Mapping the RNA-Protein Interface in Telomerase RNP

by

Christopher James Bley

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2011 by the
Graduate Supervisory Committee:

Julian Chen, Chair
James Allen
Giovanna Ghirlanda

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

In the 1970s James Watson recognized the inability of conventional DNA replication machinery to replicate the extreme termini of chromosomes known as telomeres. This inability is due to the requirement of a building block primer and was termed the end replication problem. Telomerase is nature's answer to the end replication problem. Telomerase is a ribonucleoprotein which extends telomeres through reverse transcriptase activity by reiteratively copying a short intrinsic RNA sequence to generate 3' telomeric extensions. Telomeres protect chromosomes from erosion of coding genes during replication, as well as differentiate native chromosome ends from double stranded breaks. However, controlled erosion of telomeres functions as a naturally occurring molecular clock limiting the replicative capacity of cells. Telomerase is over activated in many cancers, while inactivation leads to multiple lifespan limiting human diseases. In order to further study the interaction between telomerase RNA (TR) and telomerase reverse transcriptase protein (TERT), vertebrate TERT fragments were screened for solubility and purity following bacterial expression. Soluble fragments of medaka TERT including the RNA binding domain (TRBD) were identified. Recombinant medaka TRBD binds specifically to telomerase RNA CR4/CR5 region. Ribonucleotide and amino acid pairs in close proximity within the medaka telomerase RNA-protein complex were identified using photo-activated cross-linking in conjunction with mass spectrometry. The identified cross-linking amino acids were mapped on known crystal structures of TERTs to reveal the RNA interaction interface of TRBD. The identification of this RNA

TERT interaction interface furthers the understanding of the telomerase complex at a molecular level and could be used for the targeted interruption of the telomerase complex as a potential cancer treatment.

DEDICATION

I would like to dedicate this to my loving wife Lexie, thank you for your support over the years of my studying, thank you for keeping me sane. To my family; Mom and Dad, Brenna and DeShaun, Matt and Danielle, and of course Jacob; you guys are awesome.

ACKNOWLEDGMENTS

I would like to thank Dr. Chen for your mentoring and advice over the years of grad school. Thank you showing me the ups and downs of research and for never letting me give up.

Thank you Chen lab members past and present; Xiaodong Qi, Dustin Rand, Joshua Podlevsky, Mingyi Xie, Yang Li, Andrew Brown, Lina Franco, Tracy Niday, and Andres Morera for making the Chen lab an enjoyable experience.

I would also like to thank my graduate committee Dr. Allen, Dr. Ghirlanda, and Dr. Chen for your knowledge and advice.

Lastly, I'd like to thank for fourth grade science teacher Mr. Wollitz who first showed me that science was fun, cool, and sometimes gooey.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

Chapter 1

OVERVIEW

1.1 A history of telomerase discovery

Telomerase is a specialized enzyme responsible for maintaining the ends of linear chromosomes. The discovery of telomerase was recently celebrated with the 2009 Noble Prize in Physiology or Medicine awarded to Elizabeth Blackburn, Carol Greider and Jack Szostak (Blackburn, 2010; Greider, 2010; Szostak, 2010).

In the late 1930s, Dr. Muller and McClintock hypothesized the existence of a "terminal gene" located at the chromosome ends, which served to seal the chromosome (McClintock, 1939; Muller, 1938). When chromosomal breakage occurs, the terminal gene prevents the broken chromosome piece from being reattached to a native chromosome end (Muller, 1938). This terminal gene hypothesized by Dr. Muller was termed the telomere. In 1978 Drs. Blackburn and Gall studied the telomere fragments found on extrachromosomal DNA encoding ribosomal RNA in the ciliate *Tetrahymena thermophile* and determined the first telomere sequence. The terminal sequence was found to contain between 20-70 repeats of the sequence 5'-T_2G_4-3' (Blackburn and Gall, 1978). Shortly after, the telomere sequence from 4 additional ciliate species was found to also be T_2G_4 with a 3' single stranded extension (Klobutcher et al., 1981). The TG rich telomere was found to be more ubiquitous than a simple ciliate mechanism of chromosome protection, as the next telomere to be sequenced, from the yeast Saccharomyces cerevisiae, was also found to be TG rich with a 3' overhang (Shampay et al., 1984). Additionally, artificial linear chromosomes with ciliate

telomeres were maintained in yeast while linear plasmids capped by a hairpin, an alternate hypothesized capping mechanism, were not maintained thus providing more evidence for the crucial chromosome maintenance role of telomeres (Szostak and Blackburn, 1982).

On Christmas Day, 1984, Dr. Greider observed telomerase activity in *Tetrahymena* extracts extending telomeric repeat containing primers with a 6 nucleotide ladder pattern (Greider and Blackburn, 1985, 2004). Numerous experiments were performed to verify the observed activity was due to telomerase as hypothesized, and not a typical polymerase copying telomeres in the extract. The six nucleotide repeats were sequenced and found to be composed of the *Tetrahymena* telomere repeat. Importantly, the addition of telomeric repeats was observed when using TG rich primers containing the yeast telomeric sequence (Greider and Blackburn, 1985). The activity observed was attributed to a yet to be identified enzyme which they termed a terminal transferase.

It was not until 1987 that the term telomerase was coined by Drs. Blackburn and Greider. The telomerase activity in *Tetrahymena* extracts was further studied and found to contain an RNA component which they hypothesized may play a templating role (Greider and Blackburn, 1987). The RNA component of *Tetrahymena* telomerase was later cloned, sequenced, and found to indeed contain a template element (Greider and Blackburn, 1989). Determination of the ciliate telomerase RNA secondary structure required the cloning and sequencing of six additional ciliate Telomerase RNA (TRs), allowing for phylogenetic comparative analysis (Romero and Blackburn, 1991). Phylogenetic comparative

analysis uses co-variation to verify putative base paired sequences, and is considered the gold standard of secondary structure analysis (Pace et al., 1989).

The human telomere sequence was identified through the use of hybridization techniques in conjunction with Bal-31 digestion of terminal DNA, and found to contain repeats of TTAGGG (Moyzis et al., 1988). The similarity between human and ciliate telomere sequences provided further evidence that telomerase was not a ciliate specific telomere maintenance enzyme (Greider and Blackburn, 2004). *In situ* hybridization was used to probe the telomeres of 91 vertebrate species determining that they shared the human TTAGGG telomeric repeat (Meyne et al., 1989).

Human telomerase activity was first directly observed in HeLa cell extracts, and found to exhibit the same characteristics, a ribonucleoprotein which adds telomeric repeats to g-rich oligos, as *Tetrahymena* telomerase (Morin, 1989). The human telomerase RNA component was not cloned and sequenced until 1995 (Feng et al., 1995). The secondary structure of vertebrate telomerase RNA was determined by phylogenetic comparative analysis only after the cloning of 32 additional vertebrate TRs (Chen et al., 2000). The protein component of ciliate, yeast, and human telomerase were concurrently identified (Lingner et al., 1997; Nakamura et al., 1997). The identified Telomerase Reverse Transcriptase (TERT) proteins contained the hallmark motifs of a reverse transcriptase (RT) flanked by telomerase specific regions in the N and C-terminal (Lingner et al., 1997). The structural elements of TERT and TR will be discussed in further detail below.

1.2 Medical implications

Mutations in telomerase components and biogenesis machinery have been identified as the underlying cause of multiple human disorders. Inactivation of telomerase is often also accompanied by a predisposition to cancer. Conversely, aberrant activation of telomerase is a step of cancer progression.

