INTEGRATION MECHANISM OF THE SITE SPECIFIC NON-LTR RETROTRANSPOSON, R2BM

by

CHRISTI LYNN HULL

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN BIOLOGY

THE UNIVERSITY OF TEXAS AT ARLINGTON

MAY 2010

ACKNOWLEDGEMENTS

This research was supported by the efforts of Dr. Shawn Christensen at the University of Texas at Arlington. Lab assistance was provided by past and present members of the Christensen lab (Blaine Thompson, Dillon Cawley, Arturo Menchaca, and Melanie Verna). I would also like to thank Heiko Keller (PhD student at Max Planck Institute for Developmental Biology, Germany) for his exceptional advise on working with RNA, which ultimately led to my success in making my mini RNAs. I would like to thank my committee members for their excellent advice and support when writing my thesis. I would like to thank Dr. Laura Mydlarz and her lab members for allowing me and helping me to use the Synergy spectrophotometer. I would like to thank the University of Texas Arlington Department of Biology for providing me the opportunity to study and work with many amazing scientists. I would like to thank Dr. Raymond Jones for his continuous support concerning equipment, advice on experiment design, data analysis, resume writing, and job searching. Most importantly, I would like to thank Dr. Shawn Christensen for providing me with this exceptional learning experience.

April 13, 2010

ABSTRACT

INTEGRATION MECHANISM OF THE

SITE SPECIFIC NON-LTR

RETROTRANSPOSON,

R2

Christi Hull, M.S.

The University of Texas at Arlington, 2010

Supervising Professor: Shawn Christensen

The focus of this thesis is the mechanistic study of non-LTR retrotransposon replication using the site specific transposon R2Bm. Non-LTR retrotransposons integrate into new chromosomal sites by copying their RNA back into DNA at the site of insertion using an exposed chromosomal 3' OH to prime cDNA synthesis.[1] Although some of the early steps of an integration event have been detailed, how the element completes integration remains unknown.[1 2 3 4 5 6] Clarifying the last integration step, known as second strand synthesis, is a major goal of this research. In order to accomplish this, R2 RNA and protein substrates were generated and used in biochemical assays to explore the aspects of RNP formation and second strand synthesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

1.1 Transposons: General Background and Genome Considerations

Nancy L. Craig, in Mobile DNA II, describes transposition as the recombination reaction that mediates the movement of discrete DNA segments between many nonhomologous sites.[7] These segments have a variety of names including insertion sequences, transposons, and transposable elements.[7] Transposable elements are genomic parasites that represent a large portion of many eukaryotic genomes. For example, an estimated 45% of the human genome is comprised of mobile element derived DNA.[8] Transposon movement and accumulation provides a major contribution to the evolution of genes and genomes of most organisms studied to date.[9]

Transposable elements are highly mutagenic; causing insertions, deletions, chromosomal breaks, and ectopic recombination. These contributions to the genome can be neutral, harmful, and in some cases, helpful. For example, transposons can have a neutral impact on the host, residing in vertically inherited genomes for a very long time. Many families of eukaryotic transposable elements likely date back at least as far as the origin of the eukaryote lineage.[10] An example of one particular transposon that is harmful is LINE-1. The transposition events of LINE-1 are responsible for ~1/1,000 disease causing mutations in humans.[11] In particular, LINE-1 was identified as the likely progenitor of a mutagenic insertion in a patient with hemophilia A.[11] Transposons can also provide benefits to their host genomes. Transposon sequences have been exapted to regulate host genes, and are often the source for new host genes. In the case of V(D)J recombination, the RAG1 recombinase is an evolutionary descendent of the *Transib* transposase, arising after a transib infection 500 million years ago in the ancestor to jawed vertebrates.[12]

Transposable elements were first described by Dr. Barbara McClintock in the 1940s and 1950s.

She published many papers describing an amazing discovery of “controlling elements” which provided changes in patterns of coloration in maize kernels over generations of controlled crosses. In 1983, Dr. McClintock received the Nobel Prize for Physiology or Medicine and was credited for discovering “mobile genetic elements.” Since her discovery the field of transposons has exploded into a whole new discipline of science. As this field has progressed, many questions still lack answers: where do transposable elements come from, what is the extent of their influence on genome evolution, what are their full mechanisms of movement, and if this mechanism is understood, could it be used to manipulate genomes in a controlled manner?

1.2 Types and Mechanisms of Transposons

Transposable elements exhibit a diverse set of genetic structures and replication mechanisms.[9 13] These different genetic structures and mechanisms have been used as a means to classify transposable elements. Two major classes of transposable elements have been named: Class I Retrotransposons and Class 2 DNA transposons. First, DNA transposons have been found in bacteria, insects, worms, humans, and many other species.[9,14] DNA transposons use a single or double-stranded DNA intermediate to integrate into their host genomes. In active elements, a transposase enzyme is encoded which once translated, will bind at or near the elements inverted repeats and to the target DNA. It will then perform trans-esterification reactions to remove the transposon sequence from its old site, then inserting the transposon into its new site.[14] DNA transposable elements generally have limited sequence specificity of integration, meaning insertions can occur at a large number of genomic sites.[14] Eukaryotic DNA transposons can be further classified based on mechanism of integration. The “cut-and-paste” transposons excise themselves as double-stranded DNA and then insert elsewhere in the genome, for example, DDE-transposons (FIGURE 1a.). Another type of “cut and paste” DNA transposon, will excise itself from its original site to form a circular DNA intermediate, leaving an empty site which is repaired. A reversal of the excision steps results after capture of a target site and the transposon is inserted (FIGURE 1 e and f).[15] Lastly, are DNA transposable elements that replicate through the rolling-circle mechanism, or “copy-in or cut-out, and copy-in” mechanism (FIGURE 1g).[9 15] These elements copy their DNA directly

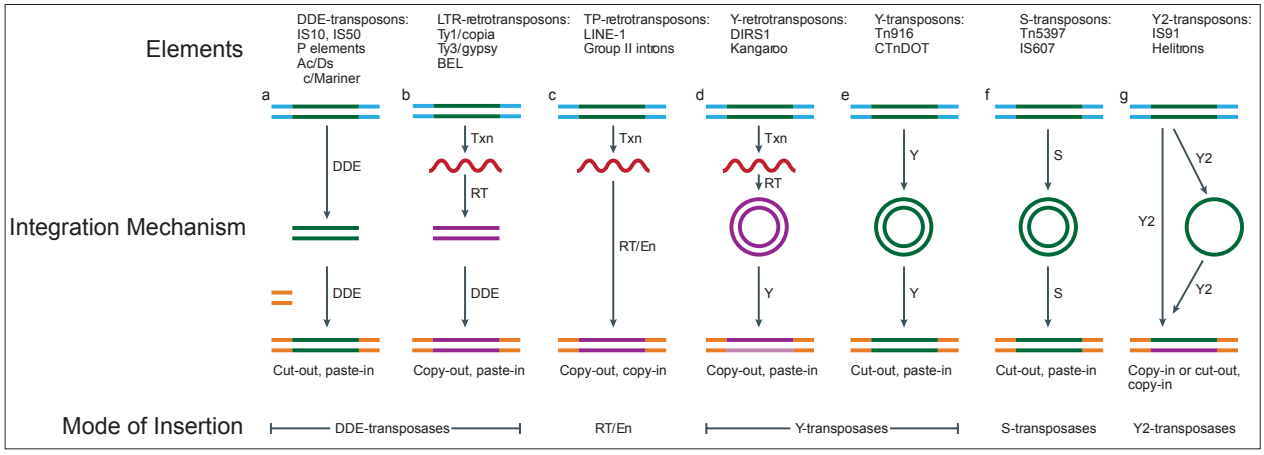


Figure 1: Transposons integrate in different ways

Five protein families dictate different transposition pathways (DDE-transposases, reverse transcriptase/ endonucleases (RT/En), tyrosine (Y)-transposases, serine (S)-transposases, and rolling-circle (RC)-or Y2-transposases. Transposons (green) can be "cut-out" or "copied-out" of the flanking donor DNA (blue). a. Generally DDE-transposons excise from the flanking DNA to generate an excised linear transposon, which is used for integration into a target site (orange). b. Retrotransposons use a copy-out mechanism by first generating an RNA template by transcription (Txn) and then generating a full copy of their RNA by reverse-transcription (RT). Long-terminal repeat (LTR) retrotransposons make a full-length cDNA copy (purple represents newly replicated DNA) from their RNA and integrate this into a target using a DDE-transposase. c. TP-retrotransposons use reverse transriptase to copy their RNA directly at a target site that has been nicked by a transposon encoded endonuclease (En). d. Y-retrotransposons are thought to generate a circular cDNA intermediate by reverse transciption and then integrate this into the target using a Y-transposase. e. and f. Y- and S- transposons encode either a tyrosine or serine transposase that mediates excision of the transposon to form a circular intermediate and a reversal of the catalytic steps reusults in integration. g. Y2-transposons "paste" one strand of the transposon into a target and use it as a template for DNA replication. Two models have been proposed for Y2-transposition integration. Representative elements of each type of transposon are listed above each pathway. Adapted from Curcio M.J. and Derbyshire K.M. (2003) Nature Reviews

into the target site by DNA replication such that each copy of the transposon (old and new) contains one newly synthesized strand. This process requires the help of the replication machinery of the host, but is initiated by a transposon-encoded protein that nicks the ends of the transposon and the target DNA. The resulting free 3' OH ends are thought to prime DNA synthesis using the transposon strand as the template.[15]

The second class of transposable elements use an RNA intermediate to integrate into the genome and are called retrotransposons. Retrotransposons use a "copy-out, paste-in" or a "copy out, copy in" mechanism (FIGURE 1 b,c,d).[15] [14] Retrotransposons with different mechanisms of replication have been found in every linage of eukaryotic organisms.[13] Not only is their diversity amazing but also their abundance within genomes. For example, retrotransposons make up ~42% of the human genome and

~75% of the maize genome.[13] Like DNA transposons, retrotransposons can be further classified based on their precise mechanism of integration. First, long terminal repeat retrotransposons and retroviruses have co-opted DDE-transposases to integrate their cDNA into a target, which is a step that is mechanistically equivalent to DDE-transposon insertions (FIGURE 1b).[15] Another family of retrotransposons called Y-transposons encodes a tyrosine (Y)-transposase, rather than a DDE transposase, and are thought to generate an excised circular cDNA copy of their transposon by reverse transcription of their RNA. It is proposed they use the Y-transposase to insert the circular DNA intermediate into the genome (FIGURE 1d).[15] The last major family of retrotransposons are called target-primed retrotransposons, or TP-retrotransposons. These elements encode an endonuclease in addition to a reverse transcriptase. The endonuclease nicks the target site, and the nick serves as a primer for reverse transcription of an RNA copy of their genome (FIGURE 1c).[15] As retrotransposons are the main focus of this thesis, these elements will be further discuss in greater detail.

1.3 LTR Retrotransposons

Long Terminal Repeat or LTR retrotransposons have a structure and use a replication mechanism similar to retroviruses.[16] Sequence analysis of these elements has revealed two open reading frames that encode a reverse transcriptase, an RNase H, an integrase, a proteinase, and nucleocapsid proteins, but lack an envelope gene (FIGURE 2).[13] Reverse transcription of LTR elements' RNA is primed by cellular tRNA molecules. Synthesis of the first and second strands results in a complete double-stranded DNA intermediate. Generation of the dsDNA is polymerized by the reverse transcriptase which is capable of jumping from the terminal repeats at the 5' end of the template to the 3' end. The dsDNA molecule is integrated into the host chromosome using a DDE integrase.[6]

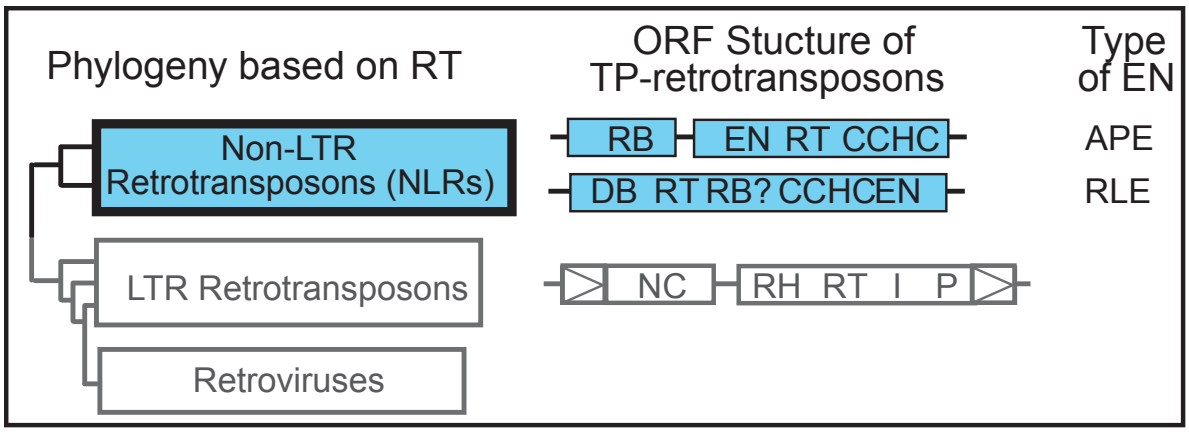
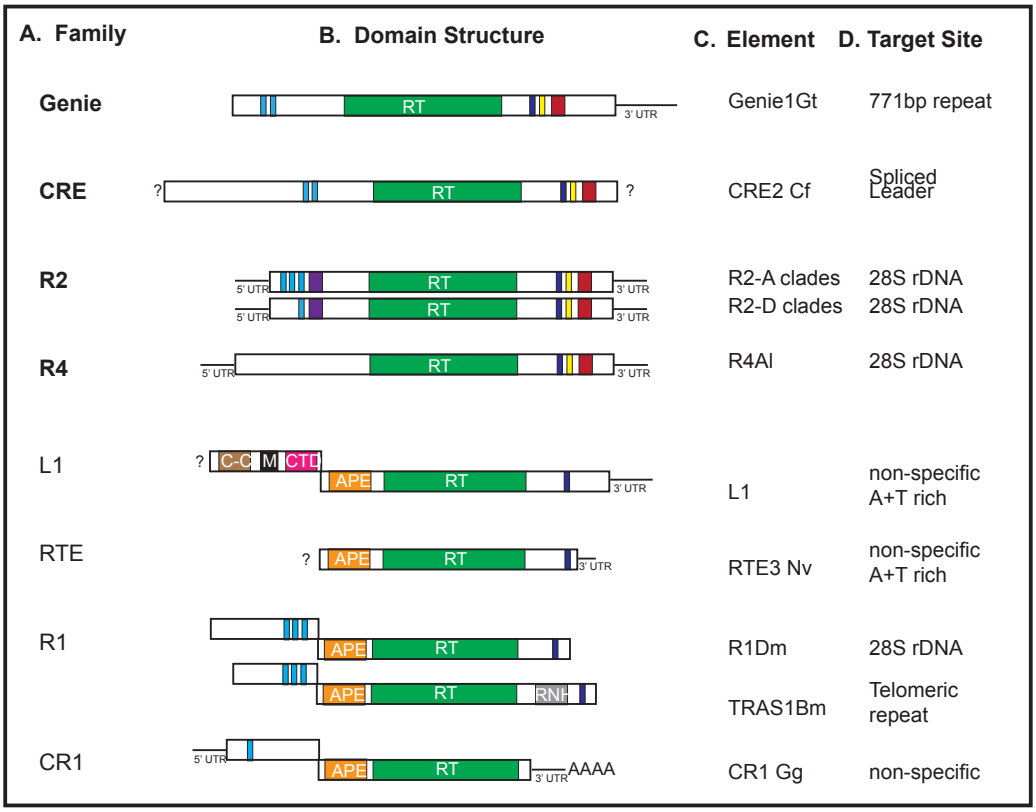


Figure 2: Phylogeny of reverse transcriptase bearing elements and structure of target primed (TP) retrotransposons.
The left column displays phylogeny based on reverse transcriptase sequence. Boxes represent all elements within a lineage (thick black lines represent lineages that replicate by target primed reverse transcription). Blue filled boxes highlight non-LTR retrotransposons (NLRs). The middle column shows the open reading frame (ORF) and most common domain structures of a given lineage. The right column represents the type of endonuclease present. Abbreviations: LTR (long terminal repeat), EN (endonuclease), RT (reverse transcriptase), RB (RNA binding), DB (DNA binding), CCHC (cys/his motif of unknown function), APE (apurinic/apyrimidinic like endonuclease), RLE (restriction like endonuclease), NC (nucleocapsid), RH (RNase H), I (integrase), P (protease), triangle (long terminal repeats). Adapted from Arkhipova I.R. et al. (2003). Nat Genet.; Eickbush T.H., and Jamburuthugoda V.K. (2008). Virus Res.

