UNIVERSITY OF CALIFORNIA
SANTA CRUZ

CONFORMATIONAL DYNAMICS IN THE STEM-TERMINAL ELEMENT OF THE *KLUYVEROMYCES LACTIS* TELOMERASE RNA

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MICROBIOLOGY and ENVIRONMENTAL TOXICOLOGY

by

Cherie Musgrove

September 2017

The Dissertation of Cherie Musgrove is approved:

_____
Professor Michael Stone, chair

_____
Professor Seth Rubin

_____
Professor Manel Camps

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

List of Figures v

List of Tables vi

Abstract vii

Acknowledgements and Dedication x

CHAPTER I SUBTLETIES OF STRUCTURE IN TELOMERASE RNA

TELOMERES AND TELOMERASE 1

TERT STRUCTURE 2

TR ORGANIZATION 3

VERTEBRATE CR4/5 STRUCTURE AND FUNCTION 6

REMAINING QUESTIONS OF STRUCTURE AND FUNCTION 14

REFERENCES 16

CHAPTER II. PROBING THE STRUCTURAL DYNAMICS OF *KLUYVEROMYCES LACTIS* CR4/5 THROUGH SINGLE-MOLECULE FRET

INTRODUCTION 21

METHODS 28

BUILDING RNA CONSTRUCTS 28

FRET MEASUREMENTS 38

RESULTS 41

FRET MEASUREMENTS 48

CONCLUSIONS AND FUTURE DIRECTIONS 64

REFERENCES 71

CHAPTER III. POTENTIAL ROLE OF CR4/5 IN THE COMPENSATION OF TORSIONAL STRAIN ACCRUED DURING TELOMERE REPEAT REPLICATION AND TRANSLOCATION

INTRODUCTION....74

THE CR4/5 AND TRANSLOCATION....74

THE CR4/5 AND TELOMERE REPEAT ADDITION....77

CURRENT MODELS OF TRANSLOCATION....79

TORQUE, STRAIN, AND HYBRID DISSOCIATION IN RNA POLYMERASE....83

EXPERIMENTAL AVENUES FOR PROBING TORSION AND ROTATION IN TELOMERASE....87

CONCLUDING REMARKS....90

REFERENCES....91

CHAPTER IV. THE INFIDELITY OF REPLICATION: STUDYING THE DETERMINANTS OF HIV-1 REVERSE TRANSCRIPTASE

INTRODUCTION....94

SEQUENCE CONTEXT AND FIDELITY....95

DRUG RESISTANCE AND THE POSITIVE SELECTION OF GENERATOR MOTIFS....97

CURRENT METHODS OF MEASURING FIDELITY....98

METHODS AND RESULTS....99

DISCUSSION....113

REFERENCES....114

LIST OF FIGURES

CHAPTER I SUBTLETIES OF STRUCTURE IN TELOMERASE RNA

FIGURE 1.1 *TR SECONDARY STRUCTURES FOR KLUYVEROMYCES LACTIS, SACCHAROMYCES CEREVISIAE, ORYZIAS LATIPES (MEDAKA FISH), AND HOMO SAPIENS* *5*

FIGURE 1.2. *COMPARISON OF SECONDARY STRUCTURE AND NMR STRUCTURES OF THE MEDAKA CR4/5 REGION* *8*

FIGURE 1.3 *U182 AND G189 MAKE SPECIFIC CONTACTS WITH TRBD* *10*

FIGURE 1.4. *CHANGES IN THE POSITIONS OF BASES IN THE MEDAKA CR4/5 JUNCTION* *12*

FIGURE 1.5 *CHANGES IN THE ARRANGEMENTS OF PAIRED AND UNPAIRED JUNCTION BASES*..... *13*

CHAPTER II. PROBING THE STRUCTURAL DYNAMICS OF *KLUYVEROMYCES LACTIS* CR4/5 THROUGH SINGLE-MOLECULE FRET

FIGURE 2.1A. *TR SECONDARY STRUCTURE FOR KLUYVEROMYCES LACTIS, SACCHAROMYCES CEREVISIAE, ORYZIAS LATIPES AND HOMO SAPIENS* *24*

FIGURE 2.1B *CR4/5 SECONDARY STRUCTURE FOR KLUYVEROMYCES LACTIS, ORYZIAS LATIPES AND HOMO SAPIENS*..... *25*

FIGURE 2.2 *COMPARISON OF LABELING EFFICIENCY* *41*

FIGURE 2.3 *COMPARISON OF THE 5-NU MODIFIED AND 5-LC-NU MODIFIED PRECURSOR INTEGRITY* *43*

FIGURE 2.4 *EXAMPLE OF TEST LIGATIONS FOR THE DELTA A60 MUTANT* *45*

FIGURE 2.5 *EXAMPLE PAGE OF TWO PREPARATIVE SCALE LIGATIONS* *47*

FIGURE 2.6 *PAGE OF ALL CONSTRUCTS*..... *48*

FIGURE 2.7 *HISTOGRAM AND TRACE EXAMPLES FOR T1 AND T2 WILDTYPE CONSTRUCTS* *49*

FIGURE 2.8 *POPULATION HISTOGRAMS AND EXAMPLE TRACES OF T3 WILDTYPE* *50*

FIGURE 2.9 *POPULATION HISTOGRAMS OF SORTED WILDTYPE CONSTRUCTS* *52*

FIGURE 2.10 *POPULATION HISTOGRAMS OF SORTED WILDTYPE T3 AT 0-40MM MAGNESIUM* *55*

FIGURE 2.11 *DISTRIBUTION OF FRET POPULATIONS FOR ALL WILDTYPE CONSTRUCTS AT 0-10MM MAGNESIUM* *56*

FIGURE 2.12 *ADDITION OF EST2 TO WILDTYPE T3* 58
FIGURE 2.13 *EMSA OF EST2 PROTEIN USED IN SMFRET* 58
FIGURE 2.14 *A60 DELETION MUTANTS AT 1MM MAGNESIUM* 59
FIGURE 2.15 *T3ΔA60 MUTANT SORTED AT 0MM, 1MM AND 10MM MAGNESIUM*..61
FIGURE 2.16 *TOTAL DATA FOR ALL ΔA60 MUTANTS AT 0-10MM MAGNESIUM* 63
FIGURE 2.17. *MODELS OF K. LACTIS CR4/5 AND CORRESPONDING DISTANCES ESTIMATED FROM FRET DATA* 66

CHAPTER IV. THE INFIDELITY OF REPLICATION: STUDYING THE DETERMINANTS OF HIV-1 REVERSE TRANSCRIPTASE

FIGURE 4.1 *RESULTS OF DIRECT COMPLEMENTATION OF HIV-RT IN JS200 CELLS* 100
FIGURE 4.2 *EXPERIMENTAL FLOWCHART FOR INDIRECT COMPLEMENTATION*.....101
FIGURE 4.3 *ESTABLISHMENT OF THE PLA23O PLASMID* 102
FIGURE 4.4 *MAINTENANCE OF THE PLA23O PLASMID* 103
FIGURE 4.5 *ESTABLISHMENT OF THE PGFPUV PLASMID* 104
FIGURE 4.6 *MAINTENANCE OF THE PGFPUV PLASMID* 105
FIGURE 4.7 *EXPERIMENTAL FLOWCHART FOR FIDELITY ASSAY*.....106
FIGURE 4.8 *INDEPENDENT, NON-CLONAL MUTATION EVENTS IN THE PGFPUV REPORTER*.....107
FIGURE 4.9 *COMPARISON OF THE MUTATIONAL SPECTRUM OF HIV-RT VERSUS POL I*.....108
FIGURE 4.10 *HOTSPOT LOCATIONS IN THE PGFPUV REPORTER* 108
FIGURE 4.11 *DESIGN DETAILS OF THE PGK REPORTER PLASMIDS* 111
FIGURE 4.12 *OVERVIEW OF SEQUENCE ARRANGEMENTS IN THE FOUR PGK REPORTER PLASMIDS* 112

LIST OF TABLES

CHAPTER II. PROBING THE STRUCTURAL DYNAMICS OF *KLUYVEROMYCES LACTIS* CR4/5 THROUGH SINGLE-MOLECULE FRET

TABLE 2.1. *NOMENCLATURE AND PROPERTIES OF RNA AND DNA OLIGONUCLEOTIDES USED IN THIS STUDY* 33
TABLE 2.2. *EXPECTED FRAGMENT SIZES OF ALL POTENTIAL LIGATION PRODUCTS* .34

Conformational Dynamics in the Stem-Terminal Element of the *Kluyveromyces lactis* Telomerase RNA

Cherie Musgrove

Abstract

Telomerase is a unique eukaryotic enzyme responsible for maintaining the genomic integrity of chromosome ends, through the addition of protective DNA sequences called telomeres. The telomerase holoenzyme is composed of a protein component called TERT, and an RNA component called TER. TERT is a well conserved protein that contains many domains, one of which is the reverse transcriptase responsible for processive addition of telomere repeats to the chromosomal DNA. TER carries several necessary functional regions, which play critical roles in the activity of telomerase. These include the RNA template that is used as the basis for telomere replication, as well as a pseudoknot fold and a three-way junction called CR4/5. Overall TER displays much less sequence conservation than its protein counterpart, but its functional domains bear significant structural similarity throughout Eukarya. The CR4/5 region plays a significant, yet enigmatic role in the mechanism of telomerase. In this work I aim to further define the functional roles of the CR4/5 region through study of its structural dynamics, using single-molecule FRET. I investigate the conformational interconversions displayed by the yeast Kluyveromyces lactis CR4/5, and the impact that magnesium and junction sequence have on these dynamics. Further, I examine how models of the K. lactis CR4/5 correspond to the solved structures of CR4/5 from other organisms, how the dynamics found in K. lactis can be related to these organisms, and what insights are

gained by real-time observations of this region. Lastly, I discuss the potential roles played by the CR4/5 region in the overall telomerase mechanism.

Acknowledgement and Dedication

I would like to thank Dr. Michael Stone and Dr. Manel Camps for their mentorship and guidance throughout my graduate career. I am sincerely grateful to each of them for their patience and generosity. I would also like to thank Dr. Fitnat Yildiz and Dr. Karen Ottemann for their support and kindness, and for encouraging me when I needed it.

I would also like to thank my friends, both old and new, who have welcomed me into their hearts. Each of these people have opened their homes to me when I would have otherwise been bereft, and without their help I would never have made it this far: Kathy Shearer and Teri Sakshaug, Alexander Shapiro, Professors Lisa Wrischnik and Gregg Jongeward, Irene Kessel and Ben Kessel. You believed I could succeed when I did not, and I am eternally grateful.

I dedicate this work to Brandon Heck and my grandfather William Musgrove. My grandfather always encouraged me in my studies, and took an active interest in my university education. His love and support allowed me to pursue a college career, and through this work I honor his memory, and his place in my life. In Brandon I have found my partner and best friend, and I look forward eagerly to our life together.

Chapter I: Subtleties of Structure in Telomerase RNA

Telomeres are sequences of repetitive DNA that are found on the ends of linear chromosomes. Exactly when telomeres arose in the timeline of life is not known, but it's likely that most eukaryotes have telomeres at the ends of their chromosomes[1–3]. Telomeres serve a critical function; they protect the ends of linear genomes from the gradual degradation that inevitably occurs when DNA is replicated. This degradation is the result of the end replication problem: DNA polymerases cannot begin replication *de novo*, they require a pre-existing primer. On the lagging strand, primers are placed on the DNA at short, regular intervals approximately 150-200[1] bases apart. After DNA is extended from a particular primer, the primer is degraded, and the resultant gap in DNA is filled in by a different DNA polymerase. But, what happens when the DNA being copied is linear instead of circular? There is a point at which there is no more DNA beyond where the primer is laid down. When the primer is degraded, there is no way to fill-in the remaining DNA. This is the problem with copying the ends of linear chromosomes; no DNA polymerase known can.

Eukaryotes have co-opted a reverse transcriptase in a clever work-around to this problem. Rather than alter the way DNA replication is done, instead simply add more DNA to the ends that need to be copied. The extra DNA must not be important coding DNA and therefore it will not be detrimental to the organism to lose it. Thus, the telomere exists. A long stretch of repetitive non-coding DNA that is added to the

[1] Eukaryotic Okazaki fragments are shorter than their counterpart in prokaryotic cells, which are 1,500 to 2,000 bases long.

ends of all chromosomes by a unique reverse transcriptase called telomerase. A simple solution, that requires a very complex protein to execute it.

Telomerase is an RNP comprised of the protein component TERT and the RNA component TR. Across phylogenies, researchers have found that both the protein and RNA of telomerase contain many functional domains that are structurally conserved[4–6].

TERT Structure

TERT contains several canonical domains: the telomerase essential N-terminal (TEN) domain[7,8], the reverse transcriptase (RT)[5,9], the telomerase RNA-binding domain (TRBD)[10], and the C-terminal extension (CTE)[11]. Within these domains are many conserved motifs; for example, the CP and T-FLY motifs in the TRBD, and the GQ motif in the TEN domain [7,8,12]. The RT domain has many motifs consistent with other reverse transcriptases, such as the 1 and 2 motifs (fingers), the 3 motif (template grip), the ABCDE and IFD motifs (palm and primer grip)[5,9,13]. The CTE domain is a homolog of the canonical thumb domain of reverse transcriptases, and as such is important for retaining telomere DNA binding[11,14]. Once TERT has found the end of a telomere, it is predicted to adopt a conformation similar to the classic hand shape common to polymerases[15]; the template and 3' end of the primer would sit in the palm where the active site is located, the fingers hold the primer/template hybrid, and the thumb closes around the single-stranded template to facilitate processive replication[6,16]. While the structure and organization of TERT is an engaging field, this work focuses on the RNA component of telomerase, TR.

TR Organization

TR has more divergent sequence than TERT (Fig 1.1, top). It is variable size across phyla, ranging from over 2000 bases in *Neurospora* to just under 150 bases in killifish[17]. Vertebrates on average range between 300-500 bases. Much of this extensive sequence is hyper-variable, and unnecessary for the strict function of telomerase[18,19]. It acts as a scaffold around which the various protein cofactors can assemble[20,21]. There are two components of TR, the core and CR4/5, that are required for telomerase to achieve its most basic function; the addition of nucleotides to a DNA strand.

The Core

The core of telomerase RNA includes the template region, the pseudoknot, the template-boundary element, and the core-enclosing helix (CEH). The template is a single-stranded region composed of 1.5 telomere repeats in humans which the reverse transcriptase uses as the template for addition of telomere repeats to the lagging strand. The template boundary element acts to stop run-through reverse transcription past the template. In *Tetrahymena* this occurs through a small helix that physically blocks the TBE from being pulled into the active site of telomerase[22], forcing a pause followed by translocation of the protein. Translocation realigns the RT active site to the end of the DNA strand, where the template is positioned through RNA: DNA hybrid formation with half of a telomere repeat[23,24]. The core-enclosing helix keeps the template region close to the pseudoknot[25], and in humans serves dual function as a template-boundary element[26]. The pseudoknot is a vital part of telomerase activity,

yet its actual function is not clear. It is required for catalysis[19,27] and assembly[28] but most of the structure is located far from the active site of TERT[29]. There are well conserved three-stranded helical regions in the pseudoknot, but the linkers and junctions around these helices can vary in size significantly[30]. Recent structural modeling and cryo-EM work suggest that the pseudoknot wraps around the outside of TERT[31,32], and may even slide back-forth across the dorsal surface of the protein as nucleotide addition proceeds[31]. *In vitro* FRET performed by Yeoman *et al.* using full-length human TR suggested that the pseudoknot could only form stably in the presence of TERT, and that free hTR took on a very different structure than TERT•TR[33]. Further evidence for this hypothesis came in the form of SHAPE data combined with dynamic molecular modeling of *in vitro* transcribed *Tetrahymena* TER, indicating that the pseudoknot was prevented from forming in a free full-length TER because the template was instead involved in double stranded interactions, and that there was an overall conformational change in *t*TR when binding to TERT[34]. This lead researchers to revive the hypothesis that the pseudoknot may act as a sensor for the presence of TERT *in vivo*; when TERT is available the pseudoknot rearranges to allow binding. However, compelling contrary evidence of this hypothesis is provided by two *in vivo* studies; live cell imaging work in HeLa cells[35], and DMS protection assays in HEK293 cells[36]. Both these studies show only small changes in the structure of the pseudoknot, and even the core itself, in the presence/absence of TERT. In addition to significant insight into the biology of the pseudoknot, these studies have

highlighted the shortcomings of relying too heavily on *in vitro* work alone; it is critical for the field to integrate *in vitro* work with thorough *in vivo* investigations.

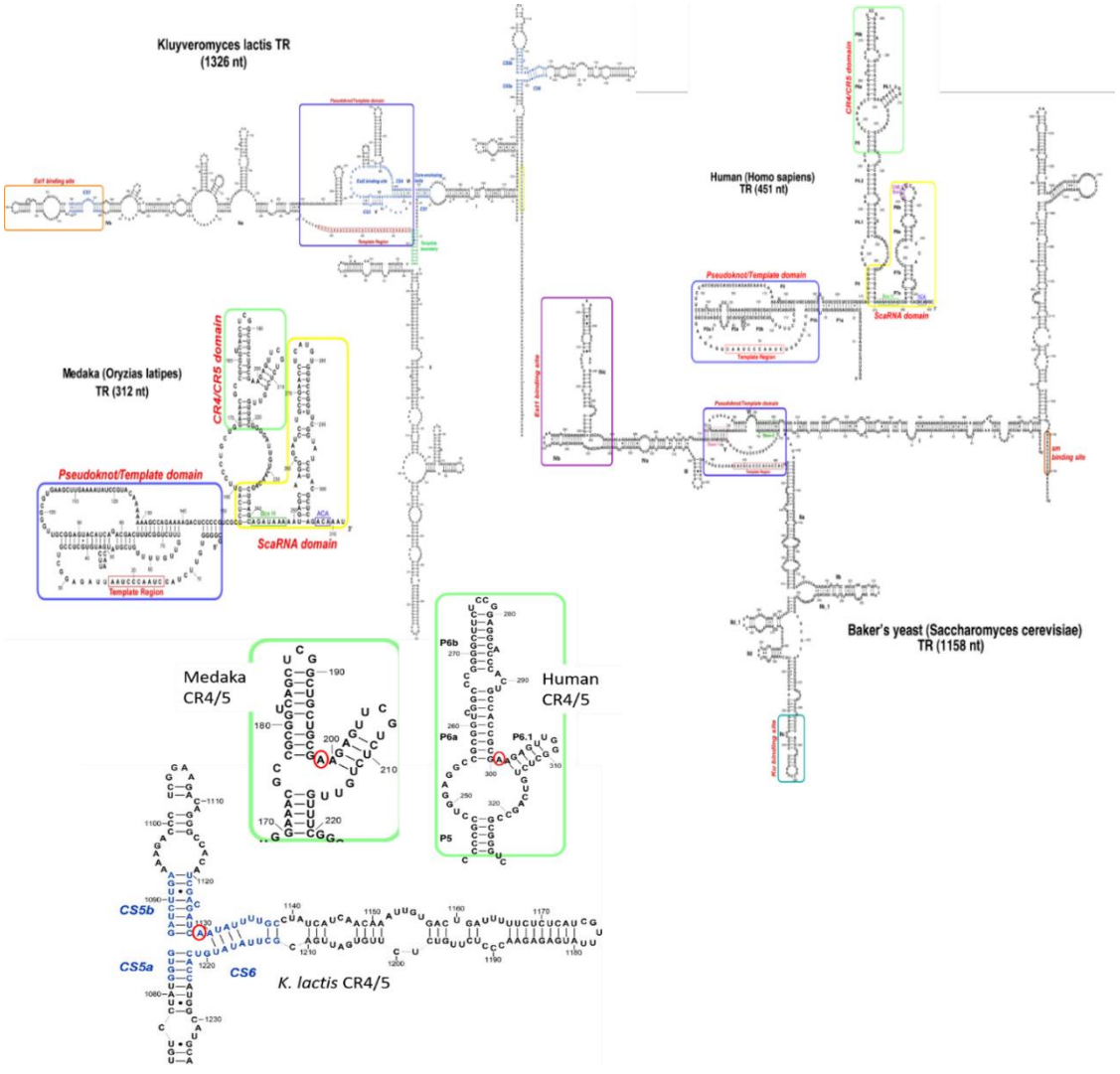


Figure.1.1 TR secondary structures for Kluyveromyces lactis, Saccharomyces cerevisiae, Oryzias latipes (medaka fish), and Homo sapiens. Top; Functional areas are boxed; blue pseudoknot, yellow/lime green H/ACA domains, green CR4/5, purple/orange Est1 binding site, red Sm binding site, and teal Ku binding site. CR4/5 domain in K. lactis nucleotides are in blue. Images of the secondary structures have been taken from the Telomerase Database, under the creative commons liscence. Bottom; CR4/5 secondary sequences. Invariant junction A nucleotides is circles in red in all structures.

CR4/5

The second region required for telomerase activity is the CR4/5 domain, also known as the three-way junction (TWJ), the stem-terminus element (STE) or in *Tetrahymena* the homologous Stem Loop IV (SL IV) (Fig 1.1, bottom). For clarity, all references in this work to the CR4/5, STE, SLIV, or TWJ will be referred to using the vertebrate nomenclature, CR4/5. In medaka fish (*oryzias latipes*), the best characterized CR4/5, much of the region makes direct contacts with the TRBD. The primary role of the CR4/5 domain is RNA-protein binding. Before the high-resolution structure of TRBD•CR4/5 was available, mutational analysis suggested that CR4/5 could play some direct role in catalysis[19,37]. Now it is clear that base substitutions in the stems of CR4/5 that impact catalytic activity are simply fatally disruptive to TRBD•RNA binding[38,39]. However, the position and structure of the loop of P6.1 is still unknown, and it is likely to impact catalysis in some way, potentially through interaction with the template or template-boundary element.

VERTEBRATE CR4/5 STRUCTURE AND FUNCTION

In most organisms CR4/5 is comprised of three helical regions of RNA joined by a central junction. For simplicity, all references in this work to the stems of CR4/5 will follow the nomenclature used for vertebrates. CR4/5 has the highest binding affinity for TERT, higher even than the pseudoknot or template[40]. The base stem, P5, is short and acts to support the other stems, it does not impact activity or binding. The second stem, P6, is longer and stacks co-axially with P5. Stem P6 has only one nucleotide important for specific binding to TRBD, while the rest of P6 provides strong non-

specific binding via the helical backbone[41]. The third stem, P6.1, is critical for activity and binding. It has many contacts with the surface of TRBD (Fig 1.2), as well as the CTE[36]. Crosslinking data shows that P6.1 sits along the interface between the TRBD and the CTE[42]. Overall, in the absence of TRBD, the stems form a Y shape, with the P5 and P6 stems stacked on top of each other with only a slight helical angle between them (12˚), while the P6.1 stem turns sharply at the junction and comes around to face the minor groove of P6 at an interhelical angle of only 35 degrees[39]. NMR data on medaka CR4/5 in complex with TRBD shows significant rearrangement upon binding. The P6.1 stem swings down like a gate-arm, opening the structure so that TRBD can fit between the P6 and P6.1 stems(Fig1.2). CR4/5 (and its homolog SLIV in *Tetrahymena*) is the only component of TR that has consistently demonstrated such large rearrangement upon binding to TERT[32,36,39,41,43,44].

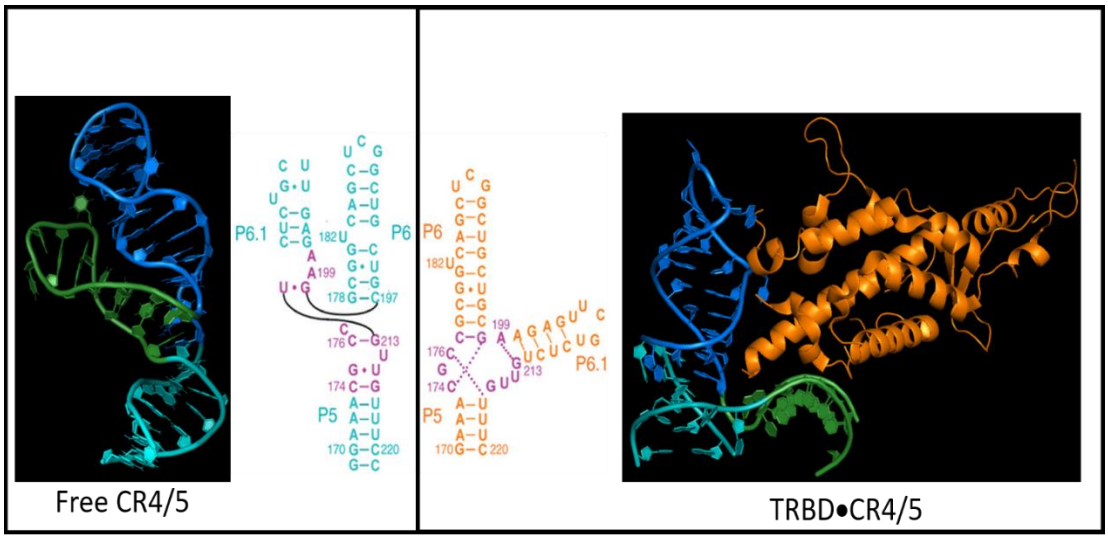


Figure 1.2 Secondary structure and NMR structures of the Medaka CR4/5 region. Left; structure of mCR4/5 in the absence of TRBD. Right; CR4/5 in complex with TRBD (orange). P5 is shown in cyan, P6 is shown in blue, and P6.1 is shown in

P6

In addition to large areas of protein binding between TRBD and the backbone of P6, the conserved pyrimidine bulge and the terminal stem loop of P6 are critical for strong, specific binding of CR4/5. In the protein-free medaka NMR structure, U182, the bulge nucleotide, sticks out perpendicularly from the major groove. This causes an S-turn in the top of the P6 stem that angles it slightly towards P6.1[39]. UV-crosslinking analysis and the NMR solution structure of *ol*CR4/5 with TRBD shows that U182 stacks specifically with Tyr503, and hydrogen bonds with Arg506[41,42] (Fig 1.3, left). Despite the interactions between TRBD and U182, transition mutation of this bulge did not impact activity in the full-length TERT/TR[39]. However, replacement of U182 with any nucleotide abrogated all binding between CR4/5 and TRBD alone[41]. This data taken together indicate that the importance of the pyrimidine

bulge for telomerase function derives from strong protein-RNA binding, and not from any direct catalytic activity. The terminal loop of the medaka P6, which is analogous with the internal loop between P6a and P6b in humans, is critical in the recognition of CR4/5 by TRBD[41]. In human CR4/5, only a short portion of the internal loop above the P6a helix is actually needed for full telomerase activity[45,46]. Consistent with this, circular permutation of *h*TR has shown that cuts to the backbone around the internal loop greatly decrease or abolish telomerase activity[43].

Examination of the loop in *ol*P6 NMR structure shows G189 stacks specifically with the aromatic sidechain of Phe376 and forms hydrogen bonds with Leu371(Fig 1.3, right), a circumstance reminiscent of U182. Again, this interaction between TRBD and G189 provides important specificity and strong binding interactions to CR4/5[41]. It is likely that nucleotides in the human P6 internal loop are forming very similar interactions with residues in *h*TERT. Mutational analysis has shown that, aside from U182, the actual sequence of P6 is not important for TRBD binding *in vitro*,

suggesting that the protein:RNA contacts are formed with the RNA backbone. This was confirmed for medaka CR4/5 in the high-resolution *ol*CR4/5•TRBD NMR structure[41] (Fig 1.2).

P6.1

The P6.1 stem contains critical protein-binding surfaces along the major groove backbone. This is demonstrated clearly in the solution structure of *ol*P6.1 (Fig 1.2), and indirectly by numerous mutational and crosslinking studies[19,37,46]. What the NMR data lacks is the pentaloop at the end of the stem, which proved too flexible to crystalize into a single conformation. Work in yeast suggests that this region may

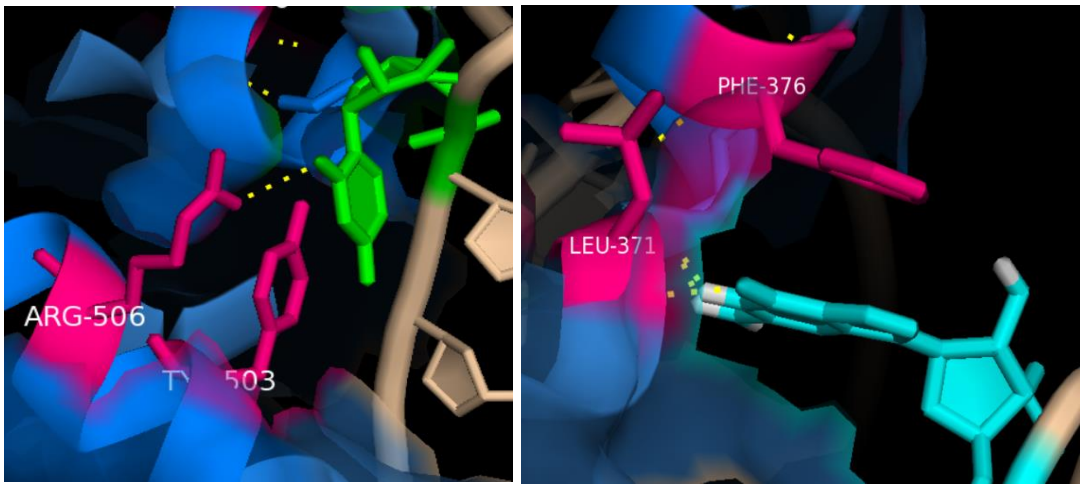


Figure 1.3 U182 and G189 make specific contacts with TRBD. Left; U182 (green) makes contacts with Arg506 (pink) and stacks with Tyr503 (pink). Right; G189 (cyan) makes specific contacts with Leu371 (pink) and stacks with Phe376 (pink, ring above G189)

interact with the stem loop terminus of the TBE helix, possibly to further stabilize the structure and prevent run-through nucleotide addition[47,48]. *In vivo* DMS protection of

*h*TR in the presence/absence of *h*TERT has shown that the central guanine of the pentaloop is strongly protected in the presence of *h*TERT, while the TBE region downstream of the template becomes more sensitive to DMS[36].