Telomerase has a direct impact on the replicative capacity of cells. Decreased cell regenerative capacity is the underlying defect in telomerase related diseases. A lack of telomerase activity in somatic cells limits their proliferative capacity. For example human fibroblasts lacking telomerase activity have steadily decreasing telomeres during serial passages (Harley et al., 1990). Telomeric DNA length directly correlates with the proliferative capacity of fibroblast cultures; cultures starting with longer telomeres are viable for more passages than those beginning with short telomeres (Allsopp et al., 1992). In a separate study, 98% of immortal human tissue culture lines were found to contain active telomerase while none of 22 non-immortal cell lines contained telomerase activity (Kim et al., 1994). Telomerase activity does not correlate with cellular TR levels, but rather with TERT expression. TR has been detected in both telomerase positive and negative tissues while TERT was only found in telomerase positive cells (Avilion et al., 1996). TERT overexpression in telomerase negative cell lines has been shown to convey an immortal phenotype upon the cells (Bodnar et al., 1998; Ramirez et al., 2001) indicating activation of TERT gene alone is capable of imparting active telomerase *in vivo* as TR is constitutively expressed. To model telomerase deficiency diseases, a mouse TR

knock out line was created (Blasco et al., 1997). The mouse telomerase null line is capable of surviving for only 4 generations before telomere defects are apparent, including telomere free ends and end to end fusion of chromosomes despite large starting telomeres (Blasco et al., 1997).

The first disease determined to be related to a telomerase defect was dyskeratosis congenita (DKC)(Mitchell et al., 1999b), a disease of regenerative tissue. Symptoms often include oral leukoplakia, abnormal skin pigmentation and nail bed defects (Vulliamy et al., 2006). Ultimately the disease results in organ failure, with bone marrow failure being the leading cause of death in DKC patients (Walne and Dokal, 2009). The first mutation characterized in DKC patients, who had an X-linked recessive form of the disease, was within the dyskerin protein gene (Heiss et al., 1998). Mutations in dyskerin were later shown to be accompanied by shortened telomeres in patient blood samples (Vulliamy et al., 2001b). Mutations in the telomerase RNA were later found in patients with the autosomal dominant form of the disease (Vulliamy et al., 2001a). Dyskeratosis congenita has since been linked to 7 genes (TERT, TR, DKC1, NOP10, NHP2, TINF2 and TCAB1) all encoding components of the telomerase holoenzyme or related to telomerase biogenesis (Walne and Dokal, 2009; Zhong et al., 2011). Patients with telomerase related mutations also present clinically with idiopathic pulmonary fibrosis and aplastic anemia (Armanios, 2009).

Telomerase is often thought of as an attractive cancer therapy target (Shay and Wright, 2002). Many cancers have developed methods to reactivate telomerase to allow unchecked proliferation. Biopsies of 90% of tumors tested

positive for telomerase activity, and telomerase activity has been observed in a high percentage of various other cancers surveyed (Kim et al., 1994; Shay and Bacchetti, 1997). It is postulated that since telomerase activation is a critical step in allowing unlimited proliferation of cancer, inhibiting telomerase could inhibit further growth or kill the cancer by shortened telomeres triggering apoptosis (Shay and Wright, 2002). The detrimental effect on cancer cells of telomerase inhibition may not be solely tied to telomere shortening. When telomerase is inhibited in some cell lines derived from cancer tissues, the increase in apoptosis is seen before a decrease in telomere length is observed (Folini et al., 2005; Saretzki et al., 2001). This telomere independent sensitivity to telomerase inhibition may be due to non-canonical telomerase functions such as involvement in DNA damage response (Masutomi et al., 2005). However, non-canonical functions of telomerase, TERT, or TR are contested. A recent study disputed the role of mouse TERT or TR as transcription factors or as members of the DNA damage response system (Vidal-Cardenas and Greider, 2010).

1.3 Structure and function of Telomerase

As previously mentioned the minimal components of active telomerase, *in vitro*, are the catalytic protein TERT and Telomerase RNA (TR). The TERT protein is composed of up to 4 domains: the Telomerase Essential N-terminal (TEN), Telomerase RNA Binding Domain (TRBD), Reverse Transcriptase Domain (RT) and C-terminal Extension (CTE) (Jacobs et al., 2006; Lai et al., 2001; Lingner et al., 1997). Though telomerase is described as having 4 domains,

they may not all fold independently, experimentally only the TEN domain can act *in trans* with the rest of TERT (TRBD-RT-CTE) (Moriarty et al., 2004).

The determination of vertebrate TR secondary structure revealed more conservation in structure than in sequence (Chen et al., 2000). Telomerase RNA shares certain structural elements conserved across all species identified, such as a template flanking pseudoknot (Chen et al., 2000; Romero and Blackburn, 1991) (Dandjinou et al., 2004). Vertebrate TR was found to have an H/ACA domain near its 3' end important for RNA processing (Mitchell et al., 1999a). One region of vertebrate TR termed conserved region (CR) 4/5 is required for telomerase activity yet the function of CR4/CR5 is not known. Point mutations of key residues within CR4/CR5 critically impair telomerase activity (Chen et al., 2002b).

Telomerase RNA has been proposed to interact with multiple regions of TERT as might be expected of two highly interacting molecules each responsible for multiple roles in telomerase activity. TEN has been shown to interact with low affinity to the TR pseudoknot (Moriarty et al., 2004). High affinity binding of TR to TERT occurs at the TRBD (Lai et al., 2001), which has been shown to interact with CR4/CR5 RNA, and with lesser affinity to the pseudoknot RNA (Moriarty et al., 2004)

Atomic level structures of telomerase have been long sought though attempts have met with limited success. Two crystal structures from the highly studied *Tetrahymena* TERT have been published consisting of the TEN and TRBD protein fragments (Jacobs et al., 2006; Rouda and Skordalakes, 2007).

The only full length TERT structure is from the beetle *Tribolium castaneum* which lacks TEN and contains truncated TERT specific motifs compared to other species (Gillis et al., 2008). A second structure of *Tribolium* TERT in complex with a DNA-RNA duplex hybrid simulating the telomere/template has also been published (Mitchell et al., 2010). While highlighting the location of the template and 3' telomere end, this structure does not provide insight into the arrangement of other RNA domains. While the field eagerly awaits the crystal structure from a species containing all domains to answer questions related to TEN, the beetle TERT structure has greatly benefited the telomerase field. For instance, mapping of residues found to impact activity on the beetle structure facilitated understanding function of a recently described motif found in the RT domain (Xie et al., 2010). Numerous NMR structures have been published for individual helices and hairpins of TR, including p6 and p6.1 which compose CR4/CR5 (Leeper et al., 2003; Leeper and Varani, 2005), though even when pieced together these structures lack a contextual element from solving the structures of protein free RNA.

1.4 Project objectives

In the absence of structural information about the interactions of telomerase RNA and TERT protein, I set out to begin to fill in the void. I embarked upon a mission to express and purify TERT fragments in *E. coli* in order to provide material for biochemical assays to determine interactions between components of telomerase. First, multiple protein fragments were screened from two vertebrate species, medaka fish and human, for overexpression

in *E. coli*. Human was chosen as it would have a more direct impact on understanding of telomerase as related to human diseases. Medaka fish TERT was also chosen as it contains all vertebrate specific motifs, and more importantly the fish TR has been identified as the smallest vertebrate TR (Xie et al., 2008). While the medaka TR is smaller than that of human it contains all of the structural elements we were interested in studying, CR4/CR5 and pseudoknot.

Chapter two describes the screening of various TERT protein fragments for overexpression in *E. coli* and solubility during purification. Native and synthetic gene sequences were used for both human and medaka TERT. The design and creation of a synthetic gene encoding for medaka TERT, codon optimized for expression in *E. coli*, is also described in chapter 2. Purification of two protein fragments from medaka TERT, TRBD and TRBD-RT as maltose binding protein fusions, are discussed in chapter three. Purification steps and conditions are discussed shortly, and reduction of ionic strength before and after RNA binding is addressed. Protein containing TRBD-RT was found to form a soluble aggregate eluting in the void volume of gel filtration column, while TRBD was a mixture of soluble aggregate and monomer. This mixture of monomer and soluble aggregate was separable and did not interconvert once the two were removed from each other.

