Biophysical Analysis of the Primary Nucleic Acid Binding Domains of HsLINE-1 ORF1p

by
Raphael Isaiah Britt
B.S., University of Connecticut, 2019

Submitted as partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Molecular Biology, Cell Biology, and Biochemistry Graduate Program in the Division of Biology and Medicine at brown University

May 2026

This dissertation by Raphael Isaiah Britt is accepted in its present form by the Department of Molecular Biology, Cell Biology, and Biochemistry as satisfying the dissertation requirement for the degree of Doctor of Philosophy

Date _____ _____
John M. Sedivy, Ph.D., Advisor

Date _____ _____
Gerwald Jogl, Ph.D., Co-Advisor

Recommended to the Graduate Council

Date _____ _____
George Lisi, Ph.D., Committee Chair

Date _____ _____
Walter Atwood, Ph.D., Reader

Date _____ _____
Erica Larschan, Ph.D., Reader

Date _____ _____
John LaCava, Ph.D., Outside Reader

Approved by the Graduate Council

Date _____ _____
David P. Lindstrom, Dean of the Graduate School

Curriculum Vitae

RAPHAEL BRITT

I am a PhD candidate in the department of molecular biology, cell biology, and biochemistry at Brown University. I am investigating the biophysical basis of nucleic acid recognition by human LINE1 ORF1p, using protein biochemistry and structural biology techniques to characterize how individual HsORF1p trimers interact with short nucleic acids. I am broadly interested in applying biophysical and biochemical approaches to the study of protein function and engineering.

EDUCATION

Ph.D., Molecular Biology, Cell Biology, and Biochemistry
Brown University, Providence, RI
Expected Spring 2026

B.S., Molecular and Cell Biology
University of Connecticut, Storrs, CT
December 2019

POSITIONS AND EMPLOYMENT

Graduate Researcher, with John Sedivy and Gerwald Jogl
Brown University, Providence, RI
2020-present

- Led biochemical characterization of human LINE1 ORF1p interactions with short nucleic acids, demonstrating new nucleic acid binding specificity and linked binding domain alanine mutations to *in vivo* retrotransposition phenotypes.
- Performed protein expression and purification in bacterial, insect cell, and mammalian systems.
- Carried out structural and biochemical characterization including X-ray crystallography, cryo-EM, MST, DSF, FP, EMSA, and analytical SEC.
- Trained and mentored junior researchers in molecular cloning, protein purification, assay development, and data analysis.

Research Assistant, Aoife Heaslip Lab
University of Connecticut, Storrs, CT
December 2019-August 2020

- Investigated the function and regulation of TgMyoF in Toxoplasma gondii using molecular cloning, cell culture, and microscopy-based approaches.
- Contributed to projects on phosphorylation effects and endogenous complementation, which were presented at undergraduate and national research conferences.

Undergraduate Researcher, Aoife Heaslip Lab
University of Connecticut
April 2018-December 2019

HONORS AND AWARDS

Fellowship, Initiative to Maximize Student Development (IMSD)
Brown University
August 2020-August 2021
Undergraduate Research Competition Winner
Minority Access Role Model Conference, Washington, D.C.
September 2019
Ronald E. McNair Scholar (TRIO Program)
University of Connecticut
2018-2019

Manuscripts in Preparation
R. Britt, T. van Eeuwen, J. Sedivy, G. Jogl, M. Taylor**. Nucleic acid specificity of human LINE1 ORF1p is mediated by distinct binding clefts.** Manuscript in preparation.
K. Kondo, A. Harms, T. Owens, H. Kodama, R. Britt… M. Taylor. **Practical Solutions to Bacteriophage Contamination for Biochemistry Labs.** Manuscript in preparation.

- I led the negative staining and TEM imaging for the bacteriophage studied in the manuscript.

PRESENTATIONS AND POSTER EXHIBITS
Biochemical Study of LINE1 ORF1p Interaction with Nucleic Acids.
Ninth Annual Northeastern Glenn Symposium on the Biology of Aging
October 2025
Current Efforts in Visualizing LINE1 ORF1p RNP.
New England Cryo-EM Symposium
November 2024
Biochemical Study of LINE1 ORF1p Interaction with Nucleic Acids.
Cold Spring Harbor Laboratory, Transposable Elements Conference
October 2024
Expression and Structural Analysis of LINE1's Retrotransposase and Endonuclease: ORF2.
Cold Spring Harbor Laboratory, Mechanisms of Aging Conference
September 2022
Impacts of Mutation and Temperature on Oligomerization of an Essential Spliceosome Assembly Protein: Survival of Motor Neuron.
IMSD Research in Progress Presentation
May 2021
Endogenous MyoF Complementation in ΔKu80-ΔMyoF *Toxoplasma gondii* Parasites.
Frontiers in Undergraduate Research Poster Presentation
April 2019
Researching the Role of an Essential Myosin Motor of *Toxoplasma gondii*, TgMyoF.
Minority Access Role Models Conference
September 2018
The Effects of Phosphorylation on the Coiled-Coil Domain of TgMyoF, a Myosin Motor of *Toxoplasma gondii*.

UConn McNair Scholars Program Summer Poster Exhibit
August 2018

PROFESSIONAL DEVELOPMENT

LBMS Cryo-EM Training Workshop Series (2024-1), April 2024
Building and Leading Inclusive Research Teams (IMSD Module), February 2024
Responsible Conduct of Research Training, November 2020

EDUCATIONAL ACTIVITIES

AspirED Minds, Mentor and Tutor, September 2021
UCAP Chemistry, Guest Lecturer, June 2018
First Star Academy, STEM Workshop Leader, July 2018

RESEARCH SKILLS

Molecular Biology and Cloning

PCR; Gibson and HiFi assembly; ligation; site-directed mutagenesis; plasmid construction; miniprep and maxiprep; bacterial plasmid amplification.

Protein Expression and Purification

Expression in *E. coli*, Sf9 baculovirus system, and mammalian cells; purification using affinity and size-exclusion chromatography; ammonium sulfate precipitation; immunoprecipitation; chemical and high-pressure lysis.

Biochemical and Biophysical Assays

MST; DSF; FP; EMSA; analytical SEC; Agilent TapeStation; ProteinSimple Jess.

Structural Biology

Cryo-EM sample vitrification, imaging, and data processing; negative stain TEM; X-ray crystallography; experience with CryoSPARC, RELION, Coot, PyMOL, and ChimeraX.

Cell Culture, Microscopy and Flow Cytometry

HEK293T/HeLa/U2OS, HFFs, Jurkat T cells; Sf9 insect cells; *Toxoplasma gondii* parasite handling; transfection; passaging; immunofluorescence assays; live-cell imaging; ImageJ; flow cytometry and FACSDiva.

Computational and Data Analysis

GraphPad Prism; Excel; basic Python and R scripting.

Acknowledgements

First, I would like to thank God, whose grace has sustained me throughout this work. My Christian faith has been a source of strength, hope, endurance, and perspective, especially during the final year of graduate school and the dissertation-writing process. I am deeply grateful for the ways God has carried me through seasons of uncertainty, fatigue, and growth, and for the people He placed around me during this journey. I would also like to thank my church community for helping keep me faithful, hopeful, and grounded. Your prayers, encouragement, friendship, and example of steadfastness helped me continue forward when the process felt overwhelming.

I would like to thank my advisors, John Sedivy and Gerwald Jogl, for their mentorship over the years. This project would not have been possible without your respective expertise, guidance, and support. Thank you for your feedback, patience, and for allowing me the freedom to learn many new skills throughout this project. I am grateful for the opportunity to work on developing skill that will undoubtably help me in future endeavors and for the ways your mentorship shaped my development as a scientist. Sedivy for giving me the broad perspective on how to think and filter out ideas to more efficiently navigate this space. Gerwald for patience and continuous on the spot guidance and help throughout the project.

I would also like to thank members of the Brown University research community and collaborators who contributed to this work and to my training. I am especially grateful to Trevor and Marty for being generous mentors and for offering their time, advice, and encouragement. I would also like to thank the Lisi lab for taking the time to help me develop and work through biophysical theory and techniques. Those conversations and guidance were invaluable as I learned how to think more carefully about the experimental and theoretical aspects of this project. Thank

you to the Neretti lab for allowing me to use their microscope and for making that resource available when it was needed.

I am deeply grateful to my family for their love, sacrifice, and encouragement. To my mother, thank you for making the ultimate sacrifice and for loving me in a way that has shaped the person I am today. To my father, thank you for working so hard to raise me and for the sacrifices you made to support me. To my stepmother, thank you for taking me into your family and making space for me with generosity and care. I would also like to thank my mother's side of the family for supporting and encouraging me despite being across the Atlantic. Your love, prayers, and belief in me have meant more than I can say.

I would also like to thank my friends, who helped make this season of life more joyful, meaningful, and bearable. Thank you to Gordo, Julia, and Collin for always being gracious hosts and intentional friends. Thank you for opening your homes, sharing yourselves with me, and creating spaces where I could rest, laugh, and feel known. I would also like to thank Myles and Elinton for being great roommates and for encouraging me during times when I might otherwise have withdrawn from the world. Your friendship and presence helped carry me through difficult stretches of this process.

Finally, I am grateful to everyone who offered encouragement, advice, patience, and kindness along the way. This dissertation represents not only years of experiments and writing, but also the support of many people who helped me endure, grow, and finish. Thank you.

Table of Contents

Signature Page iii
Curriculum Vitae iv
Acknowledgements vii
Table of Contents ix
List of Tables and Figures xi
Chapter 1 | Introduction 1
Transposable Elements: Definition and Classification 1
L1 in Genome Evolution 3
L1 Composition and Cellular Context 5
HsORF1p's Evolution and Structure 8
References 9
Chapter 2 | The Effects of Nucleic Acid Composition on HsORF1p Binding Dynamics 19
Background – HsORF1p Nucleic Acid Interactions and *Cis* Preference 19
Background – Choice of Nucleic Acid Substrates Used in This Study 21
Background – DSF and Spectral Shift 22
Materials and Methods 24
Plasmid Cloning 24
HsORF1p Expression and Purification 25
Differential Scanning Fluorimetry (DSF) Assay 26
Binding Affinity Measurements 27
Results 29
Nucleic Acid Polymer Composition Influences HsORF1p-RNP Stability 29
Nucleic Acid Polymer Composition Governs HsORF1p's Binding Affinity 33
Summary and Discussion 35
References 39
Chapter 3 | The Importance of Positively Charged Residues in HsORF1p Function 43
Background – Positively Charged Residues and Nucleic Acid Binding 43

Background – Choice of Mutant HsORF1p Constructs 45
Background – Retrotransposition Efficiency Assay and Immunofluorescence Imaging .. 48
Materials and Methods 50
Plasmid Cloning, Protein Expression and Purification 50
Cell Culture Maintenance 50
Cell Transfections for Flow Cytometry and Immunofluorescence Microscopy 51
Flow Cytometry 52
Immunofluorescence Microscopy 53
Results 54
Structural integrity and purity of mutants 54
Mutation's Effect on Thermal Stability and Ligand Binding 54
The Effect of Charge neutralization on Cellular L1 Retrotransposition 59
Summary and Discussion 61
References 65
Appendix 84

List of Tables and Figures

Figure 1.1 – Canonical L1 Retrotransposition Cycle & Domains of L1 Encoded Proteins……..4

Figure 1.2 – Domain Architecture and Uniqueness of HsORF1p's RRM and CTD……....8

Figure 2.1 – Differential Scanning Fluorimetry of HsORF1p Bound and Unbound……..25

Figure 2.2 – Differential Scanning Fluorimetry of HsORF1p vs Monomeric RRM-CTD…….27

Figure 2.3 – Biding Affinity of HsORF1p with Various Nucleic Acid Ligands…….29

Figure 3.1 – Distribution of Basic Residue Mutations Across the HsORF1p RRM and CTD……41

Figure 3.2 – Differential Scanning Fluorimetry of Mutant HsORF1p Constructs……..49

Figure 3.3 – Differential Scanning Fluorimetry Heat Map of Mutant:Ligand Stability…….51

Figure 3.4 – Retrotransposition Efficiency of HsORFeus Encoding Mutant HsORF1p……54

Figure 3.5 – Immunofluorescence Analysis of HsORFeus-Expressed HsORF1p Mutants……...55

Figure 4.1 – Proposed Stages of HsORF1p Mediated L1 mRNA RNP Formation……64

Supplemental Table 1 – Top 25 DALI Structural Matches to the HsORF1p RRM…….75

Supplemental Table 2 – Top 25 DALI Structural Matches to the HsORF1p CTD……76

Supplemental Table 3 – List of Nucleic Acids Used in Thesis…….77

Supplemental Figure 1 – SDS-PAGE of Purification Steps and HsORF1p Constructs……....78

Supplemental Figure 2 – Quality Control SEC & CD Spectroscopy of HsORF1p Constructs….79

Supplemental Figure 3 – Validation of Binding Equilibrium Prior to K_d Measurements…….80

Supplemental Figure 4 – Flow Cytometry Example Data and Gating…….81

Supplemental Figure 5 – Western Blot of HsORFeus Encoded HsORF1p Mutants…….82

Supplemental Figure 6 – Representative CellProfiler Foci Counting……...83

Chapter 1 | Introduction

Transposable Elements: Definition and Classification

Transposable elements (**transposons**) are genomic DNA sequences that encode or recruit proteins to integrate themselves, or a *de novo* copy of themselves, into other genomic locations in a process referred to as **mobilization**[1, 2]. Their biological relevance is underscored by their prevalence throughout Eukarya, their contribution to genome expansion and their role in potentially reforming genome structure within individual cells but also across evolutionary time[2-4]. Distinct transposable element lineages have evolved a diverse set of mobilization strategies and some arriving to similar strategies independently, highlighting a broad biological importance of transposons[2, 5, 6]. However, Long Interspersed Nuclear Element 1 (**LINE-1 or L1**) is a transposon that encodes an RNA-binding/nucleic acid chaperone protein, ORF1p, that is essential for efficient L1 mobilization and has unusual structural and biochemical properties relative to better-characterized nucleic acid chaperones[7-10].

Due to the wide diversity of transposable elements, they are canonically categorized into hierarchical groupings to better understand them[1, 2, 11]. **DNA transposons** (**Class II**) mobilize by excising from one locus and integrating into another, what is referred to as a 'cut and paste' mechanism, and generally only increase in copy number through indirect consequences of mobilization (i.e., replication-coupled transposition and homologous recombination)[2, 12, 13]. **Retrotransposons** (**Class I**) mobilize through a '**copy and paste**' mechanism in which the original DNA element remains unaltered but *de novo* replicate of itself integrates into another locus, increasing the copy number of the element directly[1, 2]. Retrotransposons achieve a 'copy and paste' mobilization by utilizing both transcription and reverse transcription[2, 3, 14]. Since this work focuses L1, a member of Class I, subsequent discussion centers on retrotransposons.

Within Class I, elements are divided into Long Terminal Repeat (**LTR**) and **non-LTR** retrotransposons based on their composition and mobilization strategy[1-3]. LTR retrotransposons typically encode 'retroviral-like' proteins that are involved in making retrovirus-like particles that enable housing for cytoplasmic reverse transcription to occur and subsequent genomic integration of the synthesized DNA[2, 15]. Non LTR retrotransposons lack terminal repeats and mobilize through a distinct target primed reverse transcription mechanism that occurs at the insertion site [16-18]. Because reverse transcription canonically occurs directly at chromosomal DNA rather than inside a preformed particle, non-LTR retrotransposons must stabilize, transport and protect their RNA copies long enough until their RNA can be reverse transcribed and integrated into the genome[16-20].

There are an estimated 10 total clades of non-LTR retrotransposons within the human genome, with only 3 still known be to be active, including L1, Alu and SVA[2, 16, 18]. These non-LTR retrotransposons are subdivided into LINEs (i.e., L1) and short interspersed nuclear elements (**SINEs**) (i.e., Alu and SVA)[2, 16, 18]. Active LINEs are canonically thought of as **autonomous**, meaning that they can generate the protein(s) required for their mobilization using 'host machinery'[2, 16, 18]. Interestingly, LINEs have evolved mechanisms that allow their encoded proteins to preferentially assemble on the LINE mRNA that encoded for them, a property referred to as ***cis*-preference**, to reduce the likelihood of their proteins associating with and mobilizing other cellular RNAs over their own[21-23]. In comparison, SINEs are shorter, nonautonomous elements that rely on LINE encoded protein(s), supplied ***in trans***, for their mobilization and have co-evolved strategies to overcome *cis*-preference[2, 22-26].

L1 encodes at least 2 proteins (Figure 1.1): first an endonuclease and reverse transcriptase containing protein that facilitates target site nicking, annealing and reverse transcription required

for the mobilization of both L1 and SINEs, and second a packaging/chaperone protein that assembles its RNA into a **ribonucleoprotein (RNP)** that is thought to stabilize the RNA, mediate interactions with host cellular machinery and facilitate the RNA's transport back to the genome[19, 20, 27, 28]. The endonuclease and reverse transcriptase are required for both LINE and SINE mobilization, while the RNA chaperone protein is required for successful L1 mobilization and may enhance but is not required for Alu mobilization[8, 9, 25]. Interestingly, ORF1p is structurally and mechanistically unusual relative to some other characterized nucleic acid chaperones, and its mode of nucleic acid remodeling is not fully analogous either to retroviral nucleocapsid proteins or to canonical single stranded nucleic acid binding proteins[7-10]. Although the unusual properties of ORF1p complicate experimental handling and study design, understanding its functions would broaden our knowledge of protein-nucleic acid interactions and may inform the development of detection strategies and therapeutic approaches in disease and aging contexts (see L1 Composition and Cellular Context)[8, 29, 30].

L1 in Genome Evolution

L1 has been a major contributor to genome composition[14, 31]. In humans, L1 derived sequences account for 17–20% of the reference genome and exist in 500,000 copies[14, 31, 32]. However, there are polymorphisms present across the population of copies, so L1s are typically classified into subfamilies such as L1PA and L1Hs[14, 31, 33]. These subfamilies are further divided by computed evolutionary age, where older subfamilies are generally thought of as evolutionary relics incapable of autonomous mobilization and younger subfamilies are more likely to retain autonomy[14, 31, 33]. Even within the evolutionary young L1Hs elements, only a subset retain the ability for autonomous mobilization, as the process of reverse transcription often leads to truncation or non-functional polymorph [14, 34]. So, of the 500,000 L1 derived copies, it is estimated

that the average individual's genome only has 50-100 full-length and intact L1s and usually only 5-20 of these elements are "hot", meaning they account for most novel insertions when conditions permit L1 activity[32, 35, 36].

Although LINEs like L1 are arguably positive for the development of genomes across species on the timescale of evolution, active L1 retrotransposition can be deleterious within individual cells and organisms[14, 31]. New insertions may disrupt genes, compromise genomic integrity, and activate DNA damage responses[34]. Consequently, alongside transposon evolution, host cells have co-evolved regulatory machinery that restricts transposon activity, suggesting either a complex and tentative symbiotic relationship or a 'full blown host versus parasite arms race'[6, 37, 38]. For example, L1 activity is tightly regulated at the epigenomic level by host mechanisms (i.e., DNA methylation and histone modification) and post transcriptional restriction factors (i.e., Let7, TUT4/7 and MOV10) which both may play into determining the full-length and intact L1s are 'hot' and those that are not[39-45]. However, the mechanisms of L1 regulation can deteriorate in aging, senescence and disease conditions in which it has been demonstrated that L1 expression and retrotransposition activity are frequently elevated[46-49].

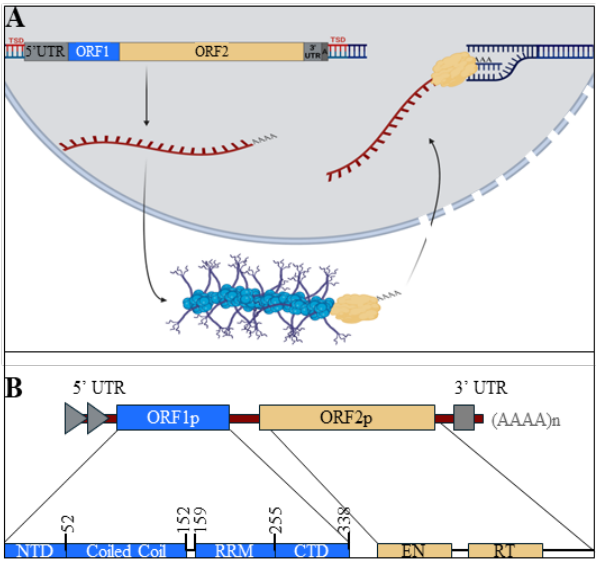


Figure 1A *Overview of the canonical L1 retrotransposition cycle. A derepressed genomic L1 is transcribed into L1 mRNA (red). ORF1p (blue) and ORF2p (orange) associate with the L1 mRNA co-translationally to form a ribonucleoprotein particle (RNP) with cis preference. After gaining access to genomic DNA, ORF2p endonuclease and reverse transcriptase mediates synthesis of a L1 copy through target-primed reverse transcription.*

Figure1B *Domains of L1 encoded proteins. ORF1p is composed of a N-terminal disordered region (NTD) responsible with phase-separation properties in vitro, a coiled-coil domain for trimerization, an RNA recognition motif (RRM) and C-terminal domain (CTD) that coordinate to form the nucleic acid binding surface. ORF2p consists of two enzymatic domains, the endonuclease (EN) that generates the nick in genomic DNA and reverse transcriptase (RT) that synthesizes L1 from mRNA template.*

L1 Composition and Cellular Context

A functional L1 is typically full length, approximately 6,000 nucleotides, and contains the features required for mobilization (Figure 1.1): a 5' UnTranslated Region (**5' UTR**) containing an internal promoter, two protein-coding Open Reading Frames, **ORF1** and **ORF2**, a **2' UTR**, and a **poly(A) tail**[34, 50]. In some younger primate L1s, an antisense promoter within the 5' UTR drives expression of ORF0p, a poorly characterized protein that is not part of the canonical L1 retrotransposition pathway[51].

As stated earlier, retrotransposons rely on both transcription and reverse transcription to complete mobilization. The canonical L1 retrotransposition cycle (Figure 1.1) begins when a derepressed, autonomous element is transcribed by RNA polymerase II; a promoter sequence in the 5' UTR recruits the polymerase and, in human L1, a YY1-binding site also within this region contributes to accurate transcription initiation[50, 52]. There is substantial indirect evidence that the L1 RNA undergoes host RNA processing associated with RNA polymerase II transcripts before export to the cytoplasm (i.e., exposure to splicing machinery, 2' polyadenylation, acquisition of a 5' cap)[53, 54]. During the interval between RNA maturation and translation, several host factors likely recognize and associate with the L1 mRNA to either promote or suppress its translation, although a fully defined pre-translation RNP has not yet been established[41, 42, 55, 56].

Once recognized by ribosome(s), the L1 mRNA undergoes a unique and still poorly understood mode of translation, likely involving to two unique features: it is a **bicistronic** RNA, meaning that ORF1 and ORF2 are arranged in-line on the same mRNA transcript, and it exhibits ***cis* preference**, in which the proteins encoded by the mRNA bind that same mRNA either during or shortly after translation[23, 57]. L1 translation appears to yield substantially more ORF1p than ORF2p (the "p" following "ORF#" is used to distinguish the protein from the DNA/RNA element). In some endogenous contexts, ORF1p is much more readily detected than ORF2p [R]. In ectopic

L1 overexpression studies, it has been reported that ORF1p trimer:ORF2p stoichiometries are on the order of ~10-100 ORF1p trimers per each ORF2p [8, 19, 58, 59]; The basis for this difference in translation efficiency, despite ORF1 and ORF2 being encoded on the same bicistronic transcript, remains unclear [8, 23]. Proposed explanations include: poor reinitiation of ribosomes after the stop codon in ORF1, host defenses (i.e., siRNA and miRNA-associated protein assemblies), and possible contributions from ORF1p association with L1 mRNA[8, 23, 30, 41, 42]. However, direct evidence for the latter two possibilities remains sparse, in part because ORF2p has historically been difficult to detect and quantify[59-62].

Due to the high ORF1p to ORF2p ratio that results from translation, together with other *in cellulo* observations and *in vitro* biochemical characteristics, ORF1p is postulated to function as a nucleic acid chaperone and a major structural component of the L1 RNP[8, 27, 63-65]. *In cellulo* data demonstrate that successful L1 mobilization depends on translation of functional ORF1p and that ORF1p plays a significant role proper RNP assembly, subcellular localization, and formation of cytoplasmic foci associated with stress granules and the nucleus[19, 20, 57, 66]. In parallel, biochemical and biophysical studies of ORF1p have shown that it has high affinity for single stranded nucleic acid, as well as nucleic acid chaperone, nucleic acid condensation, and phase separation activities *in vitro*, further supporting a role in organizing and shuttling the post-translational L1 RNP[27, 29, 30, 63-65]. Despite this central role, the molecular basis by which ORF1p recognizes and organizes RNA within the RNP remains unclear current evidence supports broad, often non-sequence-specific interactions with nucleic acids, but the extent to which HsORF1p lacks nucleotide-base preference remains unresolved[27, 63].