THE JUNCTION

At the convergence of the three CR4/5 stems lies the three-way junction. The most radical changes in base pair arrangement upon TRBD binding occur in this crucial area[36,39,41] (Figs 1.4 & 1.5), and it is from this junction that CR4/5 derives its flexibility. Comparison of the NMR structures of the native[39] (protein-free) and TRBD-bound[41] medaka CR4/5 reveals surprising changes in the bonding architecture of the junction. In its native state the junction of *ol*CR4/5 has an unusual arrangement of stacked and paired nucleotides (Figs 1.4 bottom & 1.5 left). There is a G•U wobble pair at the base of P6.1 that induces a sharp kink in the P6.1 backbone, causing it to turn, facing the minor groove of P6. Interestingly, without this kink, the face of P6 that is normally available to TRBD would be occluded. Achieving the kink requires a number of unpaired bases within the stems, including the critical A199 (Figs 1.4 bottom & 1.5 left). Two nucleotides at the top of the P5 stem are pushed out of the helix by this kink, and must stack upon themselves for stability. Two other nucleotides at the base of the P6.1 stem are also extruded from the helix, A199 and A200. This leaves G198 to form a wobble pair with U212, thus sealing the end of P6.1.

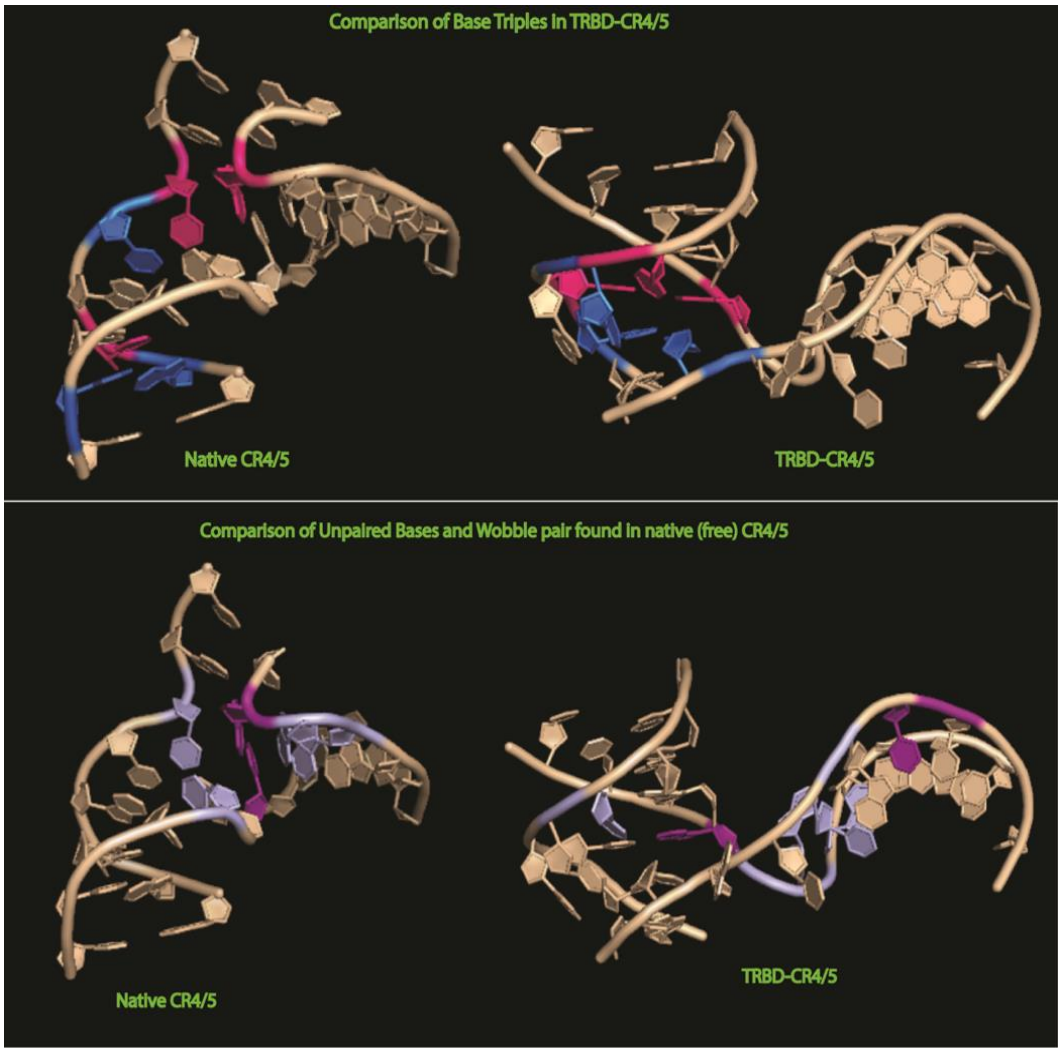


Figure 1.4 Changes in the positions of bases in the medaka CR4/5 junction. The left sides of both figures show the positions of the indicated bases in the native state, the right shows the same bases in CR4/5-TRBD. Bases that are colored the same, are the same base, in both left and right sides of the images. Top; Bases that form base-triples in TRBD-CR4/5 complex are colored in blue and pink. Bottom; Base pair positions of native CR4/5. The G198 and U212 bases (dark purple) form a wobble pair in native CR4/5. This causes a sharp kink in the backbone, resulting in several unpaired bases (lavender), including A199.

Junction flexibility is crucial to the shuffling of bonds required for medaka

CR4/5•TRBD binding. To convert to a structure competent to bind TRBD, two

canonical base pairs and two wobble base pairs are broken. The junction thus rearranges and new pairs are formed: one wobble pair, two new canonical base pairs (one including a new base triple), and second new triple formed in conjunction with a pre-existing canonical base pair (Fig 1.4, top). These base triples stabilize rearrangement by capping the ends of the P6 and P5 stems while the new wobble pair, A199•G213, ends the P6.1 stem. Binding TRBD necessitates this A199•G213 wobble pair, which is crucial for anchoring the vertex[42] of the P6/P6.1 stems onto a wedge-shaped area of TRBD. Overall this reshuffling considerably opens the junction, by reorganizing unpaired bases away from the stems (4 unpaired in P6.1 stem, one

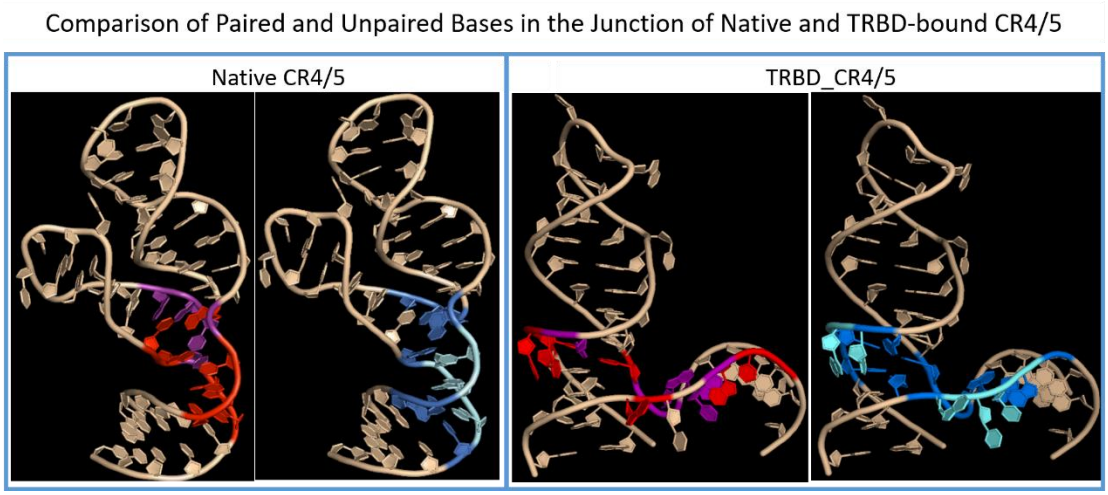


Figure 1.5 Changes in the Arrangement of Paired and Unpaired Junction Bases. Bases colored in red (paired) and purple (unpaired) highlight the position of junction bases in the native CR4/5.Bases shown in blue (paired) and cyan (unpaired) represent the positions of paired and unpaired bases in the significantly altered junction arrangement of the complexed TRBD-CR4/5. These figures illustrate the relative changes in the positions of unpaired bases in the native and bound CR4/5 structure. The unpaired bases of the native structure are found within the P6.1 and P6 stems just before the beginning of the junction; in contrast to the unpaired bases in the complexed structure, which are found only in the junction.

unpaired in P5 stem and one unpaired in J5/6 without TRBD) and into the junction (no unpaired bases in any stems, one unpaired base in J5/6, and 3 unpaired bases in J5/6.1 with TRBD, see figure 1.5). This explains why mutations disrupting the junction base pairs observed in the solution structure of native CR4/5 either had no impact or enhanced activity & TRBD binding; those base pairs would need to be broken to accommodate TRBD.

Since the ability to rearrange is crucial to CR4/5 function, it is likely that it will be retained across diverse eukaryotic lineages. Indeed, the junction A (in medaka, A199) is ***invariant*** in all organisms with a TWJ[1,49]. This particular base may be under such intense selective pressure because it is one of only three bases in CR4/5 that are known to have sequence-specific interactions with residues in TRBD[41,42], and in medaka, it is the only nucleotide of the J6/6.1[39]. As yet, data on the specific interactions between the CTE domain and CR4/5 is not available, although crosslinking data suggests that the P6.1 stem loop could interact with the CTE[42].

REMAINING QUESTIONS OF STRUCTURE AND FUNCTION

Great knowledge has been gained by the studies done so far on CR4/5, but there remain significant questions to be answered. How exactly does the CR4/5 interact with the TBE, and how does this cause changes to TBE function? Is the CR4/5 a molecular sensor for matching properly folding TERT with TR, and vice versa? Is there a conformational change between the two medaka NMR structures, or are they mutually exclusive? If both structures occur in the free RNA, is the 'open' conformation induced by TERT? If there is interconversion between the closed

CR4/5 and the open conformation seen in the complexed TRBD•CR4/5 happen incrementally? Can CR4/5 reside in a conformation somewhere between the two NMR structures? What is the purpose of the invariant junction A nucleotide? How does pseudouridylation impact CR4/5 folding, TRBD binding, and TBE interactions? In order to answer some these fundamental questions researchers need to marry *in vitro* structural and mechanistic studies to complementary *in vivo* mutational and functional assays.

REFERENCES

1. Lai, A. G. *et al.* The protein subunit of telomerase displays patterns of dynamic evolution and conservation across different metazoan taxa. *BMC Evol. Biol.* **17,** 107 (2017).
2. Podlevsky, J. D. & Chen, J. J.-L. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* **13,** 720–32 (2016).
3. Podlevsky, J. D., Li, Y. & Chen, J. J.-L. The functional requirement of two structural domains within telomerase RNA emerged early in eukaryotes. *Nucleic Acids Res.* **44,** 9891–9901 (2016).
4. Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **16,** 307–318 (2006).
5. Autexier, C. & Lue, N. F. The structure and function of telomerase reverse transcriptase. *Annu. Rev. Biochem.* **75,** 493–517 (2006).
6. Lue, N. F. & Skordalakes, E. in *Telomerases: Chemistry, Biology, and Cllinical Applications* (eds. Lue, N. F. & Autexier, C.) 53–78 (Wiley & Sons, Inc., 2012).
7. Harkisheimer, M., Mason, M., Shuvaeva, E. & Skordalakes, E. A motif in the vertebrate telomerase N-terminal linker of TERT contributes to RNA binding and telomerase activity and processivity. *Structure* **21,** 1870–1878 (2013).
8. Steczkiewicz, K. *et al.* Human telomerase model shows the role of the TEN domain in advancing the double helix for the next polymerization step. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 9443–9448 (2011).
9. Wyatt, H. D. M., Lobb, D. a & Beattie, T. L. Characterization of physical and functional anchor site interactions in human telomerase. *Mol. Cell. Biol.* **27,** 3226–40 (2007).
10. Rouda, S. & Skordalakes, E. Structure of the RNA-Binding Domain of Telomerase: Implications for RNA Recognition and Binding. *Structure* **15,** 1403–1412 (2007).
11. Hossain, S., Singh, S. & Lue, N. F. Functional Analysis of the C-terminal Extension of Telomerase Reverse Transcriptase: A PUTATIVE 'THUMB' DOMAIN. *J. Biol. Chem.* **277,** 36174–36180 (2002).
12. Wu, R. A. & Collins, K. Human telomerase specialization for repeat synthesis by unique handling of primer-template duplex. *EMBO J.* **33,** 921–935 (2014).
13. Wyatt, H. D. M., West, S. C. & Beattie, T. L. InTERTpreting telomerase

structure and function. *Nucleic Acids Res.* **38,** 5609–5622 (2010).

14. Peng, Y., Mian, I. S. & Lue, N. F. Analysis of Telomerase Processivity. *Mol. Cell* **7,** 1201–1211 (2001).

15. Mitchell, M., Gillis, A., Futahashi, M., Fujiwara, H. & Skordalakes, E. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **17,** 513–518 (2010).

16. Cech, T. R., Nakamura, T. M. & Lingner, J. Telomerase is a true reverse transcriptase. A review. *Biochem.* **62,** 1202–1205 (1997).

17. Podlevsky, J. D., Bley, C. J., Omana, R. V., Qi, X. & Chen, J. J.-L. The telomerase database. *Nucleic Acids Res.* **36,** D339-43 (2008).

18. Zappulla, D. C., Goodrich, K. & Cech, T. R. A miniature yeast telomerase RNA functions in vivo and reconstitutes activity in vitro. *Nat. Struct. Mol. Biol.* **12,** 1072–1077 (2005).

19. Mitchell, J. R. & Collins, K. Human Telomerase Activation Requires Two Independent Interactions between Telomerase RNA and Telomerase Reverse Transcriptase. *Mol. Cell* **6,** 361–371 (2000).

20. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10024–9 (2004).

21. Zappulla, D. C. *et al.* Ku can contribute to telomere lengthening in yeast at multiple positions in the telomerase RNP. *Rna* **17,** 298–311 (2011).

22. Jansson, L. I. *et al.* Structural basis of template-boundary definition in Tetrahymena telomerase. *Nat. Struct. Mol. Biol.* **22,** 883–8 (2015).

23. Parks, J. W. & Stone, M. D. Coordinated DNA dynamics during the human telomerase catalytic cycle. *Nat. Commun.* **5,** 4146 (2014).

24. Berman, A. J., Akiyama, B. M., Stone, M. D. & Cech, T. R. The RNA accordion model for template positioning by telomerase RNA during telomeric DNA synthesis. *Nat. Struct. Mol. Biol.* **18,** 1371–1375 (2011).

25. Mefford, M. A., Rafiq, Q. & Zappulla, D. C. RNA connectivity requirements between conserved elements in the core of the yeast telomerase RNP. *EMBO J.* **32,** 2980–2993 (2013).

26. Chen, J.-L., Blasco, M. A. & Greider, C. W. Secondary Structure of Vertebrate Telomerase RNA. *Cell* **100,** 503–514 (2000).

27. Theimer, C. A., Blois, C. A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* **17,** 671–682 (2005).

28. Gilley, D. & Blackburn, E. H. The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6621–5 (1999).

29. Qiao, F. & Cech, T. R. Triple-helix structure in telomerase RNA contributes to catalysis. *Nat. Struct. Mol. Biol.* **15,** 634–640 (2008).

30. Lin, J. *et al.* A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. *Proc. Natl. Acad. Sci.* **101,** 14713–14718 (2004).

31. Parks, J. W., Kappel, K., Das, R. & Stone, M. D. Single-molecule FRET-Rosetta reveals RNA structural rearrangements during human telomerase catalysis. *RNA* **23,** 175–188 (2017).

32. Cash, D. D. & Feigon, J. Structure and folding of the Tetrahymena telomerase RNA pseudoknot. *Nucleic Acids Res.* **45,** 482–495 (2017).

33. Yeoman, J. a, Orte, A., Ashbridge, B., Klenerman, D. & Balasubramanian, S. RNA conformation in catalytically active human telomerase. *J Am Chem Soc* **132,** 2852–2853 (2010).

34. Cole, D. I. *et al.* New Models of Tetrahymena Telomerase RNA from Experimentally Derived Constraints and Modeling. *J. Am. Chem. Soc.* **134,** 20070–20080 (2012).

35. Schmidt, J. C., Zaug, A. J. & Cech, T. R. Live Cell Imaging Reveals the Dynamics of Telomerase Recruitment to Telomeres. *Cell* **166,** 1188–1197.e9 (2016).

36. Zemora, G., Handl, S. & Waldsich, C. Human telomerase reverse transcriptase binds to a pre-organized hTR in vivo exposing its template. *Nucleic Acids Res.* **44,** 413–425 (2016).

37. Ueda, C. T. & Roberts, R. W. Analysis of a long-range interaction between conserved domains of human telomerase RNA. *RNA* **10,** 139–47 (2004).

38. Rosenfeld, K. K., Ziv, T., Goldin, S., Glaser, F. & Manor, H. Mapping of DNA Binding Sites in the Tetrahymena Telomerase Holoenzyme Proteins by UV Cross-Linking and Mass Spectrometry. *J. Mol. Biol.* **410,** 77–92 (2011).

39. Kim, N. K., Zhang, Q. & Feigon, J. Structure and sequence elements of the CR4/5 domain of medaka telomerase RNA important for telomerase function. *Nucleic Acids Res.* **42,** 3395–3408 (2014).

40. Xie, M. *et al.* Structure and Function of the Smallest Vertebrate Telomerase RNA from Teleost Fish. *J. Biol. Chem.* **283,** 2049–2059 (2008).

41. Huang, J. *et al.* Structural basis for protein-RNA recognition in telomerase.

Nat. Struct. Mol. Biol. **21,** 507–12 (2014).

42. Bley, C. J. *et al.* RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20333–20338 (2011).

43. Mefford, M. A. & Zappulla, D. C. Physical connectivity mapping by circular permutation of human telomerase RNA reveals new regions critical for activity and processivity. *Mol. Cell. Biol.* **36,** MCB.00794-15 (2015).

44. Akiyama, B. M., Loper, J., Najarro, K. & Stone, M. D. The C-terminal domain of Tetrahymena thermophila telomerase holoenzyme protein p65 induces multiple structural changes in telomerase RNA. *RNA* **18,** 653–660 (2012).

45. Leeper, T. C. & Varani, G. The structure of an enzyme-activating fragment of human telomerase RNA. *RNA* **11,** 394–403 (2005).

46. Zhang, Q., Kim, N.-K. & Feigon, J. Architecture of human telomerase RNA. *Proc. Natl. Acad. Sci.* **108,** 20325–20332 (2011).

47. Webb, C. J. & Zakian, V. A. Telomerase RNA stem terminus element affects template boundary element function, telomere sequence, and shelterin binding. *Proc. Natl. Acad. Sci.* **112,** 11312–11317 (2015).

48. Webb, C. J., Zakian, V. A., Webb, C. J. & Zakian, V. A. mutant Telomere les (i / s) ons from a telomerase RNA mutant. **4101,** 16–18 (2016).

49. Qi, X. *et al.* The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res.* **41,** 450–462 (2013).

50. Kabaha, M. M., Zhitomirsky, B., Schwartz, I. & Tzfati, Y. The 5' arm of Kluyveromyces lactis telomerase RNA is critical for telomerase function. *Mol. Cell. Biol.* **28,** 1875–82 (2008).

51. Dalby, A. B., Goodrich, K. J., Pfingsten, J. S. & Cech, T. R. RNA recognition by the DNA end-binding Ku heterodimer. *RNA* **19,** 841–851 (2013).

52. Chappell, A. S. & Lundblad, V. Structural Elements Required for Association of the Saccharomyces cerevisiae Telomerase RNA with the Est2 Reverse Transcriptase. *Mol. Cell. Biol.* **24,** 7720–7736 (2004).

53. Lebo, K. J. & Zappulla, D. C. Stiffened yeast telomerase RNA supports RNP function in vitro and in vivo. *Rna* **18,** 1666–1678 (2012).

54. Lubin, J. W., Tucey, T. M. & Lundblad, V. The interaction between the yeast telomerase RNA and the Est1 protein requires three structural elements. *RNA* **18,** 1597–1604 (2012).

55. Tang, W., Kannan, R., Blanchette, M. & Baumann, P. Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* **484,** 260–264 (2012).

56. Dehé, P. M. & Cooper, J. P. Fission yeast telomeres forecast the end of the crisis. *FEBS Lett.* **584,** 3725–3733 (2010).

57. Lai, C. K., Miller, M. C. & Collins, K. Template boundary definition in Tetrahymena telomerase. *Genes Dev.* **16,** 415–20 (2002).

58. Box, J. A., Bunch, J. T., Zappulla, D. C., Glynn, E. F. & Baumann, P. A flexible template boundary element in the RNA subunit of fission yeast telomerase. *J. Biol. Chem.* **283,** 24224–24233 (2008).

59. Armstrong, C. A. & Tomita, K. Fundamental mechanisms of telomerase action in yeasts and mammals: understanding telomeres and telomerase in cancer cells. *Open Biol.* **7,** 160338 (2017).

60. Webb, C. J. & Zakian, V. A. Telomerase RNA is more than a DNA template. *RNA Biol.* **6286,** 00–00 (2016).

61. Seto, A. G. *et al.* A template-proximal RNA paired element contributes to Saccharomyces cerevisiae telomerase activity. *RNA* **9,** 1323–32 (2003).

62. Cash, D. D. *et al.* Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo + Supplements. *Proc. Natl. Acad. Sci.* **110,** (2013).

63. Gunisova, S. *et al.* Identification and comparative analysis of telomerase RNAs from Candida species reveal conservation of functional elements. *RNA* **15,** 546–559 (2009).

64. Fulton, T. B. & Blackburn, E. H. Identification of Kluyveromyces lactis telomerase: discontinuous synthesis along the 30-nucleotide-long templating domain. *Mol. Cell. Biol.* **18,** 4961–70 (1998).

65. Shefer, K. *et al.* A Triple Helix within a Pseudoknot Is a Conserved and Essential Element of Telomerase RNA. *Mol. Cell. Biol.* **27,** 2130–2143 (2007).

Chapter II. Probing the Structural Dynamics of *Kluyveromyces lactis* CR4/5 through single-molecule FRET

INTRODUCTION

Telomeres are structures that protect the ends of linear eukaryotic chromosomes from inappropriately directed DNA-repair. These specialized structures are composed of many repeats of a short sequence of non-coding DNA, which is appended to the 3' end of the lagging strand by telomerase, a unique reverse transcriptase. Telomerase is the only known reverse transcriptase to carry its RNA template as an integral component. The RNA template is part of a larger functional RNA component, TR, which complements the protein component, TERT, and is required for the activity of telomerase. The precise roles played by TR in the mechanism of telomerase is not entirely known, and is the subject of extensive study in many organisms, from vertebrates to microbes.

Yet, there are serious constraints to studying human, or even vertebrate telomerase. Deleting the genomic TERT or TR genes entirely can be fatal to developing vertebrates, making it difficult and time consuming to generate knockout lineages[1]. Then there are the ubiquitous issues of genetic tractability and the ability to manipulate and monitor proteins *in vivo*. These are serious issues in telomerase biology just as in other fields. In fact, these are the very reasons why model organisms are used in biology. Telomerase is a multi-subunit RNP, which requires numerous (and not completely known) chaperones to achieve its complicated co-

translational folding[2–4]. This further complicates its study, as *in vitro* work requires a consistent method of obtaining high quality enzymes. The telomerase RNP can't be expressed and purified in bacteria, its endogenous expression levels are very low in most vertebrates, and to study it *in vitro* means seriously paring down the components to their bare minimum. These complications create the perfect storm of obstacles to both *in vitro* and *in vivo* work.

What has become clear from the extensive studies performed to date is that structural homology is the linchpin to telomerase evolution[5,6]. Arguably, the most important knowledge gained from the extensive cross-species *in vitro* work is that despite divergent sequence, the structural homology of TERT and TR is strong. Given this, the onus is on the field to develop experiments that link mechanistic studies with *in vivo* outcomes. To that end, yeast is an attractive candidate model organism. Yeast is easy and cheap to culture and has been the classic model for studying eukaryotic replication for decades. And, unlike other single-celled eukaryotes, it is possible to create non-lethal genomic mutants of telomerase[7] which allows for the observation of the *in vivo* impacts of altering telomerase biology.

TELOMERASE IN YEAST

While yeast has been used extensively to study the proteins involved in telomerase, mechanistic studies involving yeast TR have been less common. This is in part because yeast telomere maintenance seems to be different from that observed in other systems. Some yeast, such as *Saccharomyces cerevisiae*, have variable length and sequence telomeres, but show the standard mechanistic qualities of telomerase,

namely nucleotide addition processivity (NAP) and repeat addition processivity (RAP)[8–10]. Other species, such as the budding yeast *Kluyveromyces lactis*, have consistent repeat length and sequence, yet when telomerase is purified from this yeast, the enzymes show no indication of RAP as it is currently defined[11–13].

Additionally, the TR sequence in yeast is significantly larger than that of vertebrates and ciliates (Fig 2.1, top). *K. lactis* TR is 1,326 bases in length, compared to human's 451 bases, medaka fish's 312 bases, and *Tetrahymena*'s 159 bases[14]. Groundbreaking work by David Zappulla in the Cech lab demonstrated that much of this long sequence is entirely dispensable, acting merely as a scaffold around the functional regions of TR. In fact, these functional components can even be moved around within the long TR sequence with little impact on activity[10,15–17]. Most of the intervening RNA serves as a basis for protein assembly on and around the functional regions of TR[19,16,20]. Indeed, yeast is even permissive to completely altering the sequence of the RNA arms between the protein binding sites[15] with no impact on activity.

Structure of Yeast TR

Yeast shares the same basic elements of telomerase RNA; a core-enclosing helix, pseudoknot, TBE, and CR4/5. There is also extensive regions necessary for the binding of various accessory proteins[17,21], which are homologs of vertebrate proteins involved in telomere maintenance[22,2,23,24]. In line with studies in *Tetrahymena*[25,26], the yeast TBE is composed of a short helical region 5' of the pseudoknot, that forms a block preventing run-through replication[27–29]. Although, the arrangement and strength

of the closing base pairs of the TBE helix has an impact on the consistency of the repeat sequence[27,30], a phenomenon unique to yeast that will be discussed further in Chapter 3. A minimal yeast pseudoknot has been crystalized and has remarkable structural similarity to the human pseudoknot, with a conserved pyrimidine bulge in stem 2, and slight variation in the base-triplex arrangements[31].

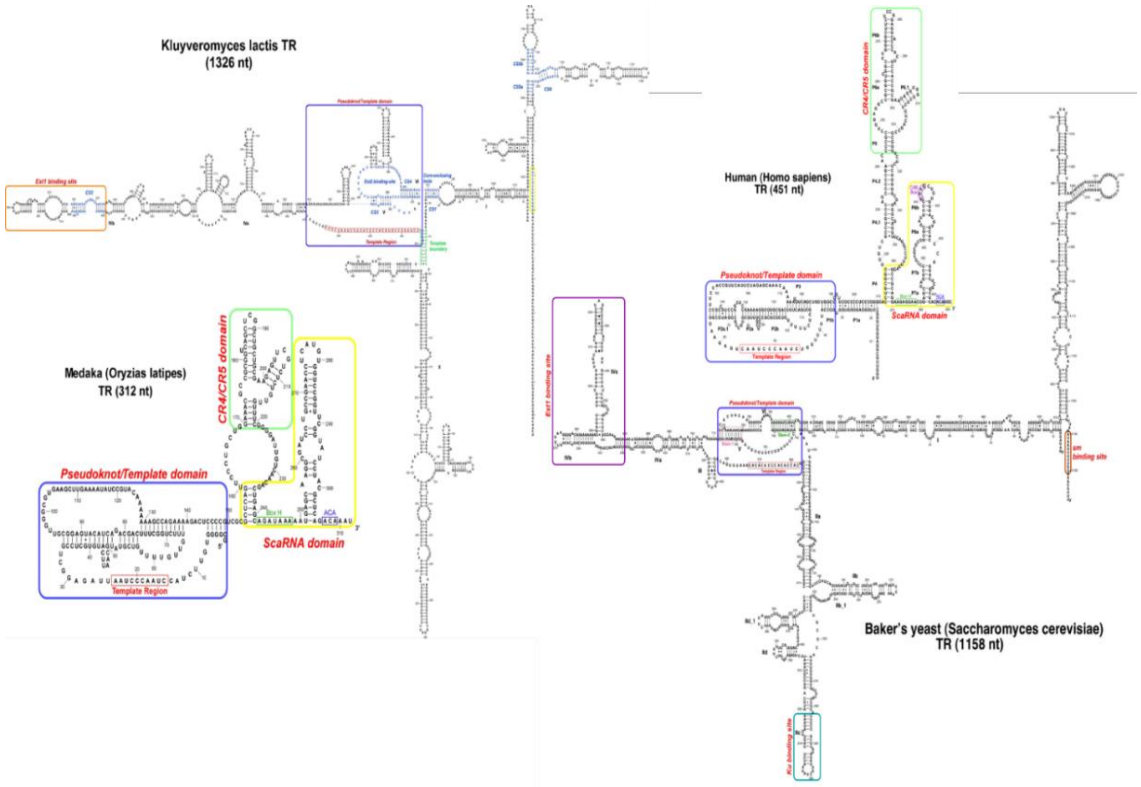


Figure.2.1a. TR secondary structures for *Kluyveromyces lactis*, *Saccharomyces cerevisiae*, *Oryzias latipes* (medaka fish), and *Homo sapiens.* Top; Functional areas are boxed; blue pseudoknot, yellow/lime green H/ACA domains, green CR4/5, purple/orange Est1 binding site, red Sm binding site, and teal Ku binding site. CR4/5 domain in *K. lactis* nucleotides are in blue. Images of the secondary structures have been taken from the Telomerase Database, under the creative commons liscence.