In chapter four the RNA binding interface on medaka TRBD occupied by CR4/CR5 RNA is identified. Mapping of the interface was achieved by identification of amino acid/nucleic acid pairs in close proximity within the RNA-protein complex. Site specific cross-linking was used to covalently bond residues

allowing them to be mapped. Limited alkaline hydrolysis identified RNA residues involved in cross-linking while MALDI-TOF and MALDI-TOF/TOF were used to identify and sequence cross-linked peptide allowing identification of precise amino acid.

REFERENCES

Allsopp, R.C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W., and Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc Natl Acad Sci U S A *89*, 10114-10118.

Armanios, M. (2009). Syndromes of telomere shortening. Annu Rev Genomics Hum Genet *10*, 45-61.

Avilion, A.A., Piatyszek, M.A., Gupta, J., Shay, J.W., Bacchetti, S., and Greider, C.W. (1996). Human telomerase RNA and telomerase activity in immortal cell lines and tumor tissues. Cancer Res *56*, 645-650.

Blackburn, E.H. (2010). Telomeres and telomerase: the means to the end (Nobel lecture). Angew Chem Int Ed Engl *49*, 7405-7421.

Blackburn, E.H., and Gall, J.G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol *120*, 33-53.

Blasco, M.A., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A., and Greider, C.W. (1997). Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. Cell *91*, 25-34.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. Science *279*, 349-352.

Chen, J.L., Blasco, M.A., and Greider, C.W. (2000). Secondary structure of vertebrate telomerase RNA. Cell *100*, 503-514.

Chen, J.L., Opperman, K.K., and Greider, C.W. (2002). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res *30*, 592-597.

Dandjinou, A.T., Lévesque, N., Larose, S., Lucier, J.-F., Abou Elela, S., and Wellinger, R.J. (2004). A phylogenetically based secondary structure for the yeast telomerase RNA. Curr Biol *14*, 1148-1158.

Feng, J., Funk, W.D., Wang, S.S., Weinrich, S.L., Avilion, A.A., Chiu, C.P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., *et al.* (1995). The RNA component of human telomerase. Science *269*, 1236-1241.

Folini, M., Brambilla, C., Villa, R., Gandellini, P., Vignati, S., Paduano, F., Daidone, M.G., and Zaffaroni, N. (2005). Antisense oligonucleotide-mediated inhibition of hTERT, but not hTERC, induces rapid cell growth decline and apoptosis in the absence of telomere shortening in human prostate cancer cells. Eur J Cancer *41*, 624-634.

Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the Tribolium castaneum telomerase catalytic subunit TERT. Nature *455*, 633-637.

Greider, C.W. (2010). Telomerase discovery: the excitement of putting together pieces of the puzzle (Nobel lecture). Angew Chem Int Ed Engl *49*, 7422-7439.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell *43*, 405-413.

Greider, C.W., and Blackburn, E.H. (1987). The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell *51*, 887-898.

Greider, C.W., and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature *337*, 331-337.

Greider, C.W., and Blackburn, E.H. (2004). Tracking telomerase. Cell *116*, S83-86, 81 p following S86.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet *19*, 32-38.

Jacobs, S.A., Podell, E.R., and Cech, T.R. (2006). Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol *13*, 218-225.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science *266*, 2011-2015.

Klobutcher, L.A., Swanton, M.T., Donini, P., and Prescott, D.M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. Proc Natl Acad Sci U S A *78*, 3015-3019.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol Cell Biol *21*, 990-1000.

Leeper, T., Leulliot, N., and Varani, G. (2003). The solution structure of an essential stem-loop of human telomerase RNA. Nucleic Acids Res *31*, 2614-2621.

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

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. Science *276*, 561-567.

Masutomi, K., Possemato, R., Wong, J.M.Y., Currier, J.L., Tothova, Z., Manola, J.B., Ganesan, S., Lansdorp, P.M., Collins, K., and Hahn, W.C. (2005). The telomerase reverse transcriptase regulates chromatin state and DNA damage responses. Proc Natl Acad Sci USA *102*, 8222-8227.

McClintock, B. (1939). The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis. P Natl Acad Sci USA *25*, 405-416.

Meyne, J., Ratliff, R.L., and Moyzis, R.K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci U S A *86*, 7049-7053.

Mitchell, J.R., Cheng, J., and Collins, K. (1999a). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. Mol Cell Biol *19*, 567-576.

Mitchell, J.R., Wood, E., and Collins, K. (1999b). A telomerase component is defective in the human disease dyskeratosis congenita. Nature *402*, 551-555.

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

Moriarty, T.J., Marie-Egyptienne, D.T., and Autexier, C. (2004). Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. Mol Cell Biol *24*, 3720-3733.

Morin, G.B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. Cell *59*, 521-529.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., and Wu, J.R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A *85*, 6622-6626.

Muller, H. (1938). The remaking of chromosomes. The Collecting Net *13*, 181-198.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. (1997). Telomerase catalytic subunit homologs from fission yeast and human. Science *277*, 955-959.

Pace, N.R., Smith, D.K., Olsen, G.J., and James, B.D. (1989). Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA- a review. Gene *82*, 65-75.

Ramirez, R.D., Morales, C.P., Herbert, B.S., Rohde, J.M., Passons, C., Shay, J.W., and Wright, W.E. (2001). Putative telomere-independent mechanisms of replicative aging reflect inadequate growth conditions. Genes Dev *15*, 398-403.

Romero, D.P., and Blackburn, E.H. (1991). A conserved secondary structure for telomerase RNA. Cell *67*, 343-353.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure *15*, 1403-1412.

Saretzki, G., Ludwig, A., von Zglinicki, T., and Runnebaum, I.B. (2001). Ribozyme-mediated telomerase inhibition induces immediate cell loss but not telomere shortening in ovarian cancer cells. Cancer Gene Ther *8*, 827-834.

Shampay, J., Szostak, J.W., and Blackburn, E.H. (1984). DNA sequences of telomeres maintained in yeast. Nature *310*, 154-157.

Shay, J.W., and Bacchetti, S. (1997). A survey of telomerase activity in human cancer. Eur J Cancer *33*, 787-791.

Shay, J.W., and Wright, W.E. (2002). Telomerase: a target for cancer therapeutics. Cancer Cell *2*, 257-265.

Szostak, J.W. (2010). DNA ends: just the beginning (Nobel lecture). Angew Chem Int Ed Engl *49*, 7386-7404.

Szostak, J.W., and Blackburn, E.H. (1982). Cloning yeast telomeres on linear plasmid vectors. Cell *29*, 245-255.

Vidal-Cardenas, S.L., and Greider, C.W. (2010). Comparing effects of mTR and mTERT deletion on gene expression and DNA damage response: a critical examination of telomere length maintenance-independent roles of telomerase. Nucleic Acids Res *38*, 60-71.

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J., and Dokal, I. (2001a). The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature *413*, 432-435.

Vulliamy, T.J., Knight, S.W., Mason, P.J., and Dokal, I. (2001b). Very short telomeres in the peripheral blood of patients with X-linked and autosomal dyskeratosis congenita. Blood Cells Mol Dis *27*, 353-357.

Vulliamy, T.J., Marrone, A., Knight, S.W., Walne, A., Mason, P.J., and Dokal, I. (2006). Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation. Blood *107*, 2680-2685.