Once ORF1p and ORF2p initially associate with their mRNA to form a nascent RNP, additional ORF1p likely binds the mRNA to further restructure it, while other host factors likely

associate to produce a mature RNP [19, 20, 57, 67]. The L1 RNP is trafficked through cellular compartments and gains access to genomic DNA. There are studies that L1 RNP condensates bind DNA and enter the nuclear compartment during mitosis, when nuclear envelope breakdown provides access to chromosomal DNA[28, 61]. Other studies argue that there is a mechanism of active transport through the nuclear membrane[68, 69]. Lastly, a new L1 sequence is synthesized and integrated by ORF2p through target primed reverse transcription[70, 71].

As L1 RNPs traverse multiple cellular environments, they interface with host machinery and biochemical pathways[20, 67]. For example, L1 expression can generate cytoplasmic L1-derived nucleic acid species, including cDNA, that engage innate immune sensing pathways classically associated with antiviral responses[30, 46, 47, 72]. This contributes to senescence-associated inflammatory signaling through cGAS-STING and interferon signaling, as well as to the broader inflammatory phenotypes associated with senescence and certain diseases[46, 47, 72, 73]. Furthermore, dysregulated L1 activity can give rise to multiple forms of genomic instability, including insertional mutagenesis, structural genome damage, and aberrant gene expression and RNA processing[34, 48, 74]. However, the consequences of L1 activity appear to be highly context dependent, as L1 is also expressed during early embryogenesis, where its activity appears better tolerated[75, 76], but is associated with genomic instability and inflammation in senescent somatic cells and disease states, including cancer and chronic inflammatory disorders[47, 48, 74]. These observations position L1 proteins both as potential therapeutic targets, as is currently being explored for ORF2p enzymatic function[47, 77-79] and potentially for ORF1p if it research can be extended to developing protein degradation technologies such as various targeted chimera treatments[80, 81], and as detectable biomarkers of disease, as is being developed for ORF1p in certain cancers[82].

HsORF1p's Evolution and Structure

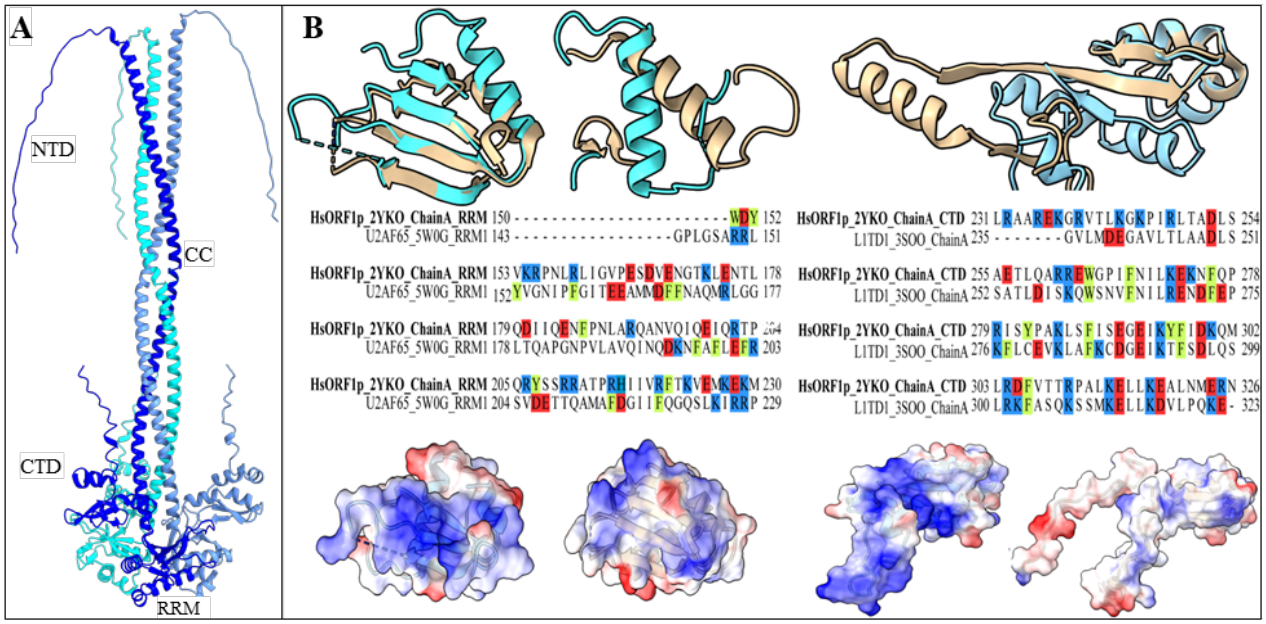


Figure 1.2 *Domain architecture and uniqueness of HsORF1p* ***A.*** *ChimeraX visualization of an Alpha Fold predicted model of full-length, trimeric HsORF1p, with each monomeric unit colored in a different shade of blue. The disordered NTD is at the top of the model followed by the central coiled-coil then the RRM and CTD at the base of the structure.* ***B.*** *Structural and sequence comparison of HsORF1p RRM (left) and CTD (right) to homologs identified by the DALI webserver. To the left are structural overlays of the each ORF1p domain (blue), derived from a monomer of PDB entry 2YKO, with their homolog (tan). To the right are primary sequence alignments and electrostatic surface maps of each domain and the indicated homolog.*

HsORF1p is an essential component of the L1 RNP and is generally described as an single-stranded nucleic acid binding protein with sequence non-specific binding properties and RNA chaperone activity[8, 10, 27, 83, 84]. Functional ORF1p and proper L1 RNP assembly are required for efficient L1 mobilization[19, 57]. Within the RNP, ORF1p is thought to stabilize the L1 RNA, promote proper RNP assembly, and support downstream steps leading to mobilization[8, 19, 27].

Structurally, HsORF1p is organized into four domains: an intrinsically disordered N-terminal domain (**NTD**), a coiled-coil domain (**CC**) responsible for trimerization, a central RNA recognition motif (**RRM**), and a C-terminal domain (**CTD**), with the RRM and CTD being the primary direct nucleic acid binding interface (Figure 1.2A)[8, 65, 83, 85, 86]. The domain architecture and amino acid sequence are more highly conserved from the base of the CC through the RRM and CTD than in the NTD of the protein (Figure 3.1A)[9, 27, 87, 88]. This is notable because although

ORF1p is generally classified as a generally sequence-independent nucleic acid binding protein with interactions strongly influenced by electrostatics[8, 27, 83], its primary sequence remains strongly conserved across active L1 orthologs[9, 87, 88]. This conservation is not only limited to functional class of amino acids used, as ORF1p appears to have preferential incorporation and conservation of arginines over lysines to form the highly basic electrostatic environment of its RRM and CTD, an essential feature that enables its proper function (Figure 1.3A). Together, these observations suggest that ORF1p function depends on underappreciated and finely tuned structural and biochemical features rather than on simple electrostatic interactions with nucleic acids alone.

Despite its essential role in L1 retrotransposition, how HsORF1p's RRM and CTD engage nucleic acids remain incompletely defined. Broadly, ORF1p is thought to bind single-stranded RNA and DNA in a largely sequence-independent manner[27, 83], although only limited work has directly compared binding across distinct nucleic acid sequences within the same class (i.e., multiple different RNAs with varying sequences)[10], while most others have examined only broad classes of substrates[8, 27, 83]. It is conceivable that nucleic acid composition may influence ORF1p RNP stability and organization, but whether different substrates differentially stabilize the protein and whether such stabilization correlates with binding affinity or retrotransposition efficiency, remains unclear. Although electrostatic interactions clearly contribute to nucleic acid engagement[8, 27, 83], how positively charged residues at different positions within the RRM and CTD influence binding affinity and support functional RNP assembly remains poorly resolved.

References

1. Wicker, T., et al., *A unified classification system for eukaryotic transposable elements.* Nat Rev Genet, 2007. **8**(12): p. 973–82.

2. Bourque, G., et al., *Ten things you should know about transposable elements.* Genome Biol, 2018. **19**(1): p. 199.

3. Hayward, A. and C. Gilbert, *Transposable elements.* Curr Biol, 2022. **32**(17): p. R904–r909.

4. Sotero-Caio, C.G., et al., *Evolution and Diversity of Transposable Elements in Vertebrate Genomes.* Genome Biol Evol, 2017. **9**(1): p. 161–177.

5. Betancourt, A.J., et al., *Causes and Consequences of Varying Transposable Element Activity: An Evolutionary Perspective.* Annu Rev Genomics Hum Genet, 2024. **25**(1): p. 1–25.

6. Zhang, X., et al., *Transposable elements employ distinct integration strategies with respect to transcriptional landscapes in eukaryotic genomes.* Nucleic Acids Res, 2020. **48**(12): p. 6685–6698.

7. Mandal, P.K. and H.H. Kazazian, Jr., *Purification of L1-Ribonucleoprotein Particles (L1-RNPs) from Cultured Human Cells.* Methods Mol Biol, 2016. **1400**: p. 299–310.

8. Martin, S.L., *The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition.* J Biomed Biotechnol, 2006. **2006**(1): p. 45621.

9. Januszyk, K., et al., *Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1.* J Biol Chem, 2007. **282**(34): p. 24893–904.

10. Kolosha, V.O. and S.L. Martin, *High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1).* J Biol Chem, 2003. **278**(10): p. 8112–7.

11. Hong, Y. and N. Liu, *Transposable elements in health and disease: Molecular basis and clinical implications.* Chin Med J (Engl), 2025. **138**(18): p. 2220–2233.

12. Zaratiegui, M., *Cross-Regulation between Transposable Elements and Host DNA Replication.* Viruses, 2017. **9**(3).

13. Feschotte, C. and E.J. Pritham, *DNA transposons and the evolution of eukaryotic genomes.* Annu Rev Genet, 2007. **41**: p. 331–68.

14. Li, X. and N. Liu, *Advances in understanding LINE-1 regulation and function in the human genome.* Trends Genet, 2025. **41**(7): p. 577–589.

15. Havecker, E.R., X. Gao, and D.F. Voytas, *The diversity of LTR retrotransposons.* Genome Biol, 2004. **5**(6): p. 225.

16. Han, J.S., *Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions.* Mob DNA, 2010. **1**(1): p. 15.

17. Jamburuthugoda, V.K. and T.H. Eickbush, *The reverse transcriptase encoded by the non-LTR retrotransposon R2 is as error-prone as that encoded by HIV-1.* J Mol Biol, 2011. **407**(5): p. 661–72.

18. Thawani, A., K. Collins, and E. Nogales, *Structural and biochemical studies of mobile retrotransposon proteins in action.* Curr Opin Struct Biol, 2025. **92**: p. 103053.

19. Doucet, A.J., et al., *Characterization of LINE-1 ribonucleoprotein particles.* PLoS Genet, 2010. **6**(10).

20. Goodier, J.L., et al., *Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion.* Hum Mol Genet, 2010. **19**(9): p. 1712–25.

21. Goodier, J.L., et al., *A potential role for the nucleolus in L1 retrotransposition.* Hum Mol Genet, 2004. **13**(10): p. 1041–8.

22. Roy-Engel, A.M., *A tale of an A-tail: The lifeline of a SINE.* Mob Genet Elements, 2012. **2**(6): p. 282–286.

23. Wei, W., et al., *Human L1 retrotransposition: cis preference versus trans complementation.* Mol Cell Biol, 2001. **21**(4): p. 1429–39.

24. Taylor, M.S., et al., *Characterization of L1-Ribonucleoprotein Particles.* Methods Mol Biol, 2016. **1400**: p. 311–38.

25. Wallace, N., et al., *LINE-1 ORF1 protein enhances Alu SINE retrotransposition.* Gene, 2008. **419**(1-2): p. 1–6.

26. Richardson, S.R., et al., *The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes.* Microbiol Spectr, 2015. **3**(2): p. Mdna3–0061–2014.

27. Naufer, M.N., A.V. Furano, and M.C. Williams, *Protein-nucleic acid interactions of LINE-1 ORF1p.* Semin Cell Dev Biol, 2019. **86**: p. 140–149.

28. Zernia, S., et al., *LINE-1 ribonucleoprotein condensates bind DNA to enable nuclear entry during mitosis.* Sci Adv, 2025. **11**(18): p. eadt9318.

29. Naufer, M.N. and M.C. Williams, *Characterizing Complex Nucleic Acid Interactions of LINE1 ORF1p by Single Molecule Force Spectroscopy.* Methods Mol Biol, 2020. **2106**: p. 283–297.

30. Xing, B., et al., *LINE-1 Retrotransposon Protein ORF1p Forms Condensates that Drive cGAS-Induced Immune Evasion in Lung Squamous Cell Carcinoma.* Cancer Res, 2026.

31. Ovchinnikov, I., A.B. Troxel, and G.D. Swergold, *Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion.* Genome Res, 2001. **11**(12): p. 2050–8.

32. Cordaux, R. and M.A. Batzer, *The impact of retrotransposons on human genome evolution.* Nat Rev Genet, 2009. **10**(10): p. 691–703.

33. Misak, M., A. Basu, and C. Niehrs, *Subfamily-selective PCR primers for the human LINE1 L1PA lineage.* Sci Rep, 2025. **15**(1): p. 32499.

34. Levin, H.L., S. Lee, and A. Anand, *The 5′ truncation of retrotransposon L1: a process of genome integrity.* Genetics, 2025. **231**(4).

35. Brouha, B., et al., *Hot L1s account for the bulk of retrotransposition in the human population.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5280–5.

36. Yang, L., et al., *Evolutionary insights from profiling LINE-1 activity at allelic resolution in a single human genome.* Embo j, 2024. **43**(1): p. 112–131.

37. Iwasaki, Y.W., et al., *Transposon-host arms race: a saga of genome evolution.* Trends Genet, 2025. **41**(5): p. 369–389.

38. Lawlor, M.A. and C.E. Ellison, *Evolutionary dynamics between transposable elements and their host genomes: mechanisms of suppression and escape.* Curr Opin Genet Dev, 2023. **82**: p. 102092.

39. Danac, J.M.C., et al., *Competition between two HUSH complexes orchestrates the immune response to retroelement invasion.* Mol Cell, 2024. **84**(15): p. 2870–2881.e5.

40. Protasova, M.S., T.V. Andreeva, and E.I. Rogaev, *Factors Regulating the Activity of LINE1 Retrotransposons.* Genes (Basel), 2021. **12**(10).

41. Warkocki, Z., et al., *Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s.* Cell, 2018. **174**(6): p. 1537–1548.e29.

42. Tristán-Ramos, P., et al., *The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition.* Nat Commun, 2020. **11**(1): p. 5712.

43. Goodier, J.L., L.E. Cheung, and H.H. Kazazian, Jr., *MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells.* PLoS Genet, 2012. **8**(10): p. e1002941.

44. Guan, Y., et al., *The MOV10 RNA helicase is a dosage-dependent host restriction factor for LINE1 retrotransposition in mice.* PLoS Genet, 2023. **19**(5): p. e1010566.

45. Choi, J., S.Y. Hwang, and K. Ahn, *Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition.* Nucleic Acids Res, 2018. **46**(4): p. 1912–1926.

46. Dou, Z., et al., *Cytosolic DNA crosstalk in senescence: a new axis of inflammatory signaling?* Embo j, 2025. **44**(19): p. 5239–5243.

47. De Cecco, M., et al., *L1 drives IFN in senescent cells and promotes age-associated inflammation.* Nature, 2019. **566**(7742): p. 73–78.

48. Kemp, J.R. and M.S. Longworth, *Crossing the LINE Toward Genomic Instability: LINE-1 Retrotransposition in Cancer.* Front Chem, 2015. **3**: p. 68.

49. Roy, N., et al., *Elevated expression of the retrotransposon LINE-1 drives Alzheimer's disease-associated microglial dysfunction.* Acta Neuropathol, 2024. **148**(1): p. 75.

50. Lavie, L., et al., *The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity.* Genome Res, 2004. **14**(11): p. 2253–60.

51. Denli, A.M., et al., *Primate-specific ORF0 contributes to retrotransposon-mediated diversity.* Cell, 2015. **163**(3): p. 583–93.

52. Athanikar, J.N., R.M. Badge, and J.V. Moran, *A YY1-binding site is required for accurate human LINE-1 transcription initiation.* Nucleic Acids Res, 2004. **32**(13): p. 3846–55.

53. Luqman-Fatah, A., et al., *Retrotransposon life cycle and its impacts on cellular responses.* RNA Biol, 2024. **21**(1): p. 11–27.

54. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells.* Cell, 1996. **87**(5): p. 917–27.

55. Dai, L., et al., *Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation.* Mol Cell Biol, 2012. **32**(21): p. 4323–36.

56. Peddigari, S., et al., *hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition.* Nucleic Acids Res, 2013. **41**(1): p. 575–85.

57. Kulpa, D.A. and J.V. Moran, *Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition.* Hum Mol Genet, 2005. **14**(21): p. 3237–48.

58. Nielsen, M.I., et al., *Targeted detection of endogenous LINE-1 proteins and ORF2p interactions.* Mob DNA, 2025. **16**(1): p. 3.

59. Ardeljan, D., et al., *LINE-1 ORF2p expression is nearly imperceptible in human cancers.* Mob DNA, 2020. **11**: p. 1.

60. Sokolowski, M., et al., *Characterization of L1 ORF1p self-interaction and cellular localization using a mammalian two-hybrid system.* PLoS One, 2013. **8**(12): p. e82021.

61. Mita, P., et al., *LINE-1 protein localization and functional dynamics during the cell cycle.* Elife, 2018. **7**.

62. Sharma, R., et al., *Immunodetection of Human LINE-1 Expression in Cultured Cells and Human Tissues.* Methods Mol Biol, 2016. **1400**: p. 261–80.

63. Callahan, K.E., et al., *Polymerization and nucleic acid-binding properties of human L1 ORF1 protein.* Nucleic Acids Res, 2012. **40**(2): p. 813–27.

64. Kolosha, V.O. and S.L. Martin, *In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition.* Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10155–60.

65. Martin, S.L., et al., *Trimeric structure for an essential protein in L1 retrotransposition.* Proc Natl Acad Sci U S A, 2003. **100**(24): p. 13815–20.

66. Goodier, J.L., et al., *LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex.* Mol Cell Biol, 2007. **27**(18): p. 6469–83.

67. Taylor, M.S., et al., *Dissection of affinity captured LINE-1 macromolecular complexes.* Elife, 2018. **7**.

68. Freeman, B.T., et al., *Identification of charged amino acids required for nuclear localization of human L1 ORF1 protein.* Mob DNA, 2019. **10**: p. 20.

69. Znaidi, R., et al., *Nuclear translocation of the LINE-1 encoded ORF1 protein alters nuclear envelope integrity in human neurons.* Brain Res, 2025. **1857**: p. 149579.

70. Ghanim, G.E., et al., *Structural mechanism of LINE-1 target-primed reverse transcription.* Science, 2025. **388**(6745): p. eads8412.

71. Feng, Q., et al., *Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.* Cell, 1996. **87**(5): p. 905–16.

72. Cao, W., *IFN-Aging: Coupling Aging With Interferon Response.* Front Aging, 2022. **3**: p. 870489.

73. Simon, M., et al., *LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation.* Cell Metab, 2019. **29**(4): p. 871–885.e5.

74. Xiao-Jie, L., et al., *LINE-1 in cancer: multifaceted functions and potential clinical implications.* Genet Med, 2016. **18**(5): p. 431–9.

75. Beraldi, R., et al., *Expression of LINE-1 retroposons is essential for murine preimplantation development.* Mol Reprod Dev, 2006. **73**(3): p. 279–87.

76. Percharde, M., et al., *A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity.* Cell, 2018. **174**(2): p. 391–405.e19.

77. Dai, L., Q. Huang, and J.D. Boeke, *Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition.* BMC Biochem, 2011. **12**: p. 18.

78. Patnala, R., et al., *Inhibition of LINE-1 retrotransposon-encoded reverse transcriptase modulates the expression of cell differentiation genes in breast cancer cells.* Breast Cancer Res Treat, 2014. **143**(2): p. 239–53.

79. D'Ordine, A.M., G. Jogl, and J.M. Sedivy, *Identification and characterization of small molecule inhibitors of the LINE-1 retrotransposon endonuclease.* Nat Commun, 2024. **15**(1): p. 3883.

80. Békés, M., D.R. Langley, and C.M. Crews, *PROTAC targeted protein degraders: the past is prologue.* Nat Rev Drug Discov, 2022. **21**(3): p. 181–200.

81. Sato, S., et al., *LINE-1 ORF1p as a candidate biomarker in high grade serous ovarian carcinoma.* Sci Rep, 2023. **13**(1): p. 1537.

82. Taylor, M.S., et al., *Ultrasensitive Detection of Circulating LINE-1 ORF1p as a Specific Multicancer Biomarker.* Cancer Discov, 2023. **13**(12): p. 2532–2547.

83. Basame, S., et al., *Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition.* J Mol Biol, 2006. **357**(2): p. 351–7.

84. Martin, S.L., *Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1.* RNA Biol, 2010. **7**(6): p. 706–11.

85. Martin, S.L., J. Li, and J.A. Weisz, *Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1.* J Mol Biol, 2000. **304**(1): p. 11–20.

86. Newton, J.C., et al., *Phase separation of the LINE-1 ORF1 protein is mediated by the N-terminus and coiled-coil domain.* Biophys J, 2021. **120**(11): p. 2181–2191.

87. Adney, E.M., et al., *Comprehensive Scanning Mutagenesis of Human Retrotransposon LINE-1 Identifies Motifs Essential for Function.* Genetics, 2019. **213**(4): p. 1401–1414.

88. Cashen, B.A., et al., *The L1-ORF1p coiled coil enables formation of a tightly compacted nucleic acid-bound complex that is associated with retrotransposition.* Nucleic Acids Res, 2022. **50**(15): p. 8690–8699.

Chapter 2 | The Effects of Nucleic Acid Composition on HsORF1p Binding Dynamics

Background – HsORF1p Nucleic Acid Interactions and *Cis* Preference

In the context of L1 retrotransposition, it is likely beneficial that HsORF1p has some mechanism of promoting ***cis* preference**, defined here as the preferred interaction with the L1 mRNA from which it was translated over other nucleic acid components and polymers present in the cellular environment during translation. Several studies support a *cis* preference model in which L1 encoded proteins preferentially associate with and mobilize the L1 mRNA from which they were translated. L1 encoded proteins only minimally support mobilization of co-expressed vectors containing retrotransposition defective mutants. In addition, HsORF1p foci show minimal mixing with other HsORF1p foci *in cellulo*, suggesting that HsORF1p preferentially assembles on the L1 mRNA from which it was translated rather than freely mixing with other L1 Ribonucleoprotein particles (**RNPs**) or nucleic acid (**NA**) substrates.

The RRM and CTD likely contain the primary binding surface through which HsORF1p associates with its encoding mRNA during *cis* preference. Both the RRM and CTD are essential for L1 mobilization, as mutations in either domain often reduce retrotransposition from an ectopically expressed L1 plasmid[1-5]. In addition to requiring an extended sugar-phosphate backbone of ~20-27-nt for stable, high affinity binding [4, 6, 7], HsORF1p also appears to prefer flexible or unstructured NAs. Although HsORF1p can engage a broad range of NA classes, current studies suggest a general binding-affinity hierarchy of ssRNA > ssDNA > mismatch-containing dsDNA >> tRNA > dsDNA[8-10]. HsORF1p having the highest affinity for ssRNA makes sense in the context of L1 retrotransposition as the L1 mRNA that encodes it is its most direct natural substrate.

Despite it being well established that these domains form the primary interaction surface for RNA[2-4, 10, 11], biophysical measurements that characterize how these binding domains interact with different ssRNA and ssDNA sequences remain limited. Most studies that have addressed this question have concluded that HsORF1p binding is largely sequence-independent, based on experiments using panels of NAs that did not vary sequence composition extensively[6, 9, 12, 13]. One study that did intentionally vary sequence composition also concluded that with the sequences tested that HsORF1p's binding affinity is largely sequence independent[6]. Together with evidence that HsORF1p can destabilize NA secondary structure at sufficient concentration, this has supported a prevailing model in which HsORF1p recognition depends primarily on proximity and access to the sugar-phosphate backbone rather than specific base identity or sequence[6, 9, 10, 14].

However, there has been research suggesting that HsORF1p may exhibit sequence preferences, and several studies indicate that more than simply flexible RNA or ssDNA may be required for high-affinity binding. For example, the original 1997 study that argued for sequence-specific binding identified two short sequences, 37- and 41-nt in length, within the L1RP RNA that were preferentially protected from RNase H1 digestion relative to the remainder of the ~6 kb transcript[15]. A follow-up study that crystallized the RRM–CTD and performed analytical size-exclusion chromatography suggested that the monomeric HsORF1p RRM–CTD domain may associate more efficiently with the 41-nt RNA sequence than with other L1-derived sequences[16]. Additionally, in the 2011 study reporting the trimeric HsORF1p structure, it was shown that HsORF1p lacked apparent affinity for poly(A) ssDNA compared to poly(T) ssDNA and poly(A) RNA. Cellular studies further show that HsORF1p is not restricted to binding its own L1 transcript[4]. HsORF1p can associate with many non-L1 RNAs, but these RNAs are not recovered uniformly across the transcriptome. Although different studies identify different enriched RNA

classes, including poly(A)-rich and/or P-body-associated RNAs in some contexts, the broader conclusion is that HsORF1p RNA binding in cells is selective rather than indiscriminate[17-20].