Extensive alignments and folding models performed in over 55 yeast species[8] have shown that the same essential elements of telomerase RNA in vertebrate and ciliate species are retained in most fungal lineages. When viewed from a wider perspective, the fact that such disparate lineages share remarkably common structural requirements for telomerase is a testament to the RNP's ancient importance. The CR4/5 region of *Kluyveromyces lactis* includes the critical pyrimidine bulge of the P6 stem, P6.1 stem, the P6.1 loop, the flexible junction region, and the conserved junction A[8,13,11].

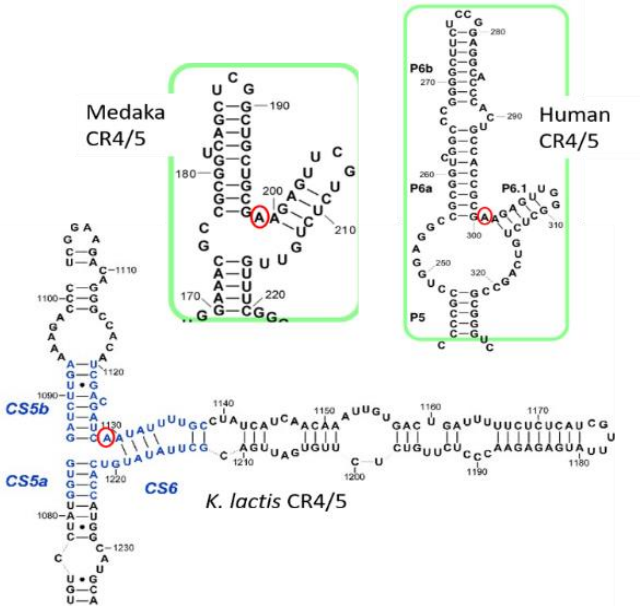


Figure.2.1b. CR4/5 secondary structures for *Kluyveromyces lactis*, *Oryzias latipes* (medaka fish), and *Homo sapiens.* Invariant junction A nucleotides is circles in red in all structures. Images of the secondary structures have been taken from the Telomerase Database, under the creative commons liscence.

KLUYVEROMYCES LACTIS CR4/5

The minimal CR4/5 required for activity in *K lactis* is quite similar to the minimal requirements of the human and vertebrate CR4/5. The length of required *K. lactis* CR4/5 is 103 bases while human minimal CR4/5 is 73 bases[32], and medaka requires only 39 nucleotides[33]. The pyrimidine bulge in P6 is also present in *K. lactis*, as is the junction A. The junction of *kl*CR4/5 is likely quite flexible, and may rearrange in a similar fashion to the medaka junction. What is different between the yeast and vertebrate CR4/5 is the length. The *K. lactis* P6 is moderately longer than the medaka P6, and has two internal loops, reminiscent of the human P6. In humans the region of P6 above the first internal loop is not needed for function. A deletion mutant of the *kl*P6 above the first internal loop has not been reported, but it would be interesting to see if the conservation of functional regions remains consistent with human P6.

The P6.1 stem is also present in *K. lactis*, although both the sequence and the structure are less conserved than the junction. The important stem-loop at the end of P6.1 in medaka, which was too unstable to crystalize, is in the form of an internal loop of 4 uracil residues in *K. lactis*. The wildtype P6.1 region of *K. lactis* is considerably longer than any vertebrate P6.1, although it can be significantly reduced with little-to-no impact on *in vitro* and *in vivo* activity[34]. A common feature of yeast functional TER regions is that the required structures are found dispersed within very long stretches of sequence. Sometimes these extended arms can have within them other important structured regions required for protein binding[19,16,20]. This may also

be the case for *kl*P6.1, as it can be reduced from 90 bases in the wildtype stem-loop, to 19 bases in the minimal CR4/5. Alternatively, the bases that extend beyond the required 19 bases of P6.1 may aid in folding the CR4/5 in the context of the full-length RNA, but are not needed when the sequence is reduced. There are many possible reasons for the large TR sequence in yeast, but it is clear that the minimal functional components are conserved between yeast and higher eukaryotes.

It is fortuitous that, despite the differences in yeast and vertebrate TR, structural conservation of the necessary functional TR regions make telomerase studies in yeast applicable to other organisms. This has provided a platform for the development of *in vivo* assays that guage the impact of telomere and telomerase aberrations in a living organism. As discussed earlier, this can be used to address the lack of uniformity between *in vivo* and *in vitro* studies, as it is possible to easily create genomic mutants of TR and TERT that can also be easily studied *in vitro*. As yet there has not been detailed structural work performed in conjunction with complementary *in vivo* studies, which is necessary for a comprehensive understanding of telomerase mechanism and telomere maintenance.

USING SMFRET TO PROBE DYNAMICS OF K.LACTIS *CR4/5*

The following chapter addresses this need by performing a mechanistic study of the *Kluyveromyces lactis* CR4/5 using single-molecule FRET. The aim of the FRET work is to observe the structural dynamics of this critical three-way junction. NMR, DMS protection, and crosslinking experiments yield valuable structural information, but these are all ensemble snapshots of the RNA, providing data only on

the most common state of the molecule. smFRET enables real-time observation of the motion of each stem of the TWJ, providing information on not only the physical state and capability of structural rearrangements in the molecule, but also insight into populations of molecules that are not in the average ensemble arrangement. Using this method, I will investigate further the nuances of the CR4/5 junction rearrangements, including how the invariant A base influences the conformation of the molecule. I will probe the movements of each arm in relation to the others, to determine if the two conformations found in the medaka CR4/5 NMR structures (native and bound) are also present in the *K. lactis* CR4/5. And, if they are, whether they exist as mutually exclusive states, or if there is the potential for interconversion between the two conformations. To further the understanding of the impact of structure on function, this work will be combined with *in vivo* experiments performed by our collaborators in the Tzfati lab at Hebrew University.

METHODS

BUILDING RNA CONSTRUCTS

DNA SPLINT DESIGN

DNA splints were designed for construct building through consideration of the three ligation site positions, annealing temperatures of the RNA precursors, and the location of internal loops in the final product. It was decided that greater control over the equilibrium of ligated products could be conferred by using two DNA splints, rather than a single large splint. In designing these splints, it was necessary to ensure that at least 4 bases before and after a ligation site would pair with the splint. It was

also determined that pairing between the splint and the RNA precursors would be best accomplished by positioning either end of the splint at a bulge in the final ligated product. By positioning the ends of the splints where there is not naturally base pairing in the final product, competition for binding is limited and thus binding between the splint and the RNA precursors is favored.

The first splint, termed GH-1, covers two RNA precursors and one ligation site. It binds from nucleotide U7 to nucleotide A29. The first ligation site is only four bases upstream from the modified U dye-label position, which could have led to problems with binding the splint. It did not, however, seem to influence the efficiency of ligation.

The second splint, termed GH-2 for the wildtype construct, and CM-06 through CM-09 for the mutants, covered three precursors and two ligation sites. Binding spanned from nucleotide G45 to A74. The dye-label position on P6.1 (Stem 3) was located two bases upstream of the 5' end of the splint, which may have had an impact on ligation efficiency. CM-06 through CM-09 differed in sequence corresponding with the mutations designed into the small, junction precursor RNA piece. See table2.1 for details on the sequences of the splints and the precursors.

PREPARATION OF RNA PRECURSORS

The RNA components needed for this project were purchased from GE-Dharmacon (Lafayette, CO). All precursors were purchased with 2'-ACE protection and were 5' phosphorylated. Precursors destined for labeling were purchased with a

5-N-U modification, and a second set was provided by Dharmacon with the 5-LC-N-U modification.

Preparation of unlabeled precursors

The small 13-nucleotide junction precursor (nucleotides C55-U67) was prepared by deprotection and desalting. Lyophilized RNA pellets were resuspended in 200µL of TEMED deprotection buffer (Dharmacon) and incubated at 60˚C for 30 minutes. Samples were then purified by fractionation over a PD-25 Desalting Column (Amersham) equilibrated with nuclease-free water (NFW), and eluted in NFW. Absorbance at 260nm was measured on the Nanodrop (Thermo). Any fractions containing greater than 70ng/µL were concentrated in a vacuum-centrifuge at 45˚C until a volume of approximately 20-50µL. All fractions were then pooled, and concentrated again to a volume of 50-100µL. The absorbance was then measured, and the concentration in µM was determined. The products were run on PAGE for confirmation of integrity.

Dye Labeling and Deprotection

Precursor fragments were labeled with mono-reactive Cy5 or Cy3 dye pack (Amersham, GE Healthcare). RNA samples to be labeled were resuspended in 100µL of NFW, and ethanol precipitated using sodium acetate. While the pellets dried, the Cy3 or Cy5 dye was resuspended in 100µL of fresh .1M sodium bicarbonate. The RNA pellet was then resuspended in the bicarbonate-dye solution. The labeling reaction took place at 37˚C in the dark for two hours. During this time, a PD-25 desalting column was equilibrated in NFW, and collection tubes were prepared. Once

the labeling reaction was complete, the sample was loaded onto the column and allowed to absorb fully into the bed. Following absorption, 400µL of NFW was added to the bed and allowed to absorb. After that, NFW was added to the bed as needed to prevent the column from running dry, until all fractions had been collected. The appropriate time to begin taking fractions was determined through observation of the progress of the dye front in the column. Once the dye front was close to the bottom of the column, two-drop fractions were collected. Collection proceeded through 25-30 fractions. The column was then capped and saved, and the absorbance of the fractions at 260 nm was measuring using the Nanodrop. Any fractions with greater than 70ng/µL of RNA were concentrated using the vacuum-centrifuge at 45˚C, until a volume of approximately 15µL. The fractions were then pooled, and concentrated again until they reached a volume of 20-30µL. Four hundred microliters of deprotection buffer was added to the pooled sample, which was then incubated at 60˚C for 30 minutes in the dark. Once the deprotection step was complete, the sample was split in half, both halves were ethanol precipitated using 5M ammonium acetate. Pellets were resuspended in 60µL of .1M TEAA/.1%TFA buffer in preparation for HPLC.

HPLC Purification

Samples were purified using the ÄKTApurifier (GE Healthcare) over a reverse-phase C8 column. Samples were loaded onto the C8 column in Buffer A (.1M TEAA/.1%TFA), after which a buffer gradient was run over 40 min from 0% to 80% Buffer B (acetone/.1% TFA). Absorbance at 260nm, 550nm, and 650 nm was

monitored throughout the procedure. Fractions were taken for peaks corresponding to unlabeled RNA (A_{260} only) and labeled RNA (A_{260} and either A_{550} or A_{650}). Fractions were concentrated in a vacuum-centrifuge at 45˚C until dry, resuspended in NFW, and ethanol precipitated using ammonium acetate. Once all ethanol had evaporated from the pellets, they were resuspended in NFW, quantitated on the Nanodrop, and run on PAGE for confirmation of integrity.

PAGE Confirmation

Precursors were checked for purity, size, and quality through PAGE analysis on denaturing 7M urea/9% polyacrylamide gels. Gel solution was made in a volume of 50mL by combining 22g of urea (Ambion) with 10mL of 5X TBE, and 11.25 mL of 40% (v/v) 19:1 Acrylamide:Bisacrylamide solution, and milli-Q water. The mixture was heated until all solid urea had dissolved, and then sonicated for 10 min. During this time the gel apparatus was cleaned and assembled. Solidification of the gel solution was initiated by the addition of 30µL of TEMED and 300µL of 10% ammonium persulfate. The solution was then gently but rapidly mixed, and poured into the gel apparatus. All gels were allowed to solidify for at least 30 min before being pre-run for 1 hour at 55mA. During the pre-run, samples were prepared for loading. Between .5µL – 1µL of RNA was combined with 4.5uL or 4 µL of 7M urea/1XTBE loading buffer, heated to 95˚C for 3 min, and snap-cooled on ice. At least 2 nanomoles of Cy-dye label are required for visualization by fluorescence, while 50ng of RNA are required for visualization by ethidium bromide. In cases where 1µL did not contain sufficient material, volumes were scaled up to ensure

proper detection. Samples were loaded into the wells by syringe. Before each sample was drawn into the syringe, the syringe was rinsed 10 times with NFW, twice with 1X TBE, and three times with 7M Urea/1XTBE loading buffer. This ensured that the void volume in the syringe would not interfere with the molarity of the loading buffer in the samples.

FOUR-WAY SPLINTED RNA LIGATIONS

Test Ligations

In order to determine the optimum relative ratios of each component in the ligation, small-scale test ligations were performed. The RNA precursors will be referred to as described in table 2.1.

Table 2.1. Nomenclature and Properties of RNA and DNA oligonucleotides used in this study.

Precursor Name	Nucleotide Position	Cy-Dye Label	Shorthand	Size	Positions Spanned by Splint	Modification Position
P5 piece	1-14	Yes	Fp5	14nt	7-14	U10
P6 piece	15-54	Yes	Fp6	40nt	15-30 (s1) 45-54 (s2)	U19 (precursor) U33 (ligation product)
Junction Piece	54-67	No	FJ(WT/M 1/M2)	13nt	54-67	Unmodified
P6.1 piece	68-103	Yes	Fp6.1	36	68-74	U5 (precursor) U72 (product)
Splint1	7-30	No	s1	24	--	NA
Splint 2	45-74	No	s2	30	--	NA

Wildtype Constructs

The wildtype (WT) T1 construct, labeled on P5 with Cy3 and P6.1 with Cy5, was made previously by Dr. Shankar Shastry.

The test ligation for the T2 WT construct, which is labeled on P6 with Cy5 and P6.1 with Cy3, was done using equimolar amounts of each precursor and splint. The test ligation was scaled to 80pmol in a final ligation volume of 20µL. All precursor components were annealed to the DNA splints prior to the ligation reaction in 10µL of .5X Ligase buffer (NEB) by heating to 95˚C for three minutes. During this time a 2X ligation solution was prepared in 10µL, omitting from the mixture any salt sensitive components (T4 DNA Ligase, RNasin). The ligation solution was prepared

Table 2.2. Expected fragment sizes of all potential ligation products. The table includes the labeled stems for each construct, and indicates whether a particular ligation product would be labeled with Cy3, Cy5, Cy3-Cy5, or neither.

Band Size	**Precursors Ligated**	**Label**		
		T1 P5(Cy3) P6.1(Cy5)	**T2** P6(Cy5) P6.1(Cy3)	**T3** P5(Cy5) P6(Cy3)
103	Fp5-Fp6-FJ-Fp6.1 (full length)	Cy3/Cy5	Cy3/Cy5	Cy3/Cy5
89	Fp6-FJ-Fp6.1	Cy5	Cy3/Cy5	Cy3
67	Fp5-Fp6-FJ	Cy3	Cy5	Cy3/Cy5
54	Fp5-Fp6	Cy3	Cy5	Cy3/Cy5
53	Fp6-FJ	Unlabeled/EtBr	Cy5	Cy3
49	FJ-Fp6.1	Cy5	Cy3	Unlabeled/EtBr
40	Fp6	Unlabeled/EtBr	Cy5	Cy3
36	Fp6.1	Cy5	Cy3	Unlabeled/EtBr
14	Fp5	Cy3	Unlabeled/EtBr	Cy5
13	FJ	Unlabeled/EtBr	Unlabeled/EtBr	Unlabeled/EtBr

in 10µL such that when the 10µL annealed RNA solution was added, the resulting 20µL would have a final concentration of 1X Ligase buffer (NEB), 2mM ATP, 1U/µL RNasin Plus (Promega) and .5µL of T4 DNA Ligase.

After the 95˚C incubation was finished, both the annealing solution and the 2X ligation solution were incubated at 30˚C for 3 minutes. The annealing solution was then added to the ligation solution and mixed, and RNasin and T4 DNA Ligase were added. The ligations were incubated for 12-24 hours, in the dark, at 30˚C. After incubation, the ligations were treated with .5µL of Turbo DNase I (Ambion) for 15 minutes at 30˚C to degrade the DNA splints. The volume was then brought up to 100µL with T50, and the RNA was phenol/chloroform extracted, and ethanol precipitated using ammonium acetate. Once dry, pelleted RNA was resuspended in 10µL of 7M Urea/1X TBE Loading buffer and run on 9% denaturing PAGE to evaluate ligation efficiency. The gels were scanned for Cy3 and Cy5 fluorescence using the Typhoon Scanner (GE Sciences). The gels were then stained with EtBr, and scanned again for ethidium bromide fluorescence. Sizes of the expected products are provided in table 2.2.

Test ligations for WT T3, which is labeled on P5 with Cy5 and P6 with Cy3, were set up using equimolar and non-equimolar ratios of RNA precursors. The first test ligation used equimolar amounts of all components, as described for the T2 WT test ligations. The second ligation condition used 80 pmol of Fp5-Cy5, Fp6-Cy3, and s1; and 160 pmol of Fp6.1, FJ, and s2. Thus there was a 1:2 ratio of the RNA

components that make up the 5' end of the final product, to the 3' components. The third test ligation used 160 pmol of Fp6.1 only, and 80 pmol of all other components. The final test ligation condition used 160 pmol of FJ only, and 80 pmol of all other components. Annealing and ligation solutions were prepared as described for the T2 WT construct, and incubated in the same conditions. The phenol/chloroform extraction was more successful in the T3 constructs when the volume of the ligation was brought to 100µL with NFW rather than T50, and heated to 55˚C for 20 minutes after phenol was added. This may be because T3 was the only construct that used unlabeled Fp6.1; in this fragment the dye-label site is only two nucleotides from the ligation site. Thus, it is possible that the ligase bound more strongly to this ligation site than it would in the T1 and T2 constructs, and therefore required the more denaturing conditions provided by low-salt and heat to be completely dissociated from the ligation products.

Mutant Constructs

Test ligations for two mutant forms of T1, T2 and T3 constructs were undertaken, using more varied ratios of precursors than the wildtype constructs. The first mutant is a deletion of A60, the universally conserved nucleotide found in the junction region. The second mutant closes an internal loop in the P6.1 branch by replacing three consecutive uracil bases with cytosines. These cytosines are then able to form wobble pairs with the uracils on the opposite strand. In addition to the 1:1 ratio of precursors used in the WT, test ligations were performed that used: .5X FJ : 1X of all other components; .5X of s1 and Fp5 : 1X of all other components; and 2X FJ : .5X

Fp6 : 1X of all other components. One additional test ligation was performed for the second mutant, using 2X FJ: .75X Fp6: 1X of all other components. For all of these test ligations 1X was 50 pmols.

The annealing and ligation solutions were prepared in the same fashion as in the WT test ligations, except for the temperature at which the annealing and ligation were incubated. The first mutant was incubated at 28˚C to compensate for the lower annealing temperature caused by the deletion in the junction precursor. The second mutant was incubated at 32˚C to favor DNA/RNA binding over RNA/RNA binding, which may occur more readily at 30˚C due to additional base-pairing in the mutant.

Preparative Scale Ligations

Preparative scale ligations were performed for the WT T2 and T3 at 1X=800pmol scale, the m1 mutants were prepared at 1X=500pmol scale, and the m2 ligations were prepared at 1X=800pmol scale. The T2 WT ligation used equimolar concentrations of all components, while the T3 WT ligation used 2X FJ: 1X all other components. The ligations for both mutants used the ratio 2X FJ: .5X Fp6:1X all other components. Ligation products were DNaseI treated, phenol/chloroform extracted, and ethanol precipitated. The products were run on a 7M Urea denaturing 6% polyacrylamide gel, visualized and bands corresponding to full-length product were excised. The full-length, 103/102-base ligation products of all forms of the T1 and T3 constructs could be distinguished from the incomplete 89/88-base product by eye alone, because the complete product would be double-labeled while the incomplete product would not (see table 2.2). The T2 constructs, however, had double-labeled full length and

incomplete products, and thus it was not possible to rely solely on visualization by eye. Any gel containing a T2 construct (mutant or WT) was scanned for Cy3 and Cy5, and the gel image was printed to scale with the actual size of the gel. The printed image was affixed securely to the benchtop, plastic foodwrap was laid down, and the gel was placed on top. The gel was aligned with the image using visual markers such as the tracking dye, edges of the gel, and the position of wells. Once the gel was carefully aligned, the full-length band was excised.

Elution

Excised bands were cut into small pieces and placed in 2mL lo-bind tubes (Eppendorf). Five hundred microliters of T50 and phenol were added to the tubes, which were then vortexed on high, in the dark, for at least 48 hours. The liquid was then pipetted off of the gel pieces into fresh tubes, and phenol/chloroform extracted and ethanol precipitated. Once the pellets were dry, the constructs were resuspended in 10µL of NFW and the A_{260}, A_{550}, and A_{650} measured on the nanodrop. Molarity of RNA, Cy3 and Cy5 dyes were calculated, and base:dye ratios were determined. Constructs were confirmed on a 7M Urea 9% PAGE.

FRET MEASUREMENTS

DETERMINING OPTIMUM BUFFER CONDITIONS

The wildtype T1 construct was tested for optimum FRET behavior using series of 5 buffers. Buffer 1 (B1) was 50mM Tris-HCl pH 7, 100mM NaCl, 1mM $MgCl_2$, and

1% glucose. Buffer 2 was the same as B1 with the addition of .1µg/µL yeast tRNA, Buffer 3 (B3) was B2 + 10% glycerol. Buffer 4 was 25mM Tris-HCl, 50mM NaCl, 1% glucose, 10% glycerol, .1µg/µL tRNA, .1mg/mL BSA, and 1mM $MgCl_2$. Buffer 5 was 50mM Tris-HCl pH 7, 100mM KCl, 1% glucose, 10% glycerol, .1µg/µL tRNA, .1mg/mL BSA, and 1mM $MgCl_2$. All imaging buffers were treated with a maximum of 2mM TROLOX, filtered and pH adjusted before use. Immediately prior to imaging, glucose oxidase and catalase were added to a final concentration of 1mg/mL and 2µg/mL, respectively. Labeled RNA molecules were heated to 70˚C for 5 minutes and allowed to cool slowly to room temperature before being diluted, and adhered to a streptavidin-coated quartz slide. Imaging was performed on a total internal reflection prism-type microscope. Data was taken using a framerate of 10Hz, with an Andor CCD camera and analyzed using in-house software run in IDL and Matlab.

FRET EXPERIMENTS

Investigation of the stability of the FRET states found in the wildtype constructs was undertaken in several ways. First, ninety-minute FRET experiments were performed using the T2 and T3 constructs imaged in buffer 1, with a framerate of 2Hz. Another approach used the addition of varying concentration of Est2 (*kl*TERT) purified from *K. lactis* and provided to the lab by our collaborators in the Tzfati lab at The Hebrew University of Jerusalem. Each wildtype construct was imaged for 2 minutes with Est2 in concentrations ranging from $1x10^{-2}$ to $1.2x10^{-3}$ units per channel, using a framerate of 10Hz. Channels were washed with incomplete imaging

buffer (lacking Est2 and glyoxy) then completed buffer, incubated for 5 minutes and imaged.

The impact of magnesium on the FRET behavior of the wildtype T2 and T3 constructs were studied by performing a titration series of 1mM, 10mM, 20mM, 30mM, and 40mM $MgCl_2$ in buffer 5. Additionally, all constructs were studied in 0mM, 1 mM, and 10mM $MgCl_2$ conditions (zero magnesium was achieved by using imaging buffer lacking added $MgCl_2$ and supplemented with 2mM EDTA).

Following data acquisition, traces were culled based on the amplitude of dye intensity of the trace (no huge spikes in fluorescence intensity for either donor or acceptor), as well as observed donor and acceptor bleach events, and a minimum pre-bleach lifetime of 5 seconds. All population histograms were made from the accepted traces. Wildtype and mutant T2 and T3 constructs were further analyzed by sorting the traces into categories based on their dynamic behavior.

RESULTS

CONSTRUCT BUILDING

PREPARATION OF PRECURSOR RNA

Precursor RNAs Fp5, Fp6, and Fp6.1 were labeled with either Cy3 or Cy5. Following the labeling and deprotection steps, all fragments were purified by HPLC. An example HPLC chromatogram is shown in Fig. 2.2. Following HPLC purification, all precursors were checked by 7M urea 9% PAGE, and example of which can be found in Figure 2.2.

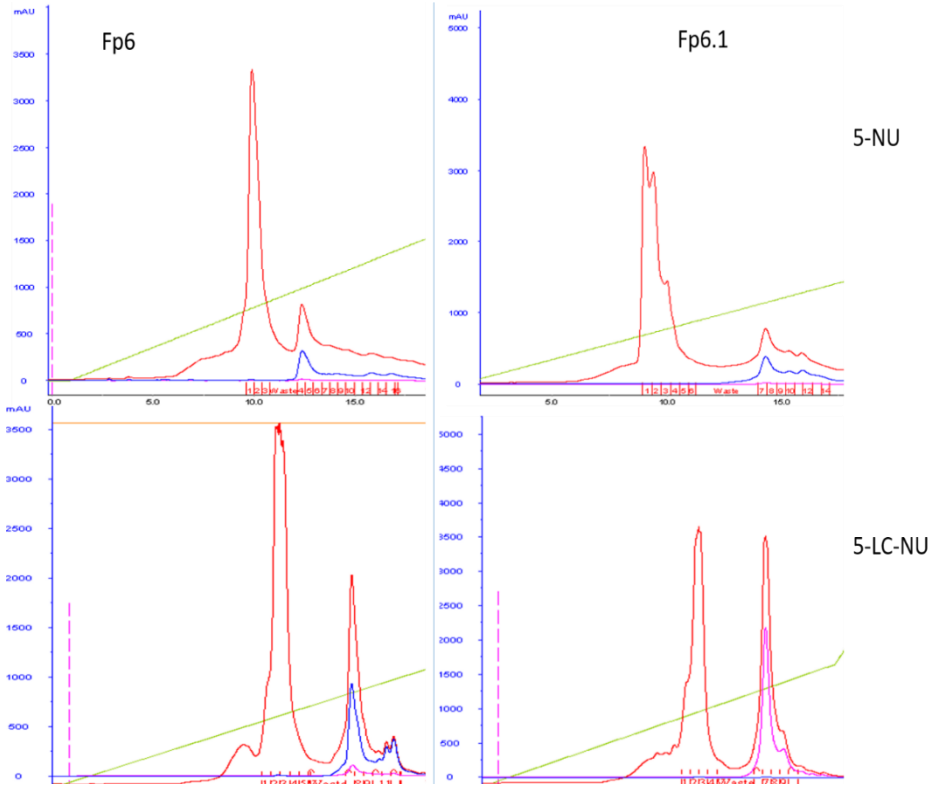


Figure 2.2 Comparison of Labeling Efficiency. The above chromatograms show the yields of HPLC purification of labeled and unlabeled precursors. From these chromatograms the differential labeling efficiency of the 5-NU modification (top row) and the 5-LC-NU modification (bottom row) can be observed. Absorbance at 260 nm (red line) represents RNA concentration, and absorbance at 650 nm (blue line) and 550 nm (pink line) represent Cy5 and Cy3 concentrations, respectively.

Precursors bearing the 5-LC-NU modification showed a much better labeling efficiency than those with the 5-NU modification, as can be seen by the difference in ratio of the peak heights in the unlabeled RNA (1st peak): labeled RNA (2nd peak) for each modification. All precursors bearing the 5-NU modification showed approximately 10% labeling, while the 5-LC-NU modification showed as much as 50% labeling efficiency.

Labeling efficiency became especially problematic when working with Fp6, the longest precursor, which can form a hairpin. We hypothesize that this piece may be dimerizing during HPLC purification, as is evidenced by consistently high base**:**dye molar ratios in the labeled 5-NU Fp6 pieces. This indicates that some unlabeled precursor was present in fractions containing the labeled precursor. Combined with the low labeling efficiency inherent with this modification, it is no surprise that ligations performed with the 5-NU modified Fp6 precursors had extremely low efficiency.