Walne, A.J., and Dokal, I. (2009). Advances in the understanding of dyskeratosis congenita. British Journal of Haematology *145*, 164-172.

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F., and Chen, J.J.-L. (2008). Structure and function of the smallest vertebrate telomerase RNA from teleost fish. J Biol Chem *283*, 2049-2059.

Xie, M., Podlevsky, J.D., Qi, X., Bley, C.J., and Chen, J.J. (2010). A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. Nucleic Acids Res *38*, 1982-1996.

Zhong, F., Savage, S.A., Shkreli, M., Giri, N., Jessop, L., Myers, T., Chen, R., Alter, B.P., and Artandi, S.E. (2011). Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. Genes Dev *25*, 11-16.

Chapter 2

IDENTIFICATION OF SOLUBLE RECOMBINANT TERT PROTEINS FRAGMENTS

2.1 ABSTRACT

Telomerase holoenzyme is composed of two parts, the protein Telomerase Reverse Transcriptase and RNA component Telomerase RNA. Many biochemical assays require a substantial amount of enzyme in order to characterize properties such as RNA protein interactions. In addition, a method to introduce mutations designed to interrupt or perturb events is advantageous. While it is possible to purify telomerase from tissue culture sample or reconstitute protein and RNA in *in vitro* systems such as rabbit reticulocyte, these methods do not yield high levels of enzyme. In order to further study telomerase protein-RNA interactions, protein fragments of vertebrate TERT were tested for overexpression and solubility using *E. coli* as a host for expression. Fragments of both human and medaka fish TERT were identified which yielded substantial overexpression. Partial purification of human C-terminal extension and medaka fragments including full length are described. A medaka protein fragment containing the RNA binding domain was identified as a likely candidate to yield useful protein for biochemical characterization.

2.2 INTRODUCTION

Many biochemical studies require more enzyme than can be purified from natively expressed sources. Over-expression in a recombinant system is a common method used for proteins. There are multiple eukaryotic and prokaryotic host systems available each with its own advantages and disadvantages. Since its discovery, labs have been trying to find useful systems to over-express TERT with varying degrees of success.

For instance, hTERT has been over expressed in human tissue culture(Cristofari and Lingner, 2006). Expression in tissue culture using cells from the native organism, human, has the advantage of providing accessory proteins which can help in the telomerase assembly. For instance it has been demonstrated that the molecular chaperones hsp90, p23 stably interact with telomerase (Holt et al., 1999) and that hsp70 interacts with naked TERT and is released upon hTR binding (Forsythe et al., 2001). Over expression in tissue culture is useful for looking at the accumulation and activity of mutants when reconstituted in vivo. This in vivo system is often used in combination with in vitro systems to verify results. However the limited amount of TERT expressed by this method does not allow purification for a more robust biochemical analysis.

Another method used for TERT expression is insect cell culture. Like human tissue culture, insect cell culture is a eukaryotic system which may provide similar molecular chaperones to promote correct TERT folding. As a eukaryotic host, insect cells have better success with multi domain proteins than prokaryotic systems. A Flag tagged hTERT was first expressed in High Five insect cells

infected with a virus carrying an hTERT over expression plasmid(Masutomi et al., 2000). When reconstituted with telomerase RNA, activity was observed from insect cell expressed hTERT. Telomerase activity was increased when mixed with human cell extract, from a telomerase negative cell line, indicating that insects lacked some components required for optimal assembly. One disadvantage of the system was that a large amount of the expressed TERT was insoluble. As with expression in human tissue culture, the expression level is relatively low not allowing purification of a significant amount of TERT protein. A more recent study improved the folding of TERT when expressed in insect cells by targeting the protein to the endoplasmic reticulum(Wu et al., 2007).

Expression of TERT has also been attempted in bacteria systems such as *E. coli*. Major advantages of *E. coli* expression include relatively quick doubling times allowing expression in a single day rather than multiple days of tissue culture systems, straightforward vector creation and transformation allowing many constructs to be tested, as well as ease and affordability of scale up. A disadvantage to *E. coli* expression is that multi domain proteins tend to misfold and form aggregates(Baneyx and Mujacic, 2004). One possible reason is that *E. coli* translation occurs faster than that of Eukaryotic translation, 12-20 amino acids per second compared to 4 amino acids per second (Dennis and Bremer, 1974; Dennis and Nomura, 1974; Pedersen, 1984).

When expressing a eukaryotic protein in a bacteria system it is not uncommon for the protein to have folding issues which lead to a decrease in the amount of soluble protein. Expression and purification schemes often increase

the aggregate protein levels to purify as inclusion body or maximize protein solubility before purification. Some purification schemes take advantage of misfolding by purifying inclusion bodies, dense regions collecting misfolded protein. The inclusion bodies are purified removing a substantial number of *E.*coli proteins, denatured and slowly refolded (Baneyx and Mujacic, 2004). Inclusion body purification is often more successful with small proteins with less complex folding. One method of increasing the yield of soluble protein is to increase the amount of *E. coli* molecular chaperones present in the bacteria(Castanie et al., 1997). Co-transformation of a set of plasmids previously described (Castanie et al., 1997) along with a plasmid encoding the target protein co-express two, pOFX-t7-KJ3, or three, pOFX-t7-KJE3, molecular chaperones. The two chaperone set is composed of dnaJ and dnaK while the three chaperone set also includes grpE. *E. coli* dnaK is a homolog of eukaryotic hsp70 family chaperones (Straus et al., 1990). In *E. coli* dnaK and dnaJ are expressed from the same operon, dnaK works in parallel with trigger factor to bind unfolded polypeptide chains as they are translated by the ribosome(Deuerling et al., 1999). DnaK and grpE are cofactors of dnaK, dnaJ facilitates dnaK binding to unfolded proteins hydrolyzing dnaK bound ATP(Baneyx and Mujacic, 2004). GrpE works to facilitate release of proteins by dnaK by exchanging ADP for ATP(Baneyx and Mujacic, 2004). Upon product release, dnaK often rebinds the protein. Cycles of protein binding and release slows down protein folding keeping the protein on the correct folding pathway (Agashe et al., 2004).

Previous studies with TERT overexpression in *E. coli* have had limited results. Jacobs et al. used a GFP reporter system to identify soluble fragments of *Tetrahymena* TERT which could be expressed in *E. coli*. Of the fragments identified one spanning the Telomerase essential N-terminal domain (TEN) was crystallized (Jacobs et al., 2006). A second successful *E. coli* expression of a TERT fragment used the TRBD of *Tetrahymena* TERT(Rouda and Skordalakes, 2007). Though crystallization was successful, the recombinant expressed TRBD fragment used was incapable of binding telomerase RNA.

Both *E. coli* expressed protein fragments came from *Tetrahymena*, a ciliate model organism from which telomerase was first identified(Greider and Blackburn, 1985). Though *Tetrahymena* has significance to the telomerase field, identifying soluble fragments of TERT from a vertebrate species would allow us to study interactions more relevant to the human form of telomerase. In order to apply biochemical experiments to telomerase we set out to find identify and soluble TERT fragments from a vertebrate species.

2.3 MATERIALS AND METHOD

2.3.1 hTERT CTE fragment expression-

All hTERT CTE fragments were cloned from the native hTERT gene sequence and placed in a modified pMAL-2CX with a his tag in-frame before the MBP tag (ASU 304). For small scale expression, 1ml culture of LB + ampicillin was grown at 37^{o}C to OD600 0.5-0.6. Initial expression tests, with or without chaperones, were induced with 1mM IPTG and continued at 37^{o}C for 2.5 hours. Protein was purified as described below.

For expression of 20 CTE fragments, 1.5ml cultures were grown at 37o to OD600 0.5-0.6. Cultures were induced with 2 mM IPTG and reduced to room temperature. Expression was allowed to continue overnight (about 22 hours) before 1ml of culture was pelleted. The pellet was used for Ni-NTA purification.