.

Background – Choice of Nucleic Acid Substrates Used in This Study

Although efforts to obtain a high resolution structure of an HsORF1p RNP using approaches such as cryo-EM are ongoing, limitations in sample homogeneity and the particle size continue to present challenges. In the absence of high resolution HsORF1p:RNA structural models, being able to measure biophysical parameters of the HsORF1p:RNA interactions can still provide insightful information of how these completely NA binding domains interact with different substrates[3, 4, 6, 12, 13]. This chapter directly tests whether different single stranded nucleic acid (**ssNA**) substrates measurably alter WT HsORF1p trimer thermal stability and binding affinity. These experiments determine whether substrate composition contributes to HsORF1p association, identify potential substrate biases, inform current models of L1 RNP formation, and support the rational design of RNP constructs that may be optimal or especially informative for structural analysis.

In the absence of structural information to contextualize the biophysical data, interpretation was simplified by restricting the NA ligand set used in this study via two ways. First, homopolymeric ligands were included to isolate the contribution of individual nitrogenous bases. Second, ligands were selected or designed to minimize secondary structure; accordingly, poly(G) substrates were avoided, and the mixed-nucleobase ligands (50% GC ssDNA and RNA) were designed to reduce predicted secondary structure. Lastly, two ligand lengths were used to assess the effect of increasing NA binding surface. The 27-nt ligand was selected because it is approximately the shortest NA length reported to support HsORF1p binding[4, 6, 7], whereas the 36-

nt ligand was included as a longer substrate to test whether additional available binding surface altered HsORF1p binding or stabilization. All the NA ligands used in this study are summarized in Supplementary Table 3.

Background – DSF and Spectral Shift

We relied on two techniques to generate two separate biophysical measurements throughout this thesis. First, thermostabilities of different unbound trimer and nucleoprotein particle (**NP**) complexes were measured and expressed as T_m, the midpoint temperature of the thermal transition detected by the assay that is often interpreted as the temperature at which half of the population of particles are unfolded or destabilized. T_m values were measured using differential scanning fluorimetry (**DSF**). Second, the binding affinity between HsORF1p trimers and NAs were measured using **spectral shift** technology and are expressed as K_d, the equilibrium concentration of trimer required to occupy half of the available NA binding sites.

DSF measurements were performed using a real-time qPCR thermocycler and SYPRO Orange, a hydrophobic fluorescent dye used to monitor protein unfolding. SYPRO Orange fluorescence increases upon binding to hydrophobic regions that become exposed during thermal denaturation. Samples were subjected to a steadily increasing temperature ramp while fluorescence intensity was monitored. The data can then be plotted as fluorescence intensity over temperature, and the resulting curves were used to determine apparent T_m values. Lastly, the difference in apparent T_m between two samples or experimental conditions, such as unbound HsORF1p versus ligand bound HsORF1p or HsORF1p bound to different ligands, can be measured and described as a **thermal shift**, or **ΔT_m**.

When interpreting the results of a DSF assay data is typically expressed as ΔT_m instead of ΔG, because the assay does not capture same thermodynamic information for ΔG calculations as

is available in calorimetry[21]. In DSF, a thermal shift of around 2°C is modest and should be interpreted cautiously unless the results are highly reproducible across replicates[22, 23]. Around a 5°C thermal shift or higher, there is stronger evidence for a robust change for a particle's stability, in this case the HsORF1p trimers that are being monitored[22].

NA binding affinity was measured using a MonolithX platform with 5' Cy5-labeled ssNAs. In these experiments, the labeled ssNA served as the tracer, with its fluorescence signal monitored as the readout of binding. Tracer concentration was held constant across a two-fold dilution series of HsORF1p trimer, ideally at a concentration ~5- to 10-fold below the apparent K_d to ensure measurement accuracy[24]. Binding of the trimer alters the local chemical environment of the conjugated fluorophore, resulting in a detectable shift in its emission spectrum[25]. Samples from the titration series were loaded into quartz capillaries, and binding curves were generated from **spectral shift** measurements by monitoring fluorescence emission at 650 nm and 670 nm and expressing the signal as a ratio (650 nm/670 nm). This response was plotted as a function of protein concentration, and the data were fit to a binding model to determine the apparent K_d as the midpoint of the binding curve.

For generating binding affinity data from the MonolithX, preliminary experiments indicated that activation of the infrared laser required microscale thermophoresis (**MST**) measurements induced aggregation or denaturation of HsORF1p. MST is distinct from spectral shift detection because it measures changes in the movement of fluorescently labeled molecules through a temperature gradient, whereas spectral shift measures binding dependent changes in fluorophore emission. In our hands, infrared laser activation compromised binding curve quality and prevented reliable reuse of capillaries for replicate measurements. Although correlation of the MST data with the spectral shift data would be useful, the complications described above

precluded reliable use of the MST function to determine binding affinity, and all measurements were performed using spectral shift detection alone.

Electrostatic interactions are central to HsORF1p:NA binding, and many of the experiments in this thesis are performed at high NaCl concentrations. High NaCl conditions may influence which molecular contacts appear most important, because salt can screen interactions between the negatively charged phosphate backbone and positively charged protein surfaces by altering the surrounding ion atmosphere. Yu and Iwahara emphasize that these ion atmospheres are central to the electrostatics of NAs, proteins, and their complexes, and that changes in the ion atmosphere can strongly shape how these molecules interact[25]. As a result, high NaCl may promote electrostatic shielding, reduce the apparent contribution of phosphate-basic surface contacts and shift the observed binding behavior toward less salt-sensitive interactions, such as those involving nucleobases, stacking surfaces, sugar contacts, or the RNA 2' hydroxyl group. Because the high NaCl conditions can affect data interpretation, some experiments were done at varying concentration of NaCl conditions.

Materials and Methods

Plasmid Cloning

10xHis-SUMO, N-terminally tagged HsORF1p constructs were generated by overlap-extension PCR and Gibson-type assembly using the pE-SUMOpro Kan backbone (LifeSensors, pE1001K). The full-length HsORF1p(1-338) plasmid used as the cloning template was provided by the Marty Taylor laboratory. Mutant constructs were designed to truncate the HsORF1p from M1-M388 to S106-N326. Four primers, two outside and two overlapping but opposite orientation, mutation harboring internal primers were used to produce two PCR fragments. The pE-SUMOpro Kan backbone was amplified separately with primers to produce a linearized vector with complementary overlaps to the ligation product. All fragments were made using Q5 High-Fidelity

DNA Polymerase (NEB). This produced three PCR products with: the first HsORF1p fragment, the second HsORF1p fragment and a linearized vectore backbone. The PCR fragments had 15-35 nt overlapping regions for assembly with KLD enzyme mix (NEB), before transformation into *E. coli* DH5α (Invitrogen) under kanamycin selection. Colony PCR and whole-plasmid sequencing (Plasmidsaurus) were used to confirm plasmid sequence and quality.

HsORF1p Expression and Purification

Rosetta (DE3) *E. coli* were transformed and used for protein expression. 50 mL of Terrific Broth (TB) (RPI) was inoculated with either a freshly transformed colony or glycerol stock and grown for approximately 16 hours at 30°C. 10-15mL of this starter culture was used to seed 1 L TB to an initia OD_{600} of 0.1. 1 L cultures were grown at at 37°C until reaching an OD_{600} of 2-3 before being cooled to 4°C. The cultures were treated with 200 uM IPTG to start protein induction and the flasks were returned to incubators to shake for 16-20 hrs at 16°C. Cultures were centrifuged at 6,000 xg at 4°C for 15 min, cell pellets collected into 50 mL conical tubes, flash frozen in liqud nitrogen and stored at -80°C.

To purify the proteins, the cell pellets were allowed to thaw slightly and were resuspended in ice cold lysis buffer (20 mM Tris pH 8.0, 2 M NaCl, 20 mM imidazole), 10 mL for every 4 or 5 g. A mixture of 50 mM PMSF and 100 mM benzamidine were added, 1 mL of mixture to 100 mL of lysis buffer to prevent serine protease activity, and the cells were lysed using a microfluidizer. The lysate was then clarified by centrifugation at 25,000 rpm for 30 minutes at 4 °C using a Ti-45 rotor. This clarified lysate was filtered through a 0.45 um PES membrane and incubated with Ni-NTA resin for up to 1 hr. After washing with lysis buffer, recombinant protein was eluted in 20 mM Tris pH 8.0, 1 M NaCl, and 500 mM imidazole, concentrated to ~5 mL using a 30 MWCO kDa centrifugal concentrator, supplemented with TCEP to 1 mM, and digested

overnight with ULP1 in 10 kDa MWCO dialysis tubing against 20 mM Tris pH 8.0, 480 mM NaCl, and 1 mM TCEP. The digested sample was filtered through a 0.22 µm system and purified by heparin affinity chromatography on two 5 mL HP Heparin columns using a gradient of 30-100% Buffer B (Buffer A: 20 mM HEPES pH 8.0; Buffer B: 20 mM HEPES pH 8.0, 2 M NaCl). Protein-containing fractions were concentrated to <1 mL, buffer-exchanged into storage buffer (10 mM HEPES pH 8.0, 500 mM NaCl) by 2 iterations of dilution to 15 mL and reconcentration to <1 mL, the last centrifugation was continued until the nanodrop-estimated protein concentration exceeded 40 mg/mL. Final preparations were aliquoted (5-20 µL), flash frozen in liquid nitrogen, and stored at -80 °C. Protein concentration was assessed by BCA assay, and purity was evaluated by SDS-PAGE with Coomassie staining (Supplemental Figure S1).

Differential Scanning Fluorimetry (DSF) Assay

DSF assays were performed on a Bio-Rad real-time PCR instrument. Plate layouts were designed in advance to define the protein, oligonucleotide, and buffer combinations for each condition. Reactions were prepared for a final buffer composition of 10 mM HEPES pH 8.0 and 300 mM NaCl. Protein stocks, oligonucleotides, and buffer diluents were thawed on ice and pre-diluted to working concentrations. To minimize aggregation in select HsORF1p variants, mutants were first adjusted to 12 mg/mL in 10 mM HEPES pH 8.0 and 500 mM NaCl, then gradually diluted to 2 mg/mL in 10 mM HEPES pH 8.0 and 300 mM NaCl using intermediate NaCl concentrations (375, 300, and 200 mM). Oligonucleotide working solutions were prepared in assay buffer at a 1.2x molar excess relative to the protein mutants. For plate assembly, 9.5 uL of oligonucleotide working solutions or assay buffer were added first, followed by 9.5 uL of protein working solutions. 1 uL of SYPRO Orange dye, thawed from frozen aliquots and diluted to 200x, was added to yield 20 uL final reaction volumes. Plates were sealed with PCR film, briefly mixed

on an IKA MS 3 vortexer, and s Sorvall ST 16R centrifuge was used at 2220 x g for 5 min at 20 °C to collect all liquid at the bottom of the wells and remove bubbles. Each plate was incubated in the dark at RT for approximately 1 hr. DSF measurements were acquired using a melt-curve program consisting of a 5 min hold at 20 °C followed by a 0.5 °C/min ramp to 90 °C. Fluorescence data were exported and analyzed using DSFworld to estimate apparent melting temperatures (Tm) by sigmoidal fitting, and results were plotted in GraphPad Prism 11.

Binding Affinity Measurements

To assess interactions between purified HsORF1p constructs and Cy5-labeled RNA, fluorescence-based binding affinity assays were performed at the Brown University BioMed Proteomics Facility using the spectral shift mode of the Monolith X (Nanotemper). Protein samples were diluted in assay buffer (10 mM HEPES pH 8.0, 500 mM NaCl, 0.1% PEG-8000, 0.05% Tween-20) to generate four independent experimental replicates at twice the highest protein concentration used in their titration series. Assay buffer was also used to dilute Cy5-labeled RNA stocks to a 40 pM working solution.

For plate setup, 10 µL of assay buffer were dispensed into each well of a 96-well plate except for the first titration-point well, which was left empty. Twenty microliters of each protein working solution were added to the appropriate first titration-point well, and 2-fold serial dilutions were generated across the plate by transferring 10 µL of each mixture into the pre-aliquoted 10 µL of assay buffer. Then, 10 µL of Cy5-RNA were added to all protein-containing wells and to a buffer-only negative control. Plates were sealed, briefly mixed on an IKA MS 3 vortexer, and centrifuged in a Sorvall ST 16R at 2220 x g for 5 minutes at 20 °C to collect liquid and remove bubbles.

After a 1-hour incubation in the dark, five technical replicate measurements were acquired from one experimental replicate using premium capillaries; at 4 hours, five technical replicates were collected from all four experimental replicates. Fluorescence at 650/670 nm was exported and plotted against protein concentration. Technical outliers (1-2 per well) and, when necessary, a full experimental replicate were removed prior to analysis. Signals were normalized by averaging the two lowest values (including the RNA-only control) and the two highest plateau values, followed by min-max normalization ($Y_{norm} = (Y - Y_{min})/(Y_{max} - Y_{min})$). Normalized data were imported into GraphPad Prism 11, where replicate means and standard deviations were calculated, and apparent dissociation constants (Kd) were determined by nonlinear regression using a one-site specific binding model ($Y = B_{max} \cdot X/(K_d + X)$).

Results

Nucleic Acid Polymer Composition Influences HsORF1p-RNP Stability

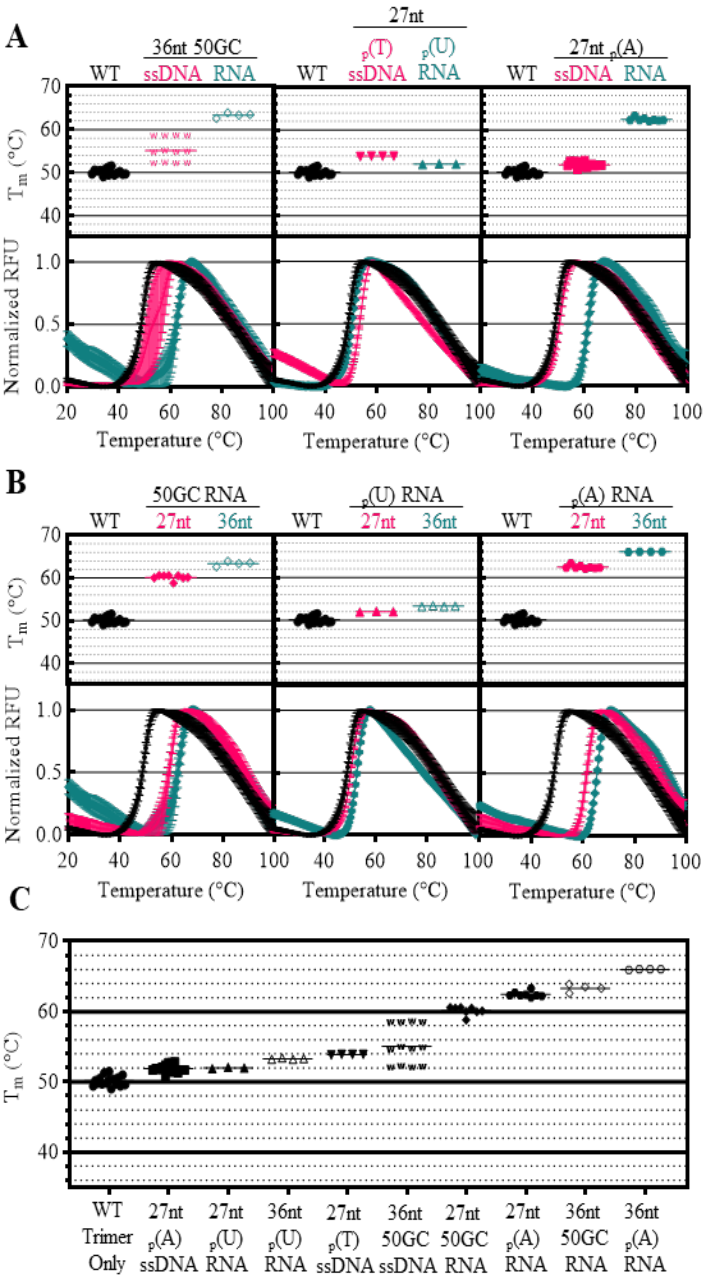


Figure 2.1 *| Differential Scanning Fluorimetry (DSF) of truncated, trimeric HsORF1p(106-326) either unbound or in complex with short nucleic acid substrates.* ***A****. Normalized melting curves (bottom) and dot-plot of derived* T_m*s (top) comparing length matched analog ssDNA (red) and RNA (green).* ***B****. Equivalent diagram set up as in A., now comparing 27-nt and 36-nt versions the similarly composed RNA oligos.* ***C****. A dot plot summary of* T_m *values across the full nucleic acid panel tested with WT HsORF1p trimer. Every horizontal line represents the average Tm of each set of independently measured data values.*

We first tested the hypothesis that nucleoprotein particles (**NPs**) that are comprised of a WT HsORF1p(S106-N326) trimer, referred to as **'trimer'** throughout this chapter, and a short oligonucleotide would exhibit different thermodynamic properties depending on the composition of the oligonucleotide sequence used. To assay bulk thermostability, Differential Scanning Fluorimetry (DSF) was used to measure apparent melting temperatures of different NPs and found detectable differences in thermodynamic coupling between NPs, when purified trimer was mixed with different oligonucleotide sequences.

We observed that, in all three cases tested, trimer experiences less thermodynamic stabilization when ssDNA was used to form single-trimer NPs compared to when length-matched,

analog sequence RNA was used (Figure 2.1 A). For example, a 36-nt 50%-GC ssDNA increased the T_m of 'HsORF1p' from 49°C to 58°C, whereas the analog 36-nt 50%-GC RNA raised the T_m to 63.5°C; this represents a 5.5°C difference in T_m between the two NPs, which is generally considered above the threshold of showing strong evidence of increased particle stability. This trend was most dramatic when 27-nt poly(A) ssDNA was used versus a 27-nt poly(A) RNA, as the ssDNA shifted the T_m by a couple of degrees to 51°C when compared to the unbound trimer, while the poly(A) RNA shifted the T_m to 65°C, that is a 14°C difference in NP thermostability in this case. In the final case studied, 27-nt poly(T) ssDNA versus 27-nt poly(U) RNA was tested and it was found that both stabilized trimer to a similar degree, but the RNA stabilized a negligible 1°C more, 54°C when the poly(T) ssDNA was used versus ~55°C from the poly(U) RNA.

The DSF data also showed, in general, that single-trimer RNPs are more thermodynamically stable when 36-nt RNA is used in place of 27-nt RNA (Figure 2.1 B). When using 50%-GC RNA, increasing the polymer's length by 9-nt caused a ΔT_m of ~4°C. Similarly, comparing 36-nt poly(A) RNA RNPs to 27-nt poly(A) RNA RNPs resulted in a ΔTm of ~4°C. However, trimers bound to 36-nt poly(U) RNA had a negligible, ~1.5-2°C higher T_m, than those bound to 27-nt poly(U) RNA which is consistent with the trend, but very low compared to the other example substrates that had longer versions compared to shorter versions used in this assay.

Lastly, it was observed that RNA composition strongly influenced the thermostability of single-trimer RNPs, even when the length of the different RNAs were held constant (Figure 2.1 C). Among the 36-nt RNAs tested, poly(U) produced the least stabilization, as it increased the T_m of the unbound trimer from 49°C to ~55°C. In contrast, a 36-nt 50%-GC RNA yielded a T_m of 63.5°C, and the 36-nt poly(A) increased the T_m further to 66.5°C. Notably, the ΔT_m associated

with 36-nt poly(U) bound RNPs relative to the unbound trimer is comparable to the ΔT_m observed between RNPs formed with 36-nt poly(A) and our 36-nt randomized 50%-GC sequence.

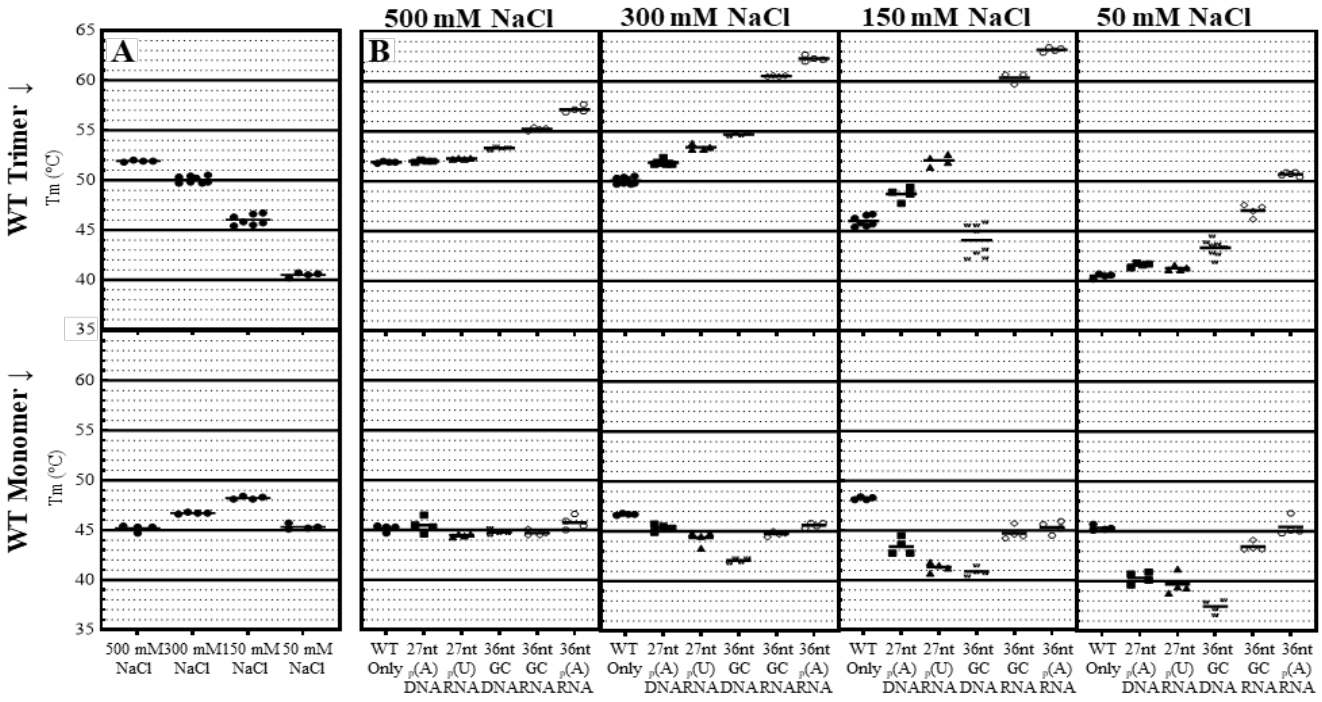


Figure 2.2 | *Differential Scanning Fluorimetry (DSF) data of truncated, trimeric HsORF1p(106-326) and truncated, monomeric HsORF1p(153_338) across different NaCl concentration and nucleic acids.* ***A.*** *Dot plot of Tm from unbound trimeric HsORF1p (top) and unbound monomeric HsORF1p (bottom).* ***B.*** *Series of dot plots of DSF data collected at different salt concentrations, 500 mM, 300 mM, 150 mM and 50 mM NaCl (from left to right), and the presence of different nucleic acid substrates within each salt condition with trimeric (top) and monomeric (bottom) HsORF1p. Every horizontal line represents the average T_m of each set of independently measured data values.*

Both the DSF and spectral shift data described in the next section were collected at relatively high salt, either 300 mM or 510 mM NaCl, respectively. These conditions were necessary because lowering NaCl below 300 mM led to aggregation in some HsORF1p constructs (introduced in chapter 3), and because the detection sensitivity of the Monolith X limited the labeled RNA concentration to ~4 nM before the signal-to-noise ratio became intolerable (data not shown). Therefore, DSF measurements were also performed with WT HsORF1p across different salt conditions to determine whether the relative ligand-dependent differences in thermal stability were preserved when NaCl was lowered. These experiments also tested whether the apparent

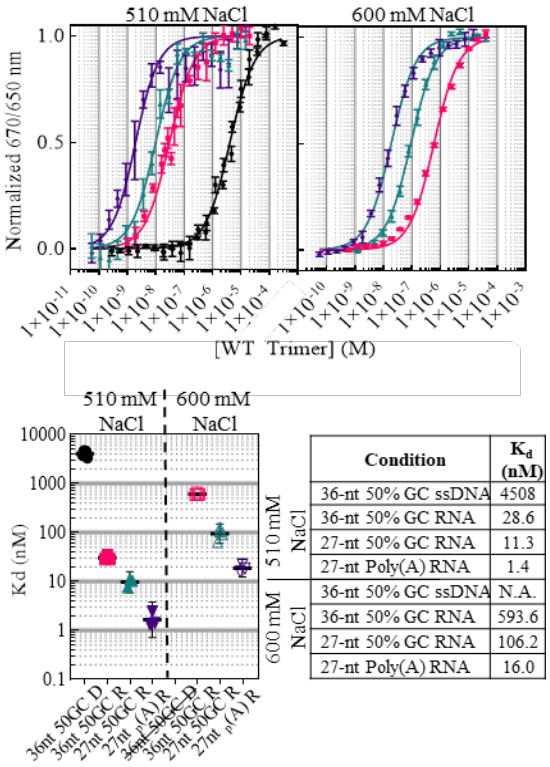
preference for adenosine containing RNAs was unique to the trimer or was also present in a monomeric context, in which a HsORF1p construct was used that only contained the RMM-CTD.