1.4 Non-LTR Retrotransposons

Non-LTR retrotransposable (NLR) elements are one of the most abundant classes of retrotransposons in mammals and use a simpler mechanism of integration.[13] Over 500,000 copies of these elements have been found in the human genome, approximating 17% of the total DNA.[17 8] NLRs are almost exclusively vertically inherited.[18] NLRs generally have internal promoters that allow the transcription of the entire element.[19] NLRs have open reading frames encoding a reverse transcriptase, various forms of endonucleases, and DNA and RNA binding domains (FIGURE 2).[13] NLRs have either a restriction-like endonuclease (RLE), or an apurinic/apyrimidinic-like endonuclease (APE).[20] NLRs that use an RLE typically have a single ORF. Some examples of RLE bearing NLRs are R2, CRE, Genie, and R4 (FIGURE 3). These elements are most often site specific, though some nonspecific ones exist. NLRs that use an APE endonuclease typically have two ORFs.[17] APE bearing elements include L1, RTE, R1, and CR1 clades (FIGURE 3). These elements typically insert at nonspecific genomic sites, although



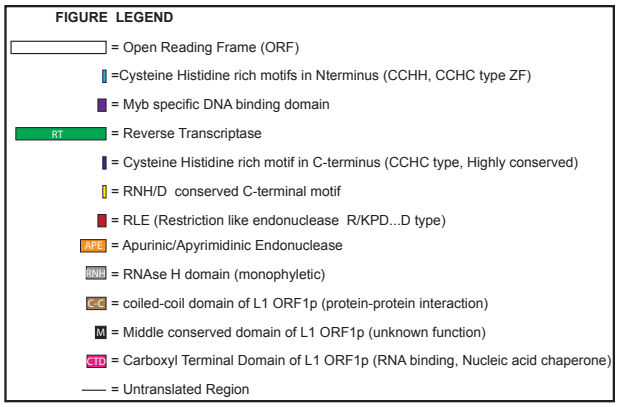


Figure 3: Structure of Non-LTR Retrotransposons

A. Family name. Bold names represent RLE bearing elements. **B.** The Domain Structures of the first or most common type of element from each major family. Element diagrams are not drawn to scale. 5' and 3' UTRs were marked for elements in which the literature stated explicitly. Elements which have a poly A or simple repeat at the 3' end have been included. **C**. Element name of element shown in panel B. **D**. Target Site states the target site of site specific elements and target preference for non-site specific elements. Figure adapted from many sources by Blaine Thompson (2010).

some site specific members exist.

While there are many different types of NLRs, it is believed they all share a basic mechanism. Non-LTR retrotransposons reverse transcribe RNA templates directly into a nick in chromosomal DNA, a process termed *t*arget-*p*rimed *r*everse *t*ranscription (TPRT).[6] TPRT is a "copy-out, copy-in" mechanism whereby element proteins bind to the mRNA from which they were translated, forming ribonucleoprotein particles (RNP). Specifically binding the RNA from which the protein was translated is known as cis-preference.[21] By an elusive process, these element specific RNPs gain access to the chromatin and use an element encoded endonuclease to cleave the host DNA at the target site, or point of insertion. The free 3' hydroxyl (OH) group exposed at the target site by the endonuclease is used to prime cDNA synthesis, thus the element RNA is integrated directly into the genome, i.e.. TPRT (FIGURE 4).[22] Integration of the 5' end of the element and how the dsDNA is created from the RNA:DNA hybrid produced during TPRT, is currently unknown. The hallmarks of this integration mechanism are apparent when observing sequence data. NLRs do not have terminal repeats. They most frequently contain a poly(A) tail at their 3' ends. NLR element 5' ends tend to contain variable deletions, because upon insertion, NLRs reverse transcribe from 3' to 5'. Stalled reverse transcription products integrate with 5' truncations which may explain some of the variation at the elements' 5' end.[13]

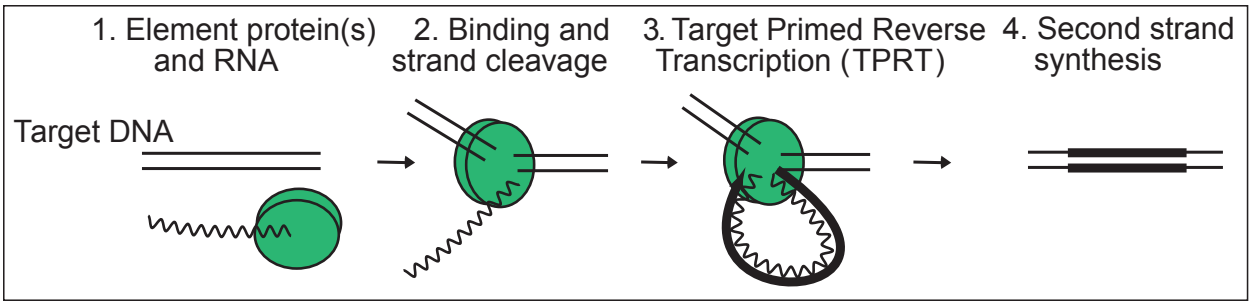


Figure 4: Target Primed Reverse Transcription.
1. Element protein (green ovals) binds to its own RNA (wavy line) forming an element specific RNA-protein particle (RNP). **2**. RNP binds to target DNA and cleaves the target DNA, exposing a free 3'OH, which is used to prime cDNA synthesis; 3. a process called target primed reverse transcription (TPRT). 4. Second strand synthesis completes integration. Adapted from Christensen & Eickbush (2005). Mol Cell Biol.

1.4.1. Structure and Function of APE bearing NLRs

APE bearing NLRs generally contain two open reading frames, the first ORF encodes a protein that binds nucleic acids.[23] In some elements such as, R1 and Jockey, this ORF contains three cysteine-histidine motifs (CCHC) (FIGURE 3). In other elements, like LINE-1, no such CCHC is present. The role of the first ORF of LINE 1 has long remained elusive because the amino acid sequence predicted lacks homology with any protein of known function.[24] Through a series of *in vivo* and *in vitro* experiments, it has been discovered that the ORF 1 protein binds both RNA and DNA, with a higher affinity for single-stranded than double stranded nucleic acids. This RNA binding function leads to RNP formation and delivery of the RNP to genomic DNA for the process of TPRT.[24] In mice, the ORF 1 protein of L1 cofractionates with full-length L1 RNA, and its expression is developmentally regulated in both spermatogenesis and oogenesis.[23] In *D. melanogaster*, the ORF 1 protein of I elements are present during oogenesis, but only when I elements are active.[23] ORF 1 protein may also act as a nucleic acid chaperone.[24] Nucleic acid chaperones are proteins that facilitate rearrangements of nucleic acids to their thermodynamically most stable form.[25] It is theorized that the nucleic acid chaperone activity of ORF 1 protein contributes to reverse transcription by facilitating the strand exchanges that place the DNA primer onto the RNA or cDNA template. The ORF 1 protein may also contribute to reverse transcription by melting secondary structure in the RNA.[24] ORF 1 of human L1 encodes an RNA binding protein with a conserved carboxyl-terminal domain that is rich in basic amino acids. Finally, the crystal structure of L1 EN in humans is a two-layered α-β sandwich with approximately 2-fold internal symmetry.[26]

The second ORF of APE bearing NLRs encodes a polypeptide which provides the two known enzymatic activities, endonuclease (EN) and reverse transcriptase (RT), that are required for TPRT.[24] Upstream of the RT domain, most NLRs encode an endonuclease domain (APE) with sequence similarity to the apurinic-apyrimidinic endonuclease involved in DNA repair.[23] The APE domains from L1, R1, Tx1, and TRAS have been expressed and shown to encode an endonuclease capable of cleaving DNA. APE domains from the site specific elements R1 and TX1L cleave the insertion site, but they also efficiently

cleave other sequences as well. The APE from the non-specific L1 element recognizes unusual structures (i.e., kinks) at AT rich regions of DNA.[17] So, the APE domain appears to provide at least some of the targeting specificity for elements with this domain. ORF2 of APE bearing elements contains a conserved cysteine-histidine-rich domain of unknown function near the carboxyl terminus. It is speculated that this area may function in nucleic acid binding as point mutations to this area reduced L1 retrotransposition by two orders of magnitude, but did not disrupt RT activity.[17] Some APE NLRs such as, I elements and LOA elements, contain an RNase H domain. Elements that lack an RNase H, either displace the RNA themselves, or use an endogenous RNase H activity (FIGURE 3).[23]

One of the best studied of the APE-like endonuclease elements is LINE 1 (Long Interspersed Nucleotide Elements). LINE-1 (L1) has been found in the human, mouse, rat, and other mammalian genomes. The integration of L1 has been shown to occur at random sites and can cause genetic diseases and cancers as well as major genome reconstruction.[27] More specifically, L1 in humans is known to make up one sixth of the total genome. The basic mechanism of LINE 1 begins with transcription of the element. The resultant RNA is transported to the cytoplasm where ORF1 and ORF2 are translated and form a ribonucleoprotein particle with the L1 RNA that encoded them. When L1 retrotransposition is activated, the RNP is imported into the nucleus and the L1 RNA is reverse transcribed at the target site via TPRT.[28] The majority of L1s are unable to retrotranspose because they are 5' truncated, rearranged, or mutated, but an estimated 30 to 100 full-length human L1s are retrotranspositionally-competent.[17] L1s are usually flanked by variable-length target-site duplications.[17]

1.4.2. Structure and Function of RLE bearing NLRs

RLE bearing NLRs generally contain one ORF and tend to be site specific, targeting a specific sequence. Most of these site specific retrotransposons target areas based on nucleotide sequence and are often areas of repetitive sequence such as microsatellites, ribosomal RNA genes, and telomeric repeats.[27] For example, R2 of *Bombyx mori* and R5 of *Dugesiid planaria*, integrate into a specific site in the 28S rDNA (FIGURE 5).[29] Other elements such as R8 of *Hydra magnipapillata* integrate at the 18S ribosomal RNA gene location (FIGURE 5). Still other elements such as heT-A and TART help maintain

the telomeres of *D. melanogaster* chromosomes, and TRAS1 and SART1 integrate into telomeric repeats of *Bombyx mori.*[14] This specificity is thought by many to be a unique strategy of the retrotransposon in which they avoid destroying their host by not inserting into essential genes, but into multi copy genes or repeats. "Flying under the radar" of the host genome and maintaining a presence but not domination. RLE bearing NLRs' one ORF also contains an endonuclease domain that is different from the APE NLRs. The active site of this EN domain is similar in sequence to that of various type II and typeIIs restriction enzymes, such as EcoRI, EcoRV, and FokI.[23] All elements like R2 that have a C-terminal EN domain also possess an N-terminal domain that frequently contains conserved zinc-finger and/or myb DNA binding motifs.[23]

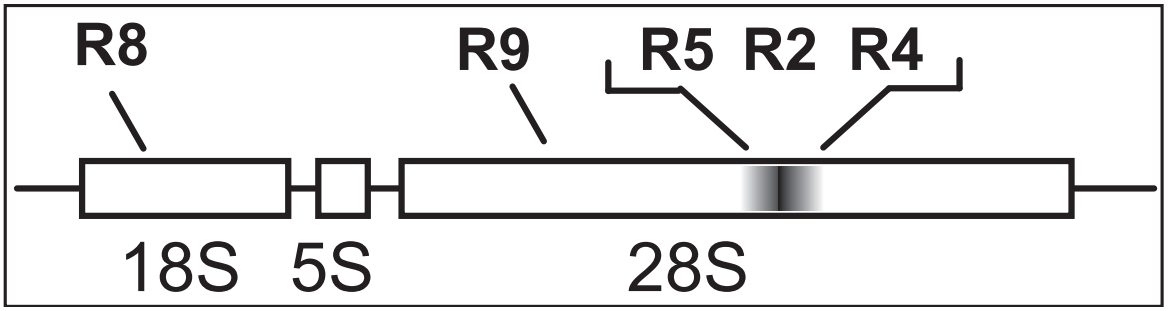


Figure 5: Ribosome locus specific non-LTR retrotransposons (NLRs).
Elements are RLE bearing elements.

1.4.2.1 Domain Structure and Function of R2 Clade Elements

Elements of the R2 group insert into a specific site in the 28S rDNA. These elements may not require their own promoter as the regions they insert into are areas that are heavily transcribed by the host. R2 was first identified as an insertion sequence in the fruit fly, *Drosophila melanogaster,* and the domestic silkworm, *Bombyx mori.*[27] The major features of the R2 group of elements is a single ORF and the presence of a C-terminal restriction-like endonuclease domain. All elements have an extensive N-terminal region that usually contains C2H2 zinc-finger and/or myb DNA binding domains.[23] Some elements also have longer N-terminal domains that extend beyond these DNA binding motifs, but the function of these domains is unknown. Kojima and Fujiwara investigated R2s distribution in many

animals and based on the number of N-terminal zinc finger motifs and phylogenetic analyses, they were led to classify R2 group elements into four clades.[30] A basic representation of the different clades can be seen in FIGURE 6. All R2 elements seem to have very similar C-terminal domains which include a cysteine-histidine rich motif (CCHC), an RNH/D motif, and an RLE domain. Elements that are R2 group elements but target a different site than R2, are given a different designation (e.g. R8 and R9 are R2-A clade members but target 18S and a non standard site in 28S, respectively) (FIGURE 5). Other R2 elements belonging to the R2-A clade, such as R2Lp (horseshoe crab) have three zinc fingers (ZF) and a myb within the N-terminal region. The R2-D clade members, such as R2Bm, have only one zinc finger and a myb in the N-terminal region. R2-C clade elements, such as R2Sm, have 2 zinc fingers and a myb. No full-length copies of R2-B clade elements, such as R2Pj-A, have been found, so little is known about their N-terminus, but they contain a reverse transcriptase and similar C-terminal domains as other R2 group elements.