2.3.2 hTERT CTE fragment, small scale Ni-NTA purification-

Pellets corresponding to 1ml of culture were resuspended in 0.75ml lysis buffer (25mM sodium phosphate pH 8.0, 300mM sodium chloride, 0.05% tween-20, 1mg/ml lysozyme). Samples were left at room temperature for 30 minutes to allow lysozyme to work then sonicated on ice 3X 20 seconds at a power level 3 and 50% duty cycle on a sonicator. Samples were centrifuged at 18K RCF for 5 minutes to remove insoluble proteins. Fifty micro liters, per sample, of Ni-NTA (Qiagen) was prewashed 3x with water and resuspended in binding buffer (lysis buffer without lysozyme). Lysate was added to resin and placed on a rotator for 60 minutes at 4^{o}C. Beads were washed twice in binding buffer and resuspended in elution buffer (binding buffer supplemented with 300mM imidazole). A 30µl

sample was added to 10µl 4X SDS loading buffer, cooked at 90^{o}C for 10 minutes and loaded to an 8% SDS-gel with a 4% stacking gel. The gel was run at 80volts until sample had passed through the stacking gel then 120 volts. Gel was stained and visualized with Page Blue Protein Staining Solution (Fermentas).

2.3.3 hTERT CTE fragment column purification-

A 500ml culture of culture containing plasmid for MBP-CTE #5 (D945-D1132) co expressed with pOFX—t7-KJ3 was induced at OD600 0.5 with 1mM IPTG overnight at room temp. Cells were harvested by spinning at 5K RCF 15 minutes, pellets washed once with 1X TBS and frozen until use. A two gram pellet was resuspened in CTE lysis buffer (20mM sodium phosphate pH 8.0, 500mM sodium chloride, 0.05% tween-20, 5mM imidazole) and sonicated 5x 30 sec power level 5 with 75% dudty cycle. Lysate was clarified by centrifugation at 17K RCF for 30 minutes. Sample was purified by Ni affinity on a 1ml HiTrap IMAC FF column (GE Healthcare) on an AKTA Purifier FPLC (GE Healthcare). After binding, sample was washed with a gradient of 1-6% buffer B (CTE lysis buffer + 500mM imidazole) over 15ml, followed by a 6% buffer B wash for 5ml, a gradient up to 26% buffer B over 20ml. Sample eluted during the second gradient was pooled and applied to a 5ml HiTrap MBP (GE Healthcare). Sample was eluted with 20mM sodium phosphate pH 8.0, 500mM sodium chloride, 10mM maltose.

2.3.4 hTERT expression test-

Nine coding segments of hTERT from a synthetic codon optimized gene amplified by PCR resulting in EcoR1 and XhoI sites at the 5’ and 3’ ends were

cloned to EcoR1 and XhoI sites of pMAL-2cx-His vector. Cloning resulted in an N-terminal MBP tag and C-terminal 6xHis tag. Regions covered for the 9 fragments are shown in figure 2.5A. Plasmids were sequenced and transformed to Bl21 along with pOFX-t7-KJ3 encoding dnaJ, dnaK and grpE. For expression single colonies were inoculated into 3ml of rich media (For one liter: 10g tryptone, 5 g Sodium Chloride, 5g yeast extract 10ml of 20% glucose) plus 50µg/ml carbenicillin and 50µg/µl spectinomycin. 100 microliters of overnight culture was used to start a 5ml culture in rich media plus antibiotics. At OD600 0.5 cultures were induced with 1mM IPTG. Expression continued at 37^{o}C for 3 hours or room temperature overnight.

2.3.5 synthetic mdTERT gene construction-

Six plasmids were created by PCR amplifying the native medaka TERT gene with primers adding a 5' EcoR1 and 3' Sal1 sites. Products were digested and inserted to EcoR1and XhoI of pMAL-2cx-His vector. Plasmid sequences were confirmed by sequencing. The six plasmids were co-transformed with pOFX-t7-KJ3 to BL21(DE3) expression cells. Single colonies were inoculated into 3ml cultures of rich media + 50µg/µl ampicillin and 50µg/µl spectinomycin and grown to OD600 0.7. A 200µl uninduced sample was removed, pelleted, resuspeded in 50µl 2X Laemmli sample buffer, and cooked at 90^{o}C for 10 minutes. The remainder of culture was induced with 1mM IPTG and incubated with shaking at room temperature overnight. Induced samples were prepared as above. Samples, 3µl uninduced and 1.5µl induced, were loaded to a 6% acrylamide SDS gel and stained with coomassie dye (Fermentas).

2.4 RESULTS

2.4.1 *Expression of CTE domain of hTERT*

Using a GFP reporter system similar to that described elsewhere (Jacobs et al., 2005) the C-terminal extension (CTE) domain of hTERT was identified as a possible soluble domain. The CTE fragment identified spanned T937-S1125, here after simply referred to as CTE fragment. The CTE fragment was sub cloned to a vector fusing an N-terminal 6xhis-MBP tag. Initially three fragments were tested for expression, ASU 306 (6xhis-MBP-T937D1132), ASU 324 (6xhis-MBP-T937-S1125) and ASU 374 6xhis-MBP-T937-S1125 containing V1035D mutation). Purification via Ni-NTA showed a minor band corresponding to the recombinant fragments (Figure 2.1). Considering the CTE fragment was identified as soluble when fused to a GFP reporter, retention of solubility when subcloned and expressed from a different fusion vector validates the GFP-screening system as capable of identifying soluble fragments.

It has previously been shown that telomerase interacts with hsp70 in rabbit reticulocyte lysate and that hsp70 may have a role in promoting telomerase RNA binding or promoting correct folding which in turn facilitates RNA binding (Forsythe et al., 2001). It was predicted that co-expression with molecular chaperones would increase the yield of soluble 6xhis-MBP-Cterm. As seen in figure 2.2 the yield of purified soluble CTE fragment increases when co-expressed with all three molecular chaperones (elution lanes). As seen in the first eluted fraction lane, a number of *E. coli* proteins bound the Ni-NTA resin even without the presence of a 6xhis-tagged protein. The strong band just below 37

kDa is believed to be dnaJ. The potential dnaJ band is more intense when co-purified with 6xhis-MBP-Cterm, co-purification of dnaJ can indicate a mis-folding in the protein of interest (Haacke et al., 2009) .

Since the GFP screening system relied on the number of transformants and colonies available to screen it was not a comprehensive library screen. It is possible that by changing the boundaries of the fragment used expression and purification yield could be improved. We set out to create a systematic serial truncation of the CTE TERT fragment by changing the N-terminal boundary. A series of 18 clones were created by shifting the N-terminal amino acid by 2 each time, this lead to start sites ranging from T937(#2)-K973(#18). In addition one fragment longer than the original was created starting at P923 (#1) located within the RT domain of telomerase as well as one fragment that also truncated the carboxyl termini of the CTE, T937-S1125 (#20). All genes were cloned to the 6xHis-MBP vector to be expressed as fusion proteins. The vectors encoding the fragments were co-transformed to BL21 expression cells along with pOFX-t7KJE3 encoding three molecular chaperones. A one milliliter culture of each was expressed overnight, lysed by sonication, and batch purified with Ni-NTA resin. A whole cell and purified sample from each was loaded on a gel to visualize expression and purification yield (Figure 2.3). For all fragments which expressed, the purification yield was similar with the protein of interest being a minor band on the gel. Three fragments, starting at K965, G967 and R971, did not have a significant band corresponding to purified 6xHis-MBP-Cterm fragment.