To test how NaCl concentration affected trimer DSF assay results and to compared these effects with those observed for monomeric HsORF1p (Figure 2.2), referred to as "**monomer**" for the remainder of this chapter. The first observation was that NaCl concentration affected the unbound monomer and trimer differently (Figure 2.2A). Increasing NaCl concentration from 50 to 500 mM steadily increased the T_m of the trimer, whereas rhe monomer experiences a T_m increase from 50 to 150 mM and subsequent NaCl concentration increases from150 mM destabilized the monomer. It is also noteworthy that from 50 to 150 mM NaCl, which includes salt concentrations often used to approximate physiological ionic strength in vitro, the monomer was more thermostable than the trimer in this assay.

When a selected panel of ssNAs was introduced into the assay, reproducible but clearly distinct patterns in the trimeric and monomeric contexts are observable (Figure 2.2B). In the trimeric context, the trend toward lower T_ms at lower salt concentration was preserved in the RNP state and was broadly comparable to the unbound trimer. A second trend was that thermostability increased progressively from when ssDNA and poly(U) RNA substrates were used and a larger upward shift occurs when 50% GC RNA and especially poly(A) RNA are used. This pattern was maintained across salt concentrations, with one exception: at 150 mM NaCl with 36-nt ssDNA, there was an unexpected drop in Tm. Because insufficient substrate remained to repeat this condition, it cannot be determine whether this represents a true deviation from the otherwise well reproduced pattern or an anomaly associated with the final aliquots of that substrate that were used. Lastly, the difference in magnitude of stabilization conferred by each ssNA when compared to one another increased as salt concentration decreased, although a reversal of that trend occurs at 50

mM NaCl. At 50 mM NaCl all conditions converged toward relatively low T_ms and magnitude T_m differences conferred between substrates were more like those observed at 300-500 mM NaCl than at 150 mM NaCl.

In the monomeric context, there was not a clear preservation between the salt dependent T_m changes of the unbound monomer and the bound monomers, although 50 mM NaCl again produced a strong overall decrease in T_m in both conditions. The monomer also showed an emergent substrate dependent pattern across ionic conditions, but unlike the trimer, the direction of the effect was mixed. ssDNA and poly(U) RNA destabilized the monomer relative to the unbound protein-only control. Similarly, 36-nt 50% GC RNA also destabilized the monomer, although to a lesser extent than the other substrates, whereas poly(A) RNA either produced a comparable mild destabilization or had a negligible stabilizing effect. As in the trimeric context, the magnitude of the substrate-dependent effect increased as salt concentration was lowered; however, in the monomer this increase continued progressively down to 50 mM NaCl.



Condition	K_d (nM)
36-nt 50% GC ssDNA	4508
36-nt 50% GC RNA	28.6
27-nt 50% GC RNA	11.3
27-nt Poly(A) RNA	1.4
36-nt 50% GC ssDNA	N.A.
36-nt 50% GC RNA	593.6
27-nt 50% GC RNA	106.2
27-nt Poly(A) RNA	16.0

Figure 2.3 | *Spectral shift binding affinity data of truncated, trimeric HsORF1p(106-326) binding to various 5'end, Cy5-labeled short nucleic acid substrates at 510 and 600 mM NaCl. Substrates tested were 36-nt 50% GC ssDNA (black), 36-nt 50% GC RNA (red), 27-nt 50% GC RNA (green), 27-nt poly(A) RNA (blue). Above are normalized binding affinity curves and below is a dot plot and table summarizing the K_d values of each condition.*

Nucleic Acid Polymer Composition Governs HsORF1p's Binding Affinity

After it was observed that NPs had dramatically different thermostabilities dependent on the composition of the NA used, to confirm if this observation could be explained in part by the trimer having different relative binding affinities for

the differently composed NA substrates. To test this, the spectral shift feature of the MonolithX was used to measure relative binding affinities of a subset of the NA panel used in the DSF assay: 27-nt 50% GC RNA, 27-nt poly(A) RNA, a 36-nt 50% GC RNA and a complementary 36-nt 50% GC ssDNA. To measure spectral shift these oligos were ordered with 5'-end cy5 label and are referred to as '**tracers**' in the rest of this chapter. This restricted panel was designed to assess if the trimer was binding poly(A) with higher affinity than the randomized RNA sequence, to confirm that the length-dependent effects on thermostability is coming from increased binding affinity, and as a validation-control if ssDNA versus RNA are binding with different apparent affinities as previously reported[8, 10].

At 450 mM NaCl and below, it was found that the apparent dissociation constant (K_d) of the trimer for RNA tracer was too close to the lowest tracer concentration that could be reliably measured by MonolithX; K_d values are only accurate if they are substantially lower than the concentration of tracer used (at least 5-fold, and ideally ≥10-fold, lower)[24, 26]. For this assay tracer concentration was limited to ~1 nM and the overall fluorescent signal and signal-to-noise ratio measured at the flat ends of the sigmoidal binding curve were not sufficient. At 450 mM NaCl and below the K_d remained as 1 nM despite lowering NaCl concentration (data not shown), which is inconsistent with what is known in literature.

We measured an 12.5-fold increase in binding affinity between the trimer and RNA tracer, when the poly(A) sequence was used compared to 50%-GC at high NaCl concentrations (Figure 2.2). At 510 mM NaCl, the average K_d of the trimer binding to 27-nt poly(A) RNA was 1.4 nM, suggesting a true binding affinity ≤ ~1 nM. In contrast the average K_ds of the trimer binding to 27-nt 50% GC RNA, 36-nt 50% GC RNA and 36-nt 50% ssDNA were 11.3 nM, 28.6 nM and 4508 nM, respectively, which are all give acceptable ratios of K_d to tracer concentration. However,

to improve the signal-to-noise ratios and enable more accurate relative K_d values for poly(A), measurements were also performed at 600 mM NaCl using 4 nM of RNA tracer. The preference for poly(A) RNA persisted at the 20% higher NaCl concentration, as the average K_ds for the 27-nt poly(A), 27-nt 50% GC RNA, and 36-nt 50% GC RNA oligos were 16 nM, 106.2 nM, 593.6 nM, respectively. The K_d was undeterminable for 36-nt 50% ssDNA because sufficient concentrations of trimer were unable to be produced to measure a binding affinity for that substrate at 600 mM NaCl.

Summary and Discussion

The observations within this chapter suggest that HsORF1p has biophysical properties that align with preferential binding to its own mRNA over other NAs. Specifically, poly(A) rich sequences, or potentially RNAs with adenosines arranged with a particular spacing, appear to promote stronger binding affinity and increased thermostability of the NPs in the context of HsORF1p. In agreement with literature, HsORF1p exhibits significantly higher binding affinity for 50% GC RNA when compared to ssDNA of analog sequence. Interestingly, poly(A) ssDNA and poly(A) RNA produced dramatically different thermal shifts in the DSF assays, the RNA homopolymer stabilizing the NP maximally while the ssDNA minimally. Additionally, the poly(U) RNA performed only marginally better than poly(A) ssDNA, meaning that HsORF1p is either has an adenosine preference or perhaps a higher avidity interaction with purine base containing RNAs in general. This is worth noting because L1 mRNA, as well as other RNAs identified to be associated with HsORF1p RNPs, determined through immunoprecipitation and sequencing, are generally AU rich (adenosine enriched specifically in the case of L1 mRNA), and/or contain a poly(A) tail [17-20, 27, 28]. This data could suggest a biophysical mechanism that helps to promote the *cis* preferential nature of HsORF1p.

However, without a structural model of HsORF1p bound to these ligands, it is not possible to distinguish whether the observed differences in stabilization and binding affinity between ssDNA and RNA arise from backbone geometry, the presence of the 2' OH, differential binding pocket orientations or preferences, avidity, or some combination of these factors. For example, the 2' OH could influence sugar puckering preferences between ssDNA and RNA, making ssDNA a poorer fit within the same binding pathway. Alternatively, 2' OH groups along the RNA polymer could serve as interaction sites that increase avidity within that same pathway and affect nucleobase orientation. Lastly, ssDNA and RNA may behave as partially distinct ligands due to differences in geometry, making them more likely to follow different binding paths across the trimer's binding domains.

The difference in ΔT_m between poly(U) and poly(A) RNA is also interesting because these are comparably flexible polymers that both contain a 2' OH. This suggests that the nitrogenous bases themselves also play an important role in stabilizing the NP, but only in the context of RNA, since poly(A) ssDNA is a poorly stabilizing ligand. With the assumption that ssDNA and RNA binding are initiated and maintained by the same residues and with the same orientations, it might appear that phosphate interactions are the primary source of binding affinity, while nitrogenous base identity, rather than direct 2' OH contacts, contributes more strongly to NP formation in the context of 1:1 binding. However, a more conservative interpretation is that nucleobase composition does contribute significantly to the NP interaction, but it cannot be confidently concluded if this is due to differences in RNA geometry, avidity, or potentially differing pathways along the binding surface without experimentally derived structures of HsORF1p bound to these different ligands.

The study of HsORF1p NP stability at different NaCl concentrations was important for showing that the thermostability differences between ligands were not an artifact of conducting measurements at high ionic strength. As explained in the background, RNA binding proteins are sensitive to ion concentration, and increasing NaCl can increase ionic shielding of the charged interactions that are important in the HsORF1p:NA interaction. Because the relative pattern of T_m shifts is preserved between ligands, these data suggest that negatively charged phosphate interactions are not the only factor driving NP stabilization. Instead, other factors, such as nitrogenous base identity, backbone geometry, and potentially 2' OH contacts, may also contribute to the observed stabilization. This suggests that this HsORF1p preferential binding could also be occurring in more biologically relevant ionic conditions especially at lower HsORF1p:RNA concentrations like in early L1 RNP formation.

Extending the salt titration DSF assay to the monomeric HsORF1p RRM-CTD yielded two main observations. First, NaCl concentration and NA binding both affected the apparent thermal stability of HsORF1p, but in a construct-dependent manner. In the trimer, increasing NaCl concentration resembled NA in that it increases stabilization of HsORF1p, whereas in the monomeric RRM-CTD, increasing either NaCl or NA concentration lowered the apparent Tm. Although these results are difficult to extrapolate beyond the DSF assay, they suggest that monovalent ionic conditions and the presence of NAs both influence HsORF1p stability. The NaCl-dependent stabilization could occur through pressuring coiled-coil formation in the trimeric construct, or through shielding and/or ionic interactions that may mimic aspects of NA binding. However, the DSF assay does not provide a mechanistic rationale for why these stability trends are the way they are.

Second, poly(A) RNA produced the highest ligand-bound T_m in both contexts. This suggests that the stabilizing contacts available to adenosine rich RNA may be at least partly retained in the monomeric RRM-CTD, which contains much of the primary binding cleft. However, because NA addition still lowers the apparent T_m of the isolated RRM-CTD, these contacts do not appear sufficient to stabilize the monomeric construct in the same way they stabilize the full trimer.

Overall, the data in this chapter suggest that HsORF1p:NA association is not driven solely by nonspecific electrostatic interactions with the phosphate backbone but is instead influenced by NA ligand sequence composition. Across both binding and thermostability assays, RNA was generally favored over ssDNA, and adenosine rich RNA produced the strongest stabilizing effects. These observations are consistent with a model in which features enriched in L1 mRNA, including adenosine content, AU richness, and the poly(A) tail, may contribute to the ability of HsORF1p to preferentially associate with its own transcript during RNP formation. At the same time, the mechanism underlying this preference cannot be assigned from this data alone. Differences between ssDNA and RNA, and between poly(A) and poly(U) RNA, could reflect changes in backbone geometry, 2' hydroxyl group mediated interactions, nucleobase-specific contacts, ligand avidity, or different binding paths across the trimer surface. The preservation of ligand-dependent thermostability trends across salt conditions supports the idea that these effects are not simply artifacts of high ionic strength, and that non-phosphate contacts contribute measurably to NP stabilization. However, the distinct behavior of the isolated RRM-CTD also suggests that the full trimeric architecture is important for converting these contacts into net stabilization. Together, these findings support a model in which HsORF1p retains broadly permissive ssNA binding

activity, while still displaying biophysical preferences that may help bias early L1 RNP assembly toward adenosine rich RNA substrates, including L1 mRNA during *cis* preference.

References

1. Adney, E.M., et al., *Comprehensive Scanning Mutagenesis of Human Retrotransposon LINE-1 Identifies Motifs Essential for Function*. Genetics, 2019. **213**(4): p. 1401–1414.
2. Martin, S.L., et al., *A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity*. Nucleic Acids Res, 2008. **36**(18): p. 5845–54.
3. Januszyk, K., et al., *Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1*. J Biol Chem, 2007. **282**(34): p. 24893–904.
4. Khazina, E., et al., *Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition*. Nat Struct Mol Biol, 2011. **18**(9): p. 1006–14.
5. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells*. Cell, 1996. **87**(5): p. 917–27.
6. Callahan, K.E., et al., *Polymerization and nucleic acid-binding properties of human L1 ORF1 protein*. Nucleic Acids Res, 2012. **40**(2): p. 813–27.
7. Zernia, S., et al., *LINE-1 ribonucleoprotein condensates bind DNA to enable nuclear entry during mitosis*. Sci Adv, 2025. **11**(18): p. eadt9318.

8. Kolosha, V.O. and S.L. Martin, *In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition*. Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10155–60.

9. Kolosha, V.O. and S.L. Martin, *High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1)*. J Biol Chem, 2003. **278**(10): p. 8112–7.

10. Martin, S.L., et al., *LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein*. J Mol Biol, 2005. **348**(3): p. 549–61.

11. Kulpa, D.A. and J.V. Moran, *Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition*. Hum Mol Genet, 2005. **14**(21): p. 3237–48.

12. Naufer, M.N., A.V. Furano, and M.C. Williams, *Protein-nucleic acid interactions of LINE-1 ORF1p*. Semin Cell Dev Biol, 2019. **86**: p. 140–149.

13. Cashen, B.A., et al., *L1-ORF1p nucleoprotein can rapidly assume distinct conformations and simultaneously bind more than one nucleic acid*. Nucleic Acids Res, 2024. **52**(22): p. 14013–14029.

14. Martin, S.L., *Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1*. RNA Biol, 2010. **7**(6): p. 706–11.

15. Hohjoh, H. and M.F. Singer, *Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon*. Embo j, 1997. **16**(19): p. 6034–43.

16. Khazina, E. and O. Weichenrieder, *Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame*. Proc Natl Acad Sci U S A, 2009. **106**(3): p. 731–6.

17. Briggs, E.M., et al., *RIP-seq reveals LINE-1 ORF1p association with p-body enriched mRNAs*. Mob DNA, 2021. **12**(1): p. 5.

18. Šulc, P., et al., *Repeats mimic pathogen-associated patterns across a vast evolutionary landscape*. Cell Genom, 2025. **5**(12): p. 101011.

19. Sun, S., et al., *Cancer cells restrict immunogenicity of retrotransposon expression via distinct mechanisms*. Immunity, 2024. **57**(12): p. 2879–2894.e11.

20. Taylor, M.S., et al., *Dissection of affinity captured LINE-1 macromolecular complexes*. Elife, 2018. **7**.

21. Gao, K., R. Oerlemans, and M.R. Groves, *Theory and applications of differential scanning fluorimetry in early-stage drug discovery*. Biophys Rev, 2020. **12**(1): p. 85–104.

22. Rosa, N., et al., *Meltdown: A Tool to Help in the Interpretation of Thermal Melt Curves Acquired by Differential Scanning Fluorimetry*. J Biomol Screen, 2015. **20**(7): p. 898–905.

23. Huynh, K. and C.L. Partch, *Analysis of protein stability and ligand interactions by thermal shift assay*. Curr Protoc Protein Sci, 2015. **79**: p. 28.9.1–28.9.14.

24. Jarmoskaite, I., et al., *How to measure and evaluate binding affinities*. Elife, 2020. **9**.

25. Dos Santos Rodrigues, F.H., et al., *Applications of fluorescence spectroscopy in protein conformational changes and intermolecular contacts*. BBA Adv, 2023. **3**: p. 100091.

26. Tso, S.C., et al., *Using two-site binding models to analyze microscale thermophoresis data*. Anal Biochem, 2018. **540-541**: p. 64–75.

27. Mandal, P.K., et al., *Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles*. Hum Mol Genet, 2013. **22**(18): p. 3730–48.

28. Taylor, M.S., et al., *Characterization of L1-Ribonucleoprotein Particles*. Methods Mol Biol, 2016. **1400**: p. 311–38.

Chapter 3 | The Importance of Positively Charged Residues in HsORF1p Function

Background – Positively Charged Residues and Nucleic Acid Binding

Protein:NA binding typically involves a combination of electrostatic interactions, hydrogen bonding, base stacking, and geometric complementarity between the protein and ligand[1, 2]. RNA and ssDNA share a negatively charged phosphodiester backbone, and proteins that bind these ligands often incorporate positively charged, or basic, residues into their binding surfaces, particularly arginine and lysine[1, 2]. These residues can stabilize contacts with the nucleic acid (**NA**) backbone and may also contribute to sequence or structure specific recognition through hydrogen bonding, cation-π interactions, and contacts with exposed bases or structural features of the ligand. Although RNA binding is rarely determined by charged interactions alone, clusters of basic amino acids provide useful candidates for probing NA binding domains. Understanding the function of the charged residues is especially relevant for HsORF1p, which is thought to interact with a broad range of NAs primarily through general electrostatic interactions and the ability of the ligand to conform to the curvature of the trimeric binding surface[3-7].

A common strategy for probing the function of amino acids along the binding surface of a protein is site-directed mutagenesis[8, 9]. To test whether amino acids with positively charged R-groups contribute to NA binding, these residues are often substituted with amino acids that either neutralize the charge or introduce a negative charge at that position. Mutant constructs can then be tested and compared to one another and to the wild-type, unmutated construct to determine whether altering specific residues disrupts NA binding. If a mutation reduces binding, this suggests that the targeted residue or region contributes directly or indirectly to NA interaction. Conversely, if a mutation does not alter NA binding, it may indicate that the targeted region is dispensable for

binding, functionally redundant with nearby residues, or involved in another aspect of protein function. Furthermore, because positively charged residues can contribute to both general electrostatic interactions and more specific recognition of NA sequence or structure, differences in mutant binding across ligands can help identify residues that broadly support NA binding versus those that contribute to ligand specific recognition.

This framework is particularly relevant for HsORF1p, as it is essential for L1 mobilization, in part because of its ability to bind the L1 mRNA that encodes it. Although it is generally known that the RRM and CTD contribute to HsORF1p function by influencing RNA binding, a clear residue level understanding of what gives HsORF1p its *in vitro* NA binding properties, how those properties relate to in cellulo observations, and vice versa is not well understood[7, 10-13]. For example, the relative 3D positioning and composition of the RRM and CTD are important for proper RNP formation, and the positively charged residues concentrated within these domains are likely major contributors to ssNA interaction[5, 11, 13, 14]. However, it remains unclear whether the positively charged, solvent exposed regions across these domains contribute equally to the HsORF1p:NA binding interaction, or whether some of these residues support other aspects of protein function, such as structural stability through salt bridges or the hydrophilic character of the solvent exposed protein surface[3, 5, 13-15]. In addition, there is limited in cellulo data linking targeted RRM and CTD surface charge neutralization to phenotypes such as ORF1p foci formation and retrotransposition efficiency[13, 15-18].

Based on this, one can reason that alanine substitutions within these charged patches, which replace positively charged side chains with small, neutral R-groups, would provide a useful way to test whether different basic surface regions contribute unequally to HsORF1p biophysical behavior during ssNA binding, and whether those effects correspond to phenotypes observed *in*

cellulo. During the planning of this research, it was hypothesized that some regions would contribute primarily to general ssNA binding, whereas others may influence ligand preference (i.e., diminish poly(A) RNA preference), binding dynamics, protein stability, or other aspects of HsORF1p function. It was expected that neutralizing distinct charged patches would produce graded effects on ssNA binding affinity rather than a uniform loss of binding across all mutants. In turn, comparing these *in vitro* binding effects with L1 retrotransposition efficiency and HsORF1p foci formation would help distinguish mutations that impair HsORF1p function through reduced ssNA binding from those that affect other aspects of the L1 mobilization pathway.

Background – Choice of Mutant HsORF1p Constructs

Available HsORF1p structural models and primary sequence data provided a framework for designing mutants to test both *in vitro* and *in cellulo*[5, 14, 18]. These structures identify basic RRM and CTD residues that cluster in 3D space to form a continuous, positively charged surface rather than a single well defined binding groove[5, 14]. To better define functional regions within this surface, mutants were designed around three potential "**binding clefts**" (Figure 3.1). Binding cleft 1 appears between the RRM and CTD within a single monomeric unit, binding cleft 2 occurs between the RRM and CTD of adjacent monomers, and binding cleft 3 is postulated to occur between the base of the coiled-coil domain and the top of the CTD. Binding cleft 3 is considered putative because charge neutralizing mutations at the base of the coiled-coil domain were previously shown to affect HsORF1p binding to both RNA and ssDNA, as determined by analytical size-exclusion chromatography[5]. This cleft based framework allowed us to ask whether different regions of the positively charged surface make distinct contributions to HsORF1p:NA binding behavior and cellular function.