Comparison of the final labeled products from 5-NU and 5-LC-NU precursors shows an additional complication. It is clear that multiple dye-labeled bands are present for nearly all products, yet this is not caused by RNase contamination, which would leave a laddering pattern. What is more likely is that the precursor is broken

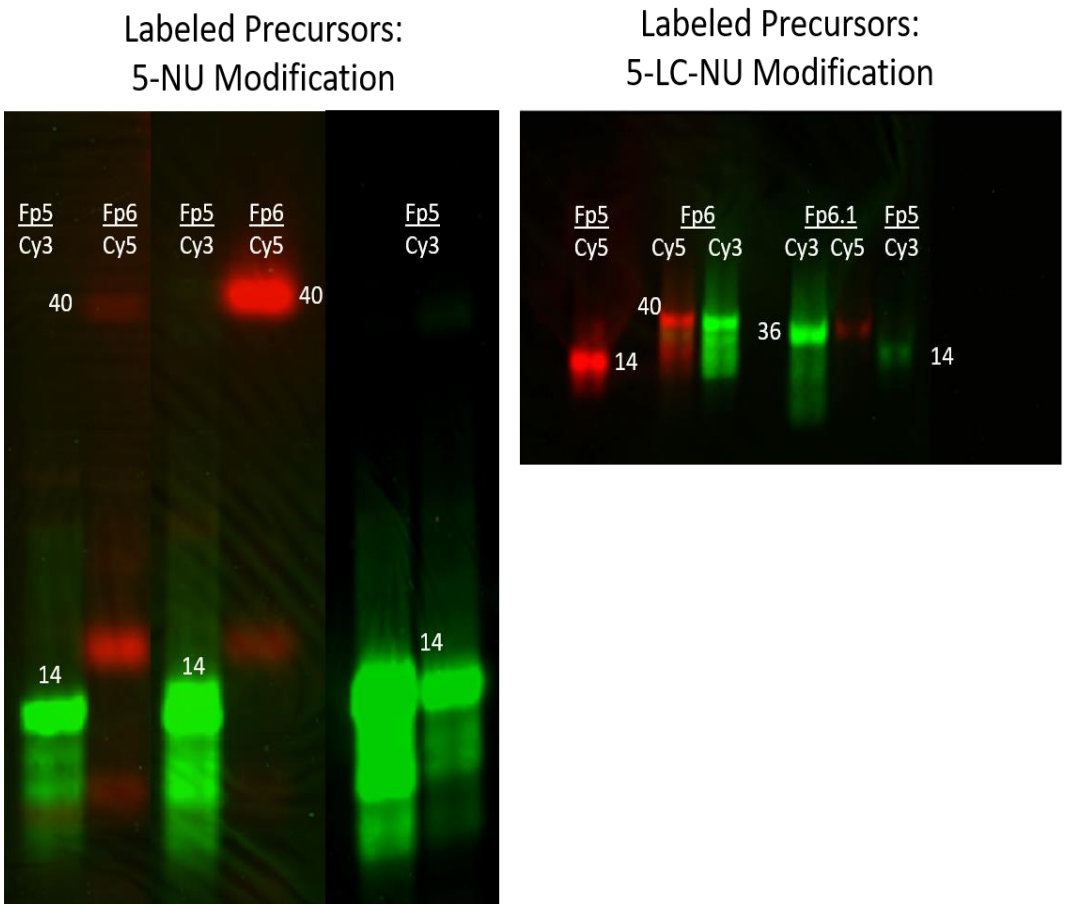


Figure 2.3. Comparison of 5-NU modified and 5-LC-NU modified precursor integrity. The gel on the left shows 5-NU modified RNA precursors following labeling and FPLC purification. The gel on the right shows the same, but using the 5-LC-NU modification. It is clear that the 5-NU was more susceptible than the 5-LC-NU to degradation during the labeling process. This is especially apparent in the Fp5 precursor. Full-length Fp5, Fp6, and Fp6.1 are 14, 40 and 36 nt, respectively.

once, at or near the modification site. This would leave three labeled bands on the gel; the full-length labeled precursor, and two labeled pieces representing each broken half of the precursor. While some degradation is difficult to avoid, the degree of degraded products is much higher in the 5-NU precursors. As apparent for Fp6-Cy5, seen in Lane 2 of figure 2.3, sometimes the degraded precursor has higher labeling efficiency than the full-length precursor.

DNA-SPLINTED FOUR-WAY RNA LIGATIONS

Test Ligations

Results of an example test ligation can be found in Figure 2.4, and a summary of the ligation ratios used can be found in table 2.3. The difficulties in labeling became problematic when generating the two mutant constructs, which due to supplies were made using the 5-NU modified precursors. Despite this, it was still possible to determine an optimum ratio of all components.

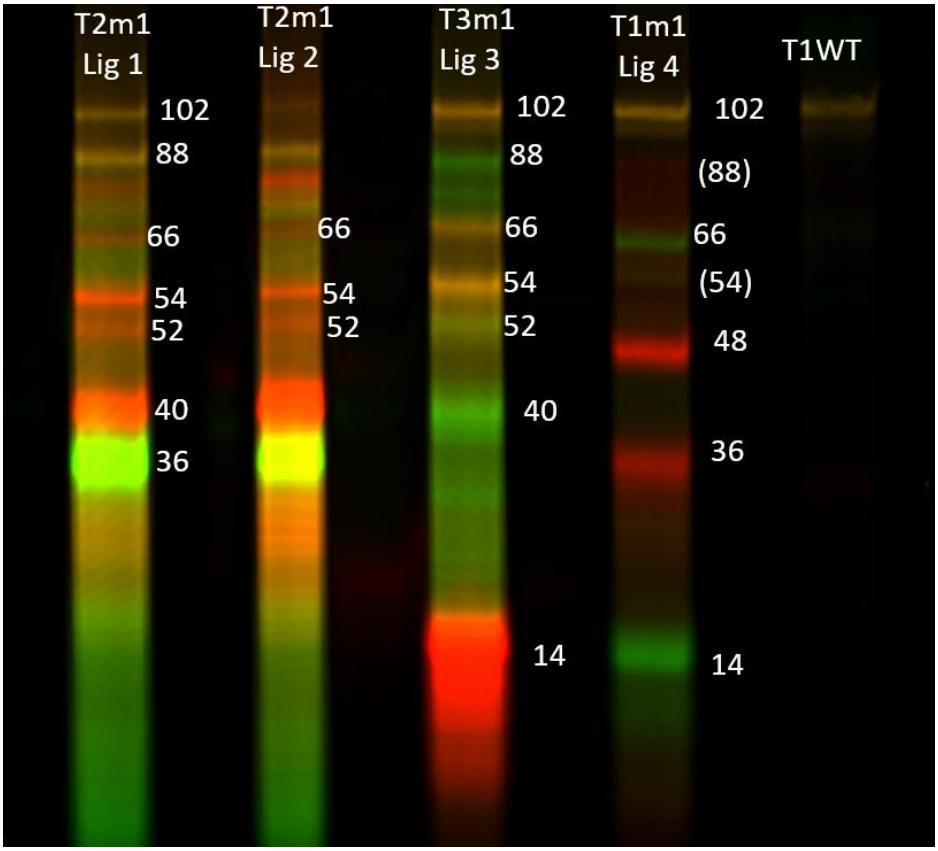


Figure 2.4. Example of Test Ligations for the delta A60 mutant. Ratios of ligation components can be found in Table 2.3. The wildtype T1 construct is used as a size standard for full-length products. Sizes of precursors and any possible labeled ligation products, full-length and incomplete, are as indicated. Sizes in parentheses represent bands that would be present as an incomplete product but are not visible.

Table 2.3. Ratios of components used in the test ligations shown in figure 2.4.

Component	Ratios			
	Lig 1: T2m1	Lig 2: T2m1	Lig3: T3m1	Lig4: T1m1
Fp5	1X	0.5X	1X	1X
Fp6	0.5X	1X	0.5X	0.5X
Splint1	1X	0.5X	1X	1X
FJ	1X	1X	2X	2X
Fp6.1	1X	1X	1X	1X
Splint2	1X	1X	1X	1X

It can be deduced from the Cy3-labeled band below 88nt in Ligation 3 that the Fp6.1 band is fragmented, as can also be concluded from the Cy3-labeled band below 40nt in that lane. This is consistent with the example gel of the labeled precursors in figure 2.3, which shows significant fragmentation of the Fp6 precursor. It is also possible to conclude that the excess of FJ, the unlabeled junction precursor, successfully drove complete ligation of the Fp6 fragment, as is evidenced by the lack of a labeled 48nt band in ligation 3, and the lack of a labeled 54nt band in ligation 4.

Preparative Scale Ligations

Preparative scale ligations for the WT T2 construct used equimolar amounts of all components, while WT T3 used 2X FJ **:** 1X all other components. Following results of the test ligations, all mutant constructs were generated using 0.5X Fp6 **:** 1X Fp5, Splint1, Fp6.1, Splint2 **:** 2xFJ. An example PAGE of T2m2 and T3m2 is shown in Figure 2.5, before the bands are excised (left) and after the bands are excised (right).

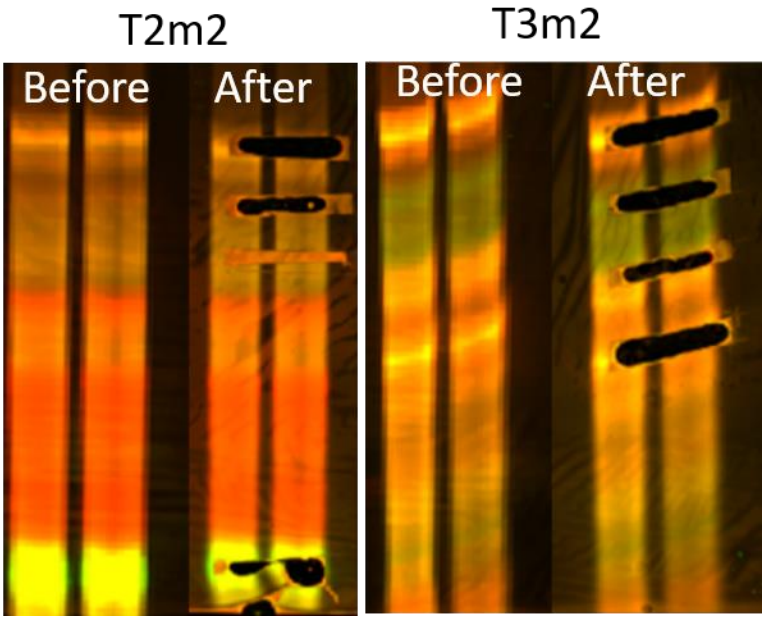


Figure 2.5. Example PAGE of two preparative scale ligations. Left side of both show the gel as it was before bands are excised. The right side shows the gel following band excision.

For supply reasons, most of the ligations using mutant junction fragments were performed using the 5-NU modified precursors. Because of the difficulty in obtaining quality labeled precursors, it was not possible to completely separate labeled Fp6 from unlabeled Fp6. Thus ligations requiring Cy3 or Cy5-labeled Fp6 would inevitably yield some full length ligation products that incorporated unlabeled

Fp6. Unfortunately, these could not be separated from products incorporating the labeled Fp6. This in turn led to difficulty obtaining a good spread of doubly-labeled molecules adhering to the slide surface during FRET experiments. It was also not possible to produce full-length T1m2 at the preparative scale. This issue is unrelated to the issues with Fp6 labeling, as the T1 constructs use the unlabeled version of this precursor. It is possible that the closed-loop mutation in P6.1 combined with the nearby label modification on Fp6.1 destabilized or disfavored complete ligation. Yields of T2m2 and T3m2 were also much lower than for the deltaA60 mutant, making it difficult to obtain sufficient FRET data. Final ligated products used for all FRET experiments are shown below in Figure 2.6

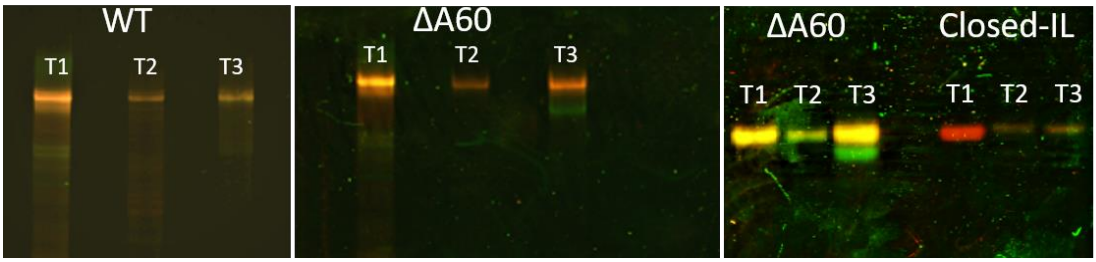


Figure 2.6. PAGE of all completed constructs. The top band of the left gel shows the wildtype constructs, while the top band of the center and right gels show the bands for the DeltaA60 and Closed internal loop mutants, respectively.

FRET MEASUREMENTS

WILDTYPE CONSTRUCTS

FRET measurements of the T1 construct show a single population of molecules at .67 FRET, as shown by the histogram in Fig 2.7. Traces were very

consistent, as shown by the example traces, with only a very small number (<1%) showing any dynamic behavior. The T2 construct yielded two populations, the major population at .17 FRET and the minor population at .67 FRET. Thus it seems that the P6-P6.1 label scheme is able to detect a conformation to which the P5-P6.1 label combination is blind. These molecules were also very consistent, with less than 10% showing FRET states outside of the two populations. These results suggest that the P6.1 stem is in close proximity with the P5 stem while the P6 stem is distant, likely from both.

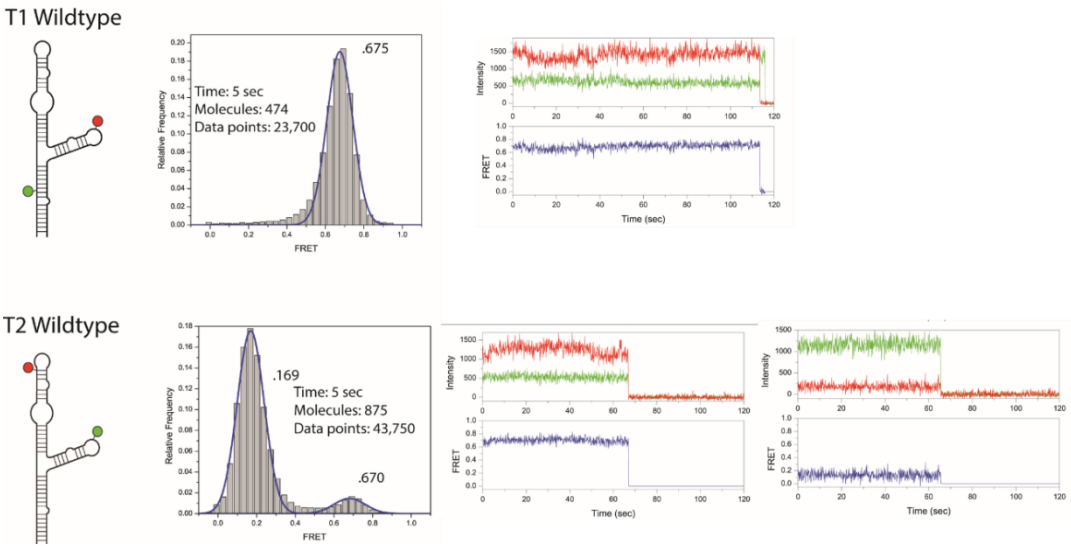


Figure 2.7. Histogram and Trace Examples for T1 and T2 wildtype Constructs. Example Traces for each construct are to the right of the population histogram of that construct.

To further probe the interactions of these three stems, the T3 wildtype construct, which is labeled on P5-P6, was examined. This construct, surprisingly,

shows 4 distinct populations; the major population is at .27 FRET, two roughly equal sized populations are found at .55 and .76 FRET, and a small population is found at .97 FRET. This is an unusual result, since the T1, T2, and T3 constructs are the same RNA with the only difference being in the label combinations. Logic would dictate that at least two of the constructs should show the same FRET populations. It was noted, when analyzing the traces from each construct, that they show an increasing amount of dynamic behavior. T1 has almost no highly dynamic molecules, T2 has dynamic molecules in approximately 5-8% of the traces, but in the T3 construct at

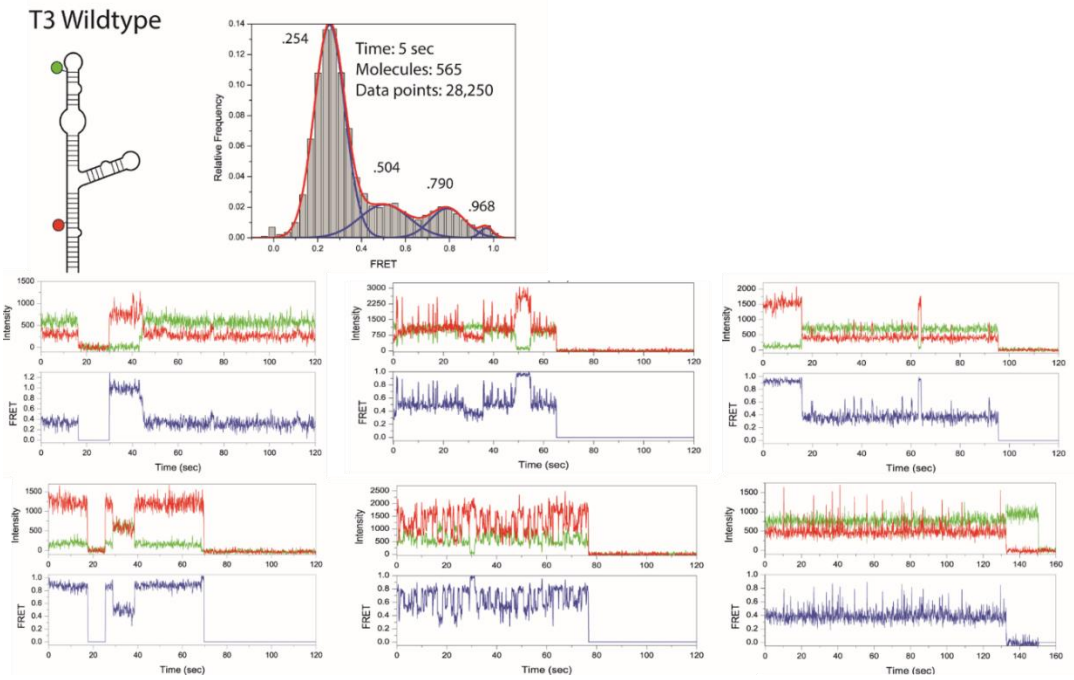


Figure 2.8. Population Histograms and Example Traces of T3 wildtype. The population histogram for all traces of the wildtype T3 construct at 1mM magnesium is shown. Below are examples of highly dynamic traces from the T3 construct, which occurred in this construct with much higher frequency than the T2 and T1 constructs.

least 20% highly dynamic molecules. Figure 2.8 shows the population histogram at 1mM Mg, and examples of the highly dynamic molecules found in the T3 construct.

In depth probing of the Dynamic Behavior Displayed in Minority Populations

Sorting molecules based on FRET behavior: Stable and Dynamic populations

To further understand the populations found in the T3 construct, the traces were separated into groups; traces that showed dynamics and those that did not. To further study this, the T3 and T2 traces were sorted into two categories; stable (non-changing) FRET of any state, or dynamic (changing) FRET states. If a given trace did not deviate greatly from the FRET state it started in (for example, ~.7(+/-.5)) then it was considered a stable trace. If there was a change in the FRET state, (for example, vacillating between .7 and .4) then it was considered a dynamic trace. Examples of each of these can be seen in figure 2.8.

The sorted populations showed an interesting result; the stable molecules found in the T3 construct had a distribution of FRET states that is almost identical to the total distribution of molecules seen in the T2 construct. The lower panel of figure 2.9 shows the population histograms of the total traces, stable traces, and dynamic traces found in T3. Stable molecules make up ~80% of all molecules, leaving approximately 20% dynamic molecules. The centers of the FRET populations found in the dynamic molecules are .31, .53, .77, and .97 FRET, values very similar to those found in the WT T3 histogram representing all molecules (see figure 2.9, compare bottom left and center histograms). This suggests that the minor peaks seen in the total trace histogram are, in fact, derived primarily from highly dynamic molecules.

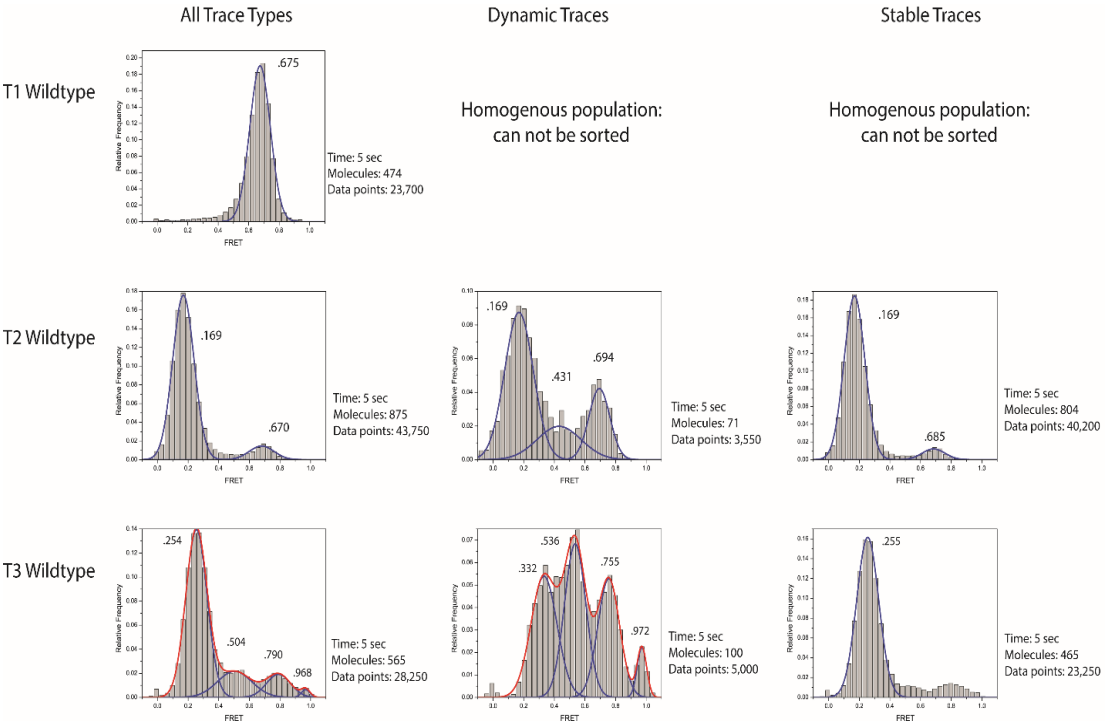


Figure 2.9. Population histograms of sorted wildtype constructs. This figure shows the histograms, number of molecules, and time interval samples to generate the histograms shown. The number of data points corresponds to the total number of molecules sampled every 100ms for the indicated time interval. Thus in a 5 second time interval, there are 50 data points recorded for every molecule.

This is further confirmed in the histogram of the stable molecules, which shows one primary population of molecules, centering at FRET of .23, consistent with the major population found in the all traces histogram.

This data could indicate that the P6 stem is in primarily one state, far from the P5 stem. Recalling the T2 data, we can conclude that the P6 stem is also quite far from the P6.1 stem. This indicates a molecule in which the P5 and P6.1 stems are nestled closely together, while the P6 stem spends most of its time far away. However, it is clear there are other possible positions for the P6 stem. It is certainly

not locked in a single rigid conformation. The dynamic behavior found in the T3 construct is evidence that the molecule is capable of sampling different conformational states, either through complete collapse and rearrangement of the molecule, or through movement or repositioning of the stems. If there was a complete collapse and rearrangement, it would be likely to impact the T1 and T2 constructs as well. It is important to note that the main population of the T2 construct is found at .2 FRET, which is quite low. This is outside the range of FRET that provides sensitive distance information. Simply put, it is too low for the molecules to be able to accurately report on changes in the conformation. T2 would only be sensitive to changes which bring the P6 and P6.1 stems closer together. Thus, a change in the conformation would have to occur that would; keep the P5 and P6.1 stems in close proximity, keep the P6/P6.1 stems far apart, and finally keep the P5/P6 stem apart but allow for movement that brings ONLY those two stems closer together. Since this seems rather difficult, other conformational possibilities yielding this result must be possible. Given the length of the P6.1 stem, and the label position on the P5 stem, it is possible for the P6.1 stem to move in the x-axis and the z-axis, and remain the same distance from the P5 label position. Thus, it is conceivable that a conformational change could occur which would move P5 and P6 closer together, but keep the P6.1 far from the P6 stem, and allow the P6.1 stem to move in relation to the P5 stem *while retaining the same FRET ratio*. This would explain why the T1 construct reports differently than the T2 and T3 constructs. The results of the T2 wildtype construct support this scenario. If the conformation that makes up the majority of molecules is

such that the label position on the P6.1 stem is essentially equidistant from the other label sites, then the T2 construct would also report a nearly homogenous population.

Probing Dynamics at Different Magnesium Concentrations

In an effort to determine whether a complete structural collapse and rearrangement, or a conformational change was responsible for the dynamic behavior found in the T3 molecules, different concentrations of Mg^{+2} were tested. The rationale behind this being that a complete collapse would be disfavored by the addition of magnesium, as it would stabilize the overall structure. On the other hand, if there is simply a conformational change in the stems, an alteration in the concentration of magnesium might shift the population to favor the less likely structure, since it is presumably higher energy. As shown in figure 2.10, a titration range of 0-40mM of $MgCl_2$ was examined. Unexpectedly, these concentrations of magnesium did little to alter the FRET states observed. The T3 construct was sorted, as shown in figure 2.10, to determine if magnesium changed the dynamics of the molecule. As shown, the most striking result of sorting was the merging of all the minor FRET populations into a single minor population. The same experiments were repeated with the T2 construct and yielded no change in the distribution of molecules in the population (data not shown). It was then considered that the range of magnesium tested was too high, so the experiments were repeated using 0mM, 1mM, and 10mM $MgCl_2$. Results of this are shown in Figure 2.11, the FRET values are plotted as a continuous line for each magnesium concentration, to illustrate the relative frequencies of populations in each condition.

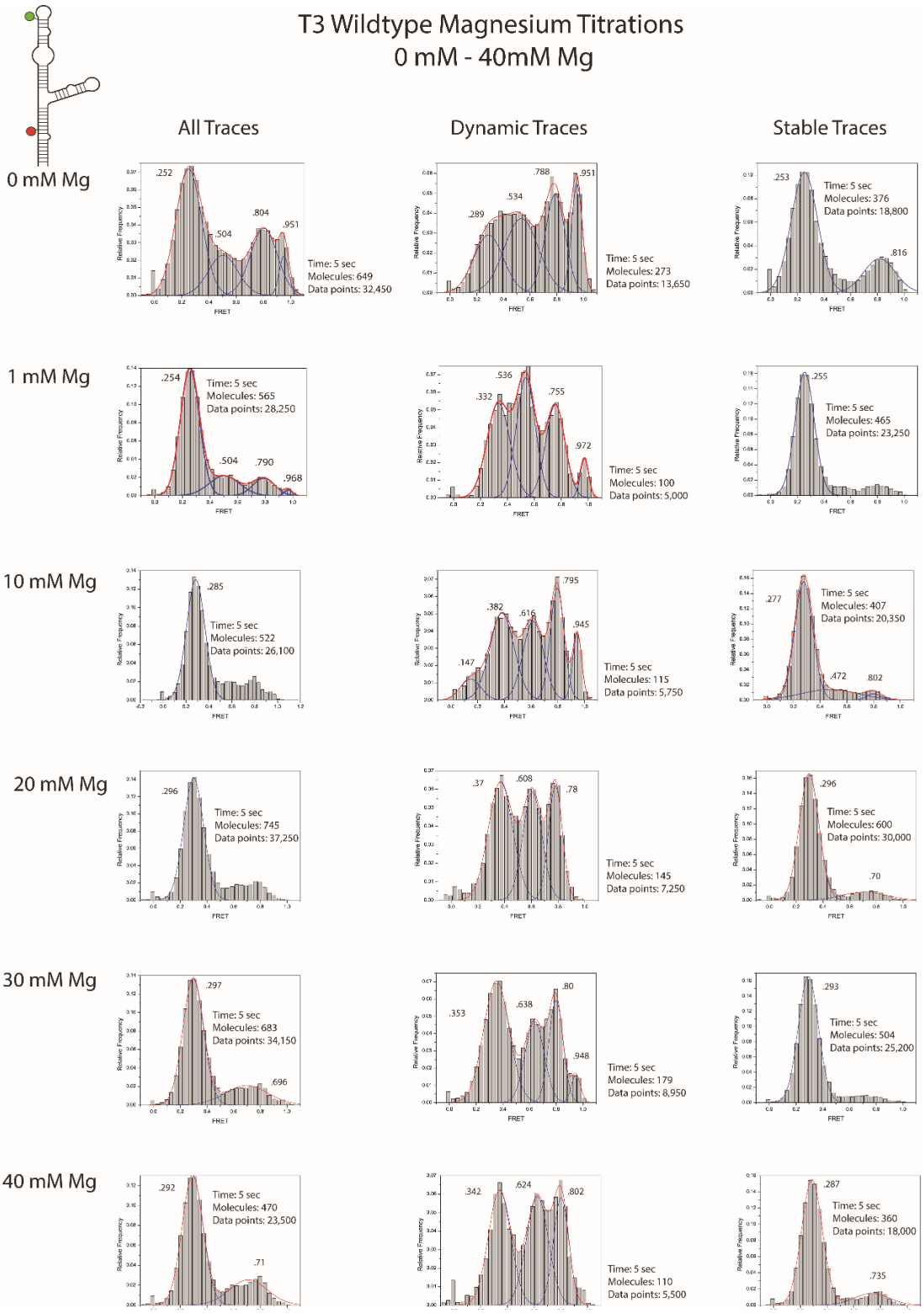


Figure 2.10. Population histograms of Sorted Wildtype T3 molecules at 0-40mM Magnesium. The left column shows the population of dynamic molecules and the right column shows the population of stable molecules.