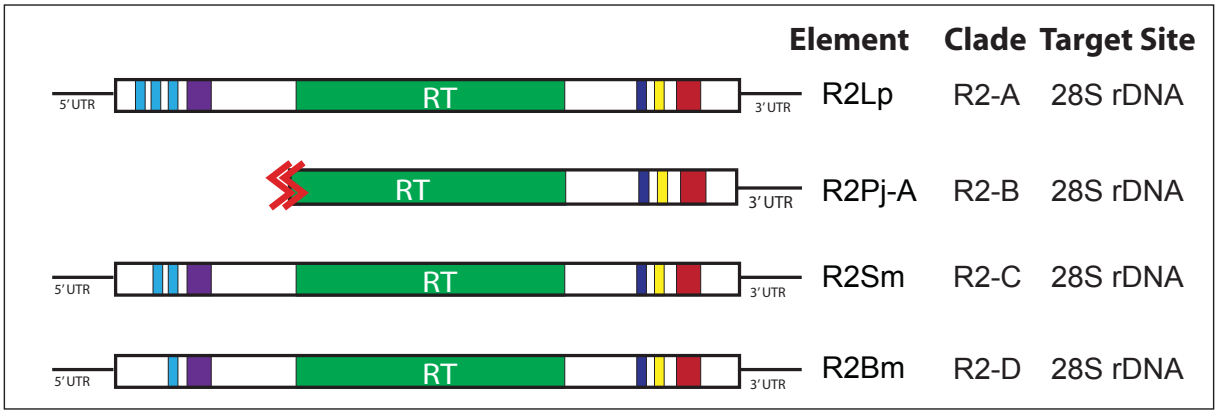


Figure 6: Domain structure of RLE bearing site specific NLRs
The amino-terminal of RLE bearing NLRs varies in size and structure: zinc fingers (light blue) and myb (purple). The carboxly-terminal domain contains a cysteine-histidine rich motif (blue), RNH/D motif (yellow), and a restriciton-like endonuclease (red). The central reverse transcriptase domain is green. R2Lp (Limulus polyphemus), R2Pj-A (Popilla japonica), R2Sm-A (Schistosoma mansoni), R2Bm (Bombyx mori). Adapted from Kojima and Fujiwara Mol. Biol. Evol. 2005.

1.4.2.2 Integration Mechanism of R2Bm: what we know now

R2 of *Bombyx mori* is one of the most investigated sequence-specific non-LTR retrotransposons that transposes exclusively into 28S rDNA. R2Bm protein is highly specific for not only the 28S DNA target site, but also the R2 RNA template from which is is made.[3]

The most current model of R2Bm integration has been formed by the work done by Shawn Christensen in the Eickbush lab and more recently in his own lab. The R2 mRNA contains two specific R2 protein binding sites that have been identified: one near the beginning of the open reading frame and the other located at the 3' untranslated region (3' UTR).[4] The protein binding motif (PBM) at the 3' end of the mRNA corresponds to the 250 nucleotide (nt) 3' UTR and consists of a set of conserved stem loops.[31] [32] The PBM near the start of the open reading frame is a highly structured, 320nt RNA segment, whose most striking feature is a conserved pseudoknot.[33] [34] R2Bm protein adopts different conformations upon binding to the 5' and 3' PBMs.[4] These different protein conformations, in turn, determine DNA binding properties and enzymatic activity.[4] The R2 protein contains an N-terminal DNA-binding domain, a central reverse transcriptase domain, and a C-terminal domain which consists of an endonuclease domain and a proposed RNA binding domain. For full integration to occur, it seems that two subunits of R2Bm protein are required (FIGURE 7). Each subunit has specific tasks in the integration reaction based on which domains are bound or exposed.[4] The subunit complexed to the 3' PBM binds to target DNA upstream of the insertion site. The subunit complexed to the 5' PBM binds to target DNA downstream of the insertion site. The 3' PBM RNA is the only RNA template capable of initiating TPRT, so the first strand DNA cleavage and TPRT are performed by the upstream subunit.[31] The downstream subunit cleaves the remaining DNA strand, but only after the 5' PBM has been removed by reverse transcription.[4] It is predicted that this downstream subunit catalyzes second strand synthesis because it is in the correct orientation and the reverse transcriptase is biochemically capable of displacing RNA from DNA:RNA heteroduplexes.[1] [35] This displacement is a necessary activity as R2Bm does not code for RNase H activity. The 5' PBM RNA thus acts to coordinate the timing of the second half of the integration reaction. Second strand synthesis has yet to be demonstrated for any element, although it is generally believed that the reverse transcriptase catalyzes this step—at least in the case of full integration events (FIGURE 7).[36] [4] [35]

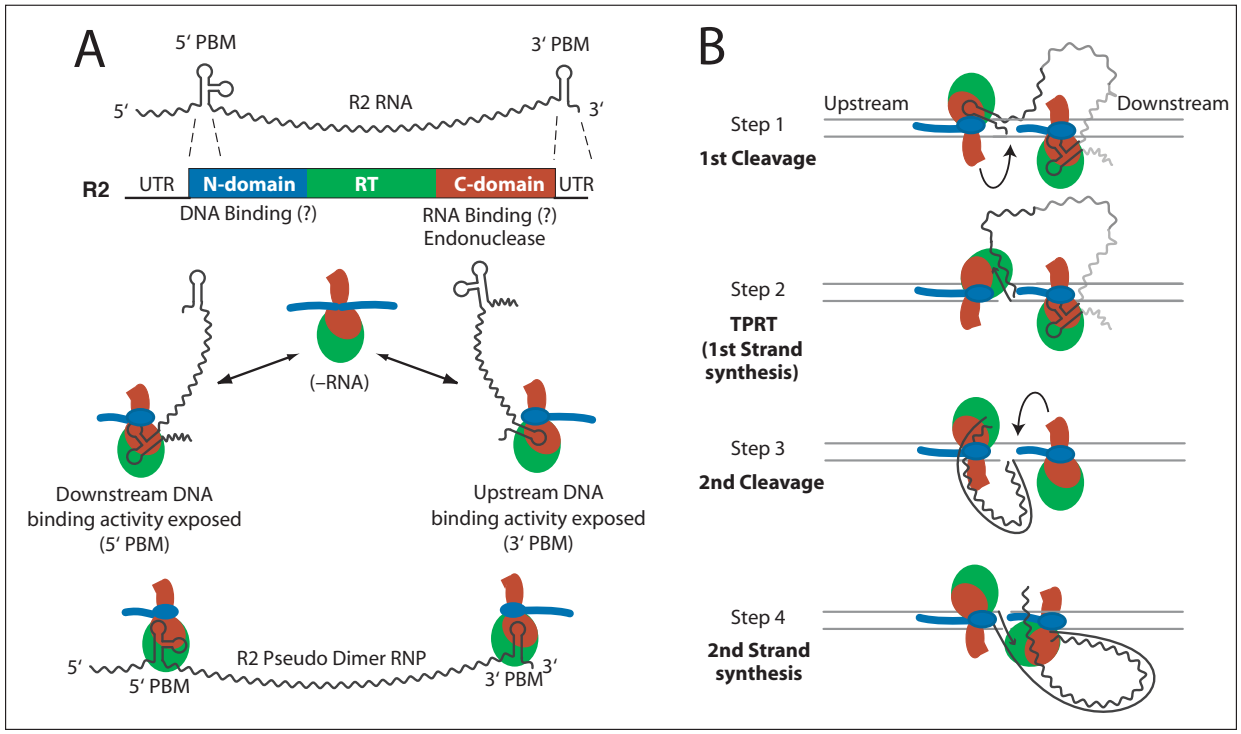


Figure 7: R2 element structure and proposed integration mechanism.
A. The R2 protein contains N-terminal DNA-binding domains (blue arms), a central reverse transcriptase (RT) domain (green), and a C-terminal domain consisting of an endonuclease domain (red oval with pick) and proposed RNA binding domain(s) (red patch on green oval). R2 protein bound to the 3' protein binding motif (PBM) sequesters one of the DNA binding domains, exposing only the domain responsible for binding to target DNA upstream of the insertion site (seen in B.). Protein bound to the 5' PBM sequesters the other DNA-binding domain, exposing the domain responsible for binding downstream of the insertion site (seen in B.). When R2 protein is bound to each 5' and 3' PBM of a single RNA, a pseudo dimer RNA and protein (RNP) complex is formed.
B. R2 integration is proposed to be catalyzed by two subunits in four steps. Step 1: The endonuclese from the upstream subunit is responsible for first strand cleavage. Step 2: The RT of the upstream subunit catalyzes first strand target primed reverse transcription (TPRT). Step 3: The downstream subunit cleaves the second DNA strand. Step 4: The downstream subunit is thought to provide the polymerase activity used to perform second strand DNA synthesis
Adapted from Christensen and Eickbush (2005) Mol.Cell.Biol.

1.5 R2 Outstanding Issues and Research Goals

The model for R2Bm integration predicts the existence of two DNA binding domains. R2Bm protein complexed with 3' PBM is known to bind to DNA sequences upstream of the insertion site, and protein complexed with the 5' PBM binds to DNA downstream of the insertion site. Downstream DNA binding activity has been mapped to the amino-terminal ZF plus myb motifs.[1 3] The protein domain used to bind the upstream subunit of the R2Bm to the DNA has yet to be identified.

The RNA binding activity of the R2Bm protein has also yet to be localized to a protein domain. The model for integration predicts the existence of two RNA binding domains (one to bind the 5' PBM and one to bind the 3' PBM). Preliminary data suggests that RNA binding activity might be located in the carboxyl-terminal region (FIGURE 8). The carboxyl-terminal polypeptide used in Figure 8 bound 5' and 3' PBM RNA. In addition weak endonuclease activity was observed as the peptide contains the endonuclease. No stable DNA complexes were found. It is tempting to speculate that the CCHC motif of the C-Terminal end is involved in RNA binding.

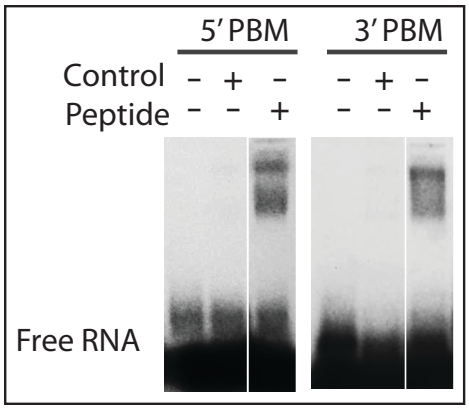


Figure 8: Electromobility Shift Assay (EMSA) of 5' and 3' PBM RNAs.
EMSA was run on a native agarose gel. PBM RNAs were 32P labeled on 5' ends. Peptide is a polypeptide corresponding to the carboxyl-terminal domain of R2Bm. Control is an empty vector expression/purification using vector DNA with no insert.

Another point of interest is the RLE endonuclease. The RLE itself is hypothesized to be essentially nonspecific, instead specificity is given by separate DNA binding domains, much like the restriction enzyme Fok 1 and Fok 1 derived chimeric endonuclease.[37] The innate specificity of RLEs from site specific non-LTRs has yet to be determined.

Demonstrating R2 catalyzed second strand synthesis in vitro has yet to be seen and is extremely fundamental to understanding the full mechanism of R2Bm integration. Past experiments done in the Eickbush lab used only the 3' PBM or the 5' PBM RNA as individual, unlinked RNA bound to R2Bm protein and exposed to target DNA. Results showed that the protein seems to take on a different structure when bound to different PBMs. These experiments also explained that the different resulting structures allowed the protein to bind and cleave the DNA target at different sites (upstream or

downstream of the insertion depending on 3' or 5'), but second strand synthesis was never observed. Therefore the major goal of my research is to elucidate second strand synthesis by reconnecting the 5' and 3' PBM RNAs.

CHAPTER 2

RATIONALE AND METHODS

2.1 Mini RNA Rationale

The 3' and 5' PBMs are very unique structures that have been shown to bind protein. As mentioned above, the 3' PBM RNA consists of a set of conserved stem loops located in the 250 nt long 3' UTR. The 5' PBM is proposed to consist of a 320nt RNA segment that folds into a complex pseudoknot. Pseudoknots are rare structural motifs that often play important roles in biological processes such as gene expression and replication.[33]

The model suggests for a full integration event to happen the 5' and 3' PBM RNA's need to be covalently linked as part of the same RNA. In vitro studies using just the 3' PBM RNA or just the 5'PBM RNA have helped to described the many activities involved in element integration, but protein dimer formation on the target DNA and second strand cleavage were always inefficient. Furthermore, second strand synthesis was never observed; possibly a result of non-covalently linked 5' and 3' PBM structures. My research involved linking the 5' and 3' PBM structures through building "mini RNAs". Mini RNA's were designed to be created *in vitro* and to contain the 3' and 5' PBMs. The RNA's were also designed to have three variable lengths of the 5' tail, which extends beyond the 5' PBM (FIGURE 9). This was done to understand the importance of the length of the 5' tail to the process of integration. The spacer RNA, either 400nt or 200nt, linking the two PBM's, is derived from sequence near the 5' and 3' PBMs to help ensure proper folding of these domains. The length of the the mini RNA spacer needs to be experimentally determined as it is unknown the length needed to allow each of the PBMs to bind an R2Bm subunit without stericly interfering with each other, hence the 400nt and 200nt spacers. A restriction site was engineered into the mini RNAs at the junction of the 5' and 3' PBMs as a means to test for eventual second strand synthesis products by restriction digest (FIGURE 9). PCR was also used to assay for the presence of second strand synthesis products. The PCR assay used mini-RNA specific primers that annealed to the spacer region at the junction of the 3' and 5' PBM sequences, so PCR will

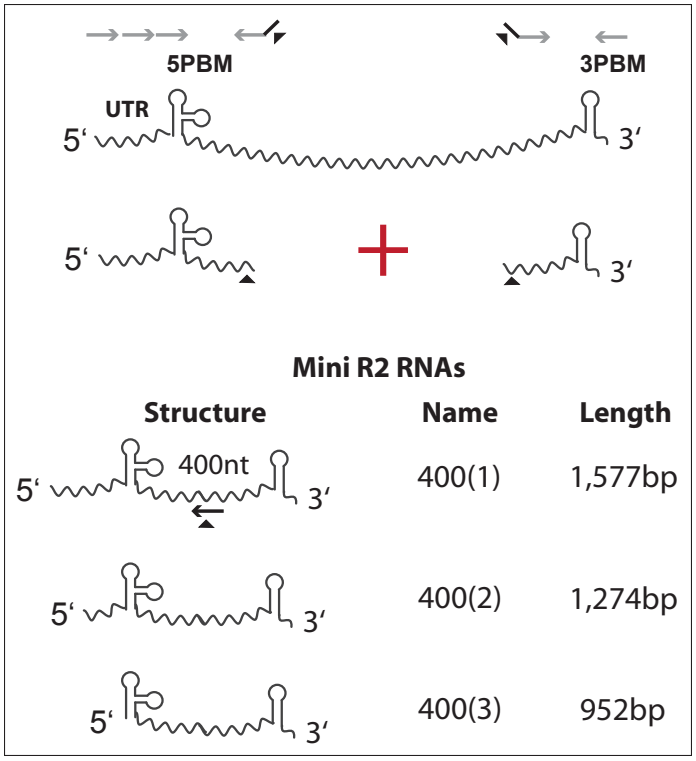


Figure: 9 Mini R2 RNA.
Templates for mini R2 RNA production are generated by recombinogenic PCR using the primers marked as gray arrows (primers are not drawn to scale). T7 promoter (not shown) is engineered into the 5' end forward primers, allowing in vitro transcription. DIfferent lengths of the 5' UTR are tested. The black triangle marks the position of the engineered restriction site. The black arrow marks the unique internal PCR primer used to assay for 5' integration events. This primer is located at the new junction between the 5' and 3' sequences.

only amplify integrated mini RNAs after an integration event. The mini RNAs were transcribed *in vitro* using template DNA containing T7 RNA polymerase promoters. Template DNA was made by recombinogenic PCR, connecting the 3' and 5' PBM motifs through a variable spacer. Two *in vitro* transcription protocols were implemented: one used the recombinogenic PCR products as DNA template (Version I), and the other used the same recombinogenic PCR derived template but in the context of a plasmid (i.e., cloned into pENTR) (Invitrogen #K2525-20) (Version II). All mini RNAs were gel purified by denaturing gel electrophoresis and refolded prior to use. After refolding the RNA quality was checked on native gels for presence of a single conformation.