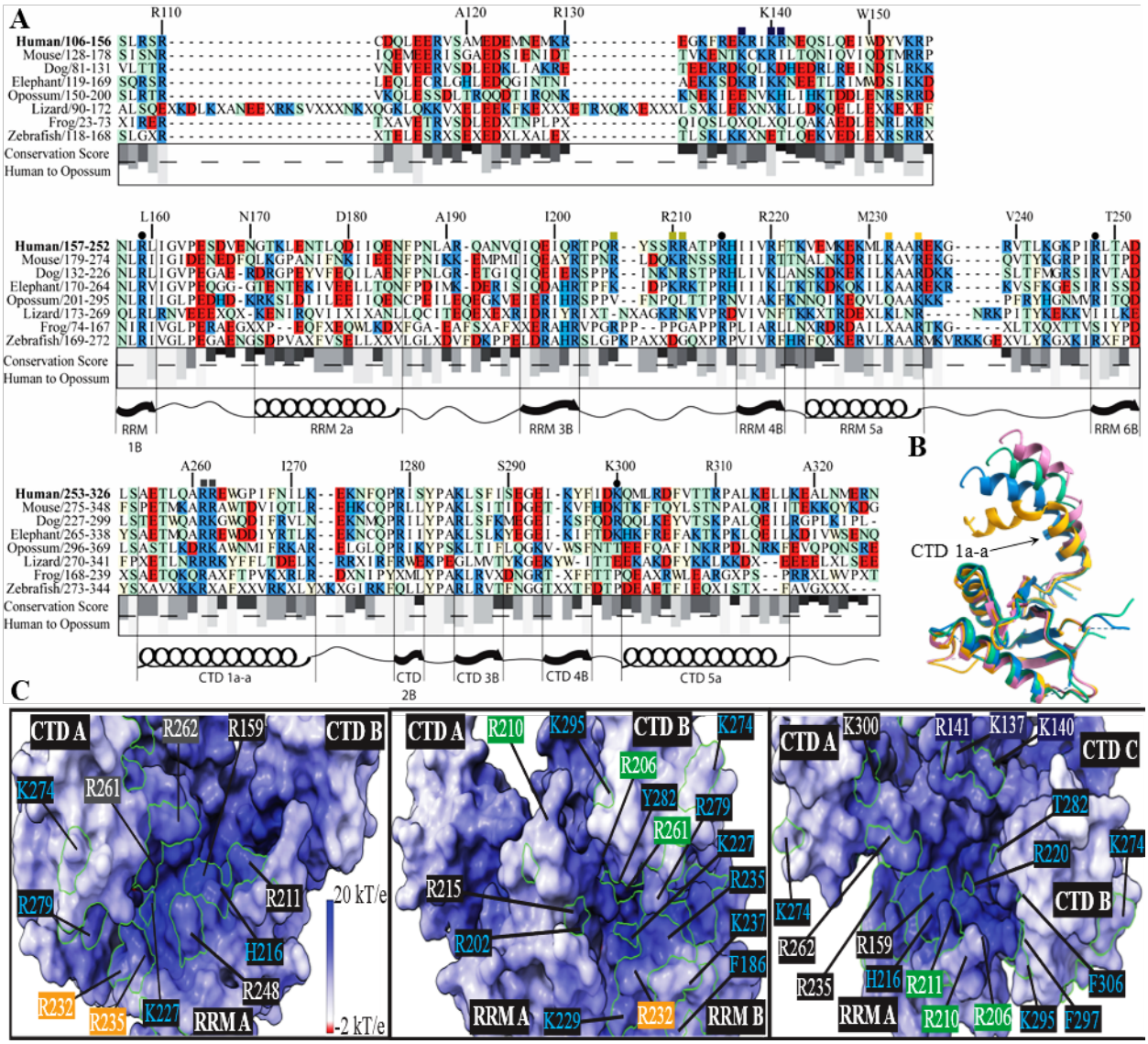


Figure 3.1 | *Sequence conservation, domain schematic and dynamics and candidate amino acid binding surfaces.* ***A****. Primary sequence alignment of ORF1p consensus sequences from multiple species, compiled from Boissinot & Sookdeo 2016. Amino acids are colored by functional group class. Conservation scores and secondary structure are mapped below the alignments. Above the alignments, symbols mark the residues examined in this study, circles are single point mutations, multi-residue mutants are marked by squares, green arrows mark residues known to impair nucleic acid binding and orange arrows are residues known to disrupt stability.* ***B****. An overlay of four selected RRM-CTD units from the three published crystal structures of trimeric HsORF1p (Khazina 2011) to show relative flexibility of RRM-CTD. Monomeric units were manipulated to only display RRM-CTD and aligned by the RRM. Flexibility captured within the crystal structure stemming from the disordered linker connecting the RRM and CTD as well as from the CTD 1a-a kink.* ***C****. Electrostatic surface maps of the HsORF1p trimer depicting the three potential binding clefts that may contribute to nucleic acid binding: the first view (left) is to the left of RRM A and shows cleft 1that forms between RRM A and CTD A. The second view (middle) is now to the right of RRM A and shows the top of cleft 2 that forms between RRM A and RRM B as well as cleft 3 that forms between the base of the coiled coil and CTD B. The third view (right) is still to the right of RRM A but now below and looking up into cleft 2 that is still between RRM A and RMM B. Most well conserved aromatics and positively charged residues from primary sequence alignment are flagged on the 3D structure. The residues that were incorporated into mutant constructs for this project have white text and residues that are not included in this study are in light blue text. Residues included in this study are color coded with symbols used in sequence alignment, so mutated residues with the same color flag were mutated together to form a single mutant construct (a single charged patch neutralization), with exception of black flags with white text, which represent single point mutations.*

Some of the mutations used in this study were based on results reported in Khazina et al. 2011 and therefore had published information regarding NA binding behavior, either through reported binding defects or more qualitative experiments showing impaired binding. Arginines 261 and 262 (binding cleft 1) were mutated together to serve as a severe NA binding defective mutant[19-22], while R235 was included as a less severe but still impaired binding mutant. R232 was included with R235 because of its proximity near the start of binding cleft 1. Also based on the 2011 study, the triple mutants K137A/K140A/R141A (binding cleft 3) and R206A/R210A/R211A (binding cleft 2) were included. Khazina et al. showed that neutralizing these positively charged residues impaired HsORF1p binding to NA by size-exclusion chromatography, but quantified binding affinity and thermostability measurements have not been reported. Lastly, the R215A mutation was inspired by the same structural work, as R215 appears to form a conserved salt bridge with E165 within the RRM domain. This mutant served as a useful positive control for the DSF measurements, since disrupting this salt bridge would be expected to alter HsORF1p thermostability.

In addition to the better characterized mutations, three additional point mutations were included based on their conservation across ORF1p consensus sequences from different phyla and their positions within the predicted binding domain. K300 was selected because it was the only relatively well conserved, positively charged residue positioned across from the base of the coiled-coil domain, making it a candidate for forming the opposing side of binding cleft 3 if the rotational flexibility observed in the crystal structure reflects a naturally accessible conformation. R159 was selected because it is well conserved and positioned relatively deep near the end of binding cleft 1, where it was expected to contribute strongly to NA binding. Finally, R248 was selected because it sits farther from the core of binding cleft 1, along the edge of the RRM, providing a useful

comparison to the more internally positioned R159. Together, these mutations provided a way to compare conserved charged residues across distinct regions of the predicted binding surface and determine whether their position corresponded to different effects on HsORF1p NA binding and cellular function.

Background – Retrotransposition Efficiency Assay and Immunofluorescence Imaging

In conjunction with the DSF and spectral shift assays used and explained in chapter two, both L1 retrotransposition efficiency and HsORF1p's localization and cytoplasmic morphology were monitored in cells. These experiments were included to test whether the effects of surface charge neutralization observed in the *in vitro* assays correlated to functional changes in a cellular context. The retrotransposition efficiency assay measured how each mutation affected a L1's overall ability to mobilize and the immunofluorescence imaging was used to evaluate if some mutant HsORF1ps altered its expected cytoplasmic localization and foci-forming phenotype. Together, these assays helped to connect the biochemical behavior of each mutant with cellular outcomes dependent on proper HsORF1p function.

The L1 retrotransposition efficiency assay is well established and has been widely used since its introduction in 1996[19]. In this context, L1 retrotransposition efficiency refers to the frequency at which an engineered L1 element completes a successful round of retrotransposition within a population of transfected cells. The output of this assay is typically measured by activation of a reporter cassette following L1-dependent mobilization and genomic insertion. Different versions of the assay use antibiotic resistance, fluorescent, colorimetric, or dual-reporter outputs, but the overall principle is the same: reporter expression occurs only after successful transcription, splicing, reverse transcription, and genomic integration of the L1 reporter construct. This study

used a reporter plasmid expressing the synthetic human L1 HsORFeus, which has been codon optimized to enhance expression of the L1-encoded proteins, followed by a GFP reporter cassette interrupted by an intron[23] (Figure 3.5). After transfection, the reporter plasmid is transcribed and the intron interrupting GFP is removed from the L1 RNA. Upon successful reverse transcription, mobilization, and integration of the reporter-containing construct, functional GFP is expressed. Retrotransposition efficiency for each construct was then measured by flow cytometry.

As a preliminary experiment, the localization and cellular phenotype of the mutant HsORF1p constructs were documented. HsORF1p is typically cytoplasmic and can appear as both diffuse signal and large cytoplasmic foci. To determine whether surface charge-neutralizing mutations altered this phenotype, the same reporter plasmids used for the retrotransposition assay, containing either WT ORF1p or a surface charge-neutralization mutation, were expressed in cells for 48 hours. HsORF1p was detected by immunofluorescence using two previously validated antibodies, mouse monoclonal 4H1 and rabbit monoclonal JH73, developed by the Burns laboratory and described in Taylor et al., which recognize an N-terminal ORF1p epitope and human ORF1p, respectively[24, 25]. The resulting staining was used to assess whether each mutant retained expected cytoplasmic localization and whether mutation altered ORF1p foci formation or overall cellular distribution. The underlying hypothesis was that HsORF1p foci formation may depend, at least in part, on its ability to bind RNA and generate locally concentrated HsORF1p:RNA assemblies, which could promote higher-order assembly, phase separation, or incorporation into cellular vesicle-like structures.

Using the spectral shift and DSF approaches described in Chapter 2, the effects of these mutations on HsORF1p's affinity for different ssNA and thermostability were measured. The same mutations were then introduced into retrotransposition reporter cassettes to compare molecular

defects with cellular outcomes, including retrotransposition efficiency and ORF1p foci formation. This strategy allowed us to ask whether reduced retrotransposition cleanly correlates with altered RNA binding or protein stability, and to distinguish mutations that primarily impair RNA engagement from those that disrupt additional aspects of ORF1p function in cells.

It was hypothesized that these mutations would have a spectrum of effects on binding affinity and retrotransposition efficiency. Instead, most mutations along the postulated binding path appear to substantially reduce both parameters, with the main exception being a previously characterized binding-deficient mutant that is clearly separated from WT and the other constructs. Interestingly, DSF data indicate differences in the thermal stability of some unbound mutant constructs relative to WT at residues not previously implicated in salt bridge formation, and it is also observed that possible differences in how individual mutants interact with distinct ssNA substrates. Lastly, fluorescence microscopy showed that HsORF1p cellular localization and foci phenotypes can remain largely preserved even when mutations severely impair NA binding affinity.

Materials and Methods

Plasmid Cloning, Protein Expression and Purification

Plasmid cloning, expression and purification strategy mirrors that used in Chapter 2. HsORFeus GFP-AI reporter plasmid was provided by the Martin Taylor lab. Plasmid DNA used for transfection was purified by Zymo Midiprep and quantified in advance for each construct.

Cell Culture Maintenance

HEK293T and **U2OS** cells were grown in 5% CO2, 37°C using a humidified NU-5800 incubator (NuAire) with **cell culture medium** (HyClone™ DMEM with high glucose (Cytiva) supplemented with 10% Tetracycline-Free FBS (GeminiBio), and 1x HyClone™ Penicillin-

Streptomycin (Cytiva)). Cells were maintained in 100 x 20 mm plates (CytoOne), and every time they reached 70-90% confluency, a **standard passaging protocol** was used: cell culture media aspirated off, cells washed with 8 mL of pre-warmed 1x HyClone™ PBS, pH 7.4 (Cytiva), 1x PBS aspirated off, 500 µL of TrypLE Express Enzyme no phenol red (GIBCO) added, and cells were placed in the incubator for 5 minutes before being resuspended in 9.5 mL of pre-warmed cell culture medium. The resuspended cells were diluted 1:10 to 1:20, and 10 mL of the dilution was used to seed each following 100 x 20 mm plate (CytoOne), to maintain healthy log-phase growth.

Cell Transfections for Flow Cytometry and Immunofluorescence Microscopy

For retrotransposition experiments, freshly thawed HEK293T or U2OS cells were allowed at least 1-2 passages prior to transfection. Once at ~80% confluency, the cells were passaged from 100 x 20 mm plates (CytoOne) into 6-well plates (CytoOne), using an amended standard passaging protocol: after resuspension with pre-warmed cell culture medium, the live cells were counted by mixing a 10 uL sample 1:1 with 0.4% Typan Blue solution and using a CellDrop FL cell counter (DeNovix). The cells were then diluted to ~80,000 cells/mL using cell culture medium. In the case of HEK293T cells, 1 mL was used to seed each well of the 6-well plate. And in the case of U2OS cells, 2 mL was used to seed each well of the 6-well plate. The cells were then allowed to rest in 5% CO2, 37°C using a humidified NU-5800 incubator (NuAire) for 24 hours.

For immunofluorescence experiments using U2OS cells, before passaging, an autoclaved 22x22mm no.1 thickness cover slip (Epredia) was added to each well of the 6-well plate. The upward facing surface of each coverslip was treated with 200 uL of a sterile, 10 ug/mL solution of fibronectin (Santa Cruz) and kept in the incubator for 2 hrs. During passaging, the fibronectin solution was then aspirated off the coverslips and U20S cells seeded into each well as described above.

After the 24hrs, transfections were performed using PEI-based delivery system. 4 ug of each plasmid DNA was diluted in freshly made, unsupplemented cell culture medium containing 12 ug of PEI MAX® (Polysciences, 24765), a 3:1 PEI:DNA mass ratio. PEI:DNA mixtures were vortexed vigorously for 10 seconds and rested at room temperature for 15 minutes. Subsequently, 1 µg of plasmid DNA was added to each well as needed. Each experiment included: 3-4 replicate wells per plasmid construct, three replicate wells of non-transfected HEK293T and one well transfected with transfection control plasmid (mSFGFP cloned into a pDARMO vector) to monitor transfection efficiency and assay performance.

After transfection, cells were returned to the incubator and allowed to recover for 48 hours. The GFP transfection control was examined by fluorescence microscopy to confirm >80% transfection efficiency. Once the transfection control was confirmed positive, the other wells, other than the negative transfection control wells, had their media exchanged. The HEK293T cells received 5 mL of cell culture medium containing a final concentration of 10ug/mL puromycin from InvivoGen. while non-transfected controls received matched medium lacking antibiotic. The U2OS cells received normal cell culture medium. Plates were incubated for an additional 3-4 days for HEK293T cells and 2 days for U2OS cells, to allow reporter expression and antibiotic selection to proceed.

Flow Cytometry

After the transfection protocol, cells were then harvested from 6-well plates by following the standard passaging protocol, but cells were resuspended by adding 900 µL of **quenching buffer** (~1x PBS, pH 7.4, 10% FBS) instead of cell culture medium, and each well had its cells transferred into individual, sterile 1.5 mL conical tubes. Each sample was immediately placed on ice after resuspension and transfer. After all samples were transferred, they were centrifuged at

400 xg 4°C for 5 minutes in a Legend Micro 21 R (ThermoScientific). The quenching buffer was aspirated off of the cell pellet and cells were resuspended in 200 µL of **sorting buffer** (~1x PBS, pH 7.4, 1% FBS, 500 µM EDTA).

Samples were transported to the Coro Center East FACS facility. Flow cytometry data was collected on a BD FACSymphony. Unstained cells and GFP-positive controls were used to set gating and detector voltages. For each sample, 5,000-30,000 single-cell events were acquired. All data were processed using FACSDiva software and graphs visualized using GraphPad Prism 11 .

Immunofluorescence Microscopy

After the transfection, U2OS cells were fixed and permeabilized with PFA and Triton X-100. Cell culture media was aspirated, and cells were gently washed with 1x PBS. Freshly prepared 4% PFA was added to each well (1-2 mL per well), and plates were incubated for 10-15 min. PFA was then aspirated, and cells were washed three times with 1x PBS. Next, 0.1% Triton X-100 prepared in 1x PBS was added to each well (1-2 mL per well), and cells were incubated for 10-15 min. The PBS wash step was then repeated. Cells were blocked using 5% BSA prepared in 1x PBS for 60 min on a rocking platform and then washed again. Anti-HsORF1p antibodies (mouse 4H1 and rabbit JH73) were diluted to ~ 0.5 ug/mL to 0.2 ug/mL in wash buffer. 50-100 µL of this dilution was added to each coverslip and incubated on the cells for 20 minutes. Cells were washed and then incubated with fluorescent secondary antibodies for 20 minutes. And washed again and incubated with 1:10,000 DAPI stain (ThermoFisher Scientific) for 10 minutes. Lastly, coverslips with cells were mounted onto glass slides using ProLong Gold (Invitrogen). Imaging was performed on an Olympus FV1000 laser scanning confocal microscope using 60x magnification. CellProfiler 4.2.8 was used to identify nuclei, bin cells as non-expressing, dim, or bright, and count foci based on area. Final violin plots were generated in GraphPad Prism 11.

Results

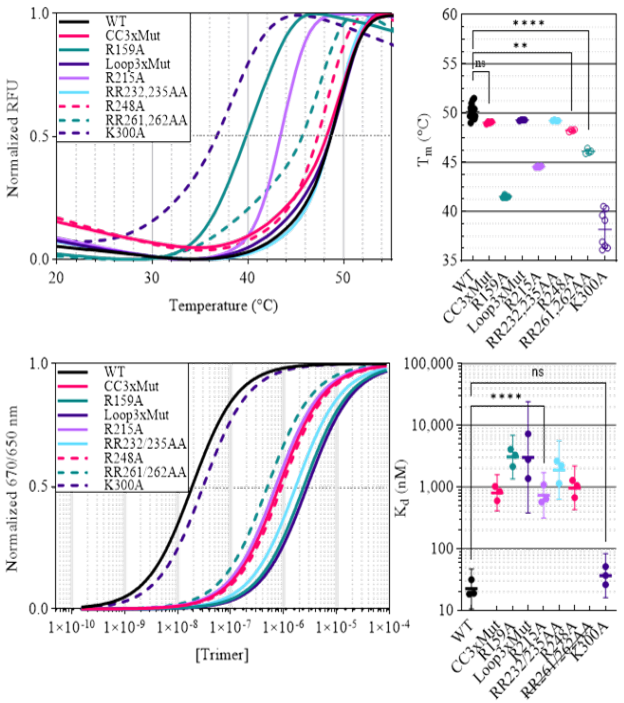


Figure 3.2. *Effects of selected HsORF1p binding domain mutants on unbound trimer stability and nucleic acid binding affinity.* ***A****. Averaged DSF melting curves of the indicated HsORF1p trimeric constructs (left) measured at 300 mM NaCl and corresponding T_m values summarized in dot plots\ (right).* ***B.*** *MST spectral shift binding curves for each mutant construct produced at 510 mM NaCl with 36-nt 50% GC RNA (left) and corresponding K_d values summarized in dot plot (right). Dot plots depict average T_m with standard deviation error bars and statistical comparison performed with one-way ANOVA with Dunnett's multiple comparisons. Significance: "ns", not significant; "****", p< 0.01; "**", p < 0.0001.*

Structural integrity and purity of mutants

Each trimer construct was purified from *E. coli* using a standardized protocol; however, some constructs, such as the K300A mutant, were inherently less stable and precipitated during overnight dialysis at 4°C. For these constructs, the dialysis buffer had a starting NaCl concentration of 450 mM, they were warmed to room temperature and carefully diluted to ~350-400 mM NaCl, after dialysis, before filtration and heparin purification. Each construct was tested for purity after processing by SDS-PAGE gel with Coomassie staining (Supplemental Figure 1). Constructs were then run on a 3.2/300 analytical Superdex column and had their CD spectra measured to ensure that the point mutations did not alter the general 3D shape and secondary structure of the trimer (Supplemental Figure 2). The instruments could not detect any alterations to HsORF1p trimer architecture as a result of introducing these mutations.

Mutation's Effect on Thermal Stability and Ligand Binding

Next, it was examined how different surface charge neutralizing alanine-mutants (Figure 3.1) perform in the biochemical assays validated in Chapter 2. The DSF and spectral shift each are

bulk measurements of protein stability (e.g,. salt-bridge formation, solubility/ aggregation, and other effects that perturb the thermodynamics of 3D organization) and ligand binding affinity for cy5-labeled 36-nt 50% GC RNA (e.g, the ratio of binding kinetics' on and off rates and the sum of different binding conformations, if occurring). It is assume that the DSF data largely represent a population in which most trimers are in a single comparable state (i.e., unbound or bound), although the assay conditions were originally optimized to ensure saturation of the WT trimer affinity for 36-nt 50% GC RNA at 510 mM NaCl. This assumption is supported by the observation that the DSF transitions generally span only ~10-14°C from onset of RFU increasing to upper bound, which is comparable to many protein-only DSF curves suggest a homogeneous population has been achieved. However, this should be taken as evidence consistent with near-saturating binding rather than as a direct measure of occupancy.

In the DSF assay, it was found that several alanine mutants reduced the thermostability of the unbound trimer state (Figure 3.2). Consistent with the observations in Chapter 2, the apparent Tm of the WT trimer measured by SYPRO Orange DSF fell within the range of 49-52°C. Three of the eight mutants tested showed no detectable effect on bulk thermostability: CC3xMut, Loop3xMut, and RR232,235AA. The R248A mutant showed a slight but reproducible decrease in stability relative to WT, with a ΔTm of -1°C. The R215A and RR261,262AA constructs, both of which include mutations at residues previously implicated in salt bridge formation in the trimer crystal structure[5, 26], showed larger decreases in apparent stability, with ΔTms of -3°C and -5°C, respectively. Surprisingly, R159A and K300A were even more destabilizing, with measured ΔTms of -8°C and -12.5°C, respectively. Consistent with this reduced stability, both mutants also appeared more prone to aggregation during

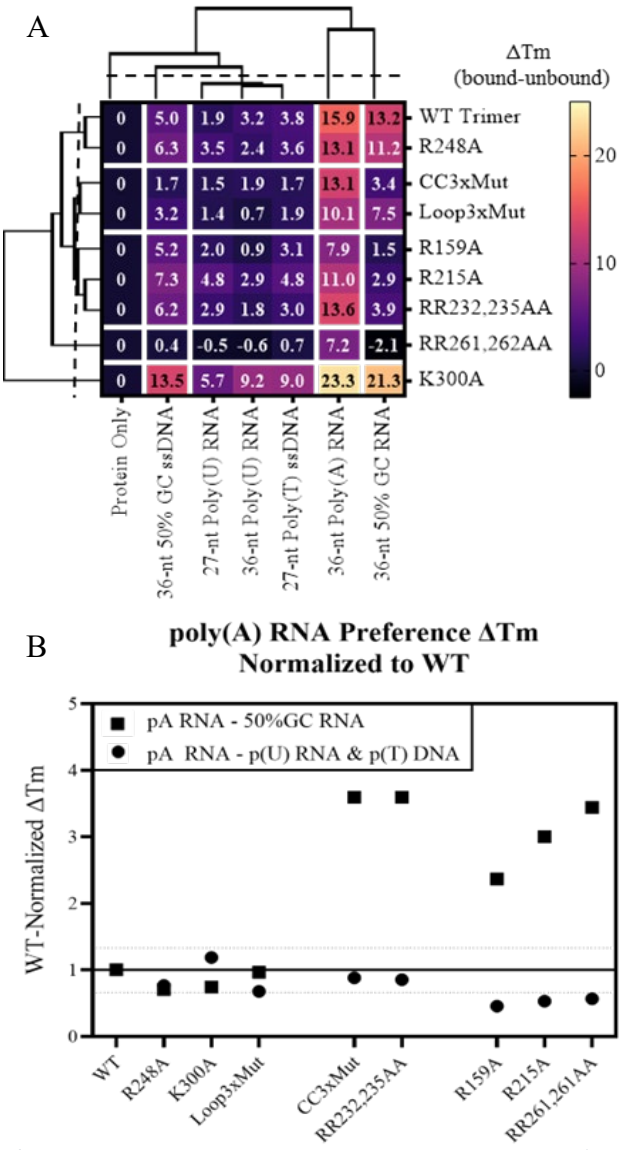


Figure 3.3.A *Heat map of ΔTm values for HsORF1p trimer constructs across different ssNA substrates. Values represent the change in apparent melting temperature relative to each unbound protein-only condition (ΔTm = bound − unbound) measured by DSF. Rows correspond to WT and mutant trimer constructs, and columns correspond to the indicated ssNA substrates. Hierarchical clustering using Euclidean distance was performed on the Broad Institute's Morpheus web server to group constructs and substrates with similar thermal stabilization patterns.* ***B.*** *WT-normalized ΔT_m values were calculated to compare the relative stabilizing effect of poly(A) RNA against other NA ligands for WT HsORF1p and each RRM-CTD charge-neutralization mutant. Squares show the relative difference between poly(A) RNA and 50% GC RNA, while circles show the relative difference between poly(A) RNA and the pyrimidine-rich substrates poly(U) RNA and poly(T) ssDNA. Values near 1 indicate WT-like relative stabilization, while values above or below 1 indicate increased or decreased poly(A)-dependent stabilization relative to WT. Dashed lines indicate range of ΔT_ms that fall 1/3 of WT value.*

purification, particularly during overnight digestion and dialysis at 4°C and upon dilution from stock solution (~500 mM NaCl) into assay buffers containing ≤300 mM NaCl (data not shown).

Quantitative binding affinity measurements in Monolith experiments were evaluated only after extended incubation times (1 h and 4 h; Supplemental Figure 3) to ensure a stable equilibrium is being reached by all mixtures by the time of measurement. Upon examining the binding affinities obtained after each protein construct was mixed with Cy5-labeled 36-nt 50% GC RNA and incubated for up to 4-hours (Figure3.2), only RR261,262AA, which is well documented to have severely impaired affinity for NA[13], failed to produce a measurable binding interaction under these assay conditions. The only mutant that performed comparably to WT was K300A, which experienced a moderate ~40% reduction binding affinity. CC3xMut, R215A, RR232,235AA, and R248A all experienced a 97-99% reduction in binding affinity when compared to WT. Lastly, R159A and Loop3xMut both experienced reductions in binding affinity greater than 99%.

DSF assays were performed using the HsORF1p constructs and a subset of ssNA panel used in Chapter 2. The data was compiled into 2-dimensional heat maps (Figure 3.3). The first insight gained was that, when the Tm of each construct:ssNA combination was evaluated relative to the unbound Tm of that construct (Figure 3.3 A), the extent of stabilization appeared to depend on both the mutation and the identity of the ssNA substrate. However, ssNA identity appeared to have the stronger overall discriminatory effect, based on the total distance between related groups in the Euclidean clustering brackets. Looking at the Euclidean clustering of the trimer constructs, it can be observe that only the R248A mutant preserves the same trends as WT trimer. In this assay, the two constructs with the most unique binding trends were the RR261,262AA and K300A mutants, where the RR261,262AA appears to have little or no stabilization induced by the presence of any ssNA other than poly(A) RNA and K300A experiences the largest increases of stabilization

when any ssNA is present. The rest of the mutants trended similarly when compared to one another, except for in the case of 36nt 50% ssDNA and the 3xCCMut and 3xLoopMut where stabilization appears lower.

