efficient reentry into the RNA folding pathway towards the final global free energy minima (Figure 4.7b). These results illustrate structural and mechanistic features likely to be common to many RNA chaperone-catalyzed processes.

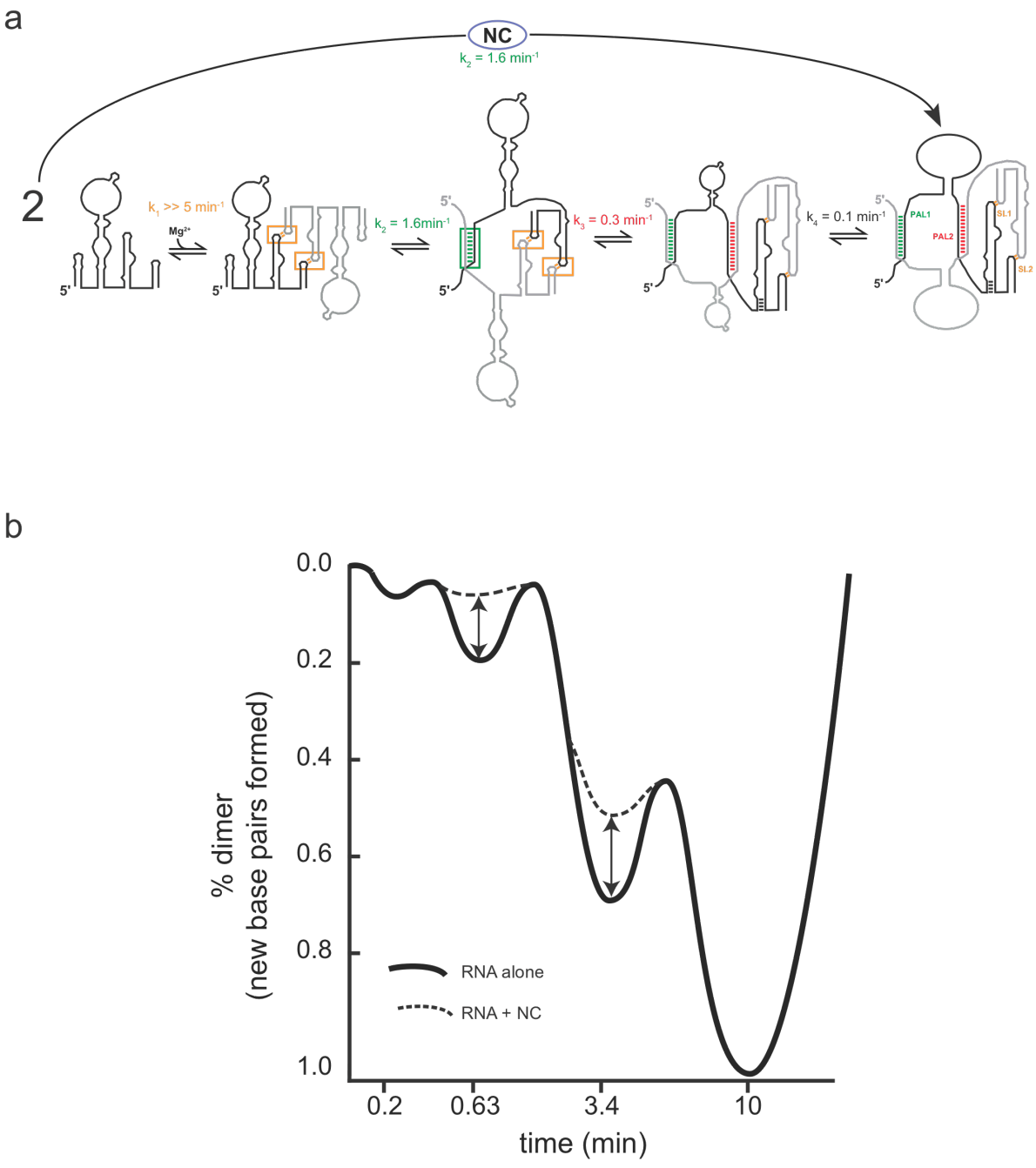


Figure 4.7 Complete assembly model of MuLV retroviral dimerization. (a) The RNA-only model is complex. In the presence of NC, a copackaged RNA chaperone protein, the multistep dimerization process now occurs in a single fast kinetic step. (b) model for the folding free energy landscape of MuLV dimerization.