2.1.1. Methods: Generate Mini RNA Template DNA

Mini RNAs composed of the 5’ and 3’ PBM joined by a variable R2Bm derived spacer-RNA, were generated using recombinogenic polymerase chain reaction (PCR) using Accuprime DNA polymerase (Invitrogen #12346-094). The 3’ PBM section was generated by PCR and contained a T7 promoter at the 3’ end and an area of 40 bases identical to the 5’ end of the 5’ PBM. The 5’ PBM section was generated by PCR and contained the 40 bases of identity at the 3’ end of the 3’ PBM section and a restriction site (FIGURE 9). Both pieces were combined by recombinogenic PCR, where the two PBM sections were allowed to anneal and primers were added to begin the extension process from both ends, meeting in the middle where the DNA was already double stranded. Thus forming a single double-stranded DNA containing a T7 promoter. Varying amounts of the 650 nt 5' UTR were included on the mini-R2 RNAs. The mini RNAs were kept as small as possible to aid in RNA stability and to ensure proper folding. Even so, the longest RNA (400(1)) was 1,577 base pairs (bp) long, the medium length RNA (400(2)) was 1,274bp long, and the shortest RNA (400(3)) was 952bp long. The spacer RNA was native sequence near the 5' and 3' PBMs to help ensure proper folding of these domains. One set of the mini RNAs (Version I) were left as PCR products. A duplicate set (Version II) was then cloned into the pENTR vector (Invitrogen) with primers containing a restriction site (BglII) at the 5’ end of the mini RNA sequence. The plasmid was then amplified, purified, and restricted.

2.1.2. Methods: In Vitro Transcription of Mini RNAs

Transcription reactions for Version I mini RNAs were 200ul and contained 1X Transcription Buffer (40mM Tris pH 7.5, 6mM MgCl2, 10mM Dithiothreitol (DTT), 10mM NaCl, 2mM Spermidine), 5mM rNTPs (each) (Epicentre #RN02825), 1U/ul RNase Inhibitor (Promega #N211A), 34mM MgCl2, 4U/ul T7 RNA polymerase (Epicentre #TU950K), 50ng/ul PCR DNA template (total). The transcription reaction was allowed to incubate at 37°C for 1.5 hours then 2U/ul T7 RNA polymerase was added and was allowed to incubate at 37°C for another 1.5 hours. The transcription reaction was then exposed to DNAseI treatment, (Promega #M6101) (1U/Ug DNA template contained within 1X buffer) in order to rid the RNA of the template DNA. Then the RNA was cleaned by phenol chloroform extraction, chloroform extraction, EtOH precipitation, and resuspended in 10ul double deionized (dd) water, and stored at -20°C. Mini RNAs were then purified by gel electrophoresis using a 3.5% polyacrylamide denaturing gel containing 8M urea, bands were visualized by staining with Sybr Green II (Lonza #50523), observed with a long

wave UV lamp, cut out of gel, and eluted overnight in elution buffer (50mM Tris, 0.1% SDS, 10mM ethylenediaminetetraacetic acid (EDTA), 0.3M NaCl). Then the RNA was cleaned by phenol chloroform extraction, chloroform extraction, and precipitated by ethanol and resuspended in 5ul dd water. The gel slices were eluted for a second time in elution buffer for 5 hours, cleaned, and resuspended in the same manner as the first. The two 5ul aliquots were added together and stored at -20°C. For sucrose gradients, the Version I RNA was P^{32} end labeled. Labeling reactions of 40ul consisted of 1X Buffer B (Fermentas #EK0031), 4.8% PEG (24% polyethleneglycol 6000) (Fermentas #EK0031), 0.25U/ul of rxn T4 PNK (Fermentas #EK0031), 0.5U/ul of rxn RNase Inhibitor (Promega #N211A), 1.75uCi/ul of rxn ATP (Perkin Elmer #NEG035C001MC), and 25pmol of RNA. This reaction was allowed to incubate at 37°C for 30 minutes. The reaction was then run over a G-25 sephadex (MP Biomedicals #195254) column and eluted in 1X TE (10mM Tris pH 8.0 and 1mM EDTA). The labeled RNA was then loaded onto a 0.8% agarose electrophoresis gel, stained with EtBr, and visualized with a long wave UV lamp. Bands were cut out of the gel and chemically eluted with Qiaquick gel extraction kit (Qiagen #28704). This product was then run through a G-25 sephadex spin column equilibrated in dd water. After the labeling and cleaning reactions, all Version I RNAs were approximately 0.2pmol/ul (~70ng/ul). Mini RNAs were refolded prior to use to ensure proper folding. The refolding reactions consisted of 3mM MgCl2, 10mM Tris pH 7.5, 100mM NaCl, and RNA. Reactions were placed in a thermocycler, heated to 65C for 3 minutes and gradually cooled to room temperature over 12 minutes. The MgCl2 was not added to the reaction until the temperature reached 40C. After refolding, RNA quality was checked on native agarose gels for presence of a single conformation. All Version I RNAs were resuspended at approximately 0.1pmol/ul. A 200ul transcription reaction yielded about 3.5ug of gel purified, labeled, and usable RNA.

Secondly, mini RNAs were transcribed using the linearized plasmid as template DNA (Version II) in 200ul reactions containing 1X Transcription Buffer (MBI Fermentas #EP0111), 5mM rNTPs (each)(MBI Fermentas #R0481), 1U/ul SUPERase In RNase Inhibitor (Ambion #AM2694), 34mM MgCl2, 3U/ul T7 RNA polymerase (MBI Fermentas #EP0111), 80ng/ul of transcription reaction restricted DNA template (16ug total). The transcription reaction was allowed to incubate at 37°C for 2 hours then 1.5U/ul T7 RNA polymerase was added (contained within 1X transcription buffer). The reactions were mixed well and allowed to incubate at 37°C for 2 more hours. The transcription reaction was then exposed to DNAseI treatment (Promega #M6101) for 30 minutes at 37°C in order to rid the RNA of the template DNA. Mini

RNAs were purified by gel electrophoresis using a 3.5% polyacrylamide denaturing gel containing 8M urea, bands were visualized by eye (density differences in gel) and by UV-shadowing (with short wave UV lamp), cut out of gel, and eluted overnight in 50mM Tris, 0.1% SDS, 10mM EDTA, 0.3M NaCl. The eluate was moved to a new tube and kept on ice until the second round of elution was finished. The gel slices were then eluted in elution buffer a second time for 6 hours and added to the first set and cleaned by phenol chloroform extraction, chloroform extraction. The RNAs were then stored at -20°C in 100% ethanol and glycogen (as a carrier). Prior to using them, the mini RNAs were precipitated and resuspended in ~40ul nuclease-free water (Teknova #W3450). Version II RNA concentrations were approximately 0.6pmol/ul (~200ng/ul). Mini RNAs were refolded prior to use in reactions to ensure proper folding of RNA structures. The refolding reactions consisted of 3mM MgCl2, 10mM Tris pH 7.5, 100mM NaCl, and RNA (~8ug). Reactions were placed in a thermocycler, heated to 65°C for 3 minutes and gradually cooled to room temperature over 12 minutes. The MgCl2 was not added to the reaction until the temperature reached 40°C. After refolding, RNA quality was checked on native agarose gels for presence of a single conformation and denaturing polyacrylamide gels for degradation. All Version II RNAs were resuspended at approximately 0.5pmol/ul. A 200ul transcription reaction yielded about 80ug of gel purified usable RNA.

2.2 R2Bm Protein Expression Construct Rationale

A past problem with R2Bm *in vitro* experiments has been that the purified R2Bm protein was contaminated with 5' PBM RNA. The mRNA generated by the original R2Bm protein expression construct contains the 5'PBM sequence as it is internal to the ORF; the first AUG codon being prior to the 5' PBM. The expressed protein binds the 5' PBM in cis and is largely protected from RNase digestion, copurifying with the R2Bm protein.[4] The remainder of the R2Bm RNA is either digested or sheared off during purification. The contaminating 5' PBM RNA inhibits the ability to form and analyze RNP complexes and their activities with the mini RNAs. This is not seen with the 3' PBM as it is not part of the ORF.

Recent data comparing the ORF and 5'PBM structure of four additional R2 elements (from silk moths related to *B. mori*) indicate the start of the R2 protein likely between stem loops III and IV of the PBM, not at the first AUG of the ORF (FIGURE 10).[33] It is hypothesized that translation of the R2 RNA

may involve an internal ribosome initiation mechanism, requiring the 5' PBM, or at least the psuedoknot.[33] It was hypothesized that by removing the area of high RNA structure conservation but low amino acid

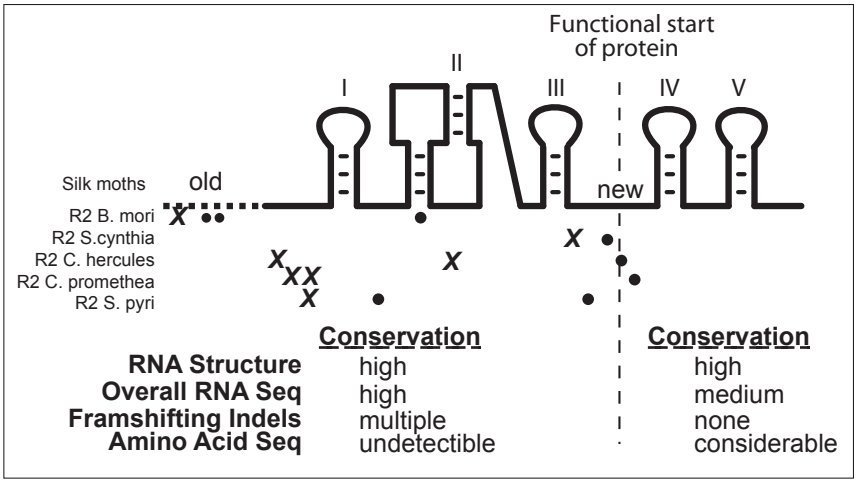


Figure 10: Structure of 5' PBM RNA and the predicted start site of functional R2 protein. The 5' PBM RNA is represented by the solid black line. Dotted line shows a portion of upstream sequence. Conserved RNA secondary structures are numbered with roman numerals. The pseudoknot is number II. Dots are in-frame methionines found in R2 elements (D clade) from five related moth species. X's are in-frame stop codons. The predicted start site of functional R2Bm protein, based on phylogenetics and RNA structure considerations, is marked. The 5' PBM is divided into two parts. Conservation of RNA structure and sequence dominates the first segment. Amino acid sequence becomes an important factor in the second segment. R2 is hypothesized to use an internal ribosome initiation, and based on this conservation information, it is further hypothesized that the 5' PBM may function as such. The start position of old R2Bm protein expression construct (old) and new construct (new) are marked. Adapted from Kierzek et al. (2009) J Mol Biol.

conservation (stem loops I-III from the protein expression construct), the problem of co-purifying the 5' RNA with the protein would be eliminated. Another aspect of my research, was to create an uncontaminated protein to bind to the mini RNAs in order to observe second strand synthesis.

Another approach to the 5' contamination problem was purchasing a codon-optimized plasmid containing the R2Bm sequence. Codon-optimized sequences are designed to produce a protein using the most common codons for a given amino acid in a particular organism (bacteria in this instance). This ensures a perfect protein that is made more efficiently. Because the codon optimized mRNA sequence is not the same as native R2 mRNA, the particular PBM structures will not form, thereby not binding to the protein during translation, and not interfering with subsequent binding reactions using the mini RNAs. Hence, a codon-optimized protein was purchased and used in my research to complex with mini RNAs and potentially observe second strand synthesis. The codon-optimized amino acid sequence is identical

to the wild type (WT) R2Bm, so all functions should be equal to that of WT R2Bm. An endonuclease (KPD) mutant version of this protein was also made. The endonuclease mutant was made as a means to understanding initial binding reactions as opposed to full integration events. Using a site mutagenesis kit (Stratagene #200523-5), a mutation (changed the D to an A (Alanine)) was made in the endonuclease region of the codon-optimized (CDN-B WT) R2Bm protein. This endonuclease mutant was named CDN-B KPD(-) R2Bm and was cloned into XL-1 Blue electrocompetent cells (Stratagene #200228).

2.2.1. Methods: Generate R2 Protein Without 5' Contaminant

Two new protein expression constructs were made. One was designed remove the area encoding most of the 5' PBM and to begin translation at the site predicted by the structural and phylogenetic data. In addition, a codon-optimized construct was purchased.

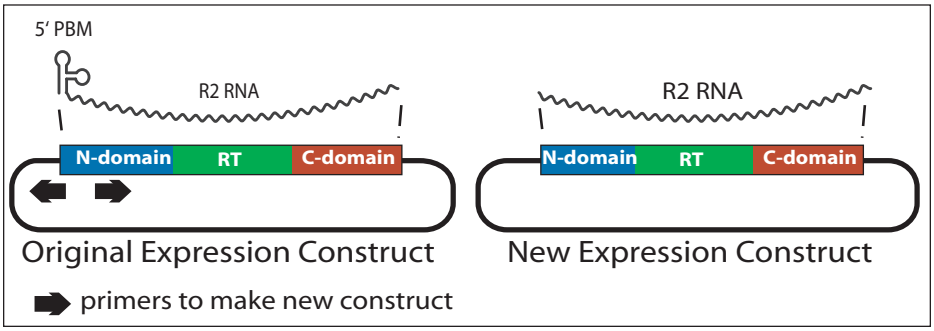


Figure 11. R2 protein lacking 5' PBM structure.
High fidelity PCR was used to make an R2Bm protein expression construct lacking the 5' RNA.