When comparing the heat map values showing how Tm changes across mutant:ssNA combinations relative to WT, two additional patterns emerge. First, the pyrimidine containing homopolymeric substrates show only marginal variation. Substituting poly(U) RNA for poly(T) ssDNA or extending poly(U) RNA from 27-nt to 36-nt, produces relatively small differences across most mutants. This suggests that HsORF1p discriminates only weakly among these substrates, both with respect to the position of surface charge neutralization and to the base composition of the ssNA. Interestingly, and somewhat counterintuitively, the shorter 27-nt poly(U) RNA appears to stabilize most constructs as well as, or slightly better than, the 36-nt poly(U) RNA, with WT and K300A being the main exceptions.

Second, when comparing 36-nt 50% GC ssDNA, 36-nt 50% GC RNA, and 36-nt poly(A) RNA more construct-specific behavior emerges. For example, R159A, R215A, and RR232,235AA appear to retain WT-like stabilization with 50% GC ssDNA, show reduced stabilization with 50% GC RNA, and yet maintain relatively strong stabilization with poly(A) RNA. Within this group, R215A and RR232,235AA remain more WT-like with poly(A) RNA, whereas R159A behaves more similarly to RR261,262AA. The two triple mutants also show broadly similar behavior: both exhibit substantial decreases in stabilization relative to WT with 50% GC ssDNA and 50% GC RNA but retain stronger responses with poly(A) RNA. Notably, CC3xMut still shows an approximately 5°C decrease relative to WT in the case of poly(A) RNA, but it is still relatively higher affinity than the other two substrates. These results suggest slight

regional biases occurring across the HsORF1p RRM-CTD when the ssNA it is binding contains a significant portion of purines.

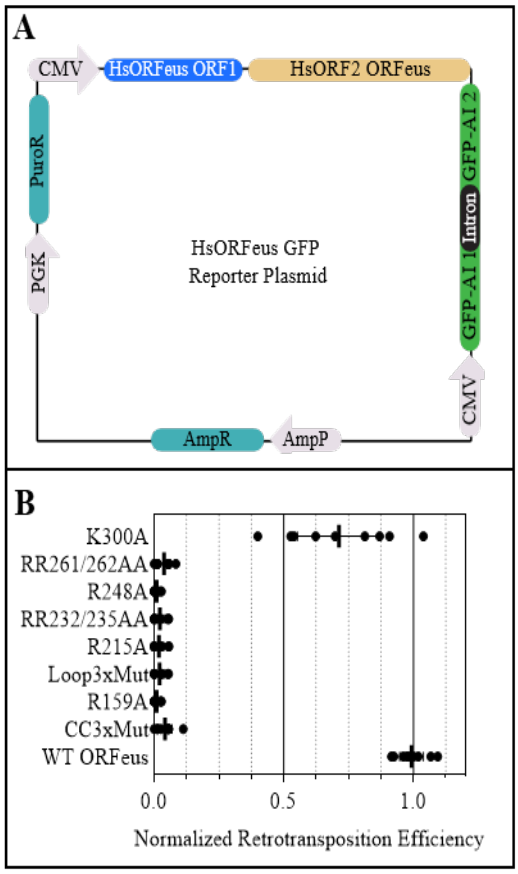


Figure 3.4 *| Overview of retrotransposition efficiency assay.* ***A.*** *Schematic of the plasmid cassette used in the GFP retrotransposition efficiency assay. After transfection, GFP signal is produced only after successful retrotransposition of the plasmid-derived L1 reporter, followed by transcription of the integrated cassette, removal of the intron, and expression of uninterrupted GFP.* ***B.*** *Dot plot showing retrotransposition efficiency of each construct normalized to WT. Each dot represents an individual replicate, and the vertical bars indicate the mean and SD for each construct.*

The Effect of Charge neutralization on Cellular L1 Retrotransposition

After characterizing the effects of binding dynamics each RRM-CTD produced, it followed that correlating the effects observed *in vitro* with *in cellulo* phenotypes could be useful in contextualizing observations back to HsORF1p within a more natural environment.

The results of the retrotransposition assay were coarser than the biophysical data and only demonstrated a binary phenotype, either successful retrotransposition efficiency or practically none (Figure 3.4B). The WT and K300A showed detectable levels of retrotransposition, where the K300A performed between 50-90% as well as the WT. All other mutants produced GFP-positive cells at a rate indistinguishable from noise, although R215A may have only expressed poorly (Supplemental Figure 5). The signal-to-noise limit for this assay may have been reached given that each replicate measurement was limited to 5,000-30,0000 events, the retrotransposition efficiency data mirrors

the trends that were observed in the binding affinity reductions caused by each mutation from spectral shift measurements (Figure 3.2 B).

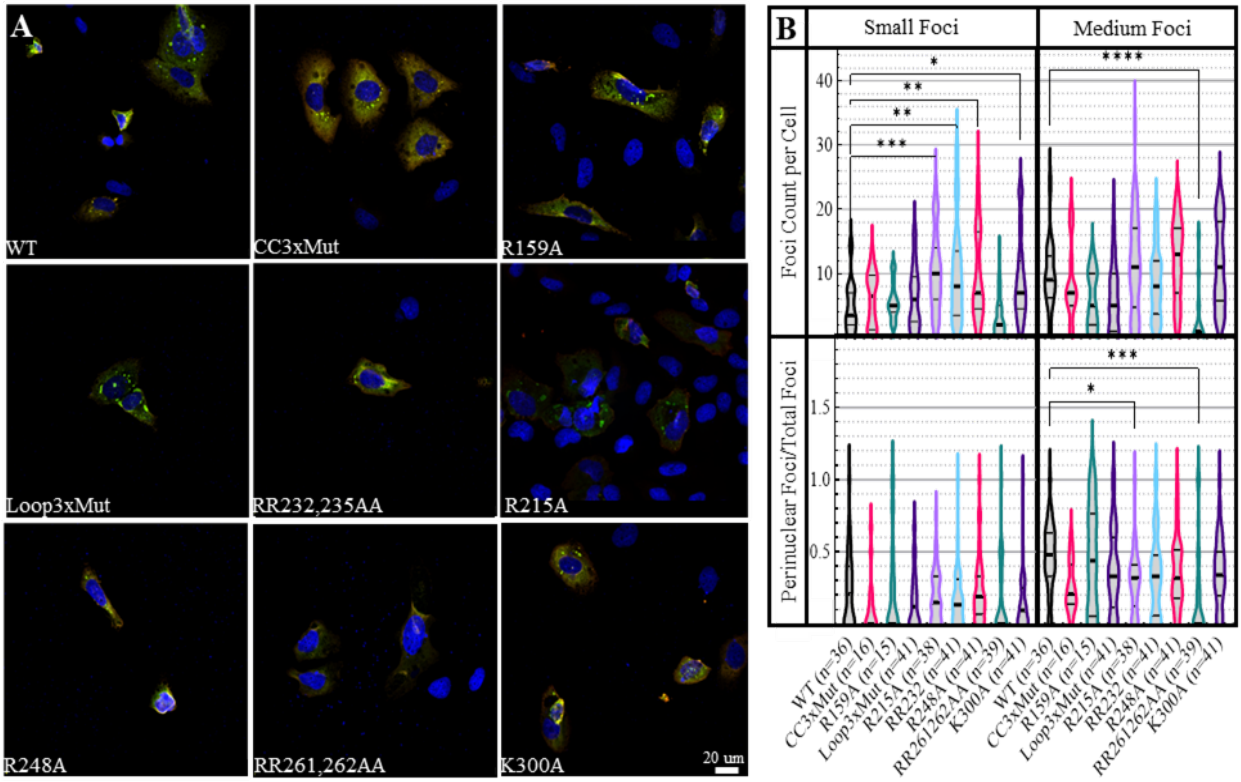


Figure 3.5 *| HsORFeus expression and immunofluorescence in U2OS cells.* ***A.*** *Representative 60x Olympus confocal microscope images of U2OS cells expressing the HsORFeus plasmid used in the earlier retrotransposition assay. Each panel shows a randomly selected field of view of cells expressing the indicated construct and was generated by compressing 7-10 image planes into maximum intensity z-stacks. The blue channel shows DAPI stained nuclei, while the green and red channels show ORF1p antibody staining (4H1 and JH73, respectively).* ***B.*** *Violin plots generated from analysis of nine fields of view for each construct. The total number of cells that reached the brightness threshold and were included in the analysis is indicated by the "n" next to each construct name on the x-axis. Foci bright enough to be distinguished from background signal were binned into three groups based on the area they covered: small and medium are shown here, whereas large foci are not shown because of their low prevalence. The two upper violin plots show the number of small foci per cell (left) and medium foci per cell (right). The two lower violin plots show the number of perinuclear small foci per cell (left) and perinuclear medium foci per cell (right). Perinuclear foci were defined as foci that overlapped the nuclear edge, as identified by CellProfiler software, within ± 3 pixels. Thick horizontal lines within each violin indicate the median, while the thinner lines above and below indicate the first and third quartiles. * = P < 0.05, ** = P < 0.01, *** = P < 0.001.*

Next, visualization of how these HsORF1p constructs appeared in cells was conducted, because cytoplasmic foci formation has been linked to productive RNA binding and higher-order RNP assembly[16, 17, 20, 27, 28]. The same GFP reporter plasmid was transfected into U2OS cells and allowed 48 h for recovery and construct expression. Cells were then fixed with PFA and stained using two primary antibodies directed against distinct HsORF1p epitopes, followed by fluorescent

secondary antibodies. After imaging, a semi-quantitative analysis pipeline was used to measure several ORF1p phenotypes, including overall cellular signal intensity, the number of detectable small, medium, and large puncta, and the number of ORF1p foci observed around the nuclear envelope per cell.

Immunofluorescence imaging of these HsORFeus ORF1p binding domain mutants showed that many mutants created ORF1p foci that are comparable to WT despite having dramatically altered *in vitro* properties and ablated retrotransposition efficiency in HEK 293T cells (Figure 3.5). The most striking observation is that the RR261,262AA mutation was the only one to reduce the median number of small foci and created a dramatic reduction in the formation of larger foci per cell. All mutants experienced a slight or modest increase in the number of smaller foci per cell detected and either had a comparable median of medium foci or a lower number, but not statistically significant. When perinuclear foci, defined as foci overlapping within ±3 pixels of the nuclear edge identified by CellProfiler, are normalized to the total number of foci per cell, there does not appear to be an obvious difference in localization behavior under these staining conditions.

Summary and Discussion

The spectral shift and DSF assays showed that HsORF1p's response to NA is highly sensitive to neutralization of basic patches around the RRM-CTD. K300A was the only mutant that retained binding affinity comparable to WT. All other mutations produced a similar magnitude of reduced binding affinity, except for RR261,262AA, which impaired binding to a more severe but immeasurable degree. This was true despite some constructs containing triple mutations and others containing only single point mutations, suggesting that the position of the targeted residue or patch may be more important than the number of residues mutated. More speculatively, this

may be consistent with the idea that L1 proteins are under pressure to maintain multiple functional interactions within a compact coding sequence, which could limit mutational tolerance across functionally dense regions such as the HsORF1p RRM-CTD surface, although this study does not directly test that model[15]. Together, these data support the interpretation that the positively charged RRM-CTD surface is broadly important for productive ssNA engagement. The K300A result weakens, but does not fully invalidate, the idea that binding cleft 3, involving the base of the CC and a rotated CTD conformation, acts as a direct NA-interaction site in the same manner as clefts 1 and 2, where charge neutralization on either side of the cleft strongly impaired binding.

The DSF screen, in which Tm values were calculated for different HsORF1p charge-neutralization mutants bound to different NA ligands, showed that NA identity was the primary determinant of the relative stabilization pattern observed across mutants. Poly(A) RNA remained the strongest stabilizing ligand regardless of which mutation was present in HsORF1p, and across most mutants the relative ΔTm values between NA-bound states remained broadly similar to WT (Figure 3.3B). Although some mutation-dependent differences appeared statistically significant, especially when comparing poly(A) RNA to 50% GC RNA or poorer ligands such as poly(T) ssDNA and poly(U) RNA, these differences should not be overinterpreted as evidence of residue-specific ligand discrimination. In these cases, it is difficult to determine whether the difference reflects mutation-specific ligand recognition, changes in NA base composition, or practical noise within the DSF assay. Overall, these data are most consistent with broad ligand-dependent stabilization patterns, suggesting that the targeted residues primarily support general and/or overlapping NA engagement functions rather than recognition of specific NA compositions. This interpretation is consistent with an avidity-based binding model in which different NA ligands interact with overlapping regions of the HsORF1p binding surface, while poly(A) RNA makes

more stabilizing contacts because of its geometry, chemistry, or both. However, without structures of HsORF1p bound to different NA ligands, it is not possible to rule out that different substrates follow distinct binding paths across the protein surface.

The distinction between reduced apparent binding affinity and retained ligand-dependent thermal stabilization becomes clearer when the DSF and spectral shift data are considered together. Although apparent binding affinities were significantly reduced for most mutants, many of these same mutants, with the exception of RR261,262AA, still trended similarly to WT once in the NA-bound state, based on the relative thermal shifts observed across different ligands (Figure 3.3B). This suggests that binding affinity and ligand-dependent thermal stabilization report on related but non-identical aspects of the HsORF1p:NA interaction. The targeted basic residues may be especially important for efficient NA binding affinity or overall kinetics, while having less effect on the relative stabilization provided by each ligand once a bound state is achieved. Because these assays do not directly measure association or dissociation rates, this interpretation should be treated as a model for how the mutations affect binding behavior rather than as direct evidence of altered kinetics. This pattern is most clearly illustrated by R248A, which had an apparent binding affinity similar to the other binding-impaired mutants but appeared among the most WT-like mutants in the 2D DSF heat map.

The retrotransposition assay produced coarse results but suggested that retrotransposition efficiency correlates broadly with binding affinity, as mutants with strongly impaired affinity for 50% GC RNA were also severely impaired in this assay compared to WT. K300A retained comparable activity in both assays. These results may reflect limitations in the sensitivity of the retrotransposition assay, potentially due to limited sample sizes, and may not capture more subtle differences in retrotransposition efficiency between mutants. However, the different mutations

also did not produce a clearly graded range of binding affinities, and most may have fallen below a functional binding threshold required for L1 mobilization. Overall, these data support the conclusion that efficient L1 RNA engagement by HsORF1p is a critical requirement for productive retrotransposition.

In contrast, HsORF1p foci formation did not correlate well with the binding affinity trends, as initially hypothesized. Most binding impaired mutants were still able to form foci at levels that appeared broadly comparable to WT, and visual inspection confirmed that no mutation caused a complete loss of foci formation. RR261,262AA was the only mutant with a severe reduction in foci-forming capability. However, it is unclear whether this phenotype reflects its severe loss of NA-binding affinity or whether mutating this region disrupts other interactions, including HsORF1p trimer-trimer interactions or interactions with surrounding host factors. These results suggest that gross foci formation is not a simple proxy for NA-binding affinity under the conditions tested.

Several factors may explain why foci formation was less sensitive to mutation than expected. The CellProfiler pipeline may require further optimization, particularly in defining the cutoff between small puncta and medium sized foci, since the current cutoff may allow smaller puncta to be counted as medium foci. In addition, HsORF1p expression from HsORFeus may produce higher protein levels than endogenous L1 expression, potentially allowing concentration-dependent foci formation even when RNA binding is impaired. Alternatively, foci formation may be less dependent on RNA binding driven by the RRM-CTD than expected and may instead reflect protein-protein interactions, recruitment to host granules, or incorporation into vesicle-like compartments. Future experiments that reduce HsORF1p expression levels, stain for L1 RNA, test colocalization with stress granule or vesicle markers, and assess ORF1p-associated RNAs would

help determine whether the foci formed by binding-impaired mutants are compositionally and functionally similar to WT foci.

Together, these results support a model in which basic residues across the HsORF1p RRM-CTD are important for HsORF1p:NA binding and downstream L1 retrotransposition. These residues may perform overlapping functions in ligand binding, as none of the mutations tested produced a severe or clearly interpretable ligand composition-dependent change in Tm. Rather than supporting residue-specific ligand discrimination, the DSF data suggest that the tested NA ligands interact with the binding domain in broadly similar ways, with poly(A) RNA producing consistently stronger stabilization potentially because of avidity effects, ligand geometry, or ligand chemistry. In cells, impaired binding generally corresponded to impaired retrotransposition, whereas gross foci formation was comparatively less sensitive to the same mutations. This distinction suggests that visible HsORF1p foci are not sufficient to predict retrotransposition competence, and that productive L1 mobilization likely depends on a more specific or functional HsORF1p:ssNA assembly state than foci formation alone.

References

1. Corley, M., M.C. Burns, and G.W. Yeo, *How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms*. Mol Cell, 2020. **78**(1): p. 9–29.
2. Walker-Gibbons, R., et al., *Sensing the structural and conformational properties of single-stranded nucleic acids using electrometry and molecular simulations*. Scientific Reports, 2024. **14**(1): p. 20582.
3. Callahan, K.E., et al., *Polymerization and nucleic acid-binding properties of human L1 ORF1 protein*. Nucleic Acids Res, 2012. **40**(2): p. 813–27.

4. Naufer, M.N., A.V. Furano, and M.C. Williams, *Protein-nucleic acid interactions of LINE-1 ORF1p*. Semin Cell Dev Biol, 2019. **86**: p. 140–149.

5. Khazina, E., et al., *Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition*. Nat Struct Mol Biol, 2011. **18**(9): p. 1006–14.

6. Rajagopalan, M., S. Balasubramanian, and A. Ramaswamy, *Insights into the RNA binding mechanism of human L1-ORF1p: a molecular dynamics study*. Mol Biosyst, 2017. **13**(9): p. 1728–1743.

7. Martin, S.L., *Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1*. RNA Biol, 2010. **7**(6): p. 706–11.

8. Al-Raawi, D. and A. Kanhere, *Site-Directed Mutagenesis Protocol to Determine the Role of Amino Acid Residues in Polycomb Group (PcG) Protein Function*. Methods Mol Biol, 2023. **2655**: p. 79–89.

9. Carrigan, P.E., P. Ballar, and S. Tuzmen, *Site-directed mutagenesis*. Methods Mol Biol, 2011. **700**: p. 107–24.

10. Martin, S.L., *The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition*. J Biomed Biotechnol, 2006. **2006**(1): p. 45621.

11. Martin, S.L., et al., *A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity*. Nucleic Acids Res, 2008. **36**(18): p. 5845–54.

12. Naufer, M.N., et al., *L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling*. Nucleic Acids Res, 2016. **44**(1): p. 281–93.

13. Martin, S.L., et al., *LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein*. J Mol Biol, 2005. **348**(3): p. 549–61.

14. Januszyk, K., et al., *Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1*. J Biol Chem, 2007. **282**(34): p. 24893–904.

15. Adney, E.M., et al., *Comprehensive Scanning Mutagenesis of Human Retrotransposon LINE-1 Identifies Motifs Essential for Function*. Genetics, 2019. **213**(4): p. 1401–1414.

16. Newton, J.C., et al., *Phase separation of the LINE-1 ORF1 protein is mediated by the N-terminus and coiled-coil domain*. Biophys J, 2021. **120**(11): p. 2181–2191.

17. Sil, S., et al., *Condensation of LINE-1 is critical for retrotransposition*. Elife, 2023. **12**.

18. Khazina, E. and O. Weichenrieder, *Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p*. Elife, 2018. **7**.

19. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells*. Cell, 1996. **87**(5): p. 917–27.

20. Kulpa, D.A. and J.V. Moran, *Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition*. Hum Mol Genet, 2005. **14**(21): p. 3237–48.

21. Doucet, A.J., et al., *Characterization of LINE-1 ribonucleoprotein particles*. PLoS Genet, 2010. **6**(10).

22. Freeman, B.T., et al., *Identification of charged amino acids required for nuclear localization of human L1 ORF1 protein*. Mob DNA, 2019. **10**: p. 20.

23. Ostertag, E.M., et al., *Determination of L1 retrotransposition kinetics in cultured cells*. Nucleic Acids Res, 2000. **28**(6): p. 1418–23.

24. Sharma, R., et al., *Immunodetection of Human LINE-1 Expression in Cultured Cells and Human Tissues*. Methods Mol Biol, 2016. **1400**: p. 261–80.

25. Taylor, M.S., et al., *Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition*. Cell, 2013. **155**(5): p. 1034–48.

26. Khazina, E. and O. Weichenrieder, *Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame*. Proc Natl Acad Sci U S A, 2009. **106**(3): p. 731–6.

27. Goodier, J.L., et al., *Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion*. Hum Mol Genet, 2010. **19**(9): p. 1712–25.

28. Goodier, J.L., et al., *LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex*. Mol Cell Biol, 2007. **27**(18): p. 6469–83.

Chapter4 | Discussion

Overview

L1 is the last remaining fully autonomous retrotransposon active in the human genome[1, 2]. The proteins encoded by L1 formHsORF1p-mediated RN Ps that interface with a wide range of fundamental cellular pathways associated with human pathologies and aging[1, 2]. Despite this broad biological relevance, substantial gaps remain in our understanding of the biophysical and biochemical properties that underlie RNP formation[1, 3]. For example, the L1-encoded proteinHsORF1p is generally described as a high-affinity, non-specific nucleic acid (NA) chaperone, yet relatively few studies have systematically defined the extent and limits of its binding capabilities[4-7]. Although recent technological advances have brought the field closer to experimentally derived, atomic-resolution structures ofHsORF1p bound to NA substrates, obtaining such structures remains challenging[8-10]. In the absence of these structures, numerous biophysical hypotheses can still be generated from existing experimental data and from crystallographic, NMR, and AlphaFold-predicted models ofHsORF1p[8-10]. In this thesis, the biophysical effects of altering single-stranded NA (**ssNA**) composition were determined in the context of nucleoprotein particles (**NPs**) composed of a single HsORF1p trimer and short ssNAs, using an N-terminally truncated HsORF1p construct. The effects of neutralizing positively charged patches across HsORF1p's trimeric binding surface through alanine substitution were tested and it was observed *in vitro* and *in cellulo* effects. In this chapter, these findings in the current context of L1 biology are interpreted, and limitations of the study are considered.

A HsORF1p:NA Binding Model and Preference for Adenosine Rich RNA

Postulated Early RNP Binding Model

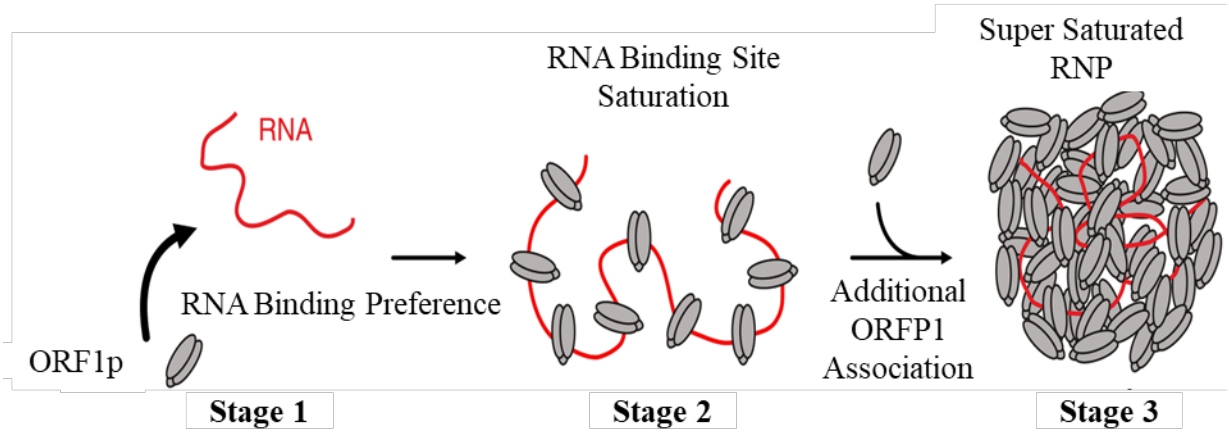


Figure 4 | This figure is adapted from Zernia et al., 2025[10], licensed under CC BY-NC 4.0. The original schematic was simplified and relabeled to illustrate three stages of HsORF1p:RNA binding. Stage 1 represents initial HsORF1p:RNA binding, when many RNA binding sites remain available and preferential binding is most likely to occur. Stage 2 represents progressive coating of the RNA by HsORF1p, during which secondary-structure melting, reannealing, and RNP condensation dynamics may occur. Stage 3 represents the supersaturation model proposed by Zernia et al., in which more HsORF1p trimers associate with the RNA than would be expected from simple stoichiometric estimates. At this stage, higher-order protein:protein interactions likely play a stronger role in recruiting excess trimers. In addition, free or partially bound RRM-CTD domains may remain available on the nucleating RNA, allowing incorporation of additional nucleic acids into the RNP. It is unclear how distinct each stage is from one another on a temporal scale but provides a framework for L1 RNP formation.