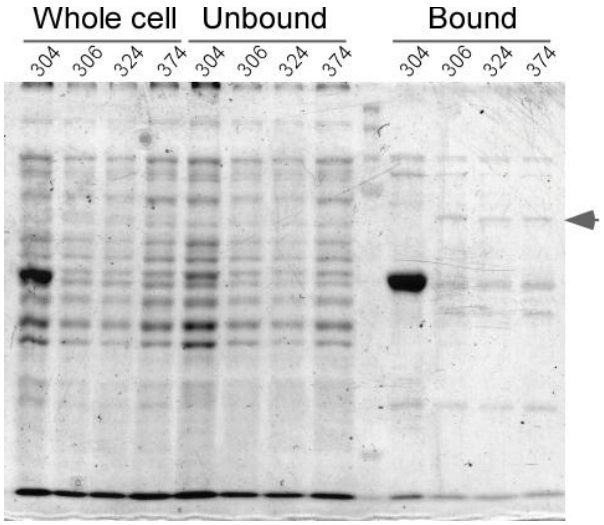


Figure 2.1- Expression and purification of three recombinant hTERT CTE fragments.
Whole cell, Ni-NTA unbound and Ni-NTA bound proteins are shown for three 6xhis-MBP-CTE fragments ASU306 (6xhis-MBP-T937D1132), ASU 324 (6xhis-MBP-T937-S1125) and ASU 374 (6xhis-MBP-T937-S1125 harboring V1035D mutation) as well as vector only control ASU-304 (6xhis-MBP only). The partially purified proteins indicated by grey arrow match the expected size ~65kDa.

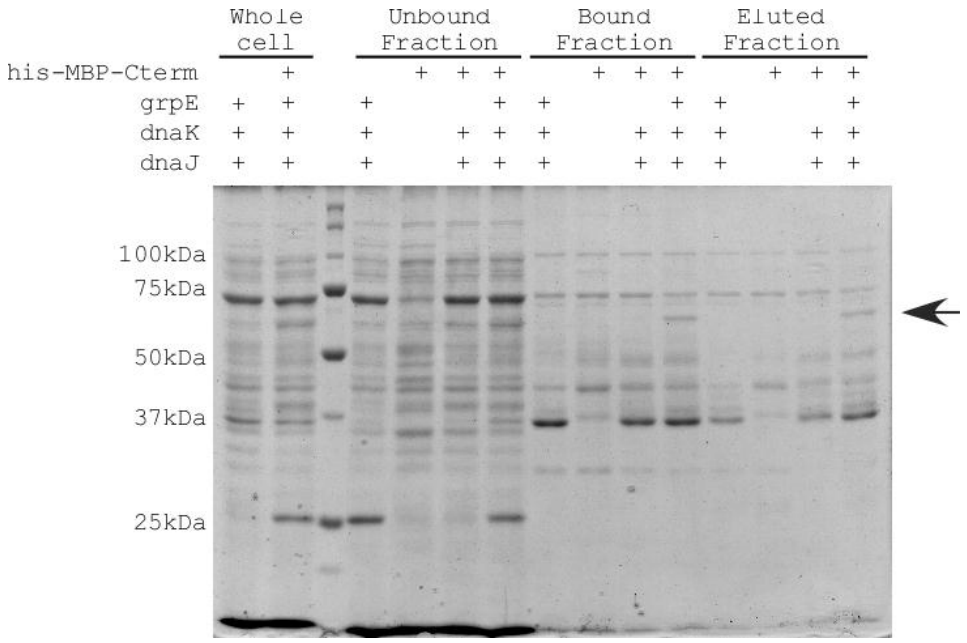


Figure 2.2- Co-expression with molecular chaperones
hTERT CTE fragment was expressed with or without co-expression of chaperone encoding vectors. Triple chaperone only control not encoding hTERT-CTE is included for comparison of banding pattern. Bound, unbound and eluted refer to fractions of Ni-NTA purification. Arrow indicates purified 6xHis-MBP-Cterm ($T_{937}D_{1132}$).

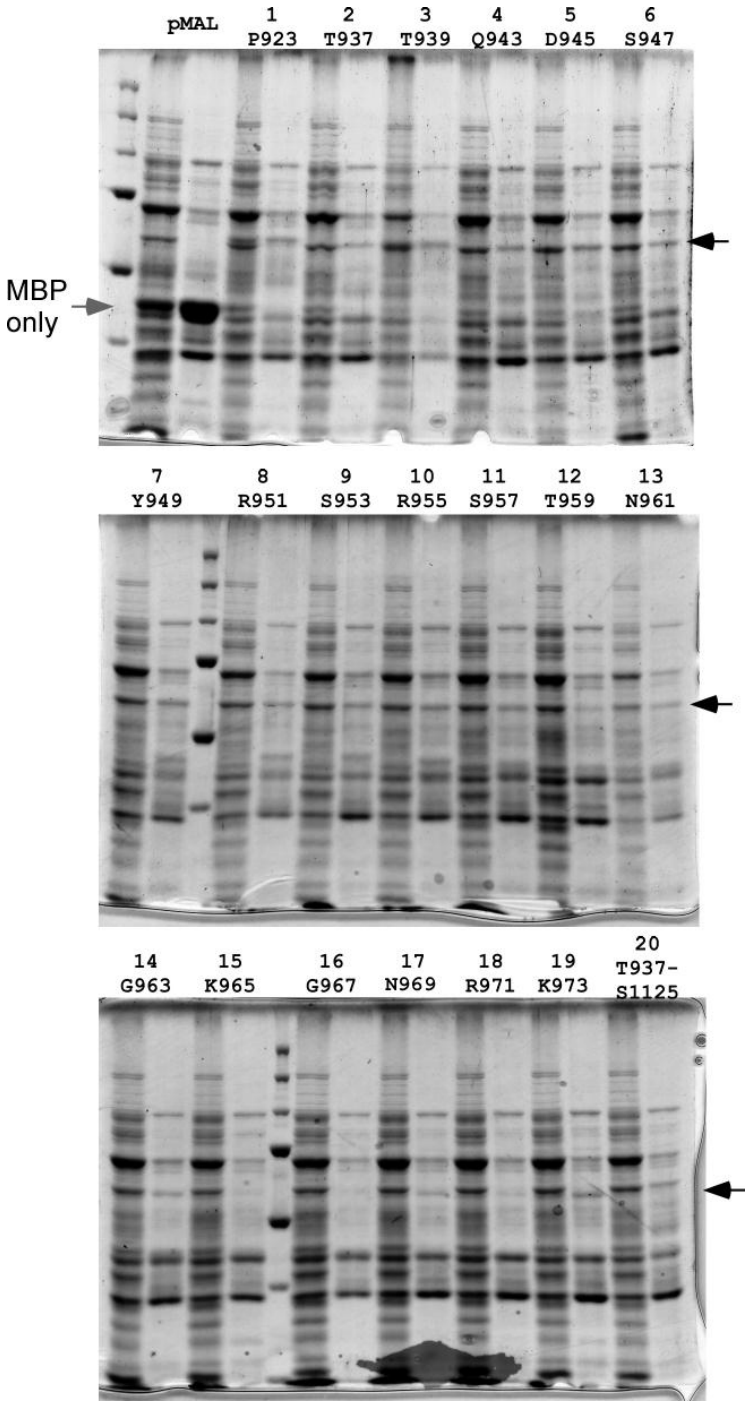


Figure 2.3- hTERT CTE fragment walk.
A series of progressively smaller CTE fragments of hTERT were transformed, expressed and partially purified by Ni-NTA batch purification. Protein fragments span from N-terminal amino acid indicated to native C-terminus of hTERT. Black arrow indicates recombinant protein band, grey arrow indicates over expression of the fusion tag only. First lane of each set is whole cell sample, second is partially purified sample.