The first R2Bm protein construct excludes a portion of the N-terminal domain that contains the 5' PBM sequence. This was accomplished through PCR (FIGURE 11). The R2Bm sequence lies within a PUC18 vector. Primers were designed just outside the 5' PBM region and included a six histidine (HIS) tag on the amino-terminal end of the ORF. PCR reactions were carried out using Pfx DNA polymerase (Invitrogen #11708-013). This new plasmid containing the new Delta 5r HIS R2Bm protein construct was transformed into JM109 competent cells (Promega L1001). These cells were grown at 37°C in LB broth (Trypton, Yeast, NaCl) until the optical density at 600nm was 1.0, then Isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 0.3mM, and the cultures were incubated for 1 hour. Cells from the 600ml culture were harvested by centrifugation, washed in cold 50mM Tris (pH 7.5), and collected by centrifugation. The cell pellets were resuspended in 2.5ml buffer A (100mM Tris pH 7.5,

5mM EDTA, and 50% glycerol) and could be stored at -80°C. All subsequent procedures were conducted at 0-4°C. The cell pellet in solution A was incubated on ice for 30 minutes in 5mM DTT, 2mM Benzamidine-HCl, and 2mg/ml lysozyme, being sure to warm the mixture with hands every 10 minutes. Then 13.2 mls of buffer B (100mM Tris pH 7.5, 1M NaCl, 5mM DTT, 0.2% Triton X-100, 10mM MgCl, 2mM Benzamidine-HCl) was added to the cell lysate and incubated on ice for 30 minutes. The lysate was centrifuged in a T865 rotor in an XL-90 Ultracentrifuge (Beckman) at 33,000 rpm for 20 hours at 2°C. The HIS tag ended up not being used to purify the protein as the tag ended up being buried (presumably) within the protein. Therefore the classical purification method was used.[22] The supernatant was diluted with dd water and 2mM DTT to lower the salt concentration to 400mM. The diluted extract was loaded onto a 7 ml Q Sepharose fast-flow column (Sigma-Aldrich #Q1126) equilibrated in 250mM NaCl buffer C (25mM Tris pH 7.5, 2mM DTT, 0.01% Triton). The column was washed with 10 ml of 250mM NaCl buffer C, and Delta 5r HIS R2Bm protein was eluted with 600mM NaCl buffer C. Fractions containing the Delta 5r HIS R2Bm protein endonuclease activity were pooled, dialyzed against 300mM NaCl buffer D (25mM Tris pH 7.5, 2mM DTT, 0.01% Triton, and 10% glycerol) and applied to a 1 ml DNA-cellulose (USB #US14394) column equilibrated with 300mM NaCl buffer D. The column was washed with 9 ml of 300mM NaCl buffer D and eluted with 800mM NaCl buffer D. Fractions containing the R2 endonuclease were pooled, dialyzed against R2 storage buffer (25mM Tris pH 7.5, 100mM NaCl, 2mM DTT, 0.01% Triton, and 50% glycerol) and 1X BSA (NEB #E9001S)) and 0.01% Triton was added to the Delta 5r HIS R2Bm protein and stored at -20°C. The new protein was free of contaminating 5' PBM and appeared to be fully functional and was at a concentration of approximately 0.27pmol/ul (~32ng/ul).

The second protein expression construct was purchased from DNA 2.0 and was expressed in BL21 cells. In the construct CDN-B WT (wild type) R2Bm, a carboxyl-terminal HIS tag was included which allowed for purification over a talon resin metal affinity column. Mutant versions of the protein were purified in the same manner. BL21 cells containing the CDN-B construct were grown and harvested in a manner similar to that described for Delta 5r HIS R2Bm protein (discussed above) with some minor changes. The cells were grown to an optical density of 0.6 at 600nm and induced with IPTG at a concentration of 0.03mM. Cells from the 600ml culture were harvested by centrifugation, washed in cold 50mM Tris (pH 7.5), and collected by centrifugation. The cell pellets were resuspended in 2.5ml buffer A

(100mM HEPES, 5mM EDTA, and 50% glycerol) and could be stored at -80°C. All subsequent procedures were conducted at 0-4°C. The cell pellet in buffer A was incubated on ice for 30 minutes in 5mM β-Mercaptoethanol (BME), and 2mg/ml lysozyme, being sure to warm the mixture with hands every 10 minutes. Then 13.2 mls of buffer B (100mM HEPES, 1M NaCl, 5mM BME, 0.2% Triton X-100, 10mM MgCl) was added to the cell lysate and incubated on ice for 30 minutes. The lysate was centrifuged in a T865 rotor in an XL-90 Ultracentrifuge (Beckman) at 33,000 rpm for 20 hours at 2°C. The extract was loaded onto a Talon resin (Clontech #635501) metal affinity column equilibrated with 3 mls buffer C (50mM HEPES, 0.05mM ZnCl2, 0.2% Triton, 2mM BME) containing 5mM Imidizole and 500mM NaCl. The column was washed with three different washes with increasing Imidizole concentration. First, 1ml of wash 1 (Buffer C plus 500mM NaCl and 10mM Imidizole). Second 1 ml of wash 2 (buffer C plus 500mM NaCl and 20mM Imidizole). Third, 600ul of wash 3 (buffer C plus 300mM NaCl and 40mM Imidizole). The CDN-B KPD(-) R2Bm protein was eluted off the column with 300ul of 300mM and 150mM Imidizole buffer C. For storage at -20°C, 300ul of 100% glycerol (final concentration=50%), 1X BSA, and 0.1% Triton was added to the eluate. The new protein was at a concentration of approximately 0.33pmol/ul (~40ng/ul).

To further purify the R2Bm protein, a size exclusion column was purchased (HiLoad 16/60 Superdex 200 prep grade) (GE Healthcare #17-1069-01). The column was equilibrated with 3 column volumes of Column buffer (50mM Tris, 100mM Nacl, 1mM DTT) at a rate of 0.5ml/minute at 4°C. The 500ul of eluate that came off of the Talon resin column was loaded onto the column. Two column volumes of buffer were washed over the column and fractions were collected. The fractions were analyzed by Sypro Orange (Sigma Aldrich #S5692-500ul) in a black 96-well plate (Costar Corning #3915) on the Synergy 2.0 BioTek spectrophotometer and analyzed by the Gene 5 software from BioTek. The wells that contained the most protein signal were concentrated by size exclusion columns (Millipore #42424). The protein was stored in 50% glycerol, 1X BSA, and 0.01% Triton at -20°C. A 10% Tris glycine gel was run to confirm quantity and correct protein size. Gels contained a titration of BSA protein standard (Bio-Rad #500-0206) to quantify samples and SDS Broad Range PAGE ladder (Bio-Rad #161--0317) for comparing protein size. This procedure was done for the CDN-B KPD (-) R2Bm protein as well as the CDN-B WT R2Bm protein.

2.3 Reconstituting Second Strand Synthesis *In Vitro*: Proof of Concept Experiment

Very preliminary integration experiments were conducted in which a set of unpurified mini R2Bm RNAs (Version I) (See section 2.2) and unpurified Delta 5r R2Bm protein (see section 2.3.1) were generated and crude experiments were performed as proof of principle (FIGURE 12). Protein subunits bound to mini RNA at different protein:RNA ratios were placed directly into *in vitro* TPRT reactions (bypassing RNP purification, section 2.4) containing an excess of 80bp target site DNA radiolabeled with P^{32}. The results were assayed by denaturing gel electrophoresis and PCR. The mini R2 RNA used in FIGURE 12 consisted of half the 5' UTR, the 5' PBM, a 400nt spacer, and the 3' PBM. As expected, TPRT is most abundant when the RNA concentration is in excess of the protein (FIGURE 12 panel A). The second panel of A of FIGURE 12, tracks the top strand reactions and shows that top strand cleavage is most efficient when the protein concentration is in excess of RNA, at ratios near 2:1. This observation is in stark contrast to cleavage observed in the presence of only the 5' PBM, where the RNA concentration had to be 10 fold higher than the protein concentration in order to achieve maximal cleavage.[4] Nevertheless, achieving efficient top strand cleavage is a significant step forward in reconstituting the integration reaction *in vitro*.

Second strand synthesis appears to be occurring in FIGURE 12, but inefficiently. PCR analysis of the middle reactions yielded 5' junctions consistent with second strand synthesis. Three of the junctions are shown in FIGURE 12 B. One junction corresponded to a full mini RNA being integrated along with four untemplated nucleotides. A second junction corresponded to a 5' truncation. These first two junctions are consistent with natural insertions. The remaining clone indicates that any nearby 3' OH can be used to prime second strand synthesis. Second strand synthesis in this clone was primed from the 3' OH located at the terminal end of the target DNA (top strand). Untemplated nucleotides and 5' truncations are also observed in these "terminal-end second strand synthesis" products. "Terminal-end second strand synthesis" dominates when there is more RNA than protein and undoubtably contributes to the high molecular weight strand smear observed in FIGURE 12 top strand data. It is important to stress that although the terminal-end second strand synthesis products used the wrong 3'-OH, they are the result of bona fide "second strand synthesis" activity by the R2Bm protein. This experiment represents real progress toward achieving integration *in vitro*. In order to make this reaction more efficient, it is

proposed that purified RNPs will allow complete integration reactions to occur with greater efficiency for study. My research continues by using the mini RNAs and R2Bm protein produced to form functional RNPs in order to more efficiently see second strand synthesis.

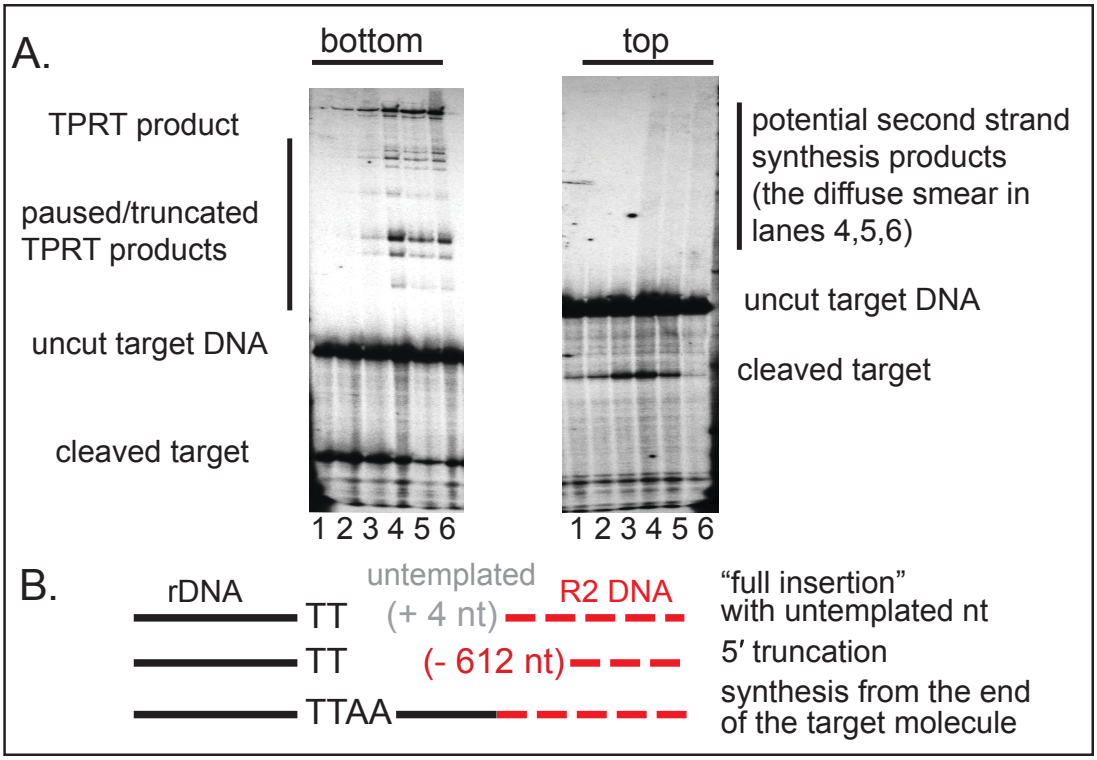


Figure 12: Analysis of transposition intermediates generated using the new R2 protein. **A.** Denaturing polyacrylamide gel analysis. DNA and protein were held constant while the mini RNA was titrated. DNA was in eight fold excess of protein concentration. The bottom strand is P32 end labeled in the first panel, allowing detection of bottom strand cleavage and TPRT. The top strand is labeled in the second panel, allowing detection of top strand cleavage and second strand synthesis. In each panel the RNA concentration was varied to yield protein: RNA ratios beginning in lane 1 and ending in lane 6 of 50:1, 10:1, 2:1, 1:2, and 1:50. **B.** 5' Junctions of possible second strand synthesis products. PCR products obtained by amplifying the DNA from reactions corresponding to the middle lanes were cloned and sequenced. Three representative clones are given. Black line is sense strand of the 28S rDNA target site. Top strand cleavage occurred after the two T residues. Red dashed line is the sense strand of the mini-RNA (i.e., the DNA complement of the cDNA).

2.4 RNP Complexes Rationale

2.4.1. Methods: Formation and purification of RNPs

According to my preliminary data (section 2.3), FIGURE 12 suggests that second strand synthesis is indeed occurring at very low levels when using low yield unpurified preparations of mini-RNA RNPs. The low activity is most likely due to competing dead-end activities from non-integration competent RNPs

and proteins not complexed with RNA. My next goal was to purify RNPs and use them in full integration reactions.

To this end, mini RNAs (section 2.1.2) were complexed with R2Bm protein (section 2.2.1) at different protein to RNA ratios. The resulting RNPs were separated on a sucrose gradient. RNPs formed on radiolabeled (Version I) RNAs were tracked by running aliquots of the sucrose fractions on slot blots and exposing to film. RNPs formed on unlabeled RNA (Version II) were tracked by assaying aliquots of sucrose fractions with Ribogreen stain (Invitrogen #R11490) and analyzing on a spectrophotometer (Synergy 2.0). "Dimeric" RNPs (two R2Bm protein subunits bound to a single RNA) were expected to dominate at high protein to RNA ratios (>>2:1), while at low protein to RNA ratios, "monomer" RNPs (a single R2Bm protein subunit bound to either the 3' or 5' PBM moiety) were expected to form.

2.4.1. Methods: Formation and Purification of RNPs

2.4.1.1. Binding Reactions With Version I Mini RNAs

Binding reactions consisted of mini RNA and a titration of protein. Binding reactions were composed of 1X TPRT buffer (3mM MgCl2, 100mM NaCl, 10mM Tris pH 7.5, 1mM DTT), 40U RNase Inhibitor (Promega), ~0.3pmol of Version I RNA, and a titration of protein at ratios of protein to RNA (PRO:RNA) of 0:1, 1:15, 1:5, 1:1, 5:1, and 15:1. Total glycerol concentration was kept at ~10% by adding R2 complete storage buffer (100mM NaCl, 25mM Tris pH 7.5, 50% glycerol, 0.01% Triton, and 1X BSA) to reactions with little or no protein. Stoichiometry of the RNA and protein complexes were calculated by gel electrophoresis of the starting material next to mass standards. Gels were stained with sybr green (to see RNA) or sypro orange (to see protein) accordingly. Reactions were incubated at room temperature (22-25°C) for 30 minutes. Then these reactions were loaded onto sucrose gradients (See 2.4.1.3).

2.4.1.2. Binding Reactions With Version II Mini RNAs

Binding reactions consisted of mini RNA and a titration of protein. Binding reactions consisted of 1X TPRT buffer (3mM MgCl2, 100mM NaCl, 10mM Tris pH 7.5, 1mM DTT), 60U SUPERase In RNase Inhibitor (Ambion), ~2.45pmol of Version II RNA, and a titration of protein at ratios of protein to RNA (PRO:RNA) of 0:1, 1:5, 1:1, 5:1, 15:1, and 1:0. Total glycerol concentration was kept at ~11% by adding R2 complete storage buffer (100mM NaCl, 25mM Tris pH 7.5, 50% glycerol, 0.01% Triton, and 1X BSA) to reactions with little or no protein. Stoichiometry of the RNA and protein complexes by gel electrophoresis of the starting material next to mass standards. Gels were stained with sybr green (to see

RNA) or sypro orange (to see protein) accordingly. Reactions were incubated at room temperature (22-25°C) for 30 minutes. Then these reactions were loaded onto sucrose gradients (See 2.4.1.3)..