Unfortunately, these too had little impact on the FRET results. Again, this was also performed in the T1 and T2 constructs, which showed no significant change from the results of 1mM magnesium, as shown in figure 2.11.

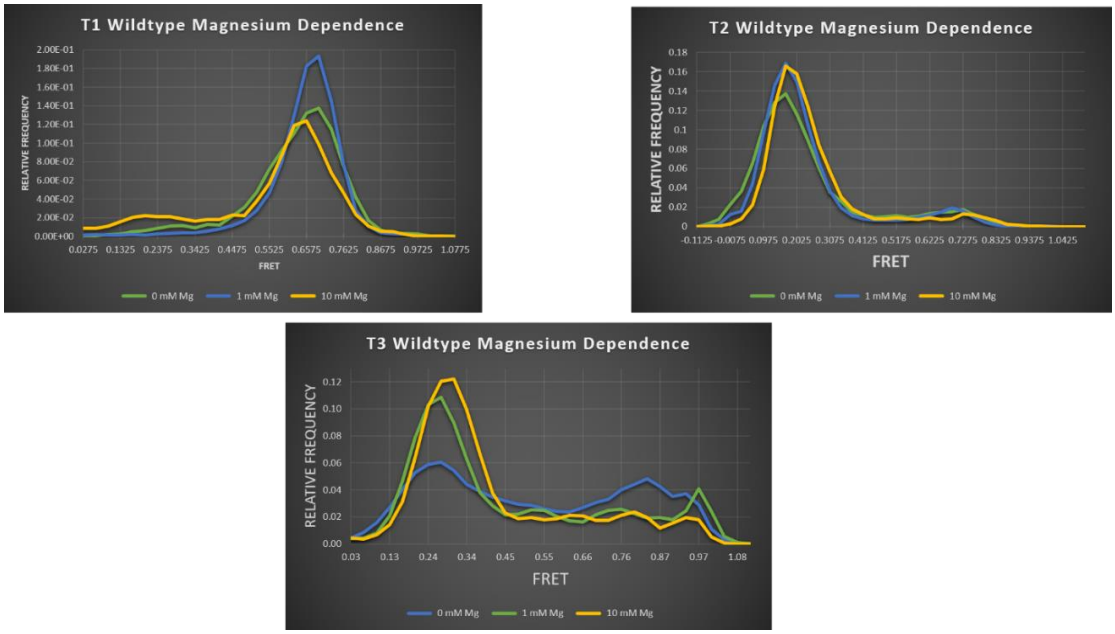


Figure 2.11. Distribution of FRET populations for all wildtype constructs at 0mM (blue), 1mM (green), and 10mM (yellow) magnesium.

Determining the Impact of Est2

Since the medaka solution NMR structure showed significantly different conformations for the TRBD-bound CR4/5 and the free CR4/5, it was important to test whether the addition of Est2 (k. lactis telomerase) would change the FRET populations. For these experiments, a unit of Est2 is defined as the amount of Est2 required to shift 1nM of RNA on a gel mobility shift assay. A titration of Est2 ranging from zero to 1.2×10^{-3} was tested. There was little change in the total population histograms of T1 and T2, and a very minor difference in the ratio of the

FRET populations of dynamic T3 molecules. Figure 2.12 shows a comparison of 0 Units Est2 with $6x10^{-2}$ Units Est2. This concentration showed the most dramatic change in the population histograms. A range of .12 to $1x10^{-3}$ units of Est2 was tested; this concentration range was based upon the binding capacity reported by our collaborators of 1U /.001pM RNA. As seen in the histograms, there is little impact on the positions of the FRET peaks with and without Est2. Data taken in the absence of Est2 does show more broad peaks than expected or found in pervious experiments. This may be due to the difference in the buffer composition between this experiment and previous experiments. For consistency with channels containing protein, the channel without Est2 was imaged in buffer containing 40µM beta-mercapto-ethanol. This may have caused the broadening of peaks observed in this data. Additionally, the transportation of the protein was not carried out under ideal conditions. Prior to use, the sample was only subjected to crude purification with nickel beads, and was not purified by size exclusion or ion exchange chromatography. This left the protein samples with degraded Est2 and extraneous cellular contamination, which hampered efforts to obtain high quality data. Therefore, there is little of consequence that can be determined from these experiments. The binding capacity of the Est2 used in this experiment was tested by EMSA, as shown in Figure 2.13. Lane 2 of this gel, which shows the binding capacity of .625U of Est2 with .01pmols of RNA, is most similar to the conditions tested with smFRET. As observed, the sample was not competent to bind at this concentration, which is consistent with the smFRET results. The EMSA was graciously performed by Christina Palka.

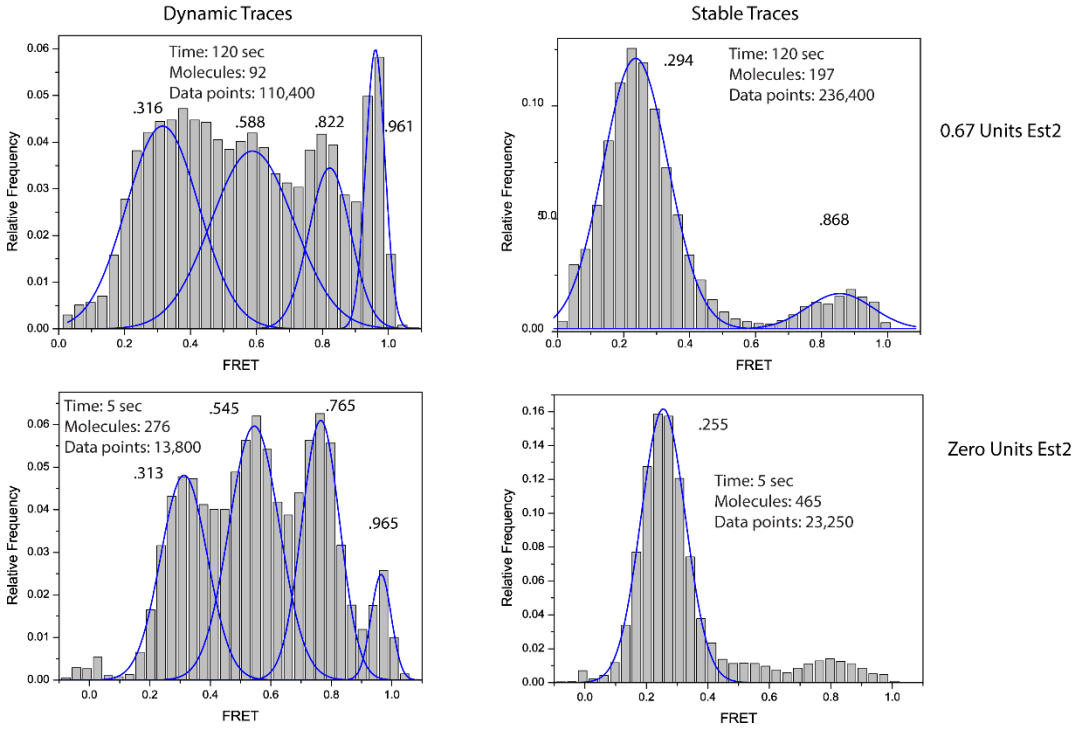


Figure 2.12. Addition of Est2 to Wildtype T3. Stable and Dynamic Traces of wildtype T3 with and without Est2 protein.

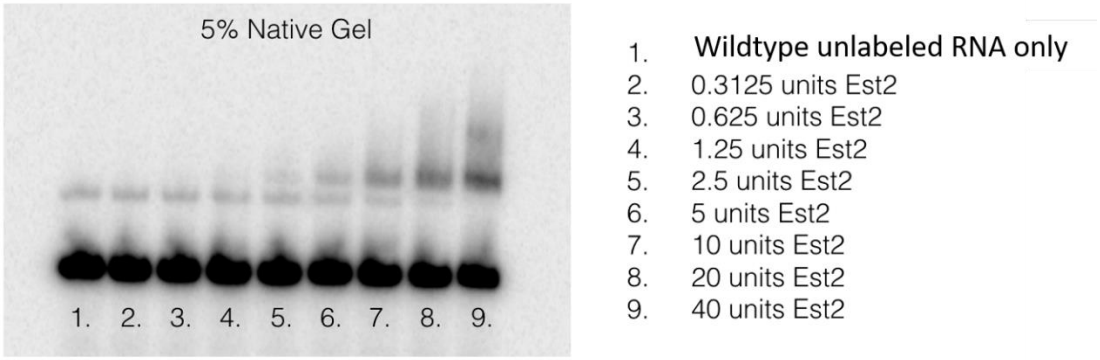


Figure 2.13. EMSA of Est2 protein used in smFRET. This figure shows the binding capability of the Est2 protein sample used in this work. The highest band represents Est2 binding to *K. lactis* CR4/5 RNA As shown in lanes 7-9, only a high concentration of this protein was competent for binding.

MUTANT TWJ CONSTRUCTS

Invariant Junction A deletion: ΔA60

To investigate the impact of alterations to the universally-conserved junction A nucleotide, mutant constructs were generated as described in the "Construct building" section of this chapter. Mutants were tested for FRET behavior at 0mM, 1 mM, and 10mM magnesium chloride. Whenever possible, traces were sorted into categories based on dynamics, as was performed for the wildtype constructs. Figure 2.14 shows the results of FRET experiments performed at 1mM $MgCl_2$ for T1ΔA60, T2ΔA60, and T3ΔA60.

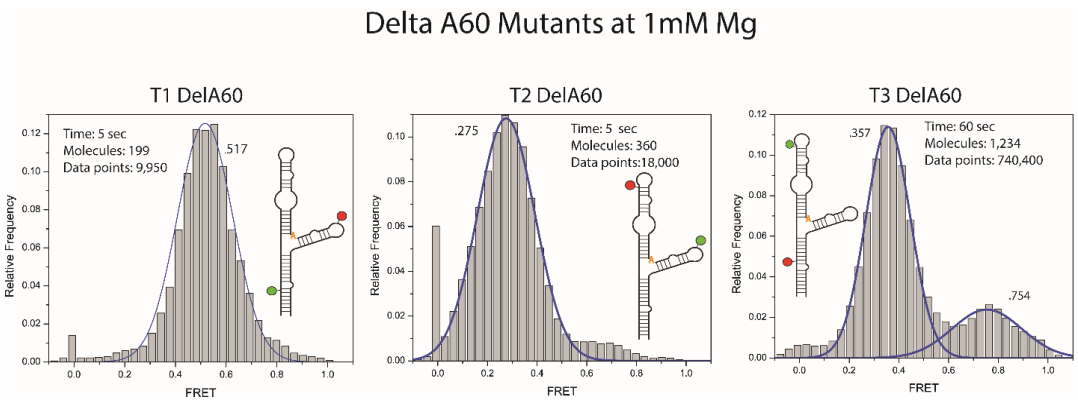


Figure 2.14 A60 deletion mutants at 1mM Mg. Above shows the population histograms for each mutant construct at 1mM $MgCl_2$. The centers of each Gaussian distribution appear next to their corresponding peaks.

As can be seen in Figure 2.14, the impact of the deletion on T1 was to broaden and shift the FRET range of the population to .517 FRET, which is lower than wildtype. The T2 deletion mutant also had a broader peak, but the population centered on a FRET of .275, which is similar to wildtype. There was also a decrease in the frequency of the minor population that was found in wildtype T2, as seen in the

central histogram of figure 2.14. The T3 deletion mutant showed a significant difference compared to the populations found in the wildtype construct. There were only two populations found for the mutant, which centered at .357 and .754 FRET. The major population, at .357, is shifted significantly from the .27 population found in the wildtype. The absence of the .55 and .97 minor populations seen in wildtype T3 is striking, and corresponds with an increase in the frequency of molecules in the ~.75 FRET peak. The .55 and .75 FRET minor populations in the T3 wildtype construct each represented approximately 10% of the total data, while the .9 FRET population represented approximately 5% of the total data. This leaves ~75% of the total data in the major FRET state at .27. The minor FRET population in the .75 FRET peak in the T3delA60 construct does not quite reach 30% of the total data, which is consistent with a circumstance in which the minor populations are less distinct and the FRET representing those molecules converge to a single, broad peak.

Dynamics of the T3 ΔA60 mutant

To determine if the dynamic behavior of the molecules in the mutant constructs was different than the behavior of the wildtype, the data for T3 were sorted.

As described earlier, the peak encompassing the major population is broader and at higher FRET in the T3delA60 mutant as compared to the wildtype. As

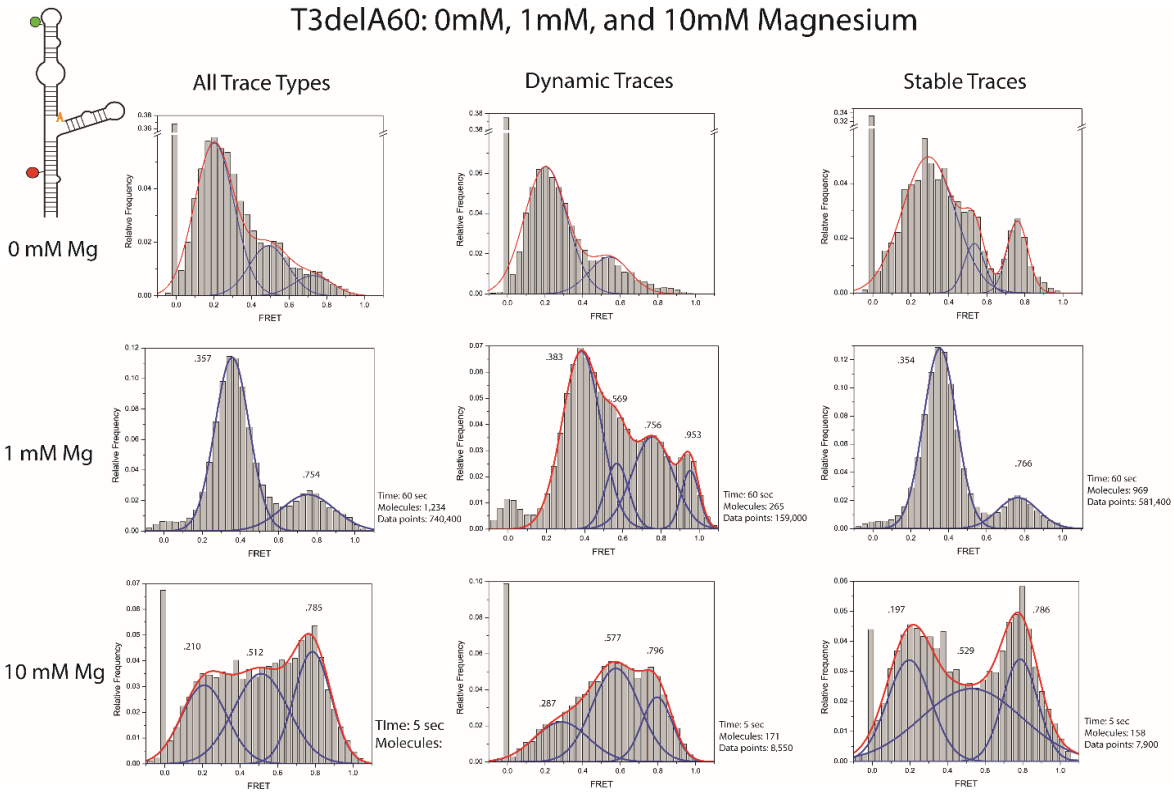


Figure 2.15. T3ΔA60 mutant sorted molecules at 0mM, 1mM, and 10 mM Magnesium. The histograms above show the distribution of molecules in the T3ΔA60 mutant at varying magnesium concentrations. The red line represents the cumulative fit, the blue lines show the fit of individual populations.

discussed above, Figure 2.15 shows that rather than a single minor population at .7, the dynamic traces show that the .7 FRET peak is a result of a widening of the ~.5 and ~.9 FRET populations. The change in the population distribution is seen most dramatically in the population centered around ~.5 FRET, which is almost reduced to a wide shoulder between the more distinct .7 and .3 FRET populations. The sorted traces also highlight the change in the relative frequencies of populations in the dynamic traces. In the wildtype construct, the .5 population had the highest frequency, followed by the .7 FRET, .3 FRET, and .9 FRET populations,

respectively. However, in the A60 deletion mutant, there is a significant decrease in the .5 FRET populations, and an increase in the .3 FRET and .7 FRET populations. The traces showed much more noise than the wildtype traces, which could be due to a general instability or breathability in the overall molecule. It seems likely from this data that removing the A60 nucleotide introduced some destabilizing forces into the molecule.

When 0mM and 10mM magnesium were investigated, the differences between wildtype and ΔA60 were even more pronounced. Rather than help stabilize the molecule, as would be expected, the 10mM magnesium buffer significantly reduced the stability of the RNA. There were far fewer molecules displaying any FRET behavior, and those that did showed significant fluctuations in the amplitude of fluorescence intensity throughout the trace. Again, this is indicative of an unstable fold, which has significant freedom of motion. Surprisingly, the 0mM magnesium condition did not cause the same degree of potential instability as high magnesium. In fact, the major population is closer to the FRET value of the wildtype than the 1mM $MgCl_2$ population.

All ΔA60 mutants at 0-10mM Magnesium

Figure 2.16 shows the aggregate data for each of the delA60 constructs tested at varying magnesium. As was seen in T3ΔA60, higher magnesium did not seem to greatly improve the stability of the molecules. In fact, unlike T3ΔA60, the T1 mutant

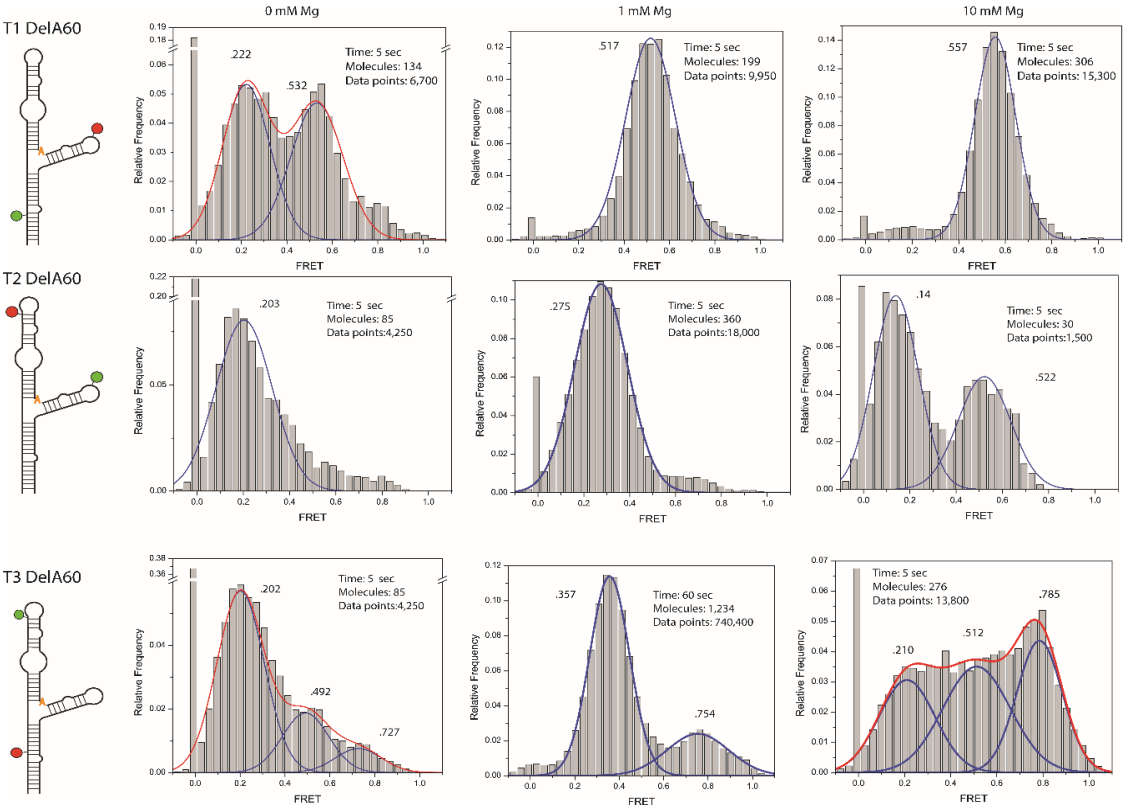


Figure 2.16 Total Data for All Delta A60 mutants at 0mM, 1mM, and 10mM magnesium. This figure shows the population histograms of all traces for each deletion mutant at 0-10mM magnesium. Top row represents the T1delA60 construct, middle row represents T2delA60, and the third row represents the T3delA60 construct. Red lines show the cumulative fit for all data, blue lines represent the fit for each population individually. Insets show the details of each dataset used to generate the histograms. Diagrams on the left are representations

did not seem to have a significant destabilization at 10mM $MgCl_2$, and in fact was more stable at this concentration than at lower magnesium. This behavior is more in

line with what the anticipated effect of high magnesium would be, to stabilize an unstable structure. T2delA60 was much less stable at 0mM and 10mM magnesium than at 1mM, as is evidenced by the low number of molecules that could be measured at those concentrations. The high bar present in many of the histograms in Figure 2.16 represent the zero FRET populations. Due to the instability of the molecules, many either bleached or collapsed in the time interval measured. Normally these would not be included in the data used to generate the histograms, however the low number of molecules that did not bleach or collapse during the time interval made it impossible to exclude those molecules. Shorter time intervals did not contain sufficient data to generate histograms, again because of the overall low number of stable molecules in these constructs. Histograms from these samples are included for a comprehensive view of the destabilizing impact of A60 deletion, but these histograms are unreliable. The FRET populations from histograms with high zero-FRET populations should not be taken to as an accurate reflection of the conformational structure of these constructs.

CONCLUSIONS AND FUTURE DIRECTIONS

In summary, the CR4/5 region of the *K. lactis* telomerase RNA displayed conformational flexibility in solution. The construct most sensitive to dynamic motion was T3, labeled on the P5 and P6 stems. None of the wildtype constructs were sensitive to changes in magnesium, which is consistent with the NMR solution

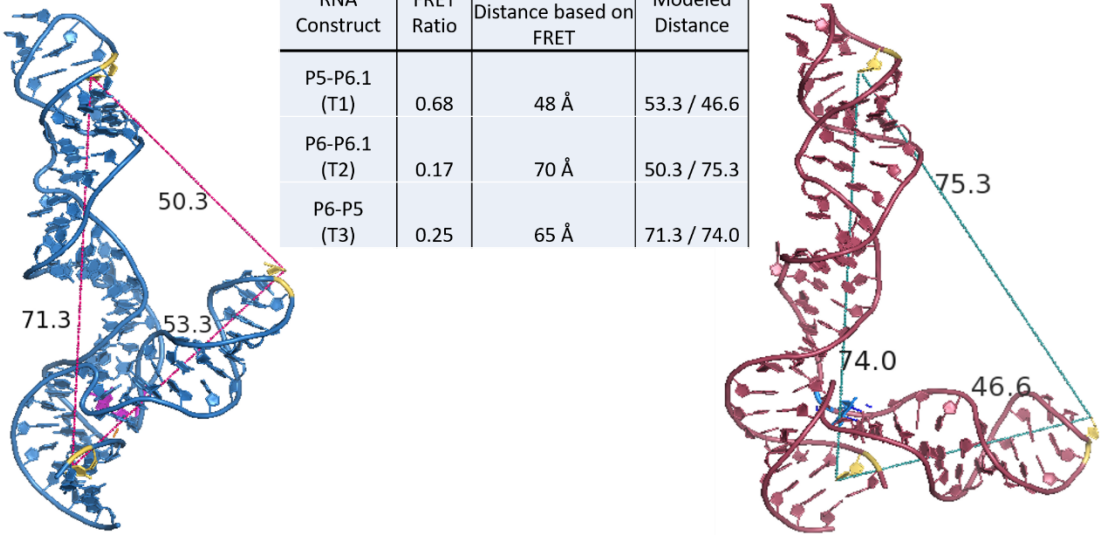
structure reported by Kim *et al.*, which also did not display changes in conformation at different magnesium levels.

The T1 construct showed a single population at .68 FRET, which would correspond to a distance that is less than 54 Å, based on an approximate Förester radius of 5.4 nm for Cy3 and Cy5. Modeling of lowest-energy structures of the *K. lactis* CR4/5 was performed by Dr. Nikoli Ulyanov, and is shown in figure 2.17. Using these, it was possible to compare the distances between the stems extrapolated from modeling with that calculated from FRET ratios. The distance between P5 and P6.1 estimated from FRET data closely resembles the distance determined from the red model shown on the right-hand side of the figure.

The major population of both the T2 and T3 wildtype constructs centered near .2 FRET, which would correspond to roughly 8-10 nm distance. The major populations of T2 and T3 constructs are both in the low FRET range, which is notably outside the range of FRET values that provide accurate correlations with distance. Despite this, values of 70 Å and 65Å between the stems were calculated for T2 and T3, respectively. Comparing these to the blue structure on the right-hand side of the figure, which is similar in shape to the closed form of the medaka CR4/5 structure, both approximated values differ by at least 5 Å from the models (the T2 differs by 20 Å). Comparison with the left-hand structure, which resembles the open conformation of medaka CR4/5, the T2 distance estimates begin to converge, at 70 Å and 75 Å. This indicates that the T2 construct could be reporting on a conformation similar to

the open conformation. The T1 data is also consistent with the molecule primarily adopting the open conformation.

The T3 distance estimates are more disparate when compared to the open-conformation model, which would seem to contradict the results of the other two constructs. However, since neither the closed or open models of *kl*CR4/5 closely match the FRET-derived distances for T3, it seems that this construct is either not a valid reporter on the structure, is reporting on a different conformational state, or the



RNA Construct	FRET Ratio	Estimated Distance based on FRET	Modeled Distance
P5-P6.1 (T1)	0.68	48 Å	53.3 / 46.6
P6-P6.1 (T2)	0.17	70 Å	50.3 / 75.3
P6-P5 (T3)	0.25	65 Å	71.3 / 74.0

Figure 2.17. Models of *K. lactis* CR4/5 and corresponding distances from FRET data. The blue structure to the left represents a model of the wildtype *kl*CR4/5, with the bases labeled in the FRET constructs highlighted in yellow. Distances between each label position are shown as measured in Angstroms. The red structure to the right represents an alternative low-energy structure of *kl*CR4/5, with distances and label positions indicated as for the blue structure.

distance estimated from the FRET ratio is not valid, because it is in the low range. It is interesting to note that none of the minor populations found in T3 fit with the modeled distances between the P5 and P6 stems. The minor populations, which are

comprised almost entirely of dynamic molecules, could represent either intermediate conformations sampled during interconversion between two states or distinct and independent conformations. The behavior observed in the center bottom trace in figure 2.8, a single molecule interconverting between the FRET states of each minor population, was common, leading this author to favor the possibility of intermediate conformations rather than individual states. Despite this, whatever intermediates present are stable enough to be sampled frequently in the 10ms timescale. While the T2 construct also displayed traces with similar dynamic behavior, they were far fewer. In combination with the results of the T1 constructs, it is likely that the length of the P6.1 stem is such that the dye-label position that was chosen is nearly equidistant from both of the other stems at all conformations. Given the high degree of structural homology displayed by other structures in the telomerase RNP, it is reasonable to assume that the *K. lactis* CR4/5 likely adopts an 'open' and 'closed' structure similar to the medaka CR4/5 reported by Kim *et. al* and Huang *et al.* In this case, the most dramatic movement between the stems occurs when the P6.1 stem swings down and around with respect to the P6 stem. If this is so, then it is expected that the T2 construct would report on this conformational change best. In the 'closed' conformation a high FRET would be expected, while a low FRET is expected in the 'open' conformation. We do observe two FRET states, high and low, in the T2 construct. The distribution of these states indicates that the molecule spends a majority of time in the 'open' conformation. What is surprising is the observation of multiple FRET states in T3, since both the 'open' and 'closed' conformations appear

to have coaxially stacked helices. It is possible that, while the junction is rearranged between the open and closed conformations, the P6 stem bends at the junction and the arm moves into distinct intermediate positions while specific rearrangements are made incrementally in the junction and P6.1 base pairs. In the medaka CR4/5, there are three critical base pairs that are found in the 'open' conformation, which are not present in the closed conformation; two base triples and one wobble pair. Formation and breakage of these triples and wobble pair could account for the three minor FRET populations seen in the T3 construct. At this point it is unclear whether the major population at .2 FRET would contain molecules that are stable in either the 'open' or 'closed' conformations, leaving the minor populations to be composed only of intermediates, or if one of the minor populations represents one of the two final conformations.