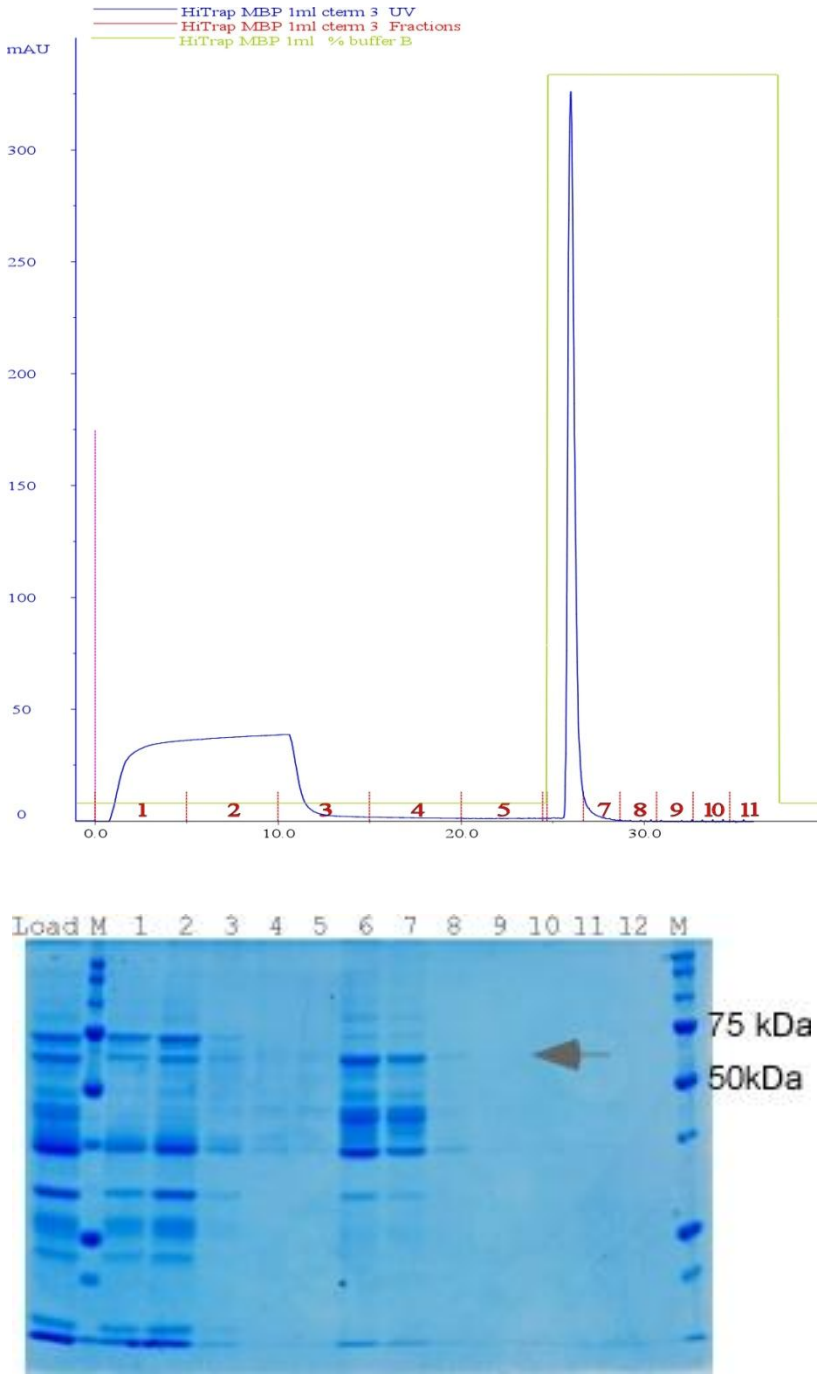


Figure 2.4- Purification of 6xHis-MBP-Cterm (T939-D1132).
Top: FPLC output showing binding of hTERT CTE fragment to HiTrap MBP column (1ml), blue line indicates $UV_{\lambda=280nm}$ absorbance. Bottom: SDS-PAGE showing protein content of fractions from above purification. Protein of interest is indicated with arrow.

2.4.2- Expression of hTERT fragments using synthetic gene.

The native hTERT coding sequence is GC rich, 66% GC, and contains a number of codons rarely used in *E. coli*. Since the *E. coli* transcriptome contains only a low number of these rare codons the cellular level of the matching tRNAs is often low. Rare codons can often lead to a number of translational problems. It has been observed that an increase in the number of rare codons in a gene is inversely proportional to the amount of protein recovered when looking at a single protein (Kane, 1995). Clusters of rare codons can lead to premature translational termination (Kane, 1995). In one case the ribosome appears to have jumped a rare codon when flanked on either side by the same high abundance codon (Kane, 1995). Vertebrate genes, coming from species with different codon bias than *E. coli*, often contain a large number of rare codons and quickly deplete the levels of corresponding charged tRNAs when translated. In order to remove rare codons an artificial hTERT gene was synthesized, at the same time GC content was decreased to near 50% to simplify PCR of the gene. As described for medaka TERT below in further detail, only the codons were changed not the resulting amino acid sequence. The synthetic gene was used to subclone a number of fragments for expression testing.

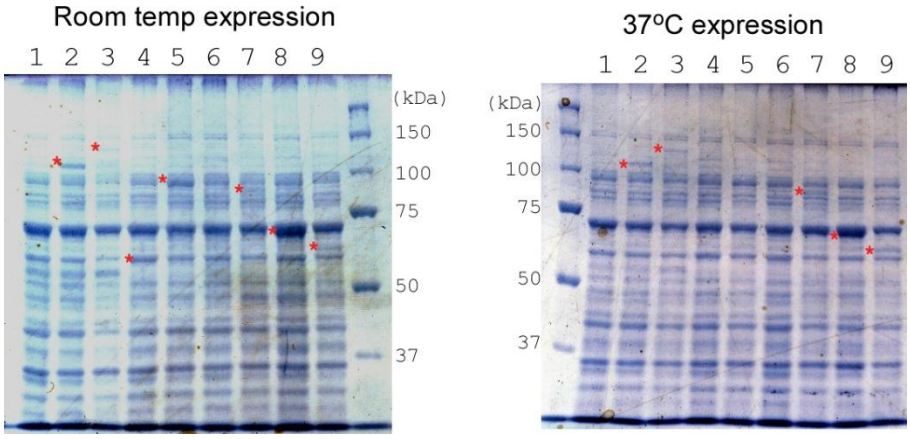
Nine truncations of hTERT were created and expressed as MBP-TEV fusion proteins containing an N-terminal maltose binding protein with a TEV protease cleavage site. The nine vectors were transformed to BL21 expression cells along with pOFX-t7-KJ3 triple chaperone plasmid. Molecular weight and encompassing amino acids are shown in figure 2.5A. Upon small scale

expression at 37^{o}C, bands corresponding to 5 of the fragments, TRBD-RT, TRBD-CTE, RT-CTE, RT only and CTE only, were observed on SDS page (figure 2.5B right). Bands corresponding to two additional fragments, partial TRBD only and partial TRBD-RT, were observed when expressed at room temperature (Figure 2.5B left).