2.4.1.3. Sucrose Gradients

Sucrose gradients consisted of sucrose solutions ranging from 20%-40% sucrose starting with a step gradient. Each percentage was made separately in 50 ml conical tubes (20%, 21%, 22%, etc. to 40% sucrose) with dry sucrose and 1X TPRT buffer (section 2.4.1.2) (DTT was added immediately before pouring the gradients). A 12ml centrifuge tube (Beckman #331372) was put on ice and each percentage of sucrose was added slowly down the side of the tube beginning with the 40% sucrose solution and working up to 20%. One gradient was poured for each titration of protein (sections 2.4.1.1 and 2.4.1.2). All subsequent procedures were conducted at 0-4°C. The binding reactions (sections 2.4.1.1 and 2.4.1.2) were added to the top of the sucrose gradients, weighed, and centrifuged in an SW 41 rotor in an XL-90 Ultracentrifuge (Beckman) at 32,000 rpm for 19 hours at 4°C. A peristaltic pump and fraction collector were used to collect 500ul fractions of each gradient. These fractions were analyzed according to the RNA they contained.

As fractions were collected from gradients containing Version I RNAs, 300ul of each fraction was run through a slot blot onto Hybond-N+ nylon (Amersham Biosciences #RPN203B). This was then cross-linked by UV-stratalinker 1800, and exposed to film (Kodak BioMax MS #8294985). The film was developed and scanned into the computer and analyzed by imal (quantifying program for Mac).

100ul of each fraction from gradients containing Version II RNA was put on black opaque 96-well plate for analysis by the Synergy 2.0 spectrophotometer. To each fraction on the 96-well plate, 100ul of Ribogreen stain at a final concentration of 0.001 was added. This was then scanned by the Synergy 2 machine and quantified by the program Gene 5. The quantification was then represented in graph form (chapter 3)

CHAPTER 3

RESULTS

3.1 Molecular Tools

3.1.1 Mini RNAs: Versions I and II

To begin, Version I RNA was made. Panel A of FIGURE 13 shows the 3.5% denaturing polyacrylamide gel electrophoresis (PAGE) that was used to band purify the Version I RNAs. It can be clearly seen that there is an abundance of degradation (smear extending below the cut out band). It is possible that a portion of the smear is due to truncated versions of the transcript, but it is not predicted that this would account for so much smear. It cannot be seen well in FIGURE 13 A, but the area cut out was not a true band, indicating very few actual full-length RNAs were produced during transcription. Panel B of FIGURE 13 shows a 0.8% agarose gel of the band purified Version I RNAs. These RNAs are estimated to be the correct size, but seem to have some degradation and possibly contain paused transcripts (the smear extending below the band (FIGURE 13 B)). These RNAs were also low in concentration (section 2.1.2). It seemed that the correct RNAs had been generated, but were experiencing some sort of RNase contamination. After trying several RNP and sucrose gradient experiments with this degraded RNA to no avail (section 3.2), two new protocols were implemented. The first new protocol concerned lab maintenance. All solutions within the lab were emptied and all glass-ware was throughly cleaned and autoclaved. All RNase containing experiments (plasmid preparations and LB media) were extracted from the lab and moved to a separate location. New protocols were put in place concerning LB and culture growing containers, for example all LB containing glass-ware and plastic-ware was washed and stored separately from non LB containing lab-ware. New reagents were made that were strictly used for RNA work only. RNaseZap was purchased and used accordingly. RNase-, DNase-, Nuclease-free water was purchased for use with RNA only. In addition to the standard protocol of wearing gloves and a lab coat, a mask was worn at all times when working with RNA. The second protocol change concerned RNA making. Version II RNA protocol was implemented in which all

new reagents (5x transcription buffer, rNTPs, T7 polymerase, RNase Inhibitor) were purchased from a different company (Fermentas) than had previously been used. A whole new protocol in terms of amounts of MgCl2, T7 polymerase, and template DNA was used. Also the RNAs were made from plasmid DNA template. Version II RNAs can be seen in FIGURE 14.

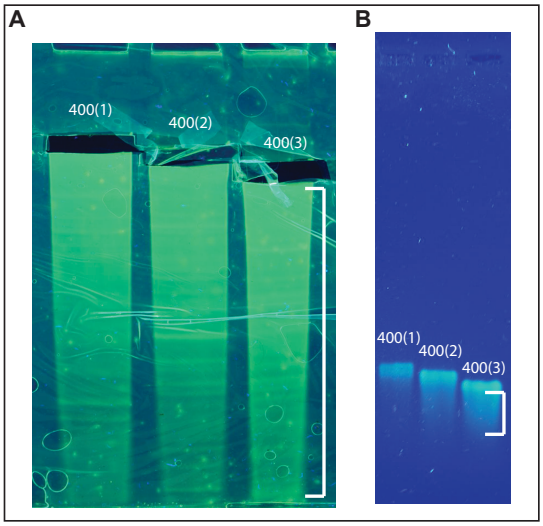


Figure 13: Version I Mini RNAs
Panel A is a 3.5% denaturing PAGE stained with Syber Green in which mini RNAs were cut out. Panel B is a 0.8% agarose gel stained with sybr Green and shows the mini RNAs after band purification (panel A). Brackets indicate mix of paused transcripts and degradation.

Panel A of FIGURE 14 shows a 3.5% denaturing PAGE that was used to purify the Version II RNAs. It can be seen that there is a certain amount of smear under the cut out areas. In this case, the smear is most likely due to paused transcripts, because when compared to FIGURE 13 Panel A, there is much less smear considering the immense increase in concentration of RNA contained within each lane in FIGURE 14 A. It cannot be seen in FIGURE 14 , but the area that was cut out was actually a band, unlike in Version I RNA, which suggests most of the transcripts are the correct size. It can also be seen in FIGURE 14 A that a second band appears under all the RNAs. This is due to an unnoticed T7 promotor internal to the pENTR vector. For future experiments, this internal promotor will be mutated or a different vector will be used. FIGURE 14 Panel B shows gel purified RNA on a 3.5% denaturing PAGE stained with Syber Green and it is clear that these RNAs are not degraded. The slight smear under the two longest RNAs (FIGURE 14 B) are most likely due to capturing a portion of the paused transcripts seen in Panel A (FIGURE 14). In the case of 400(3) RNA of FIGURE 14 A, only the top half of the large band was

extracted from the purifying gel. When this RNA was run on the gel seen in FIGURE 14 B, it shows a much crisper band than the two other RNAs, in which a much larger section was taken during gel purification. The new protocols and extreme measures appeared to pay off as the RNAs are the correct size, have very little degradation, and are far more concentrated than Version 1 RNA (section 2.1.2). Because of the increase in concentration it was unnecessary to radiolabel the Version II RNAs as there was enough RNA to been seen in a gel or by spectrophotometer after sucrose gradients. In summary, 3' and 5' PBMs were linked, differing lengths of the 5' end were made, one particular linker length was made (400nt), and mini RNAs were pure, much less degraded, and highly concentrated.

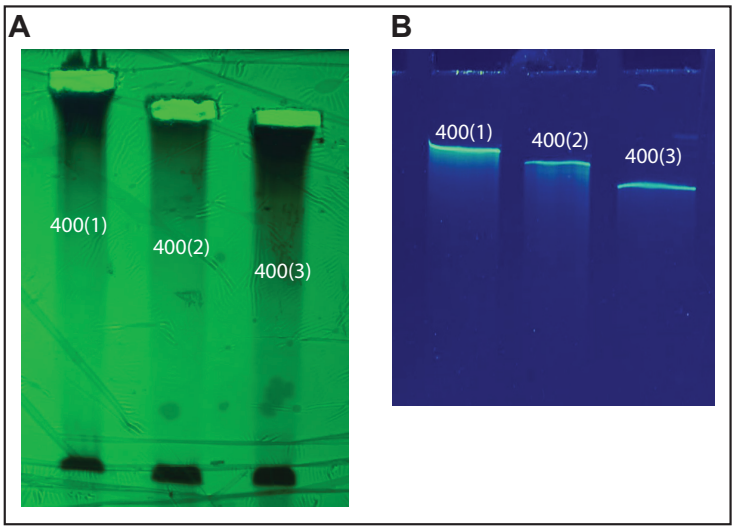


Figure 14: Version II Mini RNAs
Panel A is a 3.5% denaturing PAGE visualized by UV-shadowing with short wave UV after bands had been cut out. Note second band seen under each of the RNAs. This is a transcript that is due to pENTR having a T7 promoter internal to the vector. Panel B is a 3.5% denaturing PAGE stained with sybr Green and shows the mini RNAs after band purification (panel A).

3.1.2 R2Bm Protein: Delta 5r HIS R2Bm, CDN-B WT R2Bm, and CDN-B KPD (-) R2 Bm

FIGURE 15 shows Delta 5r HIS R2Bm, CDN-B KPD (-), and CDN-B WT proteins on 10% Tris Glycine PAGE stained with Comassie Brilliant Blue R-250 (Amresco #0472-10G) in Panel A and Sypro Orange in Panel B. All proteins seem to be the correct size (R2Bm=120kD).

FIGURE 16 shows binding capability of the Delta 5r HIS R2Bm versus the WT R2Bm. WT R2Bm is purified in the same manner as the Delta 5r HIS R2Bm. This figure shows that when the Delta 5r HIS R2Bm protein is bound to either the 3' or 5' PBM, it will bind and shift target DNA. The Delta 5r HIS

R2Bm protein's ability to bind and shift target in the presence of the 3' and 5' PBM seems comparable to the WT R2Bm protein. FIGURE 17 shows that Delta 5r HIS R2Bm protein cleaves target plasmid approximately the same as WT R2Bm. These gels are assays that are done during the protein purification process and protein concentrations are only approximately the same between gels. With these assays accomplished, RNPs were formed.

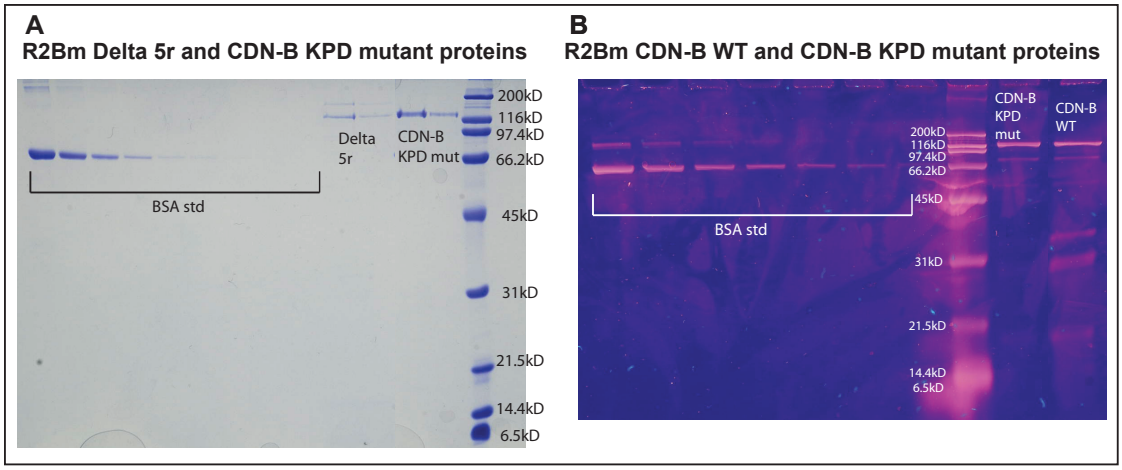


Figure 15: Versions of R2Bm Proteins on 10% Tris Glycine SDS PAGE.
BSA standard titrates by two fold dilution (2ug-62ng). Panel A is stained with Comassie Blue. Panel B is stained with Sypro Orange.

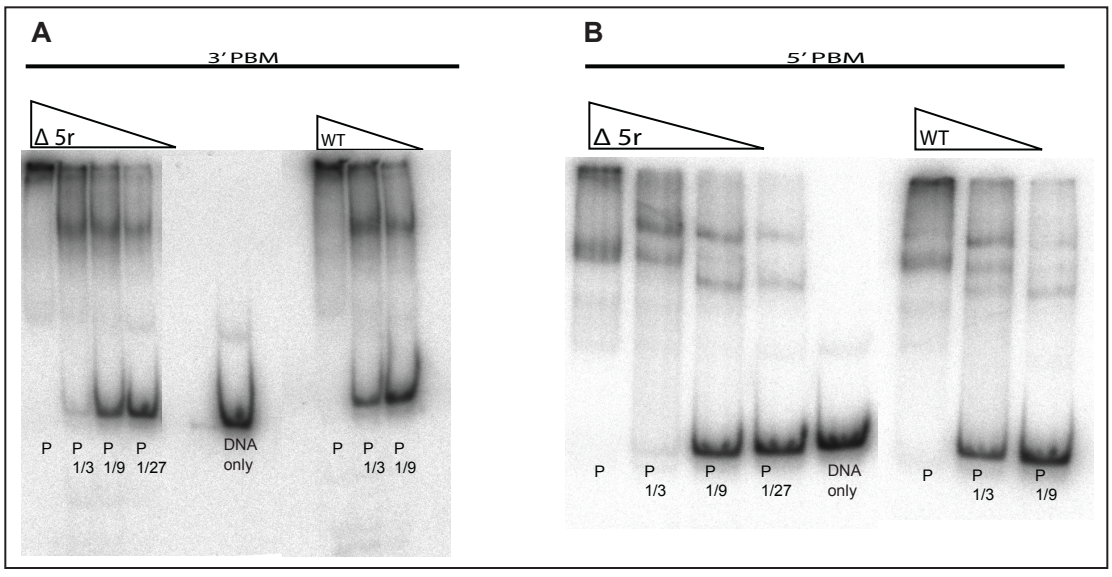


Figure 16: Native 5% TBE Bandshift gels
A 120bp DNA target with top strand radiolabeled was used. Delta 5r R2Bm (Δ 5r) protein and Wild Type R2Bm (WT) protein were used. Protein was titrated by thirds. P=straight protein from prep. Panel A shows reactions complexed with the 3' PBM structure. Panel B shows reactions complexed with the 5' PBM structure. Picture taken by phosphoimager.

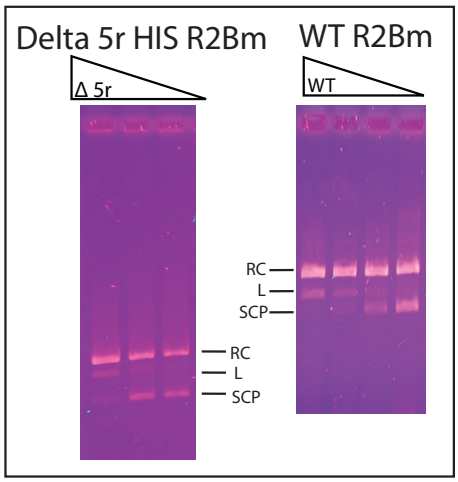


Figure 17: DNA cleavage gels
Lanes contain a portion of fractions collected during purification of protein and plasmid DNA containing target site. These assays are used during purification of the protein and protein concentrations are only approximately the same. 1% agarose gels stained with EtBr. Abbreviations: relaxed circle plasmid (RC), Linerized plasmid (L), super-coiled plasmid (SCP), triangles show concentration of protein.