The results of the A60 deletion mutant are consistent with the hypothesis that each intermediate state represents formation or breakage of one of the stabilizing base pairs. In the medaka structure A199 (A60 in klCR4/5) is not paired in the 'closed' conformation, but in the 'open' conformation it forms a crucial wobble pair with G213. When the A60 base is missing in the mutant constructs, the discrete intermediate populations become blurred. Since there are two stabilizing base triples that would still be formed in the A60 deletion construct, the complete loss of all intermediate FRET populations is not a likely result of loss of the G•A wobble. This is true for the T3ΔA60 construct, although the presence of the minor populations is less apparent in the unsorted traces as compared to the wildtype construct. It is

possible that the mutant molecule is striving to form a less stable wobble pair between G84 (homolog of G213) and A61. This would have to compete with the wildtype base pair that is predicted to form between A61 and U83. The search for a stable pair for G84 could account for the wide distribution of FRET values observed in the A60 deletion construct. Currently it is not possible to conclude whether the FRET intermediate states do indeed correspond with base pair rearrangements. An interesting question to pursue would be whether disruption of the two base triples formed in the 'open' state lead to a loss of any of the three minor populations.

While certainly not definitive, it is possible to conclude from this work that the CR4/5 region spends the majority of time in the 'open' conformation, as supported by the T1 and T2 wildtype constructs. Loss of the A60 base is destabilizing, but not debilitating, to the structure of the molecule, which is also consistent with the region being primarily in the 'open' conformation. Further work is necessary to confirm and expand upon these results. The most salient avenue to pursue would be to perform similar FRET experiments on the medaka CR4/5, so that comparisons could be made between these homologs. However, this is limited by the small size of the medaka CR4/5, which would place all the stems in the high FRET range in both conformations. Other informative work would be to generate mutant constructs of the *K. lactis* CR4/5 that destabilize either, or both, of the base triples that are predicted to form in the 'open' configuration, and observe the distribution of minor FRET populations. It would also be informative to determine the difference in

protein binding in the ΔA60 mutants, as the homolog of this base makes critical contacts with the TRBD in medaka.

References

1. Garforth, S. J., Wu, Y. Y. & Prasad, V. R. Structural features of mouse telomerase RNA are responsible for the lower activity of mouse telomerase versus human telomerase. *Biochemistry* **397,** 399–406 (2006).

2. Tang, W., Kannan, R., Blanchette, M. & Baumann, P. Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* **484,** 260–264 (2012).

3. MacNeil, D. E., Bensoussan, H. J. & Autexier, C. Telomerase regulation from beginning to the end. *Genes (Basel).* **7,** (2016).

4. Schmidt, J. C. & Cech, T. R. Human telomerase: biogenesis, trafficking, recruitment, and activation. *Genes Dev.* **29,** 1095–1105 (2015).

5. Lai, A. G. *et al.* The protein subunit of telomerase displays patterns of dynamic evolution and conservation across different metazoan taxa. *BMC Evol. Biol.* **17,** 107 (2017).

6. Podlevsky, J. D. & Chen, J. J.-L. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* **13,** 720–32 (2016).

7. Tzfati, Y. & Chen, J. J.-L. in *Telomerases: Chemistry, Biology, and Clinical Applications* (eds. Lue, N. F. & Autexier, C.) 23–51 (Wiley & Sons, Inc., 2012).

8. Gunisova, S. *et al.* Identification and comparative analysis of telomerase RNAs from Candida species reveal conservation of functional elements. *RNA* **15,** 546–559 (2009).

9. Wellinger, R. J. & Zakian, V. A. Everything you ever wanted to know about Saccharomyces cerevisiae telomeres: Beginning to end. *Genetics* **191,** 1073–1105 (2012).

10. Zappulla, D. C., Goodrich, K. & Cech, T. R. A miniature yeast telomerase RNA functions in vivo and reconstitutes activity in vitro. *Nat. Struct. Mol. Biol.* **12,** 1072–1077 (2005).

11. Fulton, T. B. & Blackburn, E. H. Identification of Kluyveromyces lactis telomerase: discontinuous synthesis along the 30-nucleotide-long templating domain. *Mol. Cell. Biol.* **18,** 4961–70 (1998).

12. Kuprys, P. V. *et al.* Identification of Telomerase RNAs from Filamentous Fungi Reveals Conservation with Vertebrates and Yeasts. *PLoS One* **8,** e58661 (2013).

13. Qi, X. *et al.* The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res.* **41,** 450–462 (2013).

14. Podlevsky, J. D., Bley, C. J., Omana, R. V., Qi, X. & Chen, J. J.-L. The telomerase database. *Nucleic Acids Res.* **36,** D339-43 (2008).

15. Lebo, K. J. & Zappulla, D. C. Stiffened yeast telomerase RNA supports RNP function in vitro and in vivo. *Rna* **18,** 1666–1678 (2012).

16. Dalby, A. B., Goodrich, K. J., Pfingsten, J. S. & Cech, T. R. RNA recognition by the DNA end-binding Ku heterodimer. *RNA* **19,** 841–851 (2013).

17. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10024–9 (2004).

18. Zappulla, D. C. *et al.* Ku can contribute to telomere lengthening in yeast at multiple positions in the telomerase RNP. *Rna* **17,** 298–311 (2011).

19. Kabaha, M. M., Zhitomirsky, B., Schwartz, I. & Tzfati, Y. The 5' arm of Kluyveromyces lactis telomerase RNA is critical for telomerase function. *Mol. Cell. Biol.* **28,** 1875–82 (2008).

20. Chappell, A. S. & Lundblad, V. Structural Elements Required for Association of the Saccharomyces cerevisiae Telomerase RNA with the Est2 Reverse Transcriptase. *Mol. Cell. Biol.* **24,** 7720–7736 (2004).

21. Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **16,** 307–318 (2006).

22. Lubin, J. W., Tucey, T. M. & Lundblad, V. The interaction between the yeast telomerase RNA and the Est1 protein requires three structural elements. *RNA* **18,** 1597–1604 (2012).

23. Webb, C. J. & Zakian, V. A. Telomerase RNA stem terminus element affects template boundary element function, telomere sequence, and shelterin binding. *Proc. Natl. Acad. Sci.* **112,** 11312–11317 (2015).

24. Dehé, P. M. & Cooper, J. P. Fission yeast telomeres forecast the end of the crisis. *FEBS Lett.* **584,** 3725–3733 (2010).

25. Jansson, L. I. *et al.* Structural basis of template-boundary definition in Tetrahymena telomerase. *Nat. Struct. Mol. Biol.* **22,** 883–8 (2015).

26. Lai, C. K., Miller, M. C. & Collins, K. Template boundary definition in Tetrahymena telomerase. *Genes Dev.* **16,** 415–20 (2002).

27. Box, J. A., Bunch, J. T., Zappulla, D. C., Glynn, E. F. & Baumann, P. A flexible template boundary element in the RNA subunit of fission yeast telomerase. *J. Biol. Chem.* **283,** 24224–24233 (2008).

28. Armstrong, C. A. & Tomita, K. Fundamental mechanisms of telomerase action in yeasts and mammals: understanding telomeres and telomerase in cancer cells. *Open Biol.* **7,** 160338 (2017).

29. Webb, C. J. & Zakian, V. A. Telomerase RNA is more than a DNA template. *RNA Biol.* **6286,** 00–00 (2016).

30. Seto, A. G. *et al.* A template-proximal RNA paired element contributes to Saccharomyces cerevisiae telomerase activity. *RNA* **9,** 1323–32 (2003).

31. Cash, D. D. *et al.* Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo + Supplements. *Proc. Natl. Acad. Sci.* **110,** (2013).

32. Leeper, T. C. & Varani, G. The structure of an enzyme-activating fragment of human telomerase RNA. *RNA* **11,** 394–403 (2005).

33. Bley, C. J. *et al.* RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20333–20338 (2011).

34. Shefer, K. *et al.* A Triple Helix within a Pseudoknot Is a Conserved and Essential Element of Telomerase RNA. *Mol. Cell. Biol.* **27,** 2130–2143 (2007).

Chapter 3: Potential Role of CR4/5 in the Compensation of Torsional Strain Accrued During Telomere Repeat Replication and Translocation

The stem terminus element (STE), also termed the TWJ in yeast, SLIV in *Tetrahymena*, or the CR4/5 in vertebrates, plays a critical yet enigmatic role in telomerase function. In the first chapter of this work I describe the structural features of the best characterized STE, the medaka CR4/5. In this chapter I will present findings from many other studies, each of which provide clues to the function of the CR4/5, and expand my analysis to propose a theoretical role for the CR4/5 in telomerase activity.

THE CR4/5 AND TRANSLOCATION

There is evidence from several model systems indicating that the CR4/5 could be involved in the mechanism of repeat replication or translocation. Physical connectivity mapping by Mefford *et al.*[1] found that when the RNA backbone connection of hTR is broken at G308, the central guanine of the UUGGG pentaloop capping the stem of P6.1, all telomerase activity is lost. This one of only two positions in the CR4/5 that caused such a drastic loss of activity. RNA-RNA crosslinking work by Ueda *et al.*[2] found that the human CR4/5 loop contacts distal regions of the telomerase RNA, between the template and the core-enclosing helix. Specifically, nucleotides in the loop of P6.1 (L6.1) and 5' of the template, near or in the P1 helix, are involved in this interaction. Sequences from any other portion of the CR4/5; P6, L6, or P6.1, had no impact on these interactions. In humans the P1 helix,

also called the core-enclosing helix, functions dually to maintain the structure of the TR core and to define the end of the template sequence. This interaction between the CR4/5 and the TBE alludes to the possibility of a functional interplay between these two regions, which is supported by the loss of activity caused by breaking the RNA backbone at G308. Mutational work in medaka has shown that, unlike the other required stems of the CR4/5, alterations to L6.1 impact telomerase activity, but not TRBD binding[3,4]. Bley *et al.*[5] demonstrate that mutations in medaka L6.1 reduced overall telomerase activity, yet increased repeat addition processivity (RAP). As in human telomerase, medaka CR4/5 requires the P6, L6 and P6.1 regions of the CR4/5 for binding to TRBD, and mutations in these regions significantly impact only TRBD binding[4]. This would indicate that the functional roles of each component of the CR4/5 is conserved, at least in vertebrates.

The most striking results from the medaka system are the drastic differences in the high-resolution structures of CR4/5; the solution structure of the RNA alone, and the high-resolution structure of CR4/5 in complex with TRBD (discussed at length in chapter 1). Of particular note is the absence of the L6.1 nucleotides from the CR4/5-TRBD structure[3,4]. These were too unstructured to be solved, and it is likely that they are not making contacts with the TRBD. Bley *et al.*[5] speculate that L6.1 is interacting with the CTE, based on modeling with the solved *Tribolium casteneum* TERT structure. Comparison of *in vivo* DMS protection in the presence and absence of hTERT confirmed that the CR4/5 undergoes significant rearrangement upon binding to the protein component of telomerase[6]. Interestingly, the central guanine in

the UUGGG pentaloop is the only base of L6.1 that has a significant change in solvent exposure when bound to TERT. While the other nucleotides of the loop remain accessible, G308 becomes unavailable. If the hypothesis is that the L6.1 loop is interacting with the TBE, or the single-stranded region 3' of the TBE, it follows that occlusion of the L6.1 nucleotides should correspond to occlusion of TBE nucleotides. However, bases in the P1 helix and template-proximal region did not show significant changes in DMS protection, indicating that their solvent exposure did not change drastically upon binding to TERT. This does not preclude an interaction between the TBE and the CR4/5, it simply indicates that the regions 5' of the template are occluded when hTR is not bound to TERT.

It is obvious that the CR4/5 is important for catalytic activity, but exactly how this structure influences catalysis is still uncertain. Aside from interacting with the TBE, it is also unclear where the L6.1 loop is situated on TERT, especially in relation to the CTE. The inability to narrow the location of the loop likely arises from the dynamic nature of telomerase RNA, which adds a layer of complexity to the telomerase mechanism. Unlike other polymerases, telomerase must carry its template with it, which significantly complicates handling of the RNA/DNA duplex. Two processes comprise the telomerase mechanism; replication and translocation. When studying telomerase activity, it can be difficult to disentangle the effects of mutations on one mechanism from the other. However, recent work from a CR4/5 mutant in *Schizosaccharomyces pombe* has helped to shed light on how the CR4/5 factors into these processes.

THE CR4/5 AND TELOMERE REPEAT FIDELITY

The fission yeast *Schizosaccharomyces pombe* has a telomere repeat that is inconsistent in length and sequence. Hiraoka *et al.*[7] investigated the nature of *S. pombe* repeat synthesis, and determined that the sequence is not as variable as once thought. In fact it is a short repeat, occasionally followed by a one or two nucleotide spacer, and sometimes also a variable length guanine stretch. What is inconsistent about these telomere sequences is: if, and how many, spacer nucleotides are present after a repeat, whether a guanine stretch is present between repeats, and the length of the guanine stretch. The reason for this unusual pattern of repeats is the combination of a structural weakness in the template boundary, and subsequent aberrant translocation events. This gives rise to a unique mechanism of telomerase replication called stuttering[8–11]. Replication stuttering is triggered by erroneous copying of a template boundary element (TBE) nucleotide into the telomere repeat. This leads to misalignment when the enzyme translocates for synthesis of the next repeat. Because the DNA does not properly match with the alignment nucleotides of the template RNA, the alignment region of the template binds mid-repeat, and adds a single G nucleotide to the 3' end of the telomere DNA. This triggers a second, premature translocation event, which is again misaligned. This second off-register alignment results in the addition of a second G nucleotide, joining that which was misincorporated in the previous round. Thus, the stuttering register is generated. Further translocation events are more likely to misalign, and again favor a register that adds guanine to the telomere DNA. At any point during the stuttering process it is

possible to spontaneously align to the correct register, and thus it is not possible to predict the length of the guanine stretch between the repeats.

To determine what influences stuttering, Webb *et.al* screened for mutants of *S. pombe* that altered the sequence of the telomere repeats added *in vivo*. This lead the authors to discover a mutant with an altered loop sequence in the stem terminal element (STE)[12,13]. This mutant had telomeres with more consistent repeat sequences than wildtype yeast, and were interrupted by infrequent and short guanine stretches. The most obvious reason behind this change would be that the mutation in the CR4/5 somehow strengthened boundary enforcement, preventing aberrant replication of the TBE, and ensuring accurate repeat synthesis prior to translocation. On the contrary, further inspection of the template and template boundary sequences revealed that rather than strengthen the boundary, the CR4/5 loop mutant in fact weakened boundary enforcement. When the CR4/5 loop nucleotides are altered, the TBE helix is weakened and run-through replication more frequently results in the addition of two bases of the TBE rather than the single base added in the stuttering mechanism. This second base, a cytosine, causes the observed change in the pattern of replication stuttering. The presence of cytosine in the repeat allows for correct alignment of the RNA/DNA hybrid after the run-through replication event, thus resulting in synthesis of an accurate repeat. This work implies that the impact of the CR4/5 on catalysis occurs through an indirect means; supporting the template boundary structure. But, it is unclear under what circumstances the TBE would require stabilization. Understanding how stress is exerted on the template boundary helix necessitates an

understanding of how telomerase handles the DNA/RNA hybrid through its catalytic cycle. To clarify this process, and the role of the CR4/5, I will describe the current models of telomerase translocation.

CURRENT MODELS OF TRANSLOCATION

There are several models that have been proposed to describe different aspects of the telomerase mechanism. The accordion model[14] focuses on how the single-stranded RNA surrounding the template is handled during NAP. The single-stranded RNA 3' of the template is fully extended when the DNA primer is aligned and annealed for the beginning of nucleotide addition. The RNA 5' of the template, on the other hand, is compressed and crowded together at the beginning of NAP. As replication proceeds, the DNA is extended upon, and the RNA template is essentially fed through the active site of telomerase like a rope through a winch. This tugs on the crowded region of RNA, pulling base-by-base through the protein until the repeat has been fully synthesized and there is no longer any excess RNA. As a result of this, the excess single stranded RNA has accumulated 3' of the template, so that the strand on that side is crowded and bunched. Following this translocation occurs, and reestablishes the original arrangement of the RNA. The best analogy for this would be that of a typewriter; at the beginning of a line the platen is in its rightmost position. As the author adds words to the page, the platen moves left. When the line is completed, the platen is in its leftmost position, the author hits the carriage return, and the platen resets to the right.

Achieving this back-and-forth requires just the right amount of ssRNA on either side of the template, to act as slack. Too much or too little ssRNA results in anomalous telomere repeat sequence. Insertions and deletions between the 5' end of the template and the TBE result in aberrant template usage[15–17], indicating that the physical limit of slack extension is likely a major force in establishing the template boundary[18,19]. But, it is imperative that the tension on the RNA strand does not overwhelm the forces that hold the tertiary structure together. In vertebrate telomerase, the template boundary is also the structure that closes and secures the catalytic core of the RNA. If the core-enclosing helix is weakened or disrupted, telomere extension is compromised[1,18,20–22]. Because of the overlapping duties of the CEH, it is difficult to determine how weakening the template boundary structure impacts translocation. Another complication is the recently discovered role of the CEH in regulation of telomerase activity. In humans, it has been shown that a competing conformation exists for the 5' end of TR, in which the sequence just past the CEH forms G-quadruplexes[23]. These structures are stronger and more stable than the helical formation of the CEH, therefore formation of the G-quadruplex competes with formation of the P1 helix[24]. Given that the P1 helix must resist both possible G-quadruplex formation and strain exerted by the action of nucleotide addition, it is likely that further structures are in place to support template boundary, especially during the rearrangement process triggered by translocation. Whether this support comes from protein binding, RNA interactions, or both, is yet to be determined.

The pseudoknot tracking model proposed by Parks *et al.*[25] is an exciting result of FRET studies incorporated into the ROSETTA modelling process. Distance constraints were determined from FRET measurements between 10 locations in the telomerase RNA core in an active, stalled, and inactive telomerase enzyme. These constraints were then incorporated into a modelling algorithm aimed at generating a complete model of human telomerase. This algorithm also used the solved structures of the human pseudoknot, *Tribolium* TERT, *Tetrahymena* TRBD and medaka TRBD-CR4/5 as guides. The FRET constraints allowed the modelling algorithm to determine that the pseudoknot moves across the external surface of the RT domain as nucleotide addition progresses. The authors hypothesize that this could be coupled with movement of the CTE domain, similar to a mechanism proposed for Pol ν replication. Lee *et al.*[26] propose a mechanism for translesion synthesis by DNA Pol ν in which damaged bases are looped out of the active site by the fingers domain, while the thumb domain allows the polymerase to bypass damaged nucleotides by stabilizing the double helix, leaving the primer strand available for loop-out. Parks *et al.* postulate that sliding of the pseudoknot could be stabilizing a similar rotation in the CTE domain.

Building off previous work by Qi *et al.*[27], which determined that the RNA/DNA hybrid vacates the active site for strand separation and translocation, Wu and colleagues[28] present the first comprehensive model of the translocation mechanism. To enable the strand separation described by Qi *et al.*, they also propose a mechanism similar to Pol ν, in which the reverse transcriptase converts from a

‘closed’ to an ‘open’ configuration, providing the space for the hybrid strands to separate. Examination of DNA/RNA hybrid handling by various mutant TERT proteins lead Wu and colleagues to the conclusion that the thumb loop and helix regions of the CTE, and motif 3 of the RT, are critical to repeat addition processivity. The CTE motifs involved in processivity are termed the single-strand DNA retention site (SRS), and are responsible for binding to the DNA primer strand and anchoring it to the reverse-transcriptase. They further demonstrate that once the hybrid reaches seven base pairs, the paired bases of on the 5’ side of the DNA primer are pulled apart. They propose that the SRS facilitates this strand separation by holding the5’ region of the DNA, and thus countering the strength of the hybrid base-pairing. As the 3’ side of the RNA template is pulled away from the thumb domain the DNA is held firmly in place, thereby allowing strands to fray apart.

The first step of the proposed translocation mechanism, the rate limiting step, is the alignment of the RNA/DNA hybrid to the proper register within the active site. Following this the RT shifts from the open to closed configuration, and a short, 3 base-pair duplex occupies the active site. During early extension, the duplex is stabilized by the SRS and the TEN domain. As extension proceeds past the mid-template position, the duplex is stabilized by the SRS alone. Once four nucleotides have been added to the DNA (thus seven bases are paired in the active site), the 5’ most end of the DNA is pulled away from the RNA template by the outward rotation of the CTE. At this point it is unclear what is stabilizing the frayed RNA strand, however the authors suggest that the TEN domain could be involved[29,30]. CTE

rotation ultimately results in the opening of the polymerase active site, at which point the ssDNA is bound by the SRS and the ssRNA is anchored by motif 3 of the RT, and possibly the TEN domain. The template strand is then realigned in some manner, and is captured by the 3' end of the DNA. The cycle then restarts with the rate-limiting step, aligning the new hybrid into the active site.

TORQUE, STRAIN, AND HYBRID DISSOCIATION IN RNA POLYMERASE

These models have much in common with other well studied polymerases, in particular RNA polymerase. DNA scrunching, similar to the mechanism described by the accordion model, is involved in promoter escape, pausing, termination, and start-site-selection in bacterial and eukaryotic RNA polymerases. Tang *et al.*[31] use the single-subunit T7 RNAP to demonstrate that scrunching is involved in both abortive initiation and successful transition from an open complex to an elongating complex. This elegantly designed work uses smFRET to discriminate between competing models of transcription initiation, and reconcile the proposed rotation of the N-terminal domain indicated by high-resolution structures. When the T7 RNAP synthesis passes the 4 nt mark, the open DNA strand must move in order for replication to proceed. To accommodate this, the RNAP both compresses the DNA behind it, and rotates the promoter in front of it, to continue synthesis. Because RNAP uses a combination of scrunching and rotation, it is able to accomplish synthesis without the dramatic +220˚ rotation previously predicted. This mechanism is essentially what has been proposed by Parks *et al.*[14,25] and Wu *et al.*[28] to explain the motion of the template, surrounding ssRNA, pseudoknot, and CTE during NAP. Tang

and colleagues estimated that following 7nt synthesis, the promoter being acted upon is rotated by approximately 18˚, although it was not possible to specify what direction rotation occurred. Because of the modeled tracking of pseudoknot motion, it was possible for Parks *et al.* to predict a right-handed rotation of the template hybrid. Given the nature of rotation, if one were to assume the same degree as that seen in RNAP, the nucleotide addition by telomerase would induce approximately 18˚ of positive rotation in the template hybrid helix.

Another exciting aspect of the RNAP mechanism relates to the buildup of torque from transcription, and the interplay between strain-induced topology changes and replication stalling. Innovative work by Ma *et al.*[32] found that RNAP is capable of producing torque up to approximately 11pN*nm, which is, unsurprisingly, sufficient force for opening a DNA helix of arbitrary sequence. This work also showed that the buildup of torque impacts the speed of transcription. If negative supercoils are induced ahead of RNAP the rate of transcription increases, whereas if the positive supercoils induced by elongation are not resolved, transcription will slow until it enters a paused state. Work by Lerner *et al.*[33] found that extension by RNAP beyond 7 nucleotides is 3.5x slower than synthesis that occurred before the 7-nucleotide mark. Ma *et al*. found that following the addition of 4-7 nucleotides, torsional strain caused by the buildup of supercoils ahead of RNA polymerase induced pausing and eventual dissociation of the polymerase. Several groups[34,35] have shown that replication-induced torque can cause protein disassociation from DNA. Brown *et al*.[36] found that there is a pause signal in the template sequence of human

telomerase RNA at the +4 nucleotide position. Following the addition of 3 nucleotides past the pause site, when a total of 7-nt have been copied, dissociation of the hybrid from the active site of telomerase is triggered. Vacation of the duplex from the active site was found to lead to either translocation or detachment from the DNA substrate. The authors propose that this pause, in addition to the physical boundary created by the P1 helix, is the method of template definition in human telomerase.

Lipfert *et al.*[37] found that when dsRNA is overwound it compacts rather than extends, as is found for dsDNA. The property called twist-stretch coupling describes how a double-stranded helix compensates for stretching forces exerted upon it. Double-stranded DNA, as has been described previously,[38,39] responds to stretch by overwinding, which in turn results in a negative twist-stretch; the length of the strand increases. Because the helical angle of B-form DNA is shallow, (the bases are nearly perpendicular to the helical axis) the stretched helix can become more narrow without deformation. This results in an increased number of helix rotations over the same distance, which is an overwound helix (a more thorough explanation and modeling of this behavior is provided by Gore *et al.*). Lipfert *et al.* discovered that, as predicted for other A-form helicies, dsRNA does not permit narrowing of the helix radius when overwound, because the plane of the basepairs is not perpendicular to the helical axis. In this case, when stretching causes positive rotation of the helix, this resulted in an increase in the helix radius. This occurred within the 4-8pN torque regime, well within the torque reported for RNAP during nucleotide addition. This has interesting implications for the torque exerted on the P1 helix during NAP, that may result from

stretching of the ssRNA flanking the template. The phenomena of decreasing helix length could help to counter the stretching exerted on the ssRNA 5' of the template hybrid, perhaps enhancing the strength of physical template-boundary enforcement in telomerase.

It is clear that stress is exerted upon the telomerase RNA as a function of both nucleotide addition and translocation. Similar forces are at work in other well-studied polymerase mechanisms, such as RNA Polymerase. The work in RNAP should be used to inform the investigation of how torque and strain impact duplex dissociation and translocation. I propose a scenario in which progressive nucleotide addition induces stretching in the ssRNA upstream of the template sequence. This stretching then exerts strain on the P1 helix, which responds by overwinding and compacting. This compaction increases the pulling force on the RNA at the active site, and contributes to the pausing observed by Brown *et al.* The resistance to stretching of the P1 helix, and the additional strain that would cause 5' of the duplex, could further enhance the dissociation of protein-nucleic acid contacts already stimulated by replication-induced torque. As NAP proceeds, the movement of the CTE closer to the active-site of TERT induces a coincidental movement of the CR4/5 loop into position between the template and the TBE to mitigate strain on the helix of the template boundary element.

The homology model presented by Parks *et al.* did not appear to position P6.1 near template or P1 helix, despite the data showing the pentaloop of the CR4/5 in contact with the ssRNA of this region. Similar homology modeling by Chan *et al.*[40]

also did not position the P6.1 pentaloop near the template or flanking regions. It is possible that this is due to the dynamic nature of the RNP throughout catalysis. Since modeling is based primarily on crystal structures, it is conceivable that in the stationary, inactive state of the RNP, P6.1 is not engaging in stabilization. As elaborated earlier, the majority of the strain experienced by the 5'ssRNA flanking the template is likely to occur near the end of NAP, when the RT has paused just prior to translocation. I propose that stabilization and contact between L6.1 and the template-surrounding RNA occurs transiently, either while extension is occurring or when the duplex has vacated the active site and the hybrid dissociates.

Experimental Avenues for Probing Torsion and Rotation in Telomerase

To further our understanding of the motion of telomerase RNA and protein domains throughout the catalytic cycle, I propose a series of smFRET experiments aimed at localizing the CTE, L6.1, and template during active extension. Labeling the CTE and L6.1 with donor-acceptor Cy-dyes, respectively, and performing alternating laser excitation smFRET (ALEX) would enable monitoring of the relative motion of these two components during a catalytic cycle. A three-color FRET experiment with a third dye positioned on the ssRNA 5' of the template would determine if motion of the CTE towards L6.1, or concerted motion of both domains towards the template occurs during catalysis. It would be especially informative to begin the telomerase with a RNA/DNA hybrid primed for translocation, to discern whether translocation also triggers a conformational reset of the mobile protein and RNA domains. This would enable confirmation of the hypothesis that CTE rotation

follows the cycle observed by Parks *et al.*, and compensates for the rotation that accrues through NAP.

The U42-U312 dye pair tested by Parks *et al.* showed a modest shift in the FRET distribution between stalled and active RNPs. This points towards a relatively consistent distance between the 5'template sequence and the P6.1 stem. The label position on the CR4/5 in this experiment is on the stem rather than the loop region, and it has been shown by structure and chemical crosslinking that this area binds to TRBD. It is also clear from mutational data that this portion of P6.1 does not impact processivity. Therefore it is possible that this area is not involved in any stabilization that may be occurring. However, since labeling of the loop may hinder stabilizing interactions, direct assessment of CR4/5 stabilization may not be possible. If labeling of L6.1 proved unfeasible, then a three-color FRET experiment using the U312 position could still determine whether P6.1 is stationary relative to the CTE, and if the CTE ever approaches the single-stranded region of the telomerase core.

Lerner *et al.*[33] describe a single-round quenched kinetic assay in which dye-quenching of a double-labeled ssDNA is monitored as a function of transcript production over time. When there are no transcripts, ALEX will show rapid changes in FRET state between the two labeled ends of the probe. When RNAP has generated a ssRNA product that is complementary to the probe, a double-stranded hybrid forms and the molecule is locked in a particular FRET state. Monitoring of dye bleaching over time allows for the direct quantification of the number of products generated by RNAP as a function of time. This assay is easily adaptable to use with telomerase.