4.8 Experimental Procedures.

4.8.1 Retroviral RNA transcripts.

The MuLV RNA construct is 331 nucleotides and spans the dimerization and packaging regulatory domain (~170 nts) including 5' and 3' flanking sequences of 46 and 115 nucleotides, respectively.[15,17] The transcript was synthesized and purified as described.[20]

4.8.2 Time-resolved SHAPE.

In vitro generated MuLV RNA monomer transcripts (60 or 30 pmol, for the 600 nM and 300 nM reactions respectively) were renatured by heating at 95 °C for 3 min in 80 µL water and cooled on ice for 3 min. Monomers were then equilibrated at 37 °C for 3 min in a folding buffer without magnesium [50 mM HEPES (pH 7.5), 200 mM potassium acetate (pH 7.5)]. The total reaction volume was 100 µL, sufficient for a 15-aliquot time course. A pre-reaction (0 sec) time point was taken before magnesium addition to capture the initial monomer state. Time-resolved SHAPE experiments were then initiated by the addition of $MgCl_2$ to 5 mM (1 µL, 50 mM) at 37 °C and the RNA structure was interrogated over a time course spanning 7s to 1h. For each time point, SHAPE modification was performed by adding 9 µL (5.4 pmol) of the evolving MuLV RNA reaction to 1 µL BzCN (200 mM in DMSO), followed by vigorous pipetting to ensure mixing. Reaction with BzCN is complete within ~1s.[22] A no-reaction control containing neat DMSO (1 µL) instead of BzCN was also performed. RNA was recovered using 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 were performed identically in the presence of a saturating amount of retroviral NCP10 protein (8 µM in 1 µL) except NC was added as a mixture with magnesium (2 µL total in mixture).

4.8.3 Primer extension.

The procedure has been described in detail previously[17,39]. MuLV primers [5'-GGUGC ACCAA AGAGU CCAAA AGC-3' (LNA nucleotides are underlined), 5'-end labeled with 5-FAM or 6-JOE] bind 3' of the the MuLV dimerization domain (nucleotides 422 to 445). Primers (1 µL; 10 pmol) were annealed to the MuLV RNA (6 µL; 5.4 pmol in 1/2× TE buffer) by heating at 65 °C and 45 °C for 5 min each and then snap-cooled on ice. Reverse transcription buffer [3 µl; 200 mM Tris (pH 8.0), 250 mM KCl, 10 mM $MgCl_2$, 2 mM each dNTP, 20 mM DTT] was added at 0 °C, and primer extension was performed with Superscript III (Invitrogen) reverse transcriptase (0.5 µl; 100 U) at 45 °C for 1 min, followed by incubation at 52 °C and 65 °C, for 10 min each. Reactions were quenched by cooling to 4 °C and addition of 3 M sodium acetate (pH 5.2). A sequencing marker was generated by adding 0.5 µL ddGTP (10 mM) to the primer extension reaction mixture using unmodified RNA. The 1M7 and DMSO control reaction mixes were each combined with equal amounts of ddGTP-terminated sequencing ladders, precipitated with ethanol, and resuspended in deionized formamide (10 µl). cDNA fragments were resolved on an Applied Biosystems 3130 capillary electrophoresis instrument.