3.2 RNP Complexes

3.2.1 Slot Blot Graphs

The first RNPs to be made used radiolabeled Version I mini RNAs and Delta 5r HIS R2Bm protein. As stated above, the fractions collected from these gradients were analyzed by slot blots. FIGURE 18 shows film developed from slot blots of these RNP reactions after sucrose gradients. These were then quantified and are graphically represented in FIGURE 19. FIGURE 20 depicts the values from FIGURE 19 normalized. From both the slot blot and the normalized graph (FIGURE 20) it is evident that RNA concentration was shifting to higher sucrose percentages based on increasing amount of protein in the reactions, suggesting the protein was binding to the RNA. In the RNA only blot, the RNA seems to peak around fraction 13 and in the protein:RNA ratio of 15:1, the RNA seems to peak around fraction 7. It is presumed that the RNA contained within the peak fraction of the protein:RNA sucrose gradient of 15:1 is forming pseudo dimers. Ideally, it was expected to see pseudo dimers at a ratio of 2:1, protein:RNA, so seeing a shift at ratios of 15:1 is a little high, but we have discovered that NaCl and $MgCl_2$ concentrations influence these ratios (data not shown). Monomers may be appearing in the gradients where protein to RNA ratios are 5:1 as the RNA peaks around fraction 10 which is a higher sucrose percentage than the

RNA only gradient. It is also very clear from the slot blot picture and the graph (FIGURE 19) that much of the RNA was degraded as observed by the broad peak (signal in every fraction) of the RNA and by the fading of RNA signal in gradients with higher concentrations of protein. The same amount of RNA was put in all reactions, so a fading of signal would suggest a degradation issue.

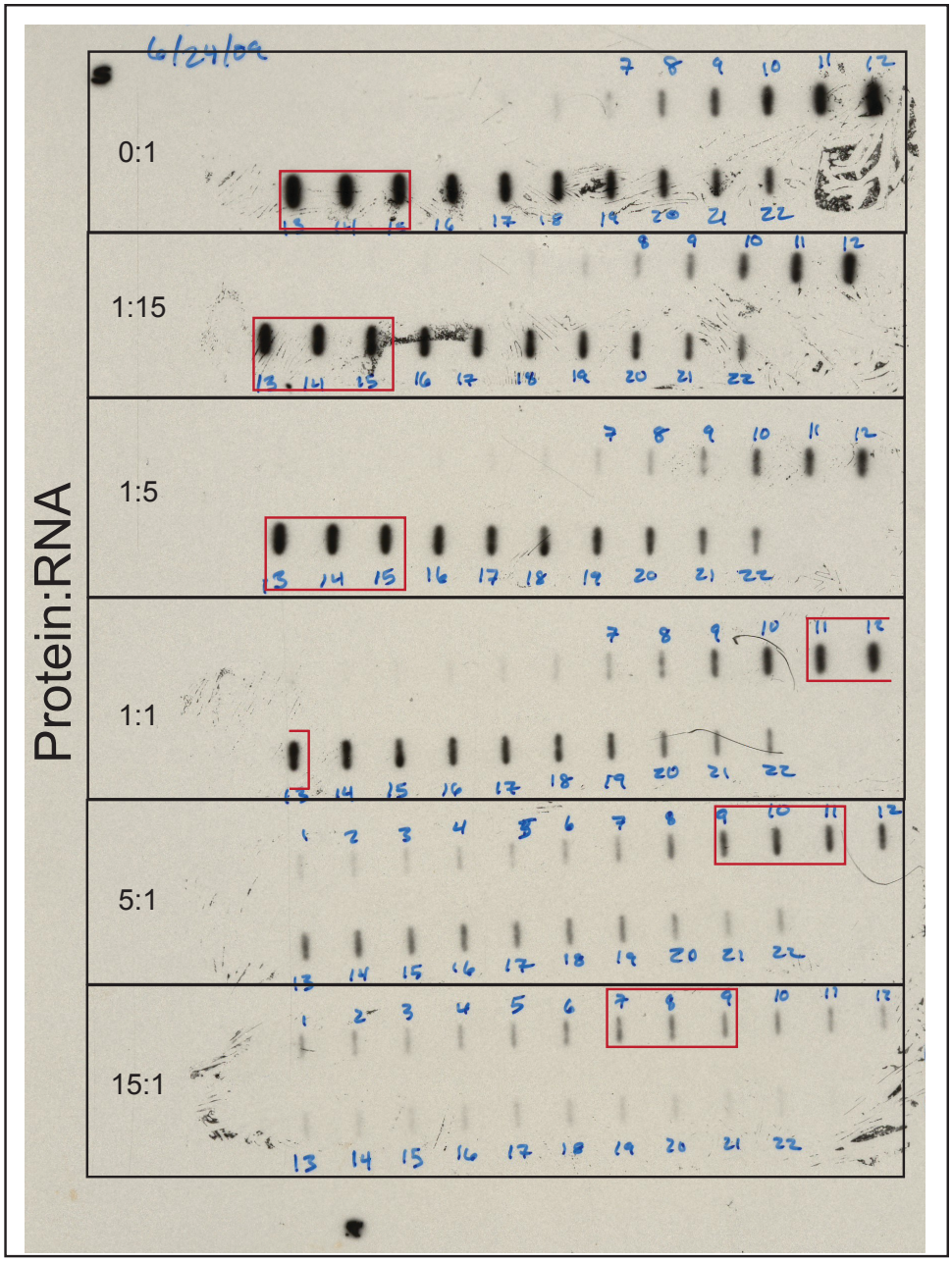


Figure 18: Slot blots of sucrose gradients using Version I RNA and Delta 5r HIS R2Bm protein. Protein:RNA ratios are 0:1, 1:15, 1:5, 1:1, 5:1, and 15:1. Mini RNA 400(1) was used. Black boxes indicate boundaries of each slot blot. Red boxes indicate fractions containing the most RNA relative to the other fractions.

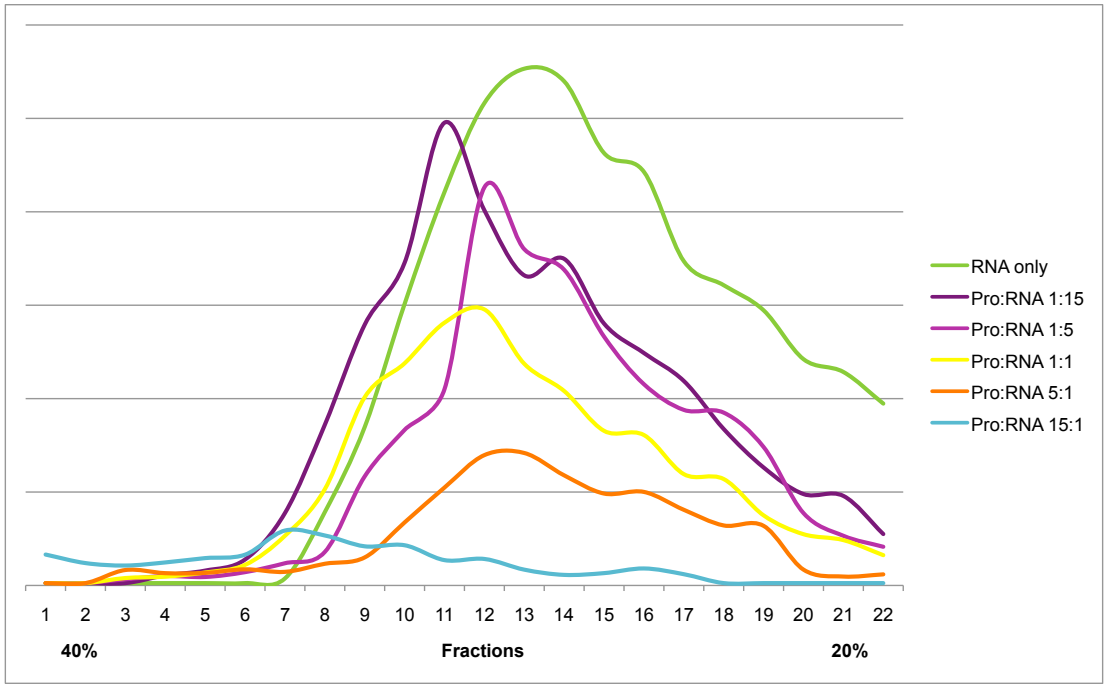


Figure 19: Sucrose gradient data.
Version I RNA and Delta 5r HIS R2Bm protein (graphical representation of Figure 18).

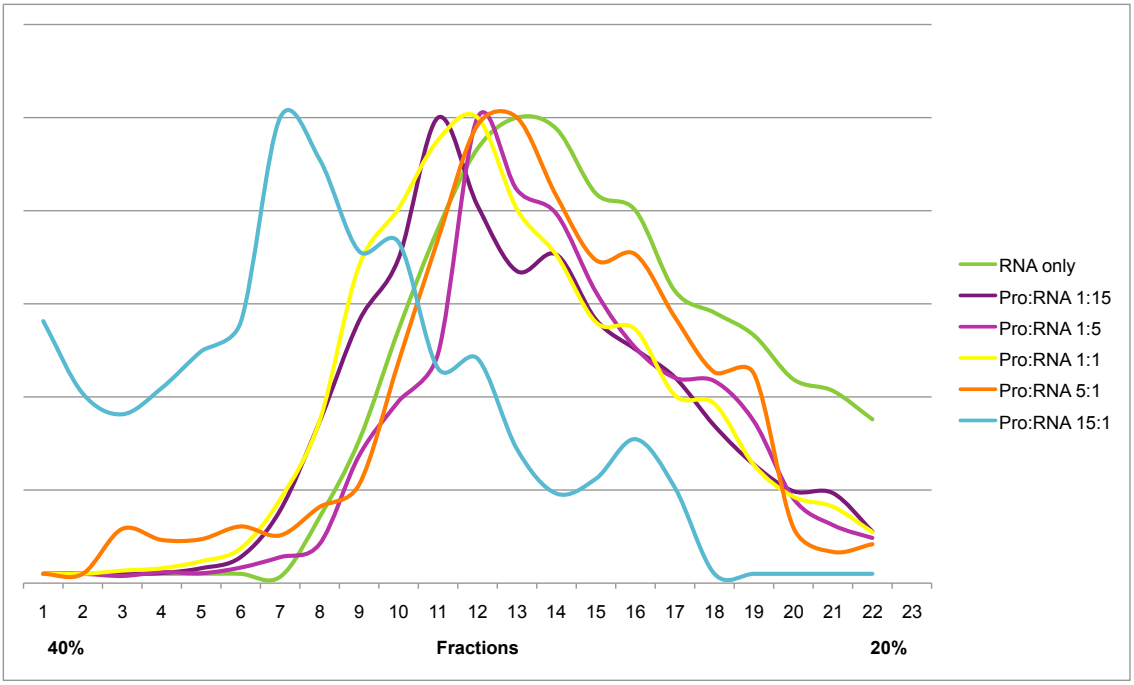


Figure 20: Sucrose gradient data normalized
Version I RNA and Delta 5r HIS R2Bm protein (normalized graph of data from Figure 19).

After many attempts at obtaining gradients that did not have such RNA degradation, the protocols were put in place for the making of Version II RNAs (section 2.1.2). It was also predicted that the protein may be contaminated with RNases as more degradation of RNA seemed to be occurring in RNP reactions that contained higher concentrations of protein. The CDN-B WT R2Bm was purchased, expressed, and purified. Because this protein was purified in a completely different manner than the Delta 5r HIS R2Bm, it was hoped that the previous RNase contaminants would not be a factor.

In the next set of experiments, Version II RNAs were used and the new CDN-B KPD(-) R2Bm protein was used. It was unnecessary to radiolabel the RNA as we were able to use stainable quantities of RNA, so slot blots were not used. Instead analysis by spectrophotometer was employed and FIGURE 21 shows the graphical representation of the data collected. The RNA appeared to be of better quality as it formed a relatively narrow peak in the RNA only gradient (green line). When the data from FIGURE 21 is normalized (FIGURE 22), it can be seen that the RNA is again binding to the protein and shifting to higher sucrose percentages. The protein:RNA ratio of 1:5 seems to peak at a slightly higher sucrose percentage (FIGURE 22) and again the greatest shift occurs in the gradient containing protein:RNA ratio of 15:1. This suggests that pseudo dimers were again being formed. It is interesting to see that in the protein:RNA ratio of 5:1 a shift below the RNA only peak is seen. It is predicted that the more protein that is added, the more binding would occur, and so the more likely that this fraction would contain more dense complexes that would shift the RNA concentration to higher sucrose percentages. The fact that this gradient has a shift below the RNA only peak is very strange and currently it is unknown as to why this fraction does not look as it is predicted. But the fact that all gradients that contain protein seem to decrease in intensity (FIGURE 21), shows that the RNA was degrading again. It seemed again that as more protein was added to the RNA, more degradation was observed.

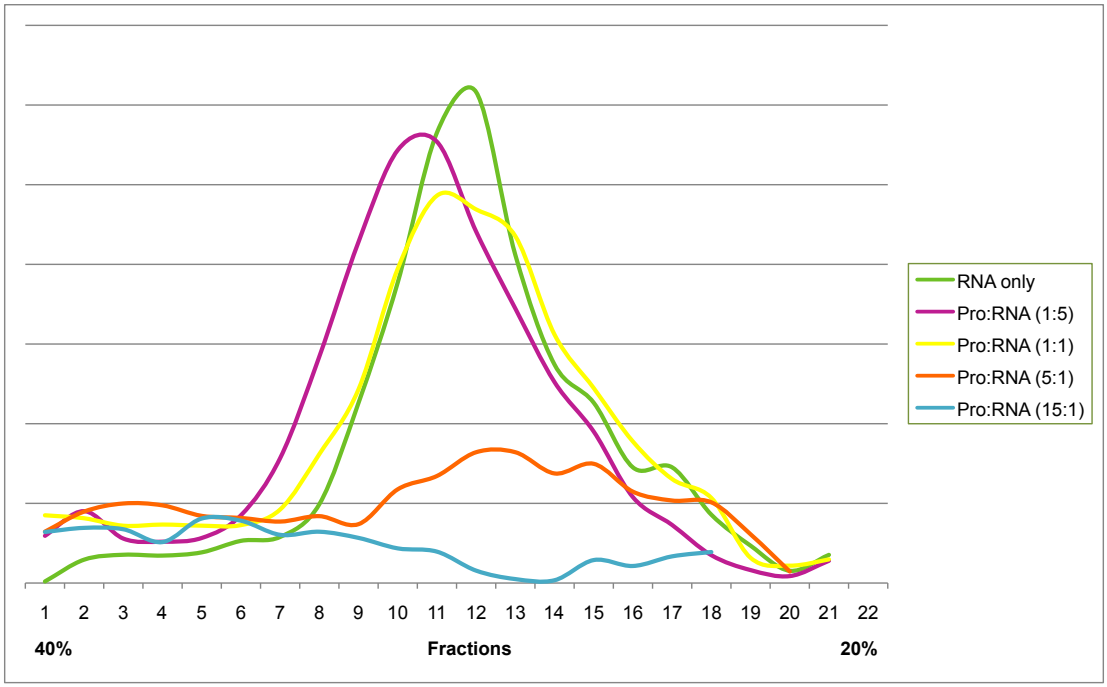


Figure 21: Sucrose Gradient Data
Version II RNA and CDN-B KPD (-) R2Bm Protein

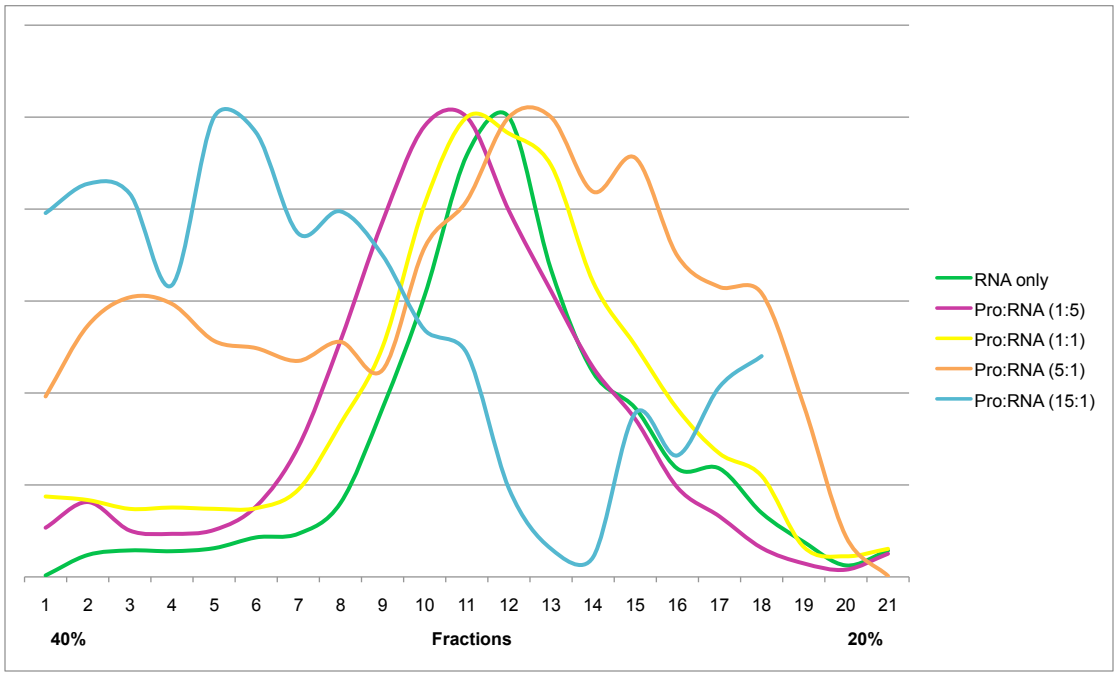


Figure 22: Sucrose gradient data normalized.
Version II RNA and CDN-B KPD (-) R2Bm protein