Short labeled DNA anchors could be used as the readout for telomerase activity. Freely diffusing telomerase would be allowed to produce unlabeled DNA product, which would find and anneal to doubly-labeled DNA anchored to the slide surface. Since sequence of the DNA primer necessary to align with the telomerase template is only 3 nucleotides, it should be possible to distinguish between an un-extended primer binding to a labeled anchor from an extended primer binding to a labeled anchor. Using this assay, direct assessment of the impact of mutations in the flanking template arms or the P1 helix on extension kinetics could be made. This would provide valuable information as to how altering the single-stranded RNA changes the kinetics of NAP. It may also be possible to adapt this assay to study the length of pausing that occurs before the hybrid is released from the active site of telomerase, and how pausing kinetics is impacted by stiffer or more flexible sequences in the single-stranded core or P1 helix.

To study how torsional strain is impacted by CR4/5 stabilization, use of the magnetic tweezers could be employed. First, it would be useful to determine how much force is exerted upon the flanking ssRNA when telomerase is extending a DNA primer. To study this, use of a telomerase template-free system could be employed. The substrate for the RNP would be an anchor composed of a double-stranded handle, followed by the complete single-stranded region of hTR core, and a second ds handle anchoring the molecule to a slide. A short DNA oligo serving as a primer would be freely floating in solution. Then template-free telomerase RNPs and all components needed for a round of synthesis would be introduced to the system, and the force

exerted on the substrate by telomerase NAP could be measured. This would provide the first assessment of the force regime exerted on hTR, and would be crucial to further studies aimed at determining how much strain is placed on the RNA through the catalytic and translocation cycles.

CONCLUDING REMARKS

The hypotheses presented in this chapter provide a glimpse into my view of polymerases and reverse transcriptases. Throughout my study of HIV-RT, DNA Pol I, and Telomerase, I have found exciting and puzzling similarities between these diverse polymerases. I am intrigued and enthused by the interplay between well-known physical properties of DNA and the mechanisms of replication, translocation, and fidelity. I am fascinated by the leap from circular to linear genomes, and how eukaryotes have stolen and adapted a low-fidelity viral polymerase to perform high fidelity, high processivity replication that is critical for survival. The simultaneous sequence diversity and structural homogeneity between telomerases of different phylogenies is a testament to the awesome forces of evolution that drive life. I am grateful for the opportunity to make even the smallest contribution to our understanding of these incredible processes.

1. Mefford, M. A. & Zappulla, D. C. Physical connectivity mapping by circular permutation of human telomerase RNA reveals new regions critical for activity and processivity. *Mol. Cell. Biol.* **36,** MCB.00794-15 (2015).

2. Ueda, C. T. & Roberts, R. W. Analysis of a long-range interaction between conserved domains of human telomerase RNA. *RNA* **10,** 139–47 (2004).

3. Kim, N. K., Zhang, Q. & Feigon, J. Structure and sequence elements of the CR4/5 domain of medaka telomerase RNA important for telomerase function. *Nucleic Acids Res.* **42,** 3395–3408 (2014).

4. Huang, J. *et al.* Structural basis for protein-RNA recognition in telomerase. *Nat. Struct. Mol. Biol.* **21,** 507–12 (2014).

5. Bley, C. J. *et al.* RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20333–20338 (2011).

6. Zemora, G., Handl, S. & Waldsich, C. Human telomerase reverse transcriptase binds to a pre-organized hTR in vivo exposing its template. *Nucleic Acids Res.* **44,** 413–425 (2016).

7. Hiraoka, Y., Henderson, E. & Blackburn, E. H. Not so peculiar : fission yeast telomere repeats. *Trends Biochem. Sci.* **4,** 126 (1998).

8. Webb, C. J. & Zakian, V. A. Identification and characterization of the Schizosaccharomyces pombe TER1 telomerase RNA. *Nat. Struct. Mol. Biol.* **15,** 34–42 (2008).

9. Leonardi, J., Box, J. A., Bunch, J. T. & Baumann, P. TER1 , the RNA subunit of fission yeast telomerase. *Nat. Struct. Mol. Biol.* **15,** 26–33 (2008).

10. Ares, M. & Chakrabarti, K. Stuttering against marginotomy. *Nat. Struct. Mol. Biol.* **15,** 18–19 (2008).

11. Webb, C. J. & Zakian, V. A. Telomerase RNA is more than a DNA template. *RNA Biol.* **6286,** 00–00 (2016).

12. Webb, C. J., Zakian, V. A., Webb, C. J. & Zakian, V. A. mutant Telomere les (i / s) ons from a telomerase RNA mutant. **4101,** 16–18 (2016).

13. Webb, C. J. & Zakian, V. A. Telomerase RNA stem terminus element affects template boundary element function, telomere sequence, and shelterin binding. *Proc. Natl. Acad. Sci.* **112,** 11312–11317 (2015).

14. Parks, J. W. & Stone, M. D. Coordinated DNA dynamics during the human telomerase catalytic cycle. *Nat. Commun.* **5,** 4146 (2014).

15. Seto, A. G. *et al.* A template-proximal RNA paired element contributes to Saccharomyces cerevisiae telomerase activity. *RNA* **9,** 1323–32 (2003).

16. Chen, J.-L., Blasco, M. A. & Greider, C. W. Secondary Structure of Vertebrate Telomerase RNA. *Cell* **100,** 503–514 (2000).

17. Moriarty, T. J., Marie-egyptienne, D. T. & Autexier, C. Regulation of 5' template usage and incorporation of noncognate nucleotides by human telomerase. *RNA* **11,** 1448–1460 (2005).

18. Mefford, M. A., Rafiq, Q. & Zappulla, D. C. RNA connectivity requirements between conserved elements in the core of the yeast telomerase RNP. *EMBO J.* **32,** 2980–2993 (2013).

19. Richards, R. J., Theimer, C. A., Finger, L. D. & Feigon, J. Structure of the Tetrahymena thermophila telomerase RNA helix II template boundary element. *Nucleic Acids Res.* **34,** 816–825 (2006).

20. Niederer, R. O. & Zappulla, D. C. Refined secondary-structure models of the core of yeast and human telomerase RNAs directed by SHAPE. *RNA* **21,** 254–261 (2015).

21. Garforth, S. J., Wu, Y. Y. & Prasad, V. R. Structural features of mouse telomerase RNA are responsible for the lower activity of mouse telomerase versus human telomerase. *Biochemistry* **397,** 399–406 (2006).

22. Chen, J. & Greider, C. W. Template boundary definition in mammalian telomerase. *Genes Dev.* **17,** 2747–2752 (2003).

23. Gros, J., Guødin, A., Mergny, J. & Lacroix, L. G-Quadruplex Formation Interferes with P1 Helix Formation in the RNA Component of Telomerase hTERC. *ChemBioChem* **9,** 2075–2079 (2008).

24. Booy, E. P. *et al.* The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.* **40,** 4110–4124 (2012).

25. Parks, J. W., Kappel, K., Das, R. & Stone, M. D. Single-molecule FRET-Rosetta reveals RNA structural rearrangements during human telomerase catalysis. *RNA* **23,** 175–188 (2017).

26. Lee, Y.-S., Gao, Y. & Yang, W. How a homolog of high-fidelity replicases conducts mutagenic DNA synthesis. *Nat Struct Mol Biol* **22,** 298–303 (2015).

27. Qi, X. *et al.* RNA/DNA hybrid binding affinity determines telomerase template-translocation efficiency. *EMBO J.* **31,** 150–161 (2012).

28. Wu, R. A., Tam, J. & Collins, K. DNA-binding determinants and cellular thresholds for human telomerase repeat addition processivity. 1–20 (2017). doi:10.15252/embj.201796887

29. Wu, R. A. & Collins, K. Human telomerase specialization for repeat synthesis

by unique handling of primer-template duplex. *EMBO J.* **33,** 921–935 (2014).

30. Steczkiewicz, K. *et al.* Human telomerase model shows the role of the TEN domain in advancing the double helix for the next polymerization step. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 9443–9448 (2011).

31. Tang, G., Roy, R., Ha, T. & Patel, S. S. Transcription Initiation in a Single-Subunit RNA Polymerase Proceeds through DNA Scrunching and Rotation of the N-Terminal Subdomains. *Mol. Cell* **30,** 567–577 (2008).

32. Ma, J. & Wang, M. Interplay between DNA supercoiling and transcription elongation. *Transcription* **5,** 28636-1–4 (2014).

33. Lerner, E. *et al.* Backtracked and paused transcription initiation intermediate of Escherichia coli RNA polymerase. *PNAS* (2016). doi:10.1073/pnas.1605038113

34. Teves, S. S. & Henikoff, S. Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* **21,** 88–94 (2014).

35. Sheinin, M. Y., Li, M., Soltani, M., Luger, K. & Wang, M. Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. *Nat. Commun.* **4,** 1–22 (2013).

36. Brown, A. F. *et al.* A self-regulating template in human telomerase. *PNAS* **111,** 11311–11316 (2014).

37. Lipfert, J. *et al.* Double-stranded RNA under force and torque: Similarities to and striking differences from double-stranded DNA. *PNAS* **111,** 15408–15413 (2014).

38. Gore, J. *et al.* DNA overwinds when stretched. *Nature* **442,** 836–839 (2006).

39. Sheinin, M. Y. & Wang, M. D. Twist–stretch coupling and phase transition during DNA supercoiling. *Phys. Chem. Chem. Phys.* **11,** 48000–4803 (2009).

40. Chan, H., Wang, Y. & Feigon, J. Progress in Human and Tetrahymena Telomerase Structure. *Annu. Rev. Biophys.* **46,** 199–225 (2017).

Chapter IV. The Infidelity of Replication: Studying the Determinants of HIV-1 Reverse Transcriptase Fidelity

INTRODUCTION

HIV is a highly adaptable virus, which has infected more than 25 million people since the disease emerged in 1981. Although rigorous scientific efforts have dramatically improved our ability to combat this pandemic, the virus evolves at such a swift rate that drug resistance has become the foremost obstacle to treatment in developed nations (Chunduri, 2011). Reverse Transcriptase (RT), the viral polymerase that copies the HIV genome upon infection of a host cell, drives this rapid evolution. *In vitro* experiments show that Reverse Transcriptase makes mistakes at a rate of between $5x10^{-4}$ and $5x10^{-5}$ mutations/base pair/replication cycle, as compared to human DNA polymerases, which have error rates ranging from $1x10^{-9}$ to $1x10^{-12}$ mutations/base pair/replication cycle.

The key to HIV's success as a pathogen lies in its enormous evolvability. This huge genetic variation allows the virus to rapidly evolve, facilitating immune system escape and resistance to drug treatments. To understand how HIV evolves, we must first understand how mutations are made. The three possible sources of mutation in HIV are human host proteins, viral proteins, and the interplay between the two [1]. The impact of host proteins on the genetic diversity of HIV is severely curtailed by the interaction between viral and host protein [2,3]. Therefore the most important source of genetic variation in HIV is the mutations made by Reverse Transcriptase (RT) during viral genome replication [4].

RT is a low fidelity polymerase, ranging from $5x10^{-4}$ to $1.5x10^{-5}$ mutations/base pair/replication cycle. The pattern of errors made by RT while copying a template is not entirely stochastic. While there are mutations made all over the template, there are also certain regions with unusually high mutation frequency, which are termed hotspots [5].

SEQUENCE CONTEXT AND FIDELITY

These mutation hotspots occur in particular positions in both viral and non-viral templates, yet there is no sequence similarity that would indicate a motif that determines the position of these hotspots. And yet, the reproducibility of the hotspot locations *in vitro* indicates that they are not randomly occurring. Hotspot positions are much harder to determine from patient samples, as these viral sequences derive from a wide range of original templates, are copied by a variety of RTs, and are placed under different selective pressures [6,7]. In order to understand the evolution of HIV, it is essential to know how mutations, and mutation hotspots, are generated. One key aspect of this is determining the distinct hotspot patterns and mutation spectra created by variant RTs.

Research has shown that although sequence context has an impact on fidelity, the sequence of the template *directly in the active site* does not have a significant influence [8–11]. However, the sequence of the primer/template pair 6-10 nucleotides upstream of the active site can alter the fidelity of polymerization by as much as 14-fold. This indicates that the impact of sequence context on fidelity is more nuanced than simple active site kinetics, and can be affected by a range of possible

polymerase-template interactions [12–14]. I hypothesize that there is an identifiable sequence motif in the sequence upstream of mutation hotspots, which drives the appearance of the hotspot. I have termed this motif a "generator sequence", and hypothesize that these sequences are located in close proximity to, but not within, mutation hotspots.

Polymerases tend to favor certain types of mutations, (transversion from A to C, or small deletions, for example) which is guided by the structure of the polymerase's active site and how the template is handled. Mutations that change the fidelity of RT are likely to change this structure, resulting in a shift in the pattern of induced mutations. Previous work has shown that resistance to Nuceloside RTI and Non-Nucleoside RTI (NRTI and NNRTI respectively) is determined by the shape of the active site. Mutant RTs with resistance to NRTIs have a more spatially constrained active site, which enables them to differentiate between the drug and a real nucleotide. This ability to more accurately distinguish between the correct unit (nucleotide) and incorrect unit (drug) is the molecular basis for fidelity. A RT that is better at choosing what nucleotides to use has a higher fidelity. Knowing how drug resistance mutations change the fidelity of viral replication gives us structural insight, and enhances our ability to design targeted RTIs. Understanding where new mutations in the HIV genome are likely to arise in a given HIV variant, based on knowledge about the pattern of mutation favored by the RT, will enable doctors to predict where mutation hotspots may arise in other viral proteins, and would help to predict which regions could evolve drug resistance. Thus, understanding the mechanisms that drive mutation and evolution

through the action of RT is essential for the development of future treatments. I investigate the replicative phenotype of HIV-1 RT by using HIV genome databases to find mutations in the RT gene that correlate with changes in drug susceptibility. I will then test and compare the fidelity of these mutants, and the pattern of mutations they induce, against that of wildtype HIV-1 RT using a unique bacterial mutagenesis assay. This assay, developed by Camps et. al, is a prokaryotic *in vivo* method that can facilitate rapid, high-throughput measurement of polymerase fidelity, in a cellular context.

DRUG RESISTANCE AND THE POSITIVE SELECTION OF GENERATOR MOTIFS

Over time, as mutations are continually introduced by reverse transcriptase replications, the combination of selective pressures exerted by the host environment, along with genetic drift, lead to variants that can have differences in some key viral characteristics. When a patient is infected with HIV, that individuals' cells are invaded not just by one, but many genetic variants of HIV. Thus, every infection is a mosaic of many versions of the same virus. When RT copies these variants, even more mutations are introduced and more variants are generated. When selective pressure is absent, a majority of the variants carried by a patient will accumulate only neutral mutations, which have no effect on the virus. However, when a selective pressure, such as treatment with reverse transcriptase inhibitors (RTI), is present the balance of the viral population shifts. This will favor viruses that have stochastically gained advantageous mutations such as those that provide drug resistance (Palmer 2005). Because the RT is copying its own gene when it replicates the HIV genome, it is also subject to selective pressures and evolution. This evolution is what generates drug resistance in patients.

Variant RTs often display different mutation spectrum, hotspot distribution, and fidelity relative to wild type RTs

Under antiretroviral drug selection, any generator sequence that emerges near a sequence space significant for drug resistance will provide the virus with a selective advantage. This advantage will act as a magnifying force, pushing the population to be dominated by viruses carrying both the generator and the drug resistance sequences. If this is so, we should be able to find mutations that are in the vicinity of drug resistance mutations but that do not have any phenotypic impact on resistance. Instead, these generator sequences will provide a selective advantage at the genome level, by inducing high mutation rates and pushing protein evolution more rapidly.

CURRENT METHODS OF MEASURING FIDELITY

Current methods for studying RT have produced results that are either inconsistent or incomparable. *In vitro* methods can be used to study a larger number of reverse transcriptases, but the results are orders of magnitude different in each study, and frequently conflicting. *In vivo* methods are more reliable between studies, but still conflicting and are severely limited in scope due to the complex nature of the methods. Thus, in order to gain useful data on how different RTs function, a method that has the consistency of mammalian *in vivo* assays, with the speed and scale of *in vitro* methods must be used [1].

To discern the impact of sequence context on the activity of RT, it is necessary to take a large scale, systematic approach involving alterations to both

template and polymerase. This would necessitate a high-throughput, cell-based method of RT replication that would eliminate, as much as possible, the effect of selective pressures and template secondary structure [1]. I have developed a bacterial assay and constructed a reporter plasmid to accomplish this, which I use to investigate how sequence context can induce local mutation hotspots. The prokaryotic system for determining the mutation spectrum and footprint of RT could be more reliable than currently used *in vitro* fidelity assays, and may provide data consistent with the more complex and expensive *in vivo* mammalian fidelity assays. Because this method would be higher throughput than current *in vivo* assays, a large number of variant RTs can be studied with results that are internally comparable.

METHODS AND RESULTS

The assay employs bacterial cells that lack Pol I, the native *E. coli* polymerase that replicates the first 1-2 kb of plasmid DNA. In this system, Pol I was replaced by HIV-RT, making RT solely responsible for plasmid replication, ensuring that all mutagenesis measured can be attributed to RT.

To determine if it was possible to use HIV-1 RT as a complement for Pol I-deficient *E. coli*, direct and indirect complementation was tested. JS200 *E. coli* cells were used for this work; these cells have a temperature sensitive Pol I allele that is functional at the permissive temperature of 30˚C, but non-functional at the restrictive temperature of 37˚C. Direct complementation refers to the capability of the

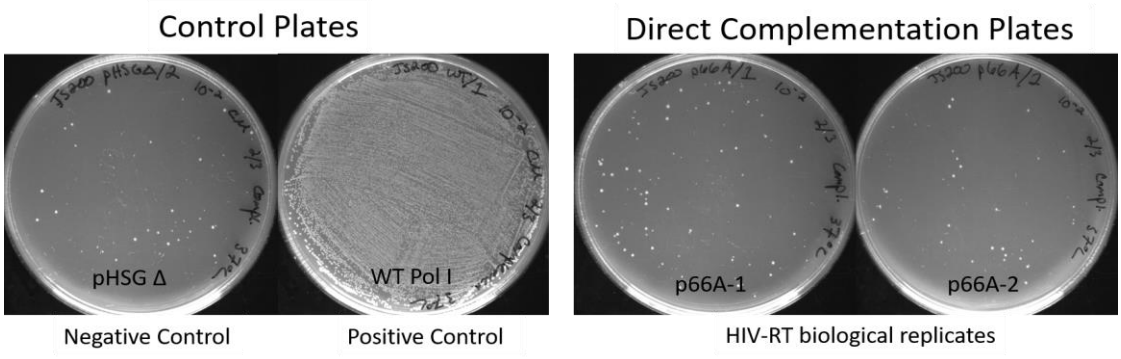


Figure 4.1. Results of Direct complementation of HIV-RT in JS200 cells. Above are the results of direct complementation at 37˚C. The left hand side shows the controls of the experiment, while the right hand side shows two independent lineages of HIV-RT expressing JS200 cells. Both biological replicates show limited complementation, not beyond what is found in the negative control.

replacement polymerase, in this case HIV RT, to facilitate growth of Pol I-deficient cells at the restrictive temperature. To test this, cells expressing HIV-1 RT are plated at the permissive and restrictive temperatures, and growth is assessed through colony counts.

To test direct complementation, JS200 cells that have successfully established the pHSG plasmid carrying the *pol* gene were plated and grown at 37˚C. If HIV RT can directly complement the loss of Pol I, and perform Okazaki fragment processing, then they would survive at the restrictive temperature. As shown in figure 4.1, HIV-RT was not able to complement growth at the restrictive temperature above what is seen in the negative control.

Indirect complementation is assessed through the establishment and maintenance of plasmids. Establishment of a plasmid refers to the cell's ability to use the transformed plasmid immediately, without significant or abnormal delay in gene

expression. The widely used standard for recovery time following transformation with plasmids is an hour in normal *E coli*. If cells utilizing HIV RT as a replacement for Pol I are not able to begin expressing selection genes in this time frame, they

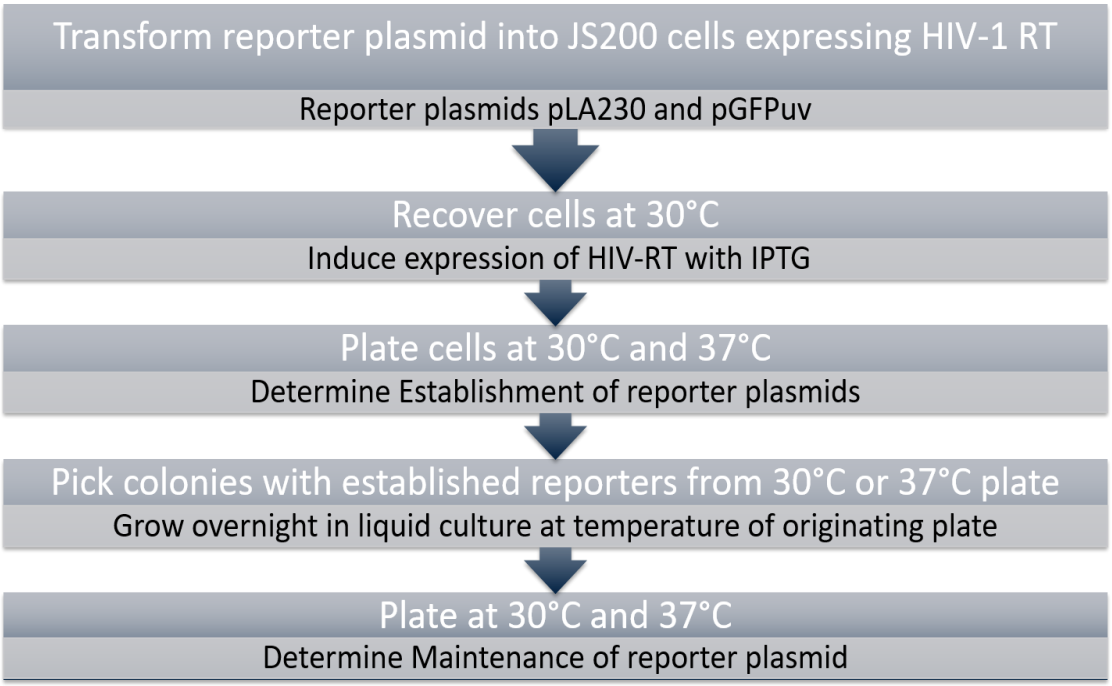


Figure 4.2. Experimental flowchart for Indirect Complementation. The chart above describes the steps that are used in testing the ability of HIV-RT to complement plasmid replication in Pol I-deficient *E. coli*.

would not be able to grow on plates containing the selection antibiotic. These are considered negative for plasmid establishment. Plasmid maintenance is the ability of cells, following plasmid establishment in permissive conditions, to propagate a plasmid to daughter cells during growth at the restrictive temperature. If daughter cells can not continue to express selection genes, and thus survive in selective conditions when shifted from continued growth at permissive to restrictive temperature, then they are negative for plasmid maintenance.

If HIV does not directly complement Pol I deficiency, it is possible that indirect complementation may be used as a means to test the HIV RT replication phenotype. To test indirect complementation, JS200 cells were transformed via electroporation with either pLA230 or pGFPuv. The pLA230 plasmid uses kanamycin resistance as a reporter. The pGFPuv plasmid uses both carbenicillin resistance and GFP expression as a reporter. The advantage of pGFPuv is that it can also be used in later experiments to measure the rate of mutagenesis, as this could easily be detected through changes in brightness of GFP fluorescence. pLA230 is used because it easier for the cells to express only one gene from the reporter plasmid, and therefore it is possible that the cells may have differential success in complementation of these two plasmids.See figure 4.2 for a summary of the method used for testing establishment and maintenance.

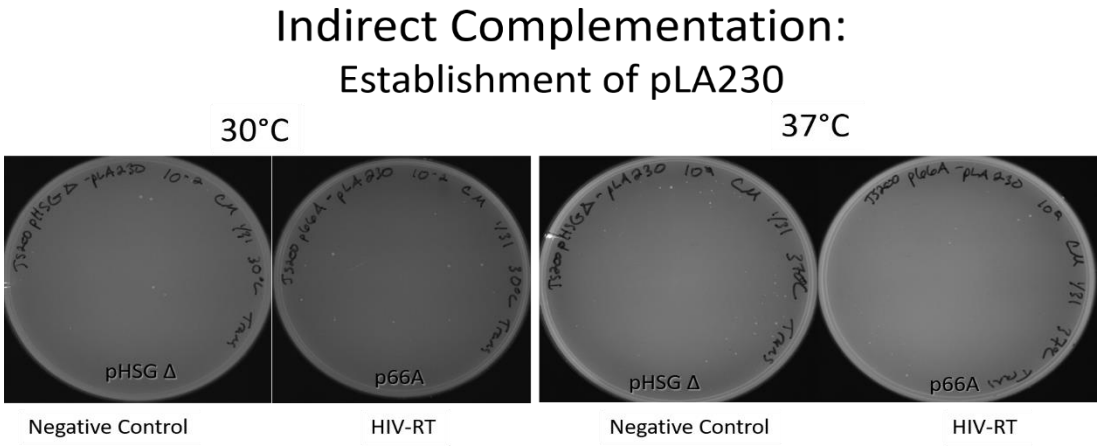


Figure 4.3. Establishment of the pLA230 plasmid. This figure displays the capability of cells to rapidly express genes from the pLA230 plasmid soon after transformation. The left side is grown at the permissive, 30˚C temperature, while the right side is grown at the restrictive 37˚C temperature.

As shown in figure 4.3, establishment of pLA230 was not successful at 30˚C or 37˚C. This means that the cells cannot immediately begin expressing selection genes following transformation. This result eliminates the possibility of using rapid plasmid establishment as a method to assess the growth capacity of these cells. However, it is still possible that they will be able to maintain a plasmid in selective conditions. To test this the cells must be transformed with the reporter plasmid and grown in non-selective conditions for an extended period of time before plated on selective media in restrictive temperature (see figure 4.2).

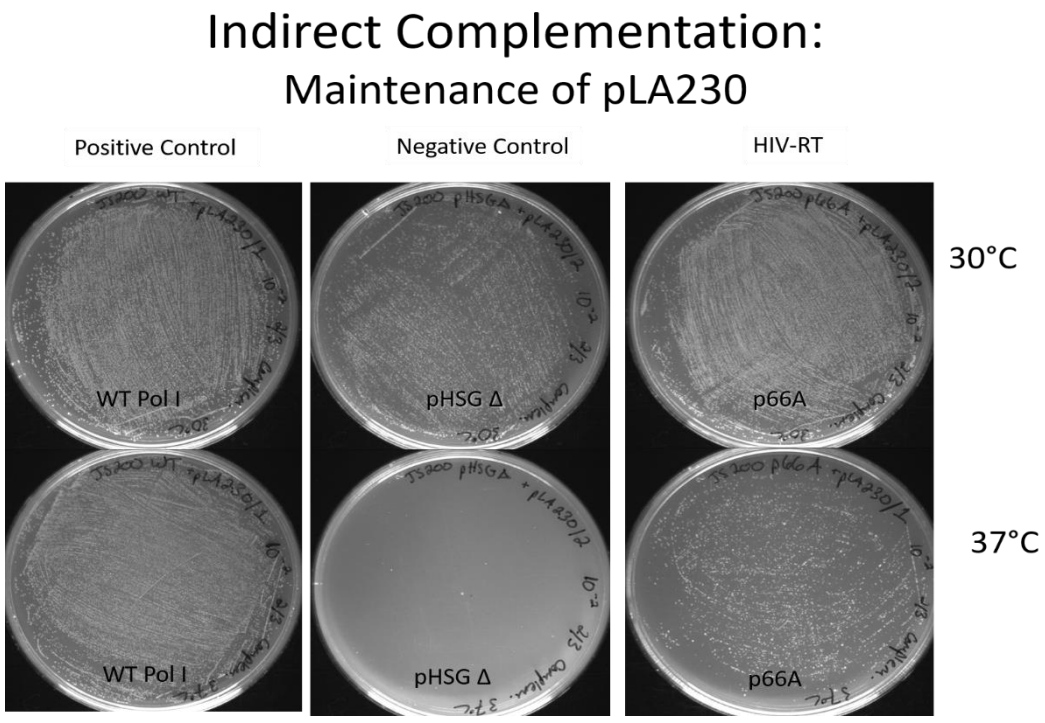


Figure 4.4. Maintenance of the pLA230 plasmid. This figure displays the capability of cells to replicate a pre-established plasmid at restrictive temperature. The top row is grown at the permissive, 30˚C temperature, while the bottom row is grown at restrictive 37˚C temperature.

Maintenance was tested with pLA230, the results of which are shown in figure 4.4. The cells bearing HIV-RT are capable of growth under selective conditions at 37˚C, at a much higher level than the negative control. This indicates that the two-plasmid complementation system can be used to investigate HIV RT replication.