4.8.4 Data processing and structure prediction.

Time-resolved SHAPE data was processed using fully automated software for the quantification of SHAPE data (QuShape). The initial raw sequence trace was corrected for dye variation and signal decay; and peak intensities were scaled, aligned, integrated, and normalized using a suite of optimized statistical algorithms. The net result of this software-facilitated analysis is an accurate and efficient quantification of SHAPE data (total analysis time for a single SHAPE trace is approximately 3 minutes). After successful analysis of this first (reference) trace, subsequent time points were analyzed automatically by alignment to the saved project file as a reference trace. Each CE separation contained a reaction performed in the presence or absence of 1M7 (labeled with FAM) and a sequencing reaction

(labeled with JOE), performed using ddGTP. 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.[40]

4.8.5 K-means clustering.

Analyzed SHAPE data were sorted by individual nucleotide kinetic behaviors in an automated way using GeneCluster software.[27] Files are uploaded as tab delimited text files using specific headings as described in the manual.[41] I have included a few nuances I used to commandeer the software for kinetic SHAPE clustering here. One, under the k-means tab, I organized the data by genes (or rows) and systematically cycle through the number of clusters (starting at 10 and reducing the cluster number by one each cycle) until a minimal set of clusters are created that define the data (i.e. you are splitting the noise). Two, under the adjust data tab, I used the normalize genes option by clicking the appropriate box and clicking apply. Three, I used the absolute correlation (uncentered) similarity metric to organize the genes. Four, click execute. The clustering analysis is performed and a file is loaded in the same folder that the data is opened from. Clusters are visualized using TreeView software.[27]

4.8.6 Non-denaturing electrophoretic mobility shift assays of MuLV dimerization.

MuLV wild type (wt) dimerization domain RNA internally labeled with [^{32}P] (~0.5 nM) was mixed with unlabeled RNA (200 or 600 nM in 20 µL). For the inosine variant RNA, synthesis by transcription was performed in the presence of inosine triphosphate instead of guanosine triphosphate. Reactions were treated exactly as per the time-resolved SHAPE experiments, outlined above. Briefly, samples were heated to 95 °C, rapidly cooled on ice, and equilibrated with a folding buffer without magnesium [50 mM HEPES (pH 7.5), 200 mM potassium acetate (pH 7.5), at 37 °C]. The reaction was initiated by addition of 1 µL of 5 mM $MgCl_2$ and placement at 37 °C. Time points (3 µL) were mixed with 3 µL of 30% (v/v)

glycerol (containing marker dyes), loaded directly onto the non-denaturing gel (5% polyacrylamide in TBE), and resolved by electrophoresis at 4 °C for 2 h at 20 W.

4.9 Acknowledgements.

This work was supported by the US National Institutes of Health (GM064803 to K.M.W.) and with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E with SAIC-Frederick, Inc.

4.10 REFERENCES.

1. Nilsen, T.W. RNA 1997-2007: a remarkable decade of discovery. *Mol Cell* **28**, 715-720 (2007).

2. Sharp, P.A. The centrality of RNA. *Cell* **136**, 577-580 (2009).

3. Coffin, J.M., Hughes, S.H., & Varmus, H.E. Retroviruses. (Cold Spring Harbor Press, Plainview, NY; 1997).

4. Darlix, J.L., Lapadat-Tapolsky, M., de Rocquigny, H. & Roques, B.P. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol* **254**, 523-537 (1995).

5. Paillart, J.C. et al. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol* **70**, 8348-8354 (1996).

6. Hibbert, C.S., Mirro, J. & Rein, A. mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J Virol* **78**, 10927-10938 (2004).

7. Bonnet-Mathoniere, B., Girard, P.M., Muriaux, D. & Paoletti, J. Nucleocapsid protein 10 activates dimerization of the RNA of Moloney murine leukaemia virus in vitro. *Eur J Biochem* **238**, 129-135 (1996).

8. Fu, W. & Rein, A. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* **67**, 5443-5449 (1993).

9. Rein, A., Henderson, L.E. & Levin, J.G. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci* **23**, 297-301 (1998).

10. Stewart-Maynard, K.M. et al. Retroviral nucleocapsid proteins display nonequivalent Levels of nucleic acid chaperone activity. *J Virol* **82**, 10129-10142 (2008).

11. Rein, A. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol* **7**, 700-705 (2010).

12. Varmus, H.E. Form and function of retroviral proviruses. *Science* **216**, 812-820 (1982).

13. Adam, M.A. & Miller, A.D. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J Virol* **62**, 3802-3806 (1988).

14. Badorrek, C.S. & Weeks, K.M. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat Chem Biol* **1**, 104-111 (2005).