It seemed evident that an RNA degrading contaminant was still purifying with the protein. A large size exclusion column was purchased and the CDN-B KPD(-) R2Bm and the CDN-B WT R2Bm proteins were expressed, purified over talon resin column, and then run through the size exclusion column. The column was designed to separate proteins based on molecular weight and shape. R2Bm is a 120kD protein and most RNases are ~30kD. It was predicted that the R2Bm protein would elute off of the column before the RNases. A 10% Tris Glycine PAGE was run of the proteins before and after the size exclusion column (FIGURE 21). It is clear from FIGURE 23 that the correct protein was seen before it was run through the exclusion column but not after. The lanes containing the supposed protein after the size exclusion large column do contain the minor variant bands (contaminants), but it does not seem to contain the 120kD band as before the column. These “after column” proteins have yet to be tested in binding or cleaving assays so it is not clear what they are. It is suspected that the R2Bm protein precipitated out inside the column for unknown reasons.

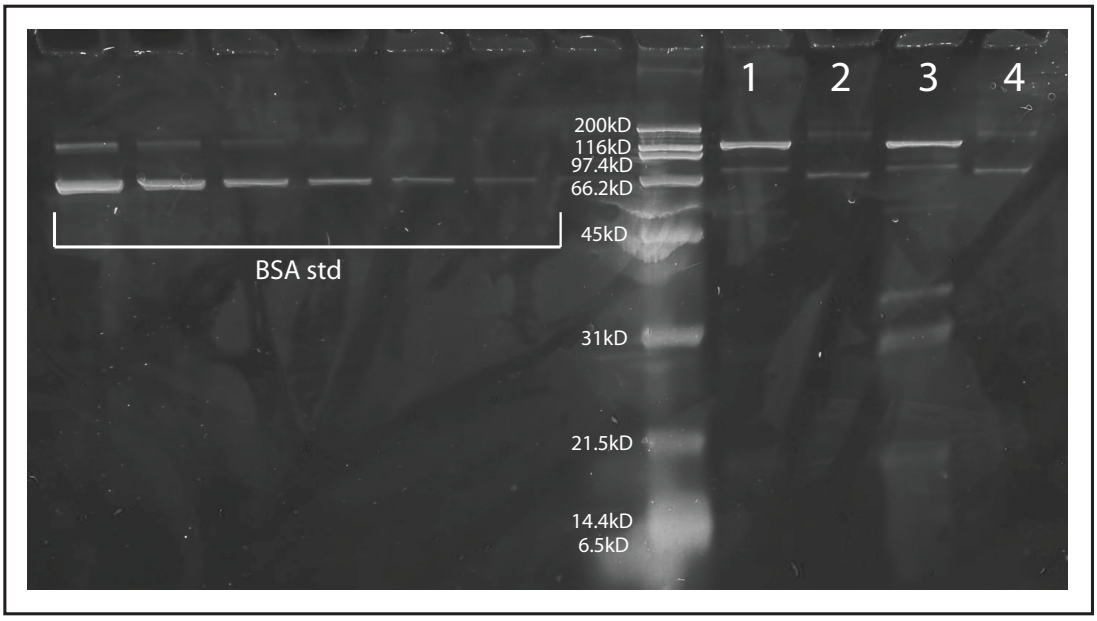


Figure 23: Versions of CDN-B R2Bm proteins before and after large size exclusion column purification.
1=CDN-B KPD(-) before Lg column, 2=CDN-B KPD(-) after Lg column, 3=CDN-B WT before Lg column, 4=CDN-B WT after Lg column. BSA std begins at 2ug and titrates by halves. 10% Tris Glycine SDS PAGE gel stained with Sypro Orange.

CHAPTER 4

DISCUSSION

4.1 Summary

Transposable elements have had a huge impact on genes and genomes of many organisms. The fact that they make up such a large portion of many organisms' genomes and can move throughout a genome relatively on their own, makes them an extremely interesting field of study. So much has been learned about these selfish elements, but so much knowledge remains to be obtained. This is true in the case of R2Bm. Much has been discovered including one DNA binding domain, the reverse transcriptase domain, and the endonuclease domain. The R2Bm protein will bind DNA either upstream or downstream of the target site based on which RNA PBM is bound by the protein. The endonuclease will cut either the top or bottom strand of DNA depending on which RNA PBM is bound by the protein. From this thesis work it is known that long Mini RNAs containing both of the RNA PBMs can be made efficiently and relatively free of contaminants. A functional protein can be produced that does not contain the contaminating 5' PBM by cutting this region out of the protein construct. Differing RNP complexes presumably can be separated by sucrose gradients. True second strand synthesis products have been seen *in vitro*. They were very inefficient when unpurified RNPs formed on poor quality RNAs were used. But with high quality RNA and uncontaminated protein constructs, purified RNPs will be separated and will be used to make second strand synthesis a very efficient *in vitro* reaction.

4.2 Future Direction

Some of the major future steps concerning this projects are to first purify the CDN-B WT and CDN-B KPD(-) R2Bm proteins away from any contaminating RNases. This will be accomplished in a very similar manner as described. Using a size exclusion column of some sort seems to the best option at this time because it is a much more straightforward approach and all the equipment has been purchased. The column was only run twice and some trouble shooting is always necessary when a new protocol is

being used. More time must be given to allow this method a chance to work. It is also possible to dialyze the protein in dialysis tubing that has a molecular cut-off weight similar to R2Bm (120kD). Any small contaminating proteins will simply flow through the membrane, leaving clean R2Bm protein. This was attempted in the lab once and the protein precipitated while in the dialysis tubing. This may have been due to incorrect buffers or too much protein contained within one dialysis tube. It is possible to change the buffer system and possibly split the protein into several tubes, thus diluting it slightly.

Once clean protein is attained, RNP reactions and sucrose gradients will be the next step. Having extremely clean substrates should allow the RNP reactions to be much more efficient and should allow the observation of dimers and monomers within the sucrose gradient to also be very easily seen.

Clearly the next step would be to collect the RNP fractions that corresponded to pseudo dimers and complex them with radiolabeled target site DNA. This reaction would be allowed to incubate for a certain amount of time and then run on a denaturing PAGE, similar to the one used in FIGURE 12.

The next major aspect of this project is the type of target site DNA to use. It was clear from FIGURE 12 that R2Bm can produce “terminal-end second strand synthesis” products. Although these are true second strand synthesis products, they are aberrant products resulting from the R2 protein acting on ends not present when integrating into the rDNA locus on a large chromosome.. It may be possible to eliminate terminal-end second strand synthesis products by using a plasmid DNA that contains the target site as opposed to the linear DNA that was originally used. Another possibility is to use chromatin. It has been shown that the R2Bm target site resides at a phased nucleosome.[38] By using histones wrapped with target DNA a more natural condition would be established for the R2 transposable element to transpose. The next step would involve obtaining target DNA bound to nucleosomes and complexing with pseudo dimer RNPs. Products would be analyzed by PCR as before. Another future direction would be to inject pure pseudo dimer RNPs into flies and then assay by PCR to see full-length integration events occurring *in vivo*.

REFERENCES

1. Christensen, S. M. & Eickbush, T. H. R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Mol Cell Biol* **25**, 6617-6628 (2005).
2. Christensen, S. & Eickbush, T. H. Footprint of the retrotransposon R2Bm protein on its target site before and after cleavage. *J Mol Biol* **336**, 1035-1045 (2004).
3. Christensen, S. M., Bibillo, A. & Eickbush, T. H. Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. *Nucleic Acids Res* **33**, 6461-6468 (2005).
4. Christensen, S. M., Ye, J. & Eickbush, T. H. RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc Natl Acad Sci U S A* **103**, 17602-17607 (2006).
5. Luan, D. D. & Eickbush, T. H. Downstream 28S gene sequences on the RNA template affect the choice of primer and the accuracy of initiation by the R2 reverse transcriptase. *Mol Cell Biol* **16**, 4726-4734 (1996).
6. Yang, J., Malik, H. S. & Eickbush, T. H. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc Natl Acad Sci U S A* **96**, 7847-7852 (1999).
7. Craig, N. L. in *Mobile DNA II* (eds Craig, NL, Craigie, R, Gellert, M & Lambowitz, A. M.) 3-11 (ASM Press, Washington, DC, 2002).
8. Deininger, P. L., Moran, J. V., Batzer, M. A. & Kazazian, H. H. J. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* **13**, 651-658 (2003).
9. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**, 331-368 (2007).
10. Corces, V. G. & Labrador, M. in *Mobile DNA II* (eds Craig, NL, Craigie, R, Gellert, M & Lambowitz, A. M.) 1008-1023 (ASM Press, Washington, DC, 2002).
11. Lutz, S. M., Vincent, B. J., Kazazian, H. H. J., Batzer, M. A. & Moran, J. V. Allelic heterogeneity in LINE-1 retrotransposition activity. *Am J Hum Genet* **73**, 1431-1437 (2003).
12. Kapitonov, V. V. & Jurka, J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* **3**, e181 (2005).
13. Eickbush, T. H. & Jamburuthugoda, V. K. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res* (2008).
14. Kazazian, H. H. J. Mobile elements: drivers of genome evolution. *Science* **303**, 1626-1632 (2004).
15. Curcio, M. J. & Derbyshire, K. M. The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* **4**, 865-877 (2003).
16. Bibillo, A. & Eickbush, T. H. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J Mol Biol* **316**, 459-473 (2002).
17. Moran, J. V. & Gilbert, N. in *Mobile DNA II* (eds Craig, NL, Craigie, R, Gellert, M & Lambowitz, A. M.) 836-869 (ASM Press, Washington, DC, 2002).
18. Burke, W. D., Malik, H. S., Rich, S. M. & Eickbush, T. H. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, Giardia lamblia. *Mol Biol Evol* **19**, 619-630 (2002).
19. Bucheton, A., Busseau, I. & Teninges, D. in *Mobile DNA II* (eds Craig, NL, Craigie, R, Gellert, M & Lambowitz, A. M.) 796-812 (ASM Press, Washington, DC, 2002).
20. Volff, J. N., Korting, C., Froschauer, A., Sweeney, K. & Schartl, M. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J Mol Evol* **52**, 351-360 (2001).
21. Kulpa, D. A. & Moran, J. V. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol* **13**, 655-660 (2006).
22. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595-605 (1993).

23. Eickbush, T. H. & Malik, H. S. in *Mobile DNA II* (eds Craig, NL, Craigie, R, Gellert, M & Lambowitz, A. M.) 1111-1146 (ASM Press, Washington, DC, 2002).
24. Martin, S. L. The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition. *J Biomed Biotechnol* **2006**, 45621 (2006).
25. Williams, M. C., Rouzina, I., Wenner, J.R, Gorelick, R.J., Musier-Forsyth, K., Bloomfield, V.A. Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *PNAS* **98**, 6121-6126 (2001).
26. Weichenrieder, O., Repanas, K. & Perrakis, A. Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon. *Structure* **12**, 975-986 (2004).
27. Kojima, K. K. & Fujiwara, H. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* **22**, 2157-2165 (2005).
28. Han, J. S. & Boeke, J. D. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays* **27**, 775-784 (2005).
29. Burke, W. D., Singh, D. & Eickbush, T. H. R5 retrotransposons insert into a family of infrequently transcribed 28S rRNA genes of planaria. *Mol Biol Evol* **20**, 1260-1270 (2003).
30. Kojima, K. K., Kuma, K., Toh, H. & Fujiwara, H. Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. *Mol Biol Evol* **23**, 1984-1993 (2006).
31. Luan, D. D. & Eickbush, T. H. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol Cell Biol* **15**, 3882-3891 (1995).
32. Ruschak, A. M. et al. Secondary structure models of the 3' untranslated regions of diverse R2 RNAs. *RNA* **10**, 978-987 (2004).
33. Kierzek, E. et al. Secondary structures for 5' regions of R2 retrotransposon RNAs reveal a novel conserved pseudoknot and regions that evolve under different constraints. *J Mol Biol* **390**, 428-442 (2009).
34. Kierzek, E. et al. Isoenergetic penta- and hexanucleotide microarray probing and chemical mapping provide a secondary structure model for an RNA element orchestrating R2 retrotransposon protein function. *Nucleic Acids Res* **36**, 1770-1782 (2008).
35. Kurzynska-Kokorniak, A., Jamburuthugoda, V. K., Bibillo, A. & Eickbush, T. H. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. *J Mol Biol* **374**, 322-333 (2007).
36. Babushok, D. V., Ostertag, E. M., Courtney, C. E., Choi, J. M. & Kazazian, H. H. J. L1 integration in a transgenic mouse model. *Genome Res* **16**, 240-250 (2006).
37. Chandrasegaran, S. & Smith, J. Chimeric restriction enzymes: what is next? *Biol Chem* **380**, 841-848 (1999).
38. Ye, J., Yang, Z., Hayes, J. J. & Eickbush, T. H. R2 retrotransposition on assembled nucleosomes depends on the translational position of the target site. *EMBO J* **21**, 6853-6864 (2002).

BIOGRAPHICAL INFORMATION

Christi Hull is receiving her Master of Science in Biology from the University of Texas at Arlington in May 2010. She also received her Bachelor of Science in Biology from the University of Texas at Arlington May 2006. She will pursue a career in the industrial sciences continuing in research. She is passionate for unique, creative, and innovative research. Christi would like to participate in research that emphasizes novel preventative medicine including gene therapy.