The same indirect complementation experiments were performed using the pGFPuv reporter plasmid. Example results are shown in Figure 4.5 and 4.6. As seen in figure 4.5, the pHSGΔ plasmid was inconsistent as a negative control. This indicated that either the cells were not being handled consistently or the temperature sensitivity of the Pol I mutant was not stringent, and functional Pol I was expressed. It

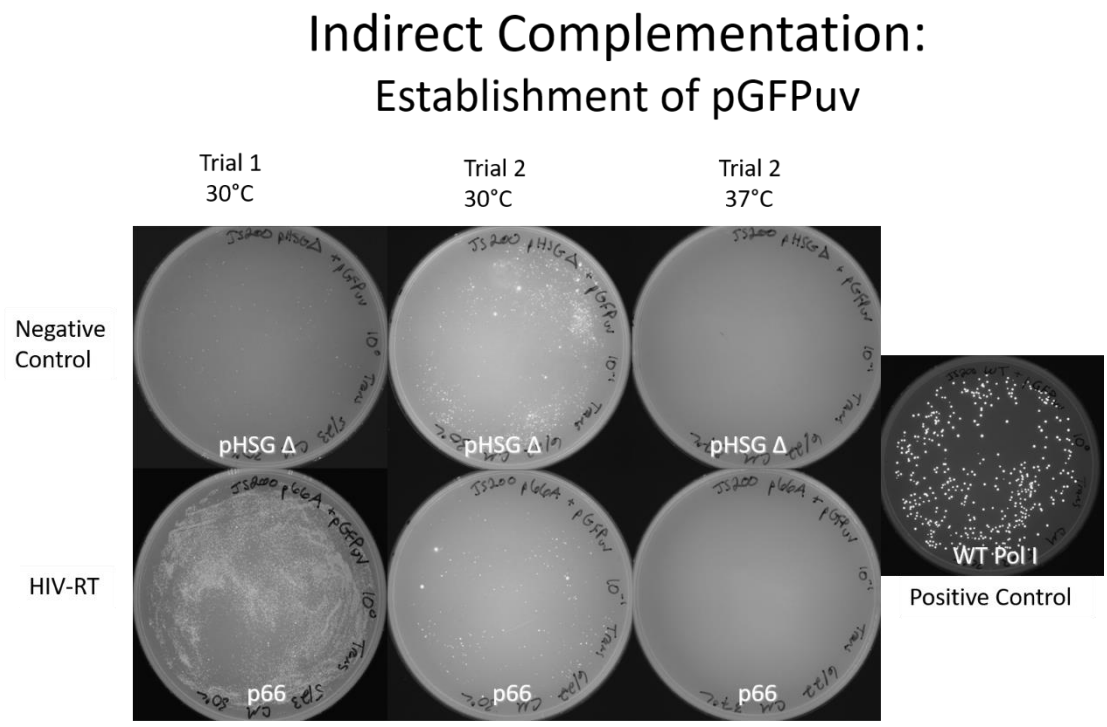


Figure 4.5. Establishment of the pGFPuv reporter plasmid. Negative controls are shown in the top row, HIV-RT is shown in the bottom row. Trials indicate technical replicates.

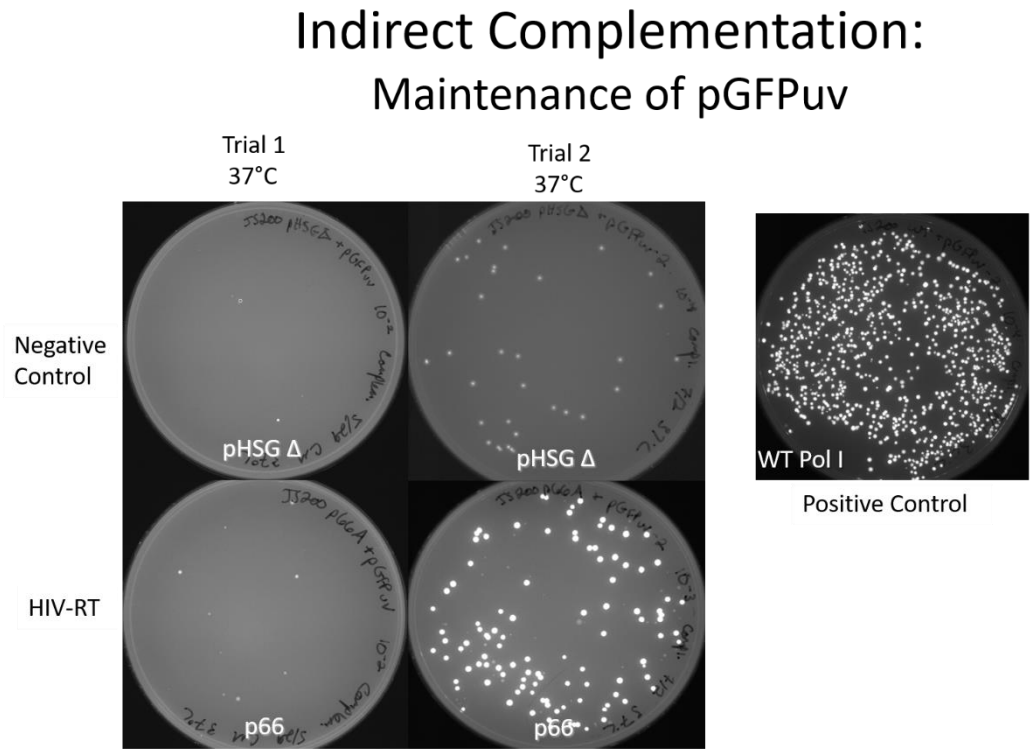


Figure 4.6. Maintenance of the pGFPuv reporter plasmid. Negative controls are shown in the top row, HIV-RT is shown in the bottom row. Trials indicate technical replicates.

is also possible that the Pol I mutation was reverted in some lineages, but this is unlikely because the temperature sensitivity involves multiple loci.

Maintenance of pGFPuv was also inconsistent, as demonstrated in figure 4.6. Because complementation was inconsistent, an alternative approach to obtaining high-throughput sequencing was performed. If the mutation pattern of HIV-RT is distinguishable from that of *E. coli* Pol I, it may be possible to distinguish plasmids mutated by one from those mutated by the other. Thus, we can allow both RT and Pol I to copy the reporter plasmid, then sequence all the replicated plasmids and disregard

those that show signs of being replicated by Pol I. A flowchart of the method is shown in figure 4.7.

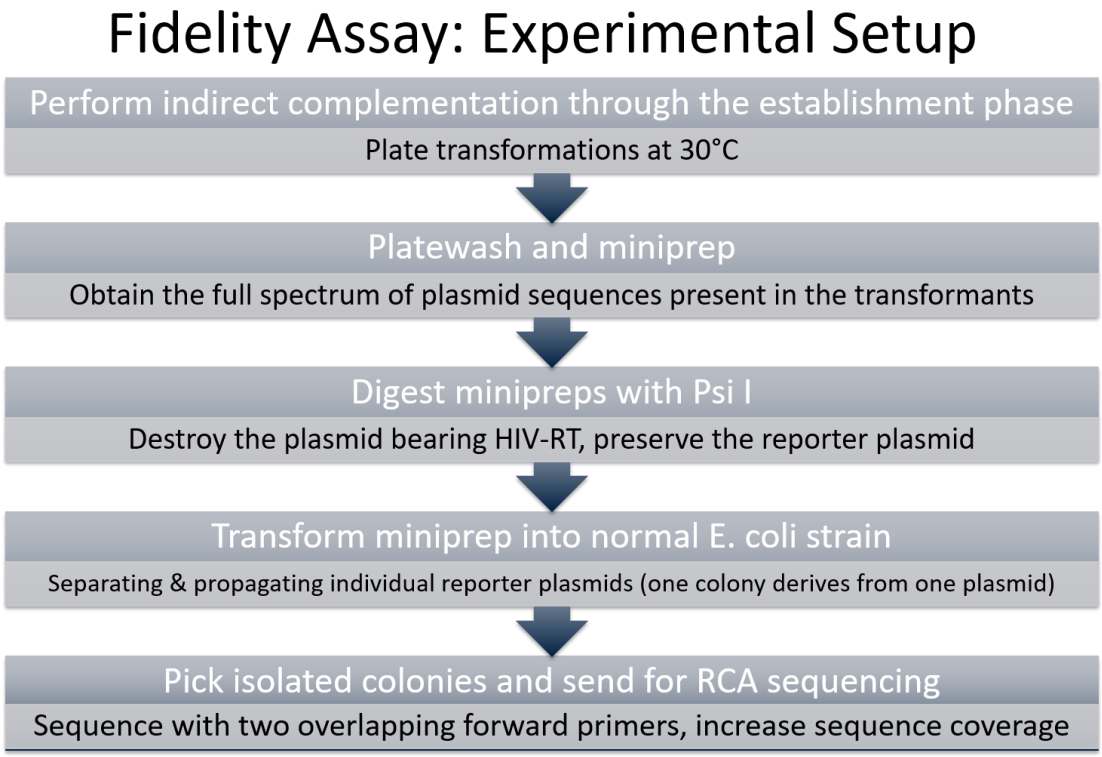


Figure 4.7. Fidelity assay performed using the pGFPuv reporter plasmid in HIV-RT expressing cells.

To test if this system can be used in a high-throughput fashion to screen and analyze RT fidelity, large-scale sequencing was undertaken using Rolling-Circle Amplification (RCA) sequencing. The results of sequencing revealed that there is indeed a distinctive mutation pattern made by HIV-RT. The comparison shown in figure 4.9 highlights the differences between the mutation phenotype of Pol I to that discovered for HIV-RT. A second important result from this was the emergence of

two mutation hotspots in the reporter. It is important to note that each of the colonies

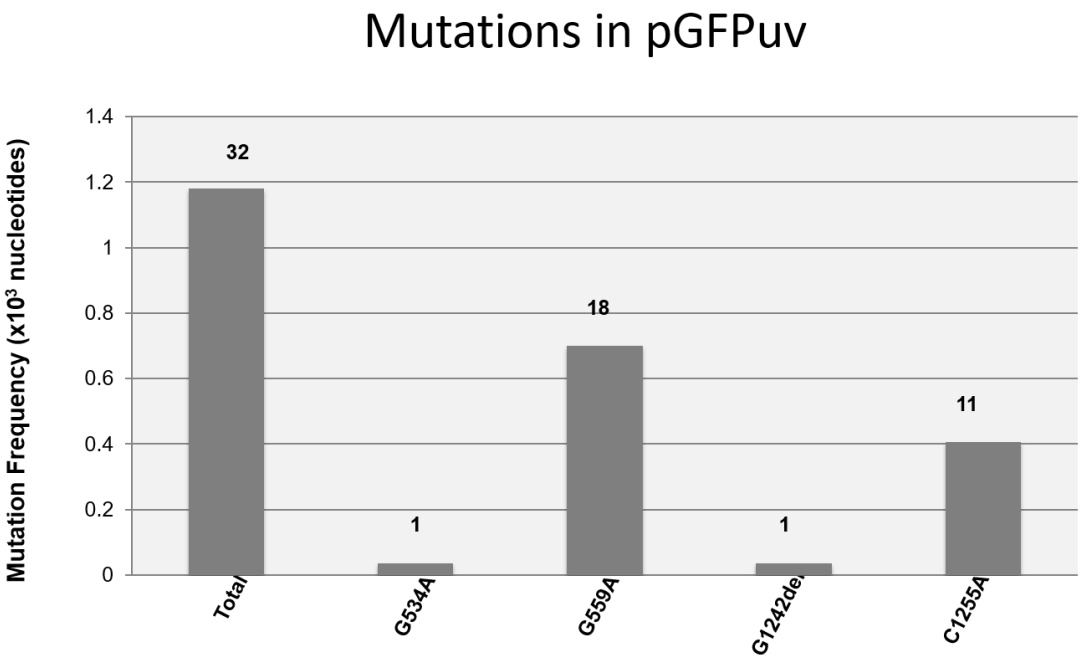


Figure 4.8. Independent, non-clonal Mutation Events in the pGFPuv induced by HIV-RT. The x-axis indicates the specific mutation being quantified, the y-axis represents the frequency of independent mutations at that location.

sequenced originated from different colonies, and from three distinct predecessor colonies from the reporter establishment phase of the assay. The same mutations arose in genetically distinct plasmids, which did not bear the mutation prior to replication in HIV-RT expressing cells (figure 4.8). This is an exciting result, as it supports the mutational hotspot phenotype. Examination of the sequence context surrounding the hotspots did not reveal an apparent convergence in sequence that could represent a generator motif. This does not discount the hypothesis, but does mean that further analytical methods are needed to understand hotspot formation.

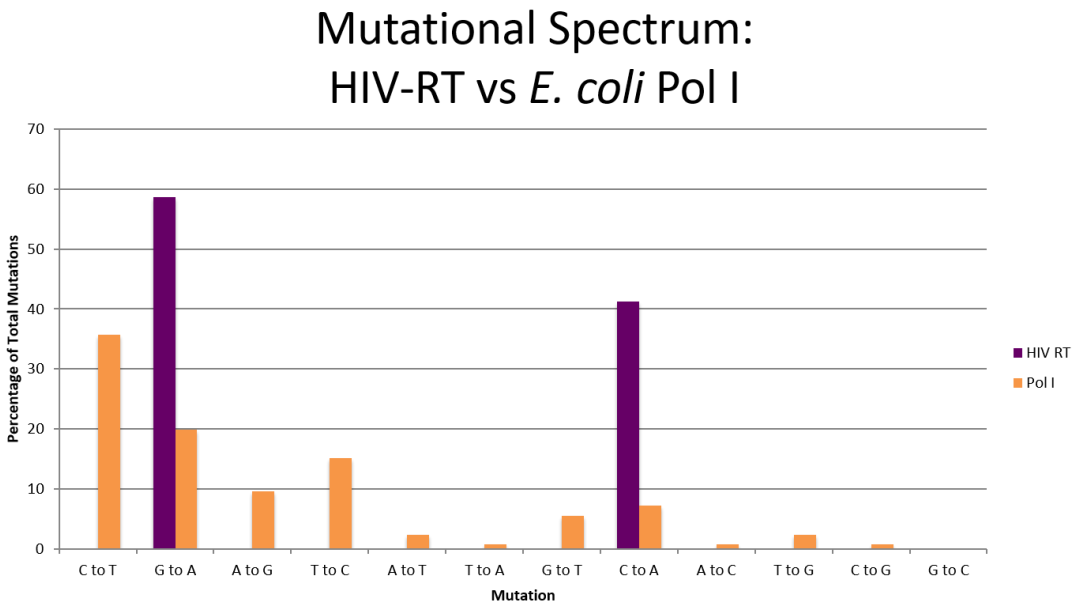
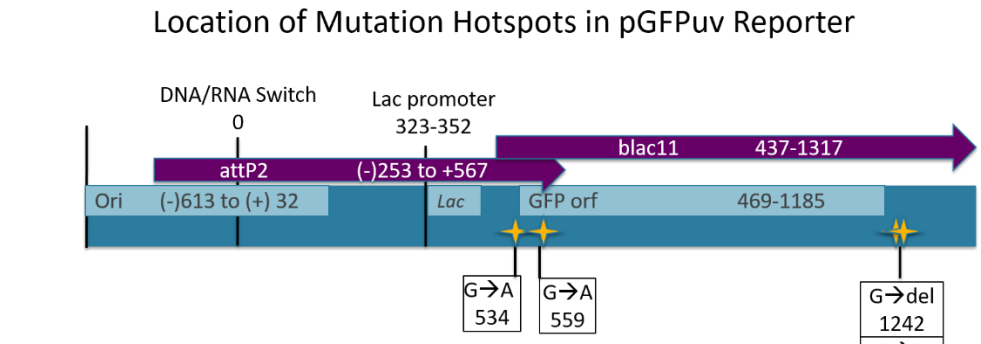


Figure 4.9. Comparison of the Mutation Spectrum determined for HIV-RT versus Pol I. Mutations made by the native Pol I are shown in orange, mutations induced by HIV-RT are shown in purple. The x-axis is organized by mutation type.



Mutation	Sequence Context
G534A	TTAATG**G**GCACAAAT
G559A	AGTGGAGA**G**GGTGAAG
G1242del	GGAAACAG_CTATGACCAT
C1255A	CTAGTCGG**C**CGTACGG

Figure 4.10. Hotspot Locations in the pGFPuv Reporter. Above is a diagram of the location of mutations induced by HIV-RT that were found in the pGFPuv reporter. The sequence context of the observed mutation hotspots is shown in the bottom left.

The fidelity assay was repeated by putting the same three predecessor ori mutants through the same process, however the second round of RCA sequencing did not yield a high number of mutations, and the mutations that were present fit the mutational pattern of Pol I and not HIV-RT. The first troubleshooting that was performed was to simply repeat with careful attention to all temperatures and timing. This did not change the negative result obtained in the second mutagenesis attempt. Subsequent troubleshooting involved generating new JS-200/p66A (RT) cell lines, using 35˚C as a restrictive temperature, using a different negative control for comparison, using varying induction timing and IPTG levels, using a different strain of *E. coli* for propagating the mutated reporter plasmids, and other slight alterations to the execution of the technique. None of these had an impact on the mutagenesis phenotype found through sequencing.

In addition to the optimization and troubleshooting steps, two mutant RTs were cloned and used to establish cell lines. The K65R and W266A reverse transcriptase variants were chosen as models for drug-resistant RTs. K65R is resistant to the chain-terminating nucleoside analog RT inhibitors (NRTIs) and W266A is resistant to the allosteric non-nucleotide RT inhibitors (NNRTIs). Of all variant RTs studied, only K65R displays a consistent fidelity relative to wild type RT, and thus is the best choice for validating that the prokaryotic system is a reliable means to compare the replication phenotypes of different RTs. K65R and W266A cover both possible outcomes for fidelity; K65R has a well documented 8-fold higher fidelity than wild type, while W266A has frequently shown to have a lower fidelity than wild

type [15,16]. These too, did not seem to display consistent complementation or mutagenesis (Data included as a supplemental.)

Another avenue of approach for exploring the fidelity of RT was to alter the sequence that was being copied. This is crucial for testing the generator hypothesis. Yet, there were no reporter plasmids available that would fit the requirements of a template for testing this hypothesis. Since the putative generator sequence motifs would be found through database searches of patient HIV genomes, any putative generator must be confirmed *in vivo* in order to establish a mechanism of hotspot development. Using a reporter plasmid with HIV viral genes would provide a biologically relevant template to use as a basis for investigating the impact of sequence context on the fidelity of RT replication. However, because such a template would be would be unique to this project, a classic sequence that is commonly used in mutation assays is also crucial for validating the system and the hypothesis. To that end, a novel reporter plasmid was constructed. I designed the pGK reporter (Figure 4.11 and 4.12) to contain four key sequence regions; a small area of native HIV sequence, an MCS that can be used to introduce more expansive areas of HIV sequence, the Kunkel sequence, and the GFP gene that can allow for rapid identifications of mutated reporters. This reporter system is unique in that it allows for the study of RT replication on a clinically relevant template; viral genes, in a high throughput manner. The MCS was designed to use only restriction enzymes that do not cut the HIV-1 genome, so that once putative generator sequences are identified they can be added into the pGK vector. By taking a putative generator sequence and

placing it amidst different surrounding genes, it is possible to test whether the motif is sufficient to induce a hotspot in non-native and native sequences. Using this reporter

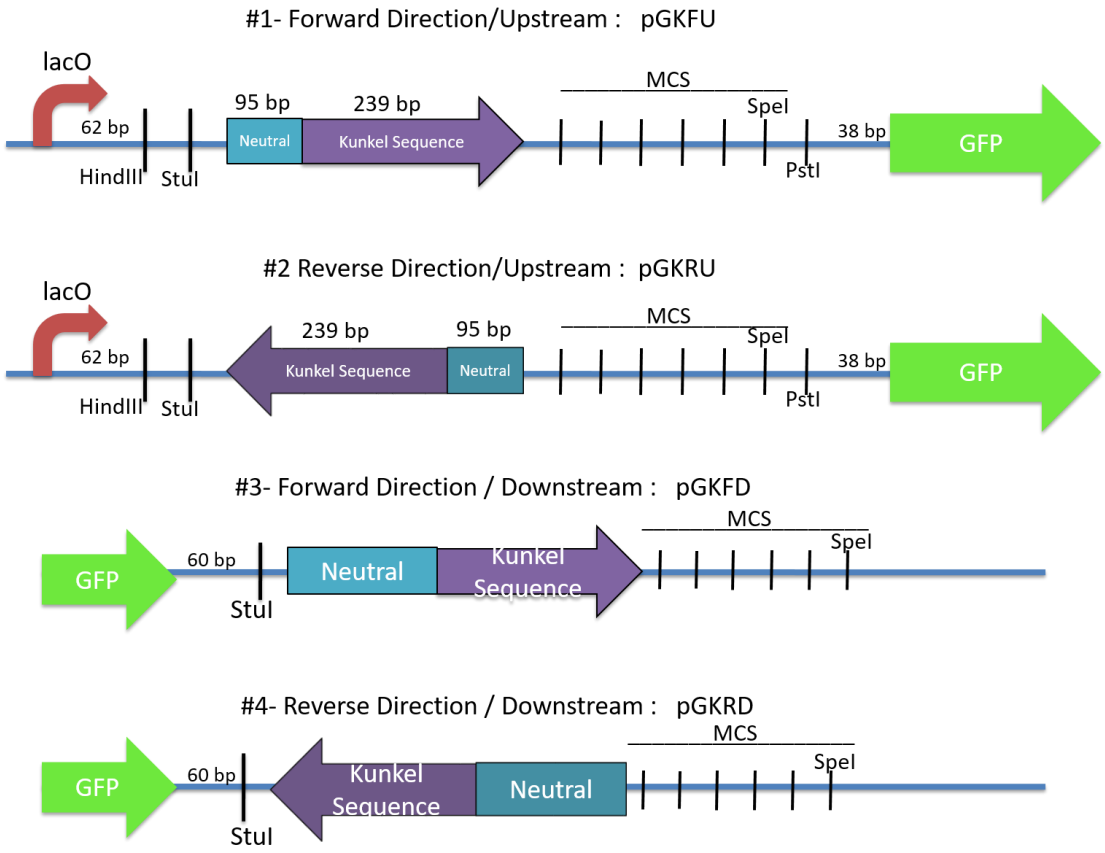


Figure 4.11. Details of the design of the pGK reporter plasmids. The region termed 'neutral' in the diagrams represents a portion of the gag gene, 100 bases surrounding a mutation hotspot.

it could be possible to determine if the regions implicated in our computational analysis are capable of inducing a change in mutation patterns *in vivo*. In addition to the various template sequences used to construct these reporters, I also placed the Kunkel sequence both upstream and downstream of the GFP gene, to determine how far along the template RT is capable of copying. Lastly, I placed the Kunkel sequence

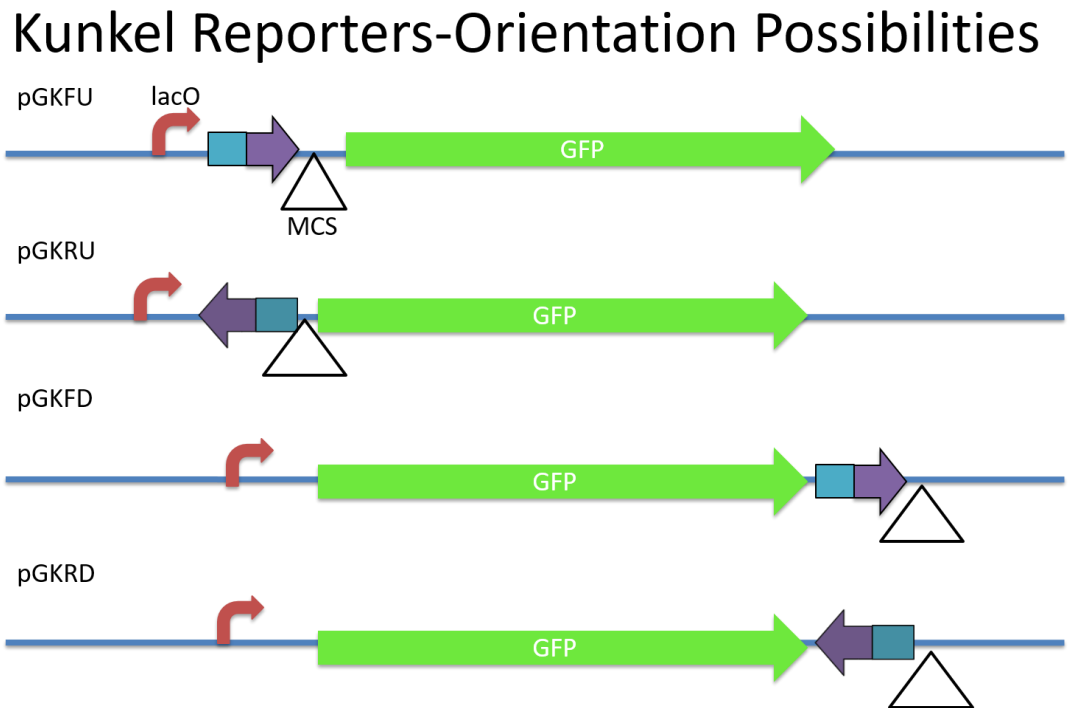


Figure 4.12. Overview of the arrangement of sequences in the four pGK reporter plasmids. This figure highlights the relative orientations of each mutation target in the pGK reporters.

in the opposite orientation in both the upstream and downstream positions, to determine if there is a strand-preference in hotspot positions. This yielded four unique reporters, as shown in figure 4.12.

Unfortunately, it was not possible to effectively troubleshoot the problems faced by the prokaryotic assay. Neither the variant RTs or the novel reporter plasmids showed any consistent phenotype. Thus, despite its promise, this project was abandoned.

DISCUSSION

Unfortunately, the bacterial system proved too inconsistent to use as a method for meaningful analysis. However, the pGK reporter and the computational search method generated by this project are useful tools in further study of polymerase fidelity, whether in a heterologous polymerase or an endogenous *E. coli* polymerase.

Other *in vitro* methods can be employed to study the replication phenotype of HIV RT. Given that the putative generators determined from the first round of RCA sequencing arose through replication of a single-stranded template, it is possible that the plasmid system employed in this work is not capable of reproducing the conditions necessary for hotspot emergence. Primer extension assays using long templates (several hundred nucleotides) can be performed, and the products sequenced to determine if hotspots are produced. If this is also unsuccessful, other *in vitro* assays using RNA templates could be performed. If needed, further analysis can be done computationally to refine the search for generator sequences. The generator hypothesis should be pursued by other means, as the implications for HIV evolution and drug development are substantial.

1. Rezende, L. F. & Prasad, V. R. Nucleoside-analog resistance mutations in HIV-1 reverse transcriptase and their influence on polymerase fidelity and viral mutation rates. *Int. J. Biochem. Cell Biol.* **36,** 1716–34 (2004).

2. Legrice, S. & Gotte, M. *Human Immunodeciency Virus Reverse Transcriptase: A Bench-to-bedside Success*. (Springer New York, 2013).

3. O'Neil, P. K. *et al.* Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. *J. Biol. Chem.* **277,** 38053–61 (2002).

4. Smyth, R. P., Davenport, M. P. & Mak, J. The origin of genetic diversity in HIV-1. *Virus Res.* **169,** 415–29 (2012).

5. Hu, W.-S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med.* **2,** (2012).

6. Ngandu, N. K. *et al.* Extensive purifying selection acting on synonymous sites in HIV-1 Group M sequences. *Virol. J.* **5,** 160 (2008).

7. Wood, N. *et al.* HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog.* **5,** e1000414 (2009).

8. Jewell, N. a, Chen, R., Raices, R. & Mansky, L. M. Nucleoside reverse transcriptase inhibitors and HIV mutagenesis. *J. Antimicrob. Chemother.* **52,** 547–50 (2003).

9. Sarafianos, S. G. *et al.* Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **385,** 693–713 (2009).

10. Götte, M., Rausch, J. W., Marchand, B., Sarafianos, S. & Le Grice, S. F. J. Reverse transcriptase in motion: conformational dynamics of enzyme-substrate interactions. *Biochim. Biophys. Acta* **1804,** 1202–12 (2010).

11. Marchand, B. & Götte, M. Site-specific footprinting reveals differences in the translocation status of HIV-1 reverse transcriptase. Implications for polymerase translocation and drug resistance. *J. Biol. Chem.* **278,** 35362–72 (2003).

12. Yamanaka, R. & Termini, J. Nucleotide sequence context influences HIV replication fidelity by modulating reverse transcriptase binding and product release. *Biosci. Trends* **1,** 52–61 (2007).

13. Garforth, S. J. *et al.* K65R and K65A substitutions in HIV-1 reverse transcriptase enhance polymerase fidelity by decreasing both dNTP misinsertion and mispaired primer extension efficiencies. *J. Mol. Biol.* **401,** 33–44 (2010).

14. Silverman, A. P., Garforth, S. J., Prasad, V. R. & Kool, E. T. Probing the active site steric flexibility of HIV-1 reverse transcriptase: different constraints for DNA- versus RNA-templated synthesis. *Biochemistry* **47,** 4800–7 (2008).

15. Svarovskaia, E., Cheslock, S., Zhang, W.-H., Hu, W.-S. & Pathak, V. Retroviral mutation rates and reverse transcriptase fidelity. *Front. Biosci.* **1,** 117–134 (2003).

16. Menéndez-Arias, L., Matamoros, T. & Cases-González, C. E. Insertions and deletions in HIV-1 reverse transcriptase: consequences for drug resistance and viral fitness. *Curr. Pharm. Des.* **12,** 1811–25 (2006).