A working model for HsORF1p:NA binding can be drawn from current structural and biophysical studies of HsORF1p:NA interactions. Across these studies, HsORF1p appears to engage NAs primarily through the negatively charged sugar-phosphate backbone, with binding shaped by both the 3D organization of the trimer and the formation of a continuous, highly basic surface across the RRM-CTD regions (Figure 3.1C)[9]. This model is consistent withHsORF1p's ability to bind a broad range of NA substrates, but it should not be interpreted as purely electrostatic. Previous SEC data suggest that nitrogenous base identity may also contribute to HsORF1p:NA interactions[9, 11]. The quantified biophysical measurements presented here reinforce the importance of these basic regions in NA binding dynamics and nucleoprotein

stability, while also showing that ligand identity influences HsORF1p stabilization in a way that likely reflects contributions from both backbone chemistry and base composition.

Several studies support a multi-step model in which HsORF1p binding, NA condensation, and higher-order RNP assembly occur progressively rather than as a single discrete event. This process can be broadly summarized in three stages (Figure 4). In the first stage, HsORF1p trimers make initial contacts with RNA, ssDNA, or mismatch-containing dsDNA substrates. This early interaction has previously been described as "caging" and may involve local engagement of one or more regions of the NA substrate by the trimer[6, 12-15]. In the second stage, additionalHsORF1p trimers associate with the substrate, promoting further NA compaction and remodeling. Depending on the substrate and experimental context, this may include condensation, partial destabilization and melting of mismatch-containing dsDNA, or rearrangement of NA secondary structure[6, 12-15]. In the third stage, additionalHsORF1p trimers can continue to associate into the NP beyond what the stoichiometry of one trimer per length of NA would permit, producing a more dynamic higher-order assembly stabilized by both protein-NA and protein-protein interactions. These interactions may involve the RRM-CTD surface as well as contributions from the NTD and coiled-coil domains[10, 14, 15]. This later state has been proposed to leave excess or transiently availableHsORF1p binding surfaces that could interact with additional NA molecules, although how often this occurs in vivo and how it contributes to chromosome association after nuclear entry or mitosis remain unresolved[10].

The experiments presented here are most directly related to the earlier stages of this model. Interactions between a single HsORF1p trimer and short ssNA substrates likely provide a useful proxy for measuring primary trimer:NA interactions, while avoiding the added complexity of long substrate architecture, extensive secondary structure, and higher-order RNP assembly. For this

reason, the spectral shift and DSF assays are best interpreted as reporting on direct or near-direct features of HsORF1p:ssNA engagement rather than the full RNP assembly pathway. Within this context, two related but still hypothetical features provide a useful framework for interpreting the data: electrostatic channeling and a non-sequence-specific melting binding mode.

In this model, electrostatic channeling would arise from the dense distribution of positively charged residues along the C-terminal face of the trimer, which could increase the local probability of NA encounter and promote engagement with the HsORF1p in a way to promote binding affinity through kinetics[8, 9]. These same residues may also contribute to the avidity of the interaction once the NA is positioned along the binding surface. In this sense, arginine-rich regions could support both early NA binding affinity and more stable ligand engagement, but the data presented here do not directly distinguish whether this occurs by increasing the apparent association rate, decreasing the apparent dissociation rate, or increasing the number of productive contacts formed in the bound state. After the initial encounter, HsORF1p may bind substrates in a manner consistent with a melting model, where the sugar-phosphate backbone is stabilized along the positively charged binding surface, disrupting secondary structure, while some of the bases remain more solvent-exposed or are directed away from the core of the cleft to allow for sampling other base pair interactions. This interpretation is consistent with prior studies showing broadHsORF1p binding across NA classes and with experiments showing thatHsORF1p can remodel imperfectly paired DNA substrates and promote annealing into more perfectly base paired duplexes under the tested conditions[6, 8, 9]. However, because high resolution structures of HsORF1p bound to different NA ligands are still lacking, this model should be viewed as a useful interpretation rather than a directly validated structural mechanism.

NA Composition Preferences

The data presented in this thesis suggest that HsORF1p engages NA in a ligand dependent manner, with differences observed based on both backbone composition, such as RNA versus ssDNA, and nitrogenous base composition, such as poly(A) RNA versus poly(U) RNA. This preference for RNAs with higher adenosine content was observed not only for WT HsORF1p, but also across the mutant constructs tested in the DSF thermal shift experiments. Poly(A) RNA specifically had an approximately five-fold higher affinity for HsORF1p than 50% GC RNA. Although these data do not directly explain the structural basis of this binding preference, they raise the possibility that the HsORF1p binding clefts engage certain NA substrates more favorably than others. Because mutant ΔTm patterns were generally comparable to WT, and because only one mixed-composition RNA was tested, these data cannot distinguish residue-specific ligand discrimination from broader sequence- or geometry-dependent stabilization effects. The strong stabilization observed with poly(A) RNA may therefore reflect more favorable engagement of purine-rich RNA, RNA specific backbone geometry, or both.

One possible explanation for this increased binding affinity and thermal stabilization is that arginine-rich regions within the HsORF1p RRM-CTD binding surface provide higher avidity interactions with adenine-rich RNA. Once the RNA backbone is positioned along the basic surface, arginine side chains may support additional stabilizing contacts with exposed bases through cation–π interactions, hydrogen bonding, or other geometry-dependent interactions. Under this model, poly(A) RNA would be more stabilizing because it can make more favorable or more frequent contacts across the binding surface, whereas pyrimidine-rich substrates may contribute fewer contacts beyond backbone engagement. Importantly, however, none of the arginine mutations tested here caused a complete loss of poly(A) associated stabilization or produced a clear switch in ligand preference. Therefore, the arginines tested in this study are better interpreted

as contributing to overlapping binding and avidity functions rather than acting as single-residue determinants of adenine specificity.

This interpretation remains indirect. The substrate panel used here does not fully separate base identity, backbone chemistry, sequence context, and ligand geometry, and these experiments do not directly identify arginine:adenine contacts. It also remains possible that non-basic residues contribute to ligand-dependent stabilization. For example, aromatic residues located along the CTD side of cleft 2 or near the base of the coiled-coil region may be positioned close enough to the predicted binding domain to contribute to NA engagement, even if their R-groups do not appear consistently oriented for direct RNA contact in available static structures. Without HsORF1p:NA bound structures, these potential contacts cannot be ruled out. Therefore, the stronger stabilization observed with poly(A) RNA should be interpreted as consistent with an arginine-supported avidity model, but not as direct proof that arginines alone drive adenine-specific recognition.

Backbone identity also appeared to influence thermal stabilization independent of nitrogenous base composition, particularly when comparing poly(A) RNA and poly(A) ssDNA with WT HsORF1p. This is consistent with the binding affinity observations and with published measurements suggesting that NA backbone properties influenceHsORF1p binding, including findings that RNA is often preferred over ssDNA despite both classes of ligand showing relatively high affinities[6, 9, 12]. In the DSF dataset, poly(A) ssDNA performed poorly as a stabilization factor relative to poly(A) RNA, whereas poly(T) ssDNA performed more similarly to poly(U) RNA. This difference may reflect loss of favorable RNA-specific contacts in the absence of the 2′ hydroxyl group, differences in preferred sugar pucker, differences in phosphate spacing between RNA and ssDNA, or some combination of these features[9, 12]. Taken together, these observations suggest

that both ssNA backbone geometry and nucleobase composition contribute toHsORF1p:NA interaction, although the relative contribution of each feature remains unresolved.

The possibility of base-sensitive behavior is also consistent with the enrichment of conserved arginines, rather than lysines, at several highly conserved positively charged positions across theHsORF1p RRM-CTD (Figure 3.3C)[8, 9, 11]. Although both lysine and arginine can make general ionic contacts with phosphates in the NA backbone, arginine contains a planar guanidinium group capable of donating multiple hydrogen bonds and supporting interaction modes that differ from the primary ammonium group of lysine. These chemical differences are consistent with the stronger phosphate-associated and condensation-promoting interactions often observed for arginine relative to lysine. In the context ofHsORF1p, conserved arginine-rich surfaces may therefore support both general backbone engagement and additional stabilizing interactions with favorable RNA geometries or base compositions. However, the data presented here do not identify which specific arginines, if any, directly mediate preferential stabilization of poly(A) RNA. Instead, the data are more consistent with a model in which multiple basic residues contribute overlapping functions to NA recruitment, avidity, and ligand-dependent stabilization.

The Importance of Adenosine Rich Sequences in L1 Biology

There are several possible connections between these *in vitro* observations and cellular L1 biology, but these connections should be interpreted cautiously. The most direct interpretation is that HsORF1p may possess biochemical properties that favor *cis* preference engagement with its own L1 RNA, beyond the advantage conferred by proximity to the transcript during translation[16-18]. Native L1 mRNA is adenosine-rich across the ORF regions, contains a long 3′ poly(A) tail, and includes a GC-rich 5′ UTR. Therefore, the strong in vitro stabilization observed with poly(A)

RNA raises the possibility that adenosine-rich regions of L1 RNA, could contribute to favorableHsORF1p engagement during RNP formation.

The L1 poly(A) tail remains biologically relevant for retrotransposition because L1 mobilization decreases when the tail is substantially shortened, although this effect could reflect impaired ORF2p recruitment, altered RNA stability, reduced translation, or changes in host protein association[19]. Thus, enhancedHsORF1p binding to poly(A) RNA does not by itself establish thatHsORF1p directly recognizes the poly(A) tail as a dedicated sequence element. Importantly, this should not be taken to mean thatHsORF1p has been shown here to bind primarily or exclusively to the poly(A) tail in cells. The data instead supports that poly(A) RNA is a favorable ligand in vitro, and this may be relevant to howHsORF1p engages adenosine-rich features of L1 RNA in cells.

This interpretation also needs to be considered alongside other cellular proteins that influence L1 RNA abundance and translation. HsORF1p's association with regions such as the poly(A) tail could also be influenced by the activity of poly(A)-binding proteins (PABPs). PABPs are particularly relevant because they bind poly(A) tails and promote mRNA stability and translation, meaning that any model involvingHsORF1p interaction with the L1 poly(A) tail would need to account for potential cooperation, competition, or indirect effects mediated through PABP-associated RNP organization. Therefore, future experiments could involve distinguishing directHsORF1p binding to poly(A)-rich regions and monitoring the presence of other L1 mRNA binding proteins such as PABPs.

A potential counterargument against the biological relevance of HsORF1p's binding preferences for adenosine rich RNA sequences is that HsORFeus mobilizes efficiently despite extensive re-coding of both ORFs, which reduces adenosine content from 40% to 27%[20].

However, HsORFeus is not a straightforward counterexample because it retains a poly(A) tail and was designed to produce higherHsORF1p and ORF2p expression, which may compensate for reduced binding efficiency or any potential effects of its altered RNA composition[20]. For this reason, future cellulo studies may benefit from revisiting naturally occurring L1 elements, particularly now that methods for detecting ORF proteins have improved[21]. For exampleHsORF1p occupancy, L1 RNA half-life, HsORF2p translation stoichiometries relative to HsORF1p, and retrotransposition efficiency across native and re-coded L1 RNAs would help determine whether RNA composition directly contributes to L1 RNP formation or whether increased protein expression can overcome otherwise unfavorable RNA features.

The apparent preference of HsORF1p for adenosine rich RNA should not be interpreted as evidence for a simple poly(A) specific binding motif, nor should it be considered only in the context of *cis* preferential binding to L1 RNA. Rather, it suggests that RNA composition is one feature that can bias HsORF1p engagement, which is relevant because HsORF1p associates with many non-L1 RNAs in cells, including polyadenylated transcripts and highly repetitive RNAs[22-26]. Although HsORF1p may exhibit binding preferences, cellular RNPs also contain non-ideal ligands, including inverted-repeat Alu containing RNAs. Their recovery suggests that HsORF1p association in cells is not limited to optimal single-stranded substrates and may occur through exposed single-stranded regions, local duplex character, or additional RNP-associated feature[16, 22, 27]. It also remains unclear when and where theseHsORF1p-associated RNAs enter the RNP cycle, whether co-translationally at the ribosome, through free cytosolicHsORF1p, or within stage 3 supersaturated RNPs. Together, these observations suggest that non-L1 RNAs may enter HsORF1p RNPs through multiple modes of association that are not fully captured by simple *in vitro* binding preferences.

Taken together, these considerations place the in vitro poly(A) preference of HsORF1p within a broader cellular framework. Adenosine-rich RNA may be one feature that favorsHsORF1p engagement, but its biological impact is likely modulated by translation-coupled cis preference, ORF protein abundance, RNA stability, PABP-associated organization, and the competing or cooperative effects of other RNA-binding proteins. This model can accommodate both the potential contribution of L1 RNA composition to RNP assembly and the ability of non-L1 RNAs, such as Alu RNAs and processed pseudogene transcripts, to access the L1 machinery. Therefore, HsORF1p binding preference is best viewed as a context-dependent contributor to L1 RNP composition and function, rather than as evidence for direct recognition of a single poly(A)-specific motif.

The Roles of the Mutated Residues Tested in this Study

These experiments were designed to test whether charge-neutralizing mutations across the HsORF1p CTD-RRM surface would produce graded or substrate-specific effects on ssNA binding and ssNA induced thermal stabilization. The original expectation was that mutations at different positions would distinguish regions that contribute unequally to nucleic-acid engagement and potentially reveal surfaces responsible for HsORF1p sequence preference. Instead, most mutants affected binding, thermal stabilization, and retrotransposition efficiency to broadly comparable extents. This pattern argues against a model in which HsORF1p sequence preference is dictated by a small number of uniquely important charged residues.

Rather, the data support a model in which productive HsORF1p:ssNA association depends on distributed and partially redundant interactions across the CTD-RRM surface. Substrate composition had a stronger effect on thermal stabilization trends than mutation position alone, suggesting that ssNA flexibility, base composition, and local structural context collectively shape

RNP stability. Thus, HsORF1p sequence preference is more likely to emerge from the combined properties of the nucleic-acid substrate and the multivalentHsORF1p trimer surface than from discrete residue-level recognition sites.

Two exceptions provided useful boundaries for this interpretation. K300A behaved distinctly from the other constructs, but its phenotype was most consistent with altered protein solubility or unbound trimer stability rather than a direct role in nucleic-acid preference. In contrast, RR261,262AA showed the strongest disruption of productive nucleic-acid engagement, consistent with prior work identifying R261 and R262 as especially important forHsORF1p binding and RNP formation. Even here, however, the data do not indicate that these residues alone encode sequence preference; instead, they appear to support broad and productive ssNA engagement.

Additional residue-specific trends further suggest that the relationship between CTD-RRM position and HsORF1p:ssNA binding is not explained by a simple cleft-based model. Mutations that destabilized the trimer in the DSF assay were located near the proposed cores of binding clefts 1 and 2, contrary to the expectation that peripheral cleft residues would be more likely to affect solubility or folding, as observed for K300A. The behavior of the third proposed binding cleft was also distinct from that of clefts 1 and 2. If this region rotates into an open conformation and contributes to nucleic-acid engagement, the data suggest that it may not function through the same mode of charge-mediated binding observed for the other proposed clefts. For example, the CC3xMut constructed both binding affinity and ssNA-induced thermal stabilization, but this could reflect altered stabilization of a specific CTD-RRM conformation rather than direct nucleic-acid contact. Structural studies of HsORF1p bound to RNA will be needed to distinguish between these possibilities.

Overall, these findings refine the view of HsORF1p:ssNA recognition by showing that the CTD-RRM surface contains residues with distinct biophysical contributions while providing limited support for a model based on many discrete sequence-specifying sites. Instead, HsORF1p sequence preference appears more likely to emerge from the combined effects of substrate composition, nucleic-acid flexibility, RNP assembly state, and distributed interactions across theHsORF1p trimer. In this model, individual residues or residue clusters can influence binding affinity, protein stability, and substrate-dependent thermal stabilization, but productive NA recognition and binding is best understood as the sum result of each of these sites that perform functions that more like one another than different.

References

1. Li, X. and N. Liu, *Advances in understanding LINE-1 regulation and function in the human genome*. Trends Genet, 2025. **41**(7): p. 577–589.
2. Hong, Y. and N. Liu, *Transposable elements in health and disease: Molecular basis and clinical implications*. Chin Med J (Engl), 2025. **138**(18): p. 2220–2233.
3. Martin, S.L., *The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition*. J Biomed Biotechnol, 2006. **2006**(1): p. 45621.
4. Martin, S.L., *Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1*. RNA Biol, 2010. **7**(6): p. 706–11.
5. Kolosha, V.O. and S.L. Martin, *High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1)*. J Biol Chem, 2003. **278**(10): p. 8112–7.
6. Martin, S.L., et al., *LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein*. J Mol Biol, 2005. **348**(3): p. 549–61.

7. Naufer, M.N., A.V. Furano, and M.C. Williams, *Protein-nucleic acid interactions of LINE-1 ORF1p*. Semin Cell Dev Biol, 2019. **86**: p. 140–149.

8. Januszyk, K., et al., *Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1*. J Biol Chem, 2007. **282**(34): p. 24893–904.

9. Khazina, E., et al., *Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition*. Nat Struct Mol Biol, 2011. **18**(9): p. 1006–14.

10. Zernia, S., et al., *LINE-1 ribonucleoprotein condensates bind DNA to enable nuclear entry during mitosis*. Sci Adv, 2025. **11**(18): p. eadt9318.

11. Khazina, E. and O. Weichenrieder, *Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame*. Proc Natl Acad Sci U S A, 2009. **106**(3): p. 731–6.

12. Kolosha, V.O. and S.L. Martin, *In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition*. Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10155–60.

13. Basame, S., et al., *Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition*. J Mol Biol, 2006. **357**(2): p. 351–7.

14. Cashen, B.A., et al., *The L1-ORF1p coiled coil enables formation of a tightly compacted nucleic acid-bound complex that is associated with retrotransposition*. Nucleic Acids Res, 2022. **50**(15): p. 8690–8699.

15. Naufer, M.N., et al., *L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling*. Nucleic Acids Res, 2016. **44**(1): p. 281–93.

16. Kroutter, E.N., et al., *The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition*. PLoS Genet, 2009. **5**(4): p. e1000458.

17. Kulpa, D.A. and J.V. Moran, *Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition*. Hum Mol Genet, 2005. **14**(21): p. 3237–48.

18. Wei, W., et al., *Human L1 retrotransposition: cis preference versus trans complementation*. Mol Cell Biol, 2001. **21**(4): p. 1429–39.

19. Janecki, D.M., et al., *LINE-1 mRNA 3' end dynamics shape its biology and retrotransposition potential*. Nucleic Acids Res, 2024. **52**(6): p. 3327–3345.

20. Adney, E.M., et al., *Comprehensive Scanning Mutagenesis of Human Retrotransposon LINE-1 Identifies Motifs Essential for Function*. Genetics, 2019. **213**(4): p. 1401–1414.

21. Nielsen, M.I., et al., *Targeted detection of endogenous LINE-1 proteins and ORF2p interactions*. Mob DNA, 2025. **16**(1): p. 3.

22. Mandal, P.K., et al., *Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles*. Hum Mol Genet, 2013. **22**(18): p. 3730–48.

23. Briggs, E.M., et al., *RIP-seq reveals LINE-1 ORF1p association with p-body enriched mRNAs*. Mob DNA, 2021. **12**(1): p. 5.

24. Taylor, M.S., et al., *Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition*. Cell, 2013. **155**(5): p. 1034–48.

25. Goodier, J.L., L.E. Cheung, and H.H. Kazazian, Jr., *Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition*. Nucleic Acids Res, 2013. **41**(15): p. 7401–19.

26. Goodier, J.L., et al., *LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex*. Mol Cell Biol, 2007. **27**(18): p. 6469–83.

27. Ahmad, S., et al., *Breaching Self-Tolerance to Alu Duplex RNA Underlies MDA5-Mediated Inflammation*. Cell, 2018. **172**(4): p. 797–810.e13.

Appendix

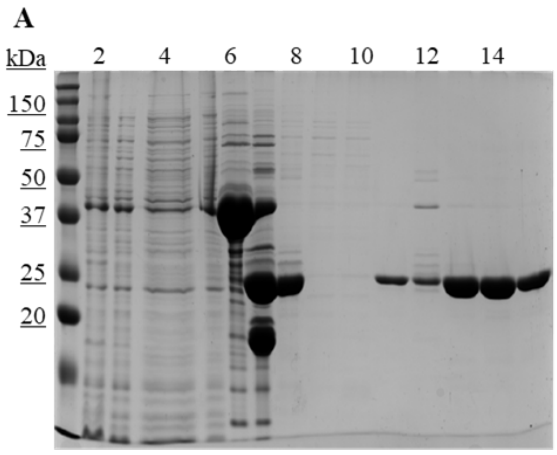
PDB	Z-score	Alignment Length	rmsd	Subject Length	Sequence identity	Description
8bwfB	7.6	65	2.2	77	11	POLYPYRIMIDINE TRACT-BINDING PROTEIN 1
6n7rC	7.2	75	3	133	15	U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KDA HOMOLOG
3r27A	7.2	66	2.3	84	17	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN L
7ntnA	7	67	2.3	77	22	TRNA (URACIL-5-)-METHYLTRANSFERASE HOMOLOG A
5vsuA	6.9	71	2.7	387	18	U4/U6 SNRNA-ASSOCIATED-SPLICING FACTOR PRP24
7q4lA	6.9	68	2.7	227	18	DEAD END PROTEIN HOMOLOG 1
6tznA	6.8	77	2.9	115	12	PROTEIN POF8
8xmdM	6.8	71	2.7	1052	17	DNA-DIRECTED RNA POLYMERASE IV SUBUNIT 1
3smzA	6.8	65	2.4	280	12	RIBONUCLEOPROTEIN PTB-BINDING 1
2hzcA	6.7	72	2.9	87	10	SPLICING FACTOR U2AF 65 KDA SUBUNIT
4wikB	6.7	69	2.7	225	16	SPLICING FACTOR, PROLINE- AND GLUTAMINE-RICH
5d77A	6.7	66	2.2	82	14	RNA-BINDING PROTEIN MIP6
6iczO	6.7	65	2.4	285	14	PROTEIN MAGO NASHI HOMOLOG 2
7w59O	6.7	65	2.4	285	14	PRE-MRNA-PROCESSING-SPLICING FACTOR 8
4zkaF	6.7	65	2.4	79	14	RNA BINDING PROTEIN FOX-1 HOMOLOG 1
4p6qA	6.7	65	2.5	286	25	MSX2-INTERACTING PROTEIN
9n93A	6.6	71	2.9	650	15	CSC1-LIKE PROTEIN 1
3sdeB	6.6	69	2.7	239	19	PARASPECKLE COMPONENT 1
5ca5A	6.6	69	2.7	248	17	NONO-1
7pu5G	6.6	69	2.8	237	19	NON-POU DOMAIN-CONTAINING OCTAMER-BINDING PROTEIN
6sxwA	6.6	66	2.6	74	15	ZINC FINGER PROTEIN 638
3h2uB	6.6	65	2.4	279	12	VINCULIN
7oqcA	6.5	76	3	132	13	PROTEIN NAM8
3jb9k	6.5	68	2.8	89	12	PRE-MRNA-SPLICING FACTOR SPP42
2vooB	6.5	67	2.7	177	19	LUPUS LA PROTEIN

PDB	Z-score	Alignment Length	rmsd	Subject Length	Sequence identity	Description
5vzlC	4.8	56	2.2	87	7	SINGLE GUIDE RNA (116-MER)
8fg6B	4.4	58	3.1	151	12	AMYLOIDOGENIC PEPTIDE
9qbqC	4.4	55	3	423	11	DUF4365 DOMAIN-CONTAINING PROTEIN
5wi2B	3.9	54	2.7	93	7	CDNA FLJ56409, HIGHLY SIMILAR TO SERINE/THREONINE
6jsxA	3.9	53	3.9	73	9	FLAGELLAR BIOSYNTHESIS PROTEIN FLAG
3sooA	3.9	40	3.3	87	35	LINE-1 TYPE TRANSPOSASE DOMAIN-CONTAINING PROTEIN
4woeA	3.8	58	2.9	716	10	TAUROCYAMINE KINASE
3ln7A	3.8	49	2.6	747	16	GLUTATHIONE BIOSYNTHESIS BIFUNCTIONAL PROTEIN GSH
3pq1B	3.8	44	2.1	321	9	POLY(A) RNA POLYMERASE
7waeA	3.7	54	3.7	877	6	CYANOPHYCIN SYNTHASE
6m1uA	3.6	57	2.8	122	11	RNA N6-ADENOSINE-METHYLTRANSFERASE METTL16,RNA N6
6tv6A	3.5	58	3	362	10	PROTEIN-ARGININE KINASE
2ehbD	3.5	57	2.9	126	11	CALCINEURIN B-LIKE PROTEIN 4
8optA	3.5	43	2.2	783	9	TERMINAL URIDYLYLTRANSFERASE 7
7wrgA	3.4	56	2.8	812	4	KINESIN-LIKE PROTEIN
4jpdA	3.3	60	3.1	109	12	PROTEIN CYAY
7bgsB	3.3	52	2.6	129	17	HOLLIDAY JUNCTION RESOLVASE
2hjjA	3.3	49	2.8	66	12	HYPOTHETICAL PROTEIN YKFF
6jomA	3.2	57	3.2	331	14	LIPOATE--PROTEIN LIGASE
3osmA	3.1	60	3.8	119	5	SERINE/THREONINE-PROTEIN KINASE KCC4
1sgoA	3.1	58	3.1	139	17	PROTEIN C14ORF129
1dq3A	3.1	55	3	454	9	ENDONUCLEASE
8ovwP	3.1	53	2.8	257	8	CENTROMERE-BINDING PROTEIN 1
2ltrA	3.1	51	2.7	105	10	PROTEIN RDE-4
1ob8B	3.1	51	2.9	127	10	HOLLIDAY-JUNCTION RESOLVASE