15. Badorrek, C.S., Gherghe, C.M. & Weeks, K.M. Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *Proc Natl Acad Sci U S A* **103**, 13640-13645 (2006).

16. Badorrek, C.S. & Weeks, K.M. Architecture of a gamma retroviral genomic RNA dimer. *Biochemistry* **45**, 12664-12672 (2006).

17. Gherghe, C. & Weeks, K.M. The SL1-SL2 (stem-loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer. *J Biol Chem* **281**, 37952-37961 (2006).

18. D'Souza, V. & Summers, M.F. How retroviruses select their genomes. *Nat Rev Microbiol* **3**, 643-655 (2005).

19. Gherghe, C., Leonard, C.W., Gorelick, R.J. & Weeks, K.M. Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. *J Virol* **84**, 898-906 (2010).

20. Gherghe, C. et al. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci U S A* (2010).

21. Vo, M.N., Barany, G., Rouzina, I. & Musier-Forsyth, K. Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J Mol Biol* **363**, 244-261 (2006).

22. Mortimer, S.A. & Weeks, K.M. Time-resolved RNA SHAPE chemistry. *J Am Chem Soc* **130**, 16178-16180 (2008).

23. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129**, 4144-4145 (2007).

24. Mortimer, S.A. & Weeks, K.M. Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc* **4**, 1413-1421 (2009).

25. Grohman, J.K., Kottegoda, S., Gorelick, R.J., Allbritton, N.L. and Weeks, K.M. Attomole SHAPE reveals regulatory structures in the authentic XMRV RNA genome. *J Am Chem Soc* (2011).

26. McGinnis, J.L., Duncan, C.D. & Weeks, K.M. High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol* **468**, 67-89 (2009).

27. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-14868 (1998).

28. Schuler, W., Dong, C., Wecker, K. & Roques, B.P. NMR structure of the complex between the zinc finger protein NCp10 of Moloney murine leukemia virus and the single-stranded pentanucleotide d(ACGCC): comparison with HIV-NCp7 complexes. *Biochemistry* **38**, 12984-12994 (1999).

29. Cruceanu, M., Gorelick, R.J., Musier-Forsyth, K., Rouzina, I. & Williams, M.C. Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J Mol Biol* **363**, 867-877 (2006).

30. Richardson, N.A., Wesolowski, S.S. & Schaefer, H.F., 3rd Electron affinity of the guanine-cytosine base pair and structural perturbations upon anion formation. *J Am Chem Soc* **124**, 10163-10170 (2002).

31. Vendeix, F.A., Munoz, A.M. & Agris, P.F. Free energy calculation of modified base-pair formation in explicit solvent: A predictive model. *RNA* **15**, 2278-2287 (2009).

32. Fairman, M.E. et al. Protein displacement by DExH/D "RNA helicases" without duplex unwinding. *Science* **304**, 730-734 (2004).

33. Russell, R. RNA misfolding and the action of chaperones. *Front Biosci* **13**, 1-20 (2008).

34. Fedorova, O., Solem, A. & Pyle, A.M. Protein-facilitated folding of group II intron ribozymes. *J Mol Biol* **397**, 799-813 (2010).

35. Zhao, X. & Jain, C. DEAD-box proteins from Escherichia coli exhibit multiple ATP-independent activities. *J Bacteriol* **193**, 2236-2241 (2011).

36. Laederach, A., Shcherbakova, I., Jonikas, M.A., Altman, R.B. & Brenowitz, M. Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc Natl Acad Sci U S A* **104**, 7045-7050 (2007).

37. Lee, M.K., Gal, M., Frydman, L. & Varani, G. Real-time multidimensional NMR follows RNA folding with second resolution. *Proc Natl Acad Sci U S A* **107**, 9192-9197 (2010).

38. Woodson, S.A. RNA Folding Pathways and the Self-Assembly of Ribosomes. *Acc Chem Res* (2011).

39. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610-1616 (2006).

40. Mathews, D.H. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101**, 7287-7292 (2004).

41. Eisen, M. Cluster and TreeView manual. *available online* (1998).