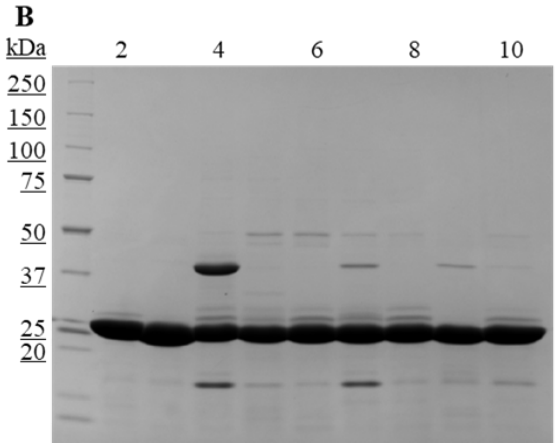
Supplemental Tables 1 & 2 | Top 25 DALI results from searching in silico truncations of the HsORF1p crystal structure (PDB 2YKO). *All results were first filtered by highest Z-score and then by highest alignment length, defined as the number of residues structurally aligned between the query and hit. The results were then edited to remove HsORF1p PDB hits and duplicate descriptions. The top table shows results associated with the isolated RRM (150-230) domain, and the bottom table shows results associated with the CTD (231-326).*

Broad Categorization	Nucleic Acid Name	Nucleic Acid Sequence
Poly(A) homopolymers	27-nt poly(A) ssDNA	AAAAAAAAAAAAAAAAAAAAAAAAAAA
	27-nt poly(A) RNA	rArArArArArArArArArArArArArArArArArArArArArArArArArArArA
	36-nt poly(A) RNA	rArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArArA
pyrimidine Homopolymers	27-nt poly(T) ssDNA	TTTTTTTTTTTTTTTTTTTTTTTTTTT
	27-nt poly(U) RNA	rUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrU
	36-nt poly(U) RNA	rUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrU
GC-balanced Heteropolymers	27-nt 50% GC RNA	rCrArCrUrArArArGrArCrGrUrArArArGrGrUrArGrArCrUrCrCrArU
	36-nt 50% GC RNA	rUrGrCrCrCrArCrUrArArArGrArCrGrUrArArArGrGrUrArGrArCrUrCrCrArUrGrGrCrCrU
	27-nt 50% GC ssDNA	CACTAAAGACGTAAAGGTAGACTCCAT
	36-nt 50% GC ssDNA	TGCCCACTAAAGACGTAAAGGTAGACTCCATGGCCT

Supplemental Table 3 *| A list of the RNA and ssDNA sequences that were used in this study.*

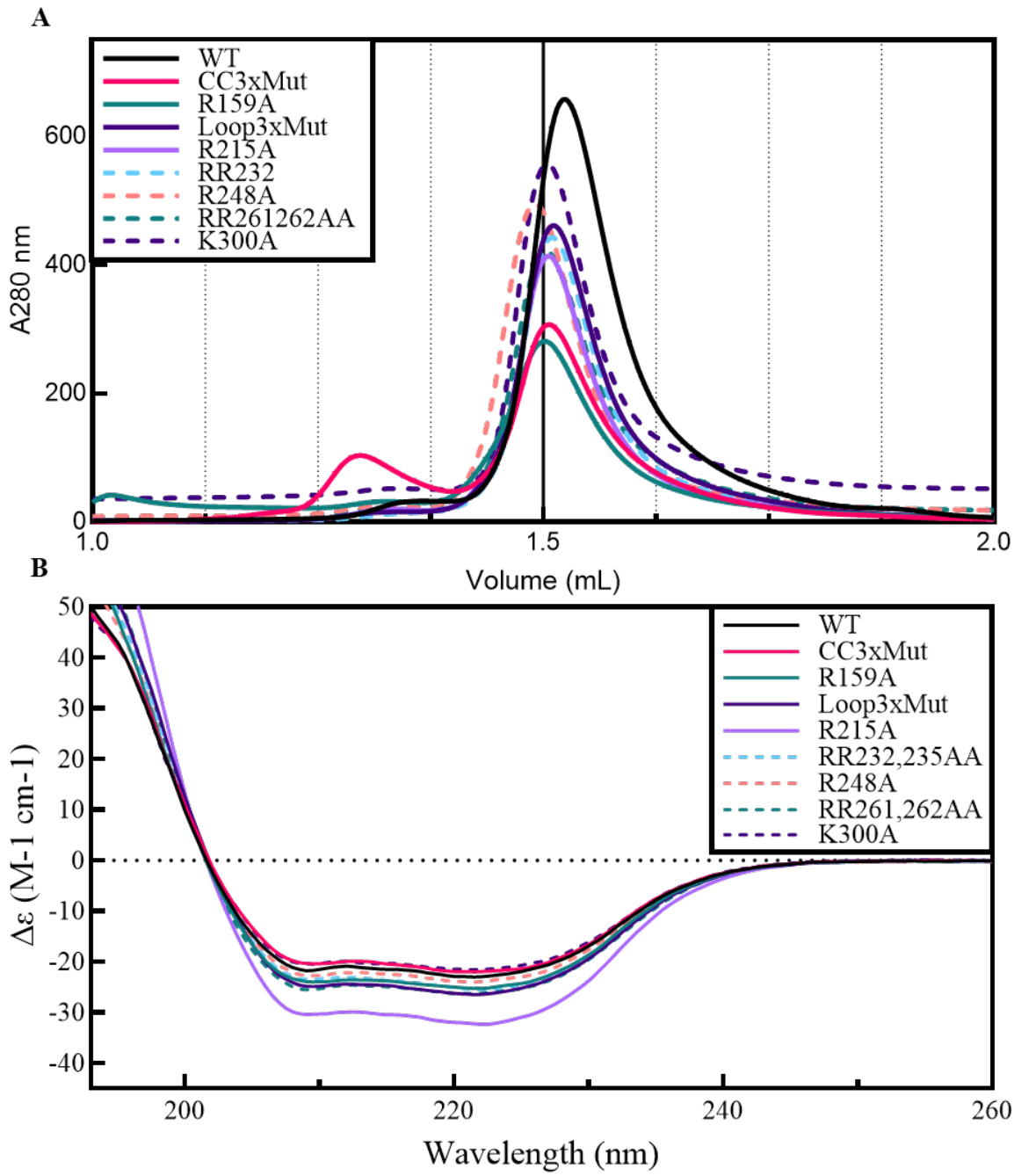


Lane	Sample
1	Kaleidoscope Ladder
2	Whole Cell Lysate
3	Nickel Resin Flowthrough
4	Nickel Resin Wash
5	Lysate Supernatant
6	After Nickle Elution, Before ULP1 Digest
7	After ULP1 Digest
8	After ULP1 Subtraction
9	Heparin Flow Through Fraction
10	Heparin Flow Through Fraction
11	Heparin Elution Fraction
12	Heparin Elution Fraction
13	Heparin Elution Fraction
14	Heparin Elution Fraction
15	Heparin Elution Fraction

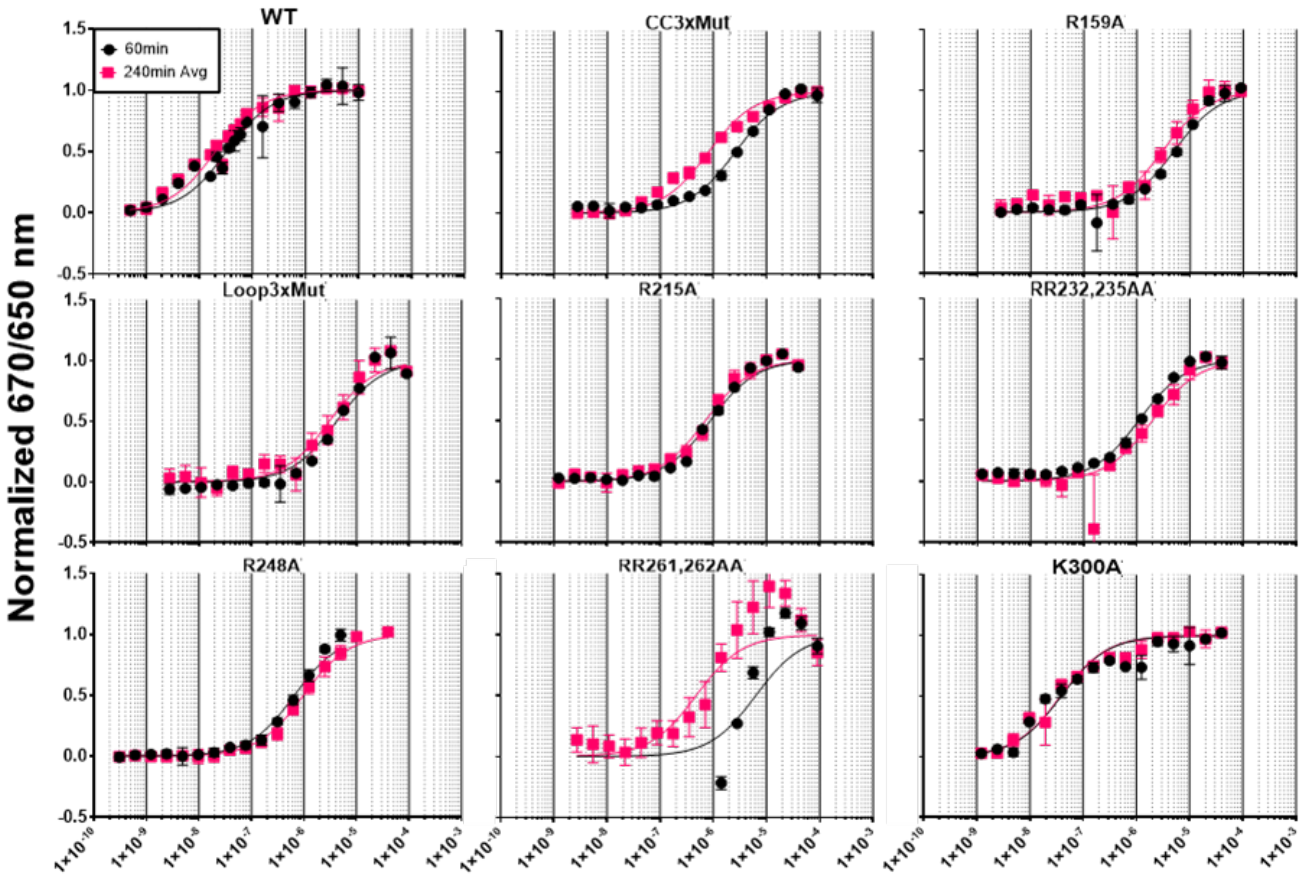


Lane	Sample
1	Kaleidoscope Ladder
2	HsORF1p(106_326)WT
3	HsORF1p(106_326)CC3xMut
4	HsORF1p(106_326)R159A
5	HsORF1p(106_326)Loop3xMut
6	HsORF1p(106_326)R215A
7	HsORF1p(106_326)RR232
8	HsORF1p(106_326)R248A
9	HsORF1p(106_326)RR261262AA
10	HsORF1p(106_326)K300A

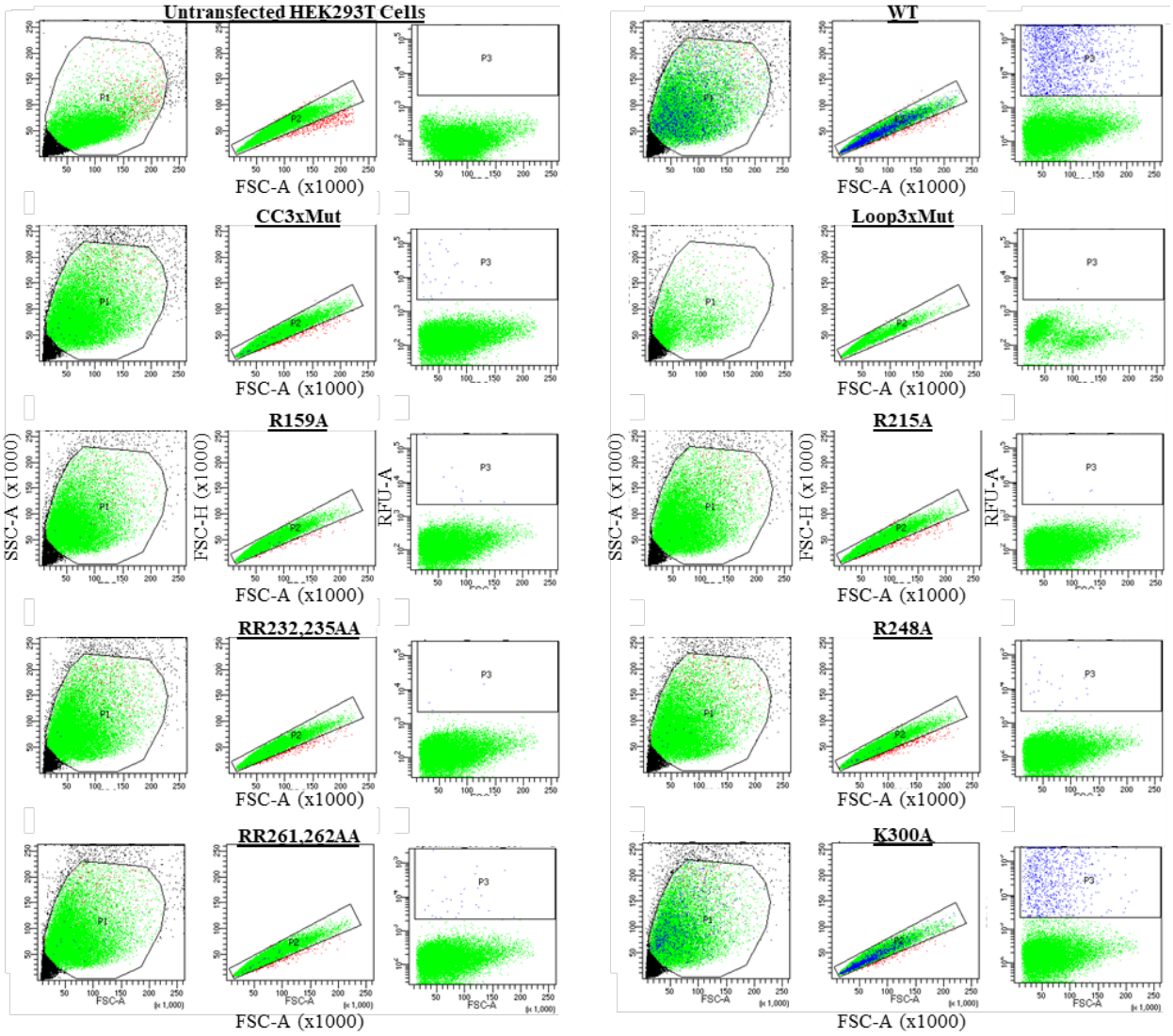
Supplemental Figure 1 | Coomassie stained SDS-PAGE gels of HsORF1p(106-326) constructs used for biophysical measurements. ***A.*** *Gel of samples collected throughout the purification protocol of WT HsORF1p(106–326). Sample prep order was from: cell lysis, lysate clarification by ultracentrifugation, filtering, nickel resin affinity purification, overnight dialysis with ULP1 digestion to remove his-tag, filtering, subtraction by incubating overnight digest with cleaned ni resin to remove undigested protein, heparin affinity purification and then centrifugal concentration. All lanes are indicated by the adjacent table and were loaded in this sequential order of the purification, except lane 5 which contains clarified lysate supernatant loaded after whole-cell lysate was applied to the nickel resin.* ***B.*** *Protein gel of all the HsORF1p(106-326) constructs after purification.*



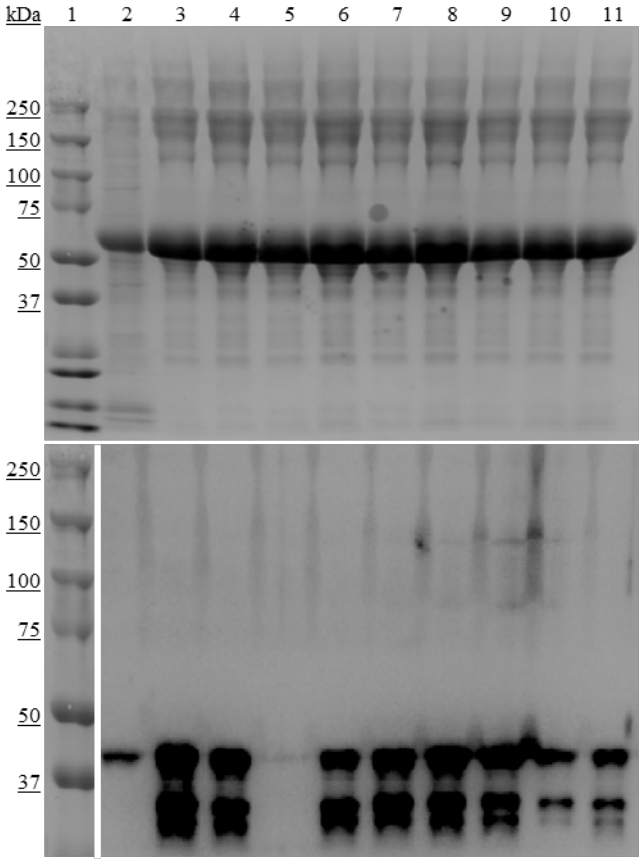
Supplemental Figure 2 | Quality control analysis of HsORF1p(106–326) constructs used for biophysical measurements. ***A.*** *Analytical size exclusion chromatography (SEC) of purified proteins performed on a Superdex 200 Increase 3.2/300 GL column in 500 mM NaCl and 10 mM HEPES pH 8. Elution profiles are plotted as absorbance at 280 nm (A280) versus elution volume.* ***B.*** *Circular dichroism (CD) spectra collected in 50 mM sodium phosphate buffer (pH 7.4). All proteins were analyzed at ~3 μM trimer concentration. Data suggests that the secondary structure and overall shape of the trimers are not compromised by the introduction of these mutations.*



Supplemental Figure 3 | Ensuring enough time has been allotted for equilibrium to be achieved in Monolith X spectral shift experiments. *~5 nM 36-nt 50% GC RNA was mixed with each construct titration in 510 mM NaCl and incubated for 1 or 4 hours (black and pink, respectively). The lack of differences between time points across most constructs suggests that equilibrium is achieved in less than 1 hour.*

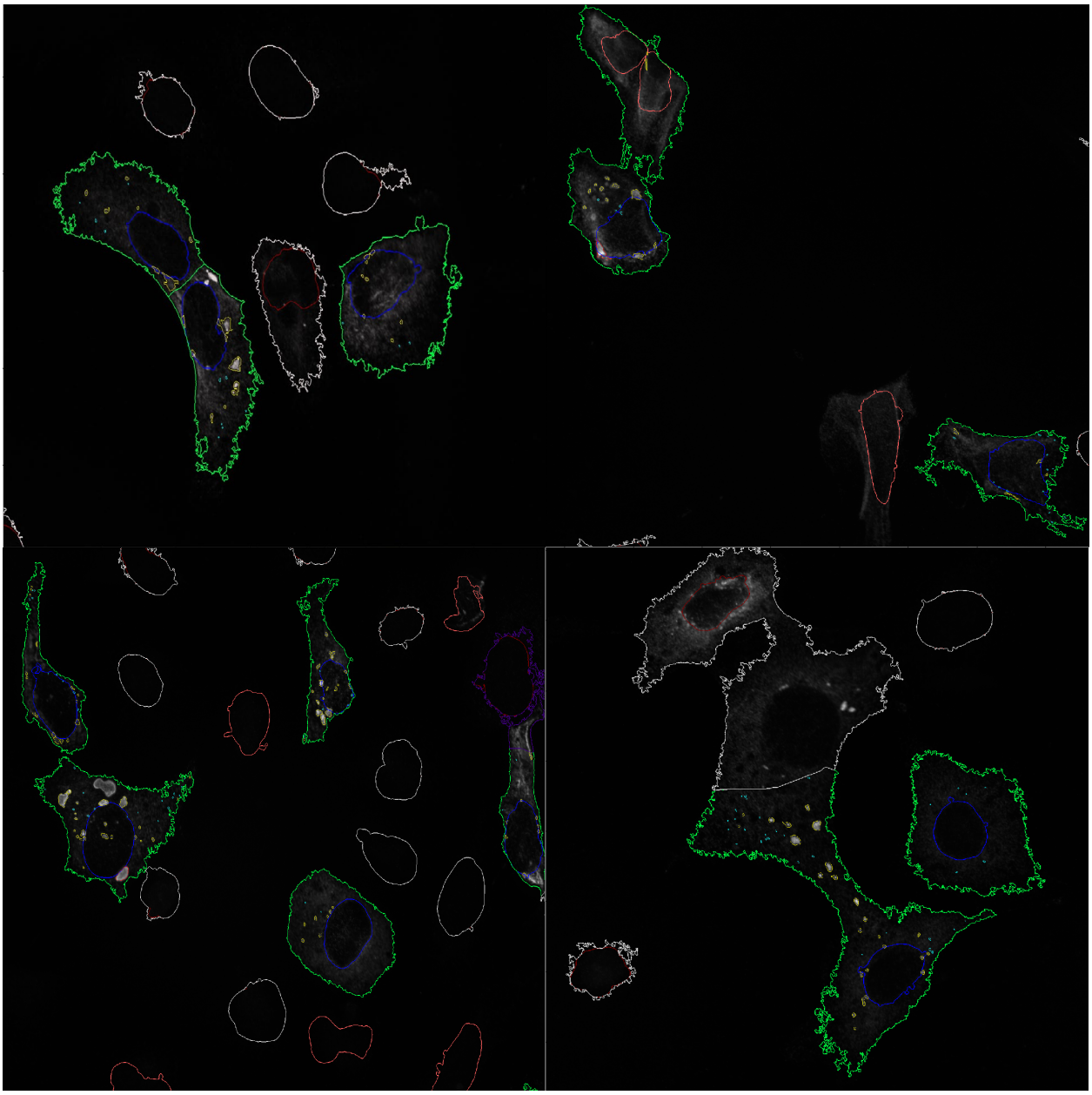


Supplemental Figure 4 | Representative flow cytometry gating output from the retrotransposition assay. *Events were gated by forward scatter area and side scatter area measurements to remove potential debris and/or potentially unhealthy cells (left). Then a forward side scatter area forward side scatter height gate was used to select singlets (middle) and a fluorescence detector channel was used to identify the cells that produced enough fluorescent green signal (right, blue).*



Lane	Sample
1	Kaleidoscope Ladder
2	HEK293T Non-Transfected Control ½ Conc.
3	HEK293T **WT** ORFeus Transfection
4	HEK293T CC**3xMut** ORFeus Transfection
5	HEK293T **R159A** ORFeus Transfection
6	HEK293T **Loop3xMut** ORFeus Transfection
7	HEK293T **R248A** ORFeus Transfection
8	HEK293T **R215A** ORFeus Transfection
9	HEK293T **RR232,235AA** ORFeus Transfection
10	HEK293T **RR261,262AA** ORFeus Transfection
11	HEK293T **K300A** ORFeus Transfection

Supplemental Figure 5 | Relative HsORFeus expression in HEK293T cells. *Remaining HEK293T cells from the third retrotransposition assay were collected after flow cytometry. Cells were pelleted by centrifugation at 4 °C and 600 × g for 10 min, resuspended in RIPA buffer, and quantified by BCA assay. Samples were then diluted to equal total protein concentration, separated by SDS-PAGE, and transferred to a PVDF membrane. Total protein loading was assessed by Ponceau S staining of the membrane (top). After removal of the Ponceau stain with DI water, ORF1p was detected by Western blot using an Abcam anti-HsORF1p primary antibody (EPR21844-108) and HRP-conjugated secondary antibodies (bottom). The ladder from the Ponceau-stained membrane was transposed next to the Western blot to indicate approximate molecular weights.*



Supplemental Figure 6 | Foci Counting with CellProfiler. *Four fields of view from different construct expressing U2OS cells were randomly selected and run through the CellProfiler pipeline. The field of views were then cropped to show 204 cells that were used for quantification. For the pipeline, after collecting fields of view, images were compressed into maximum intensity z-stacks, and only the DAPI and 4H1 (488 nm channel) channels were used for analysis in CellProfiler. Healthy nuclei are outlined in blue, while misshapen or touching nuclei are outlined in gray and pink. Healthy cells are outlined in green, and those filtered out for insufficient area are outlined in purple. Identified small foci are outlined in cyan, and identified medium foci are outlined in yellow.*