A

glycerol stock #	Sample #	Start AA	End AA	# AA	MW (kDa)	MBP+X MW (kDa)
ASU 493	1	H330	E605	276	32.2	77.0
ASU 494	2	H330	D945	616	70.7	115.5
ASI 495	3	H330	D1132	803	91.6	136.4
ASU 496	4	L448	E605	158	18.9	63.7
ASU 497	5	L448	D945	498	57.4	102.2
ASU 498	6	L448	D1132	685	78.3	123.1
ASU 499	7	R657	D1132	476	53.4	98.2
ASU 500	8	R657	D945	289	32.5	77.3
ASU501	9	P923	D1132	210	23.4	68.2

B



C

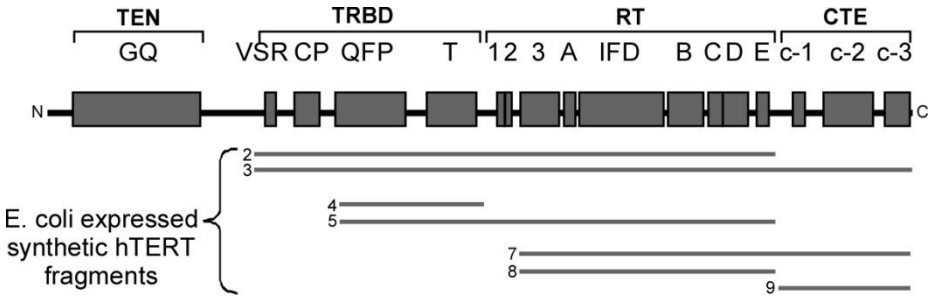


Figure 2.5- Summary of expressed hTERT fragments
Nine fragments of hTERT were expressed as MBP-fusion proteins in BL21 *E. coli* harboring a triple chaperone plasmid to co-express dnaJ, dnaK, and GRPE. (A) ASU glycerol stock numbers, start and end amino acids corresponding to full length hTERT, and the molecular weight in kDa of the fusion fragment as well as full fusion protein is shown. (B) SDS gels showing whole protein content of cultures expressed at room temperature (left) and at 37°C (right), asterisk indicating over expression band of correct size. (C) Top-domain structure of TERT, bottom visual representation regions which were over expressed in B.

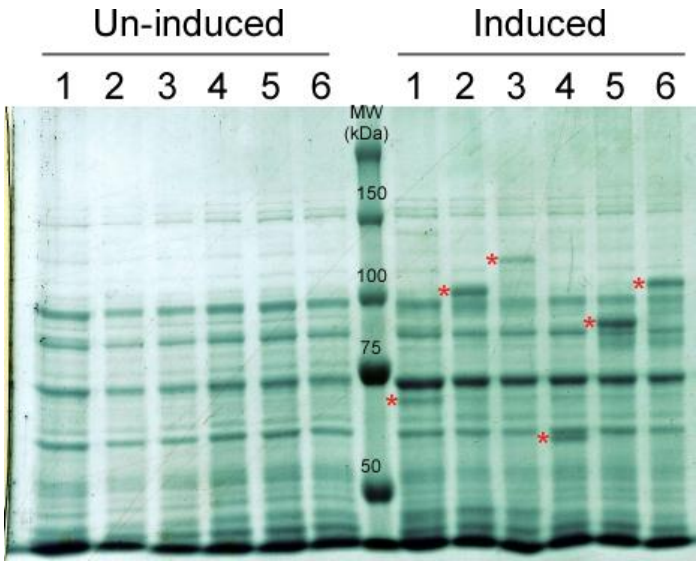
2.4.3 Expression of mdTERT fragments.

When a recombinant protein has low expression, as seen with the hTERT fragments, one common method used to circumvent this obstacle is to use the protein from a closely related species. Six TERT protein fragments, from the Japanese rice fish Oryzias Latipes or medaka fish, similar to constructs 1-6 of hTERT (Figure 2.5) were expressed in *E. coli*. The telomerase RNA component from medaka fish was recently discovered(Xie et al., 2008). The medaka TR is the smallest vertebrate TR known to date. Despite smaller than average length, the medaka TR contains all conserved motifs except the cab box(Xie et al., 2008). Use of a telomerase with a small RNA component would be beneficial to structural studies as many flexible variable regions are removed. Of the 6 constructs tested for medaka TERT (mdTERT) all six were visible on SDS PAGE of whole cell lysates including fragment 1 (G318-Q572) corresponding to the TRBD. Over expression of fragment #1 was not observed with human TERT (Compare Figure 2.5B lane one with Figure 2.6B lane 1 induced). Fragments 1 and 3 (G318-Q572, G318-M1090) covering just TRBD and TRBD-CTE, had lesser expression while fragments 2, 4, and 6 (G318-D903, L415-Q572, L4150M1090) exhibited intermediate expression. Fragment 5, spanning mid TRBD-through the end of the RT (L415-D903, motifs QFP-E) had the highest level of over expression. While fragment 5 had the highest expression, fragment 2 was chosen for further examination as it contained the full RNA binding domain.

A

glycerol stock #	Sample #	Start AA	End AA	# AA	MW (kDa)	MBP+X MW (kDa)
ASU 436	1	G318	Q572	255	29.8	74.6
ASU 437	2	G318	D903	286	66.4	111.2
ASU 438	3	G318	M1090	773	87.9	132.7
ASU 439	4	L415	Q572	158	18.5	63.3
ASU 440	5	L415	D903	489	55.1	99.9
ASU 441	6	L415	M1090	676	76.6	121.4

B



C

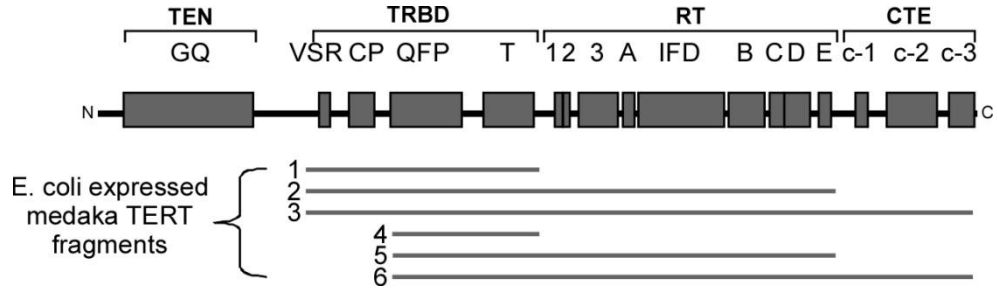


Figure 2.6-Summary of expressed medaka TERT fragments
Six protein fragments of mdTERT were cloned as MBP fusion proteins and co-expressed with dnaj, dnaK, and grpE. (A) Table summarizing fragments cloned including the amino acid junction, numbers correspond to full length mdTERT. ASU glycerol stock numbers are indicated as well as molecular weight of the fragments and fusion proteins. (B) Whole cell lysates of un-induced and induced cultures were loaded onto an 8% SDS gel. Red asterisk indicate bands matching to the predicted molecular weight not present in un-induced samples. (C) Domain structure of TERT is shown, lines indicate regions of mdTERT overexpressed.

MBP-mdTERT spanning from the TRBD through the RT-domain was transformed to both BL21 and Rosetta2, a modified BL21 strain harboring genes for rare tRNAs. The construct was expressed overnight with or without three molecular chaperones (Figure 2.7). Un-induced, induced whole cell samples as well as purified protein eluted from Amylose beads was loaded to an SDS gel in order to compare expression levels as well as amount and purity of recovered protein (Figure 2.7). Co-expression of the three molecular chaperones greatly increased the expression level in both BL21 and Rosetta 2 cells (Figure 2.7 lanes 2 vs 5 and 8 vs 11). Without co-expression of chaperones the full length fusion protein constituted only a small portion of the total purified protein (Figure 2.7 lane 3). In BL21 cells co-expression of the chaperones increased the level of full length fusion protein but also increased a number of other smaller bands (Figure 2.7 lane 6). The other bands are hypothesized to be truncation products likely due to ribosome stalling at rare codons as when the protein is expressed in Rosetta 2 cells these bands disappear (Figure 2.7 lane 12). When MBP-mdTRBD-RT was co-expressed with chaperones in Rosetta 2 the full length fusion protein is estimated to be 50% pure from single step amylose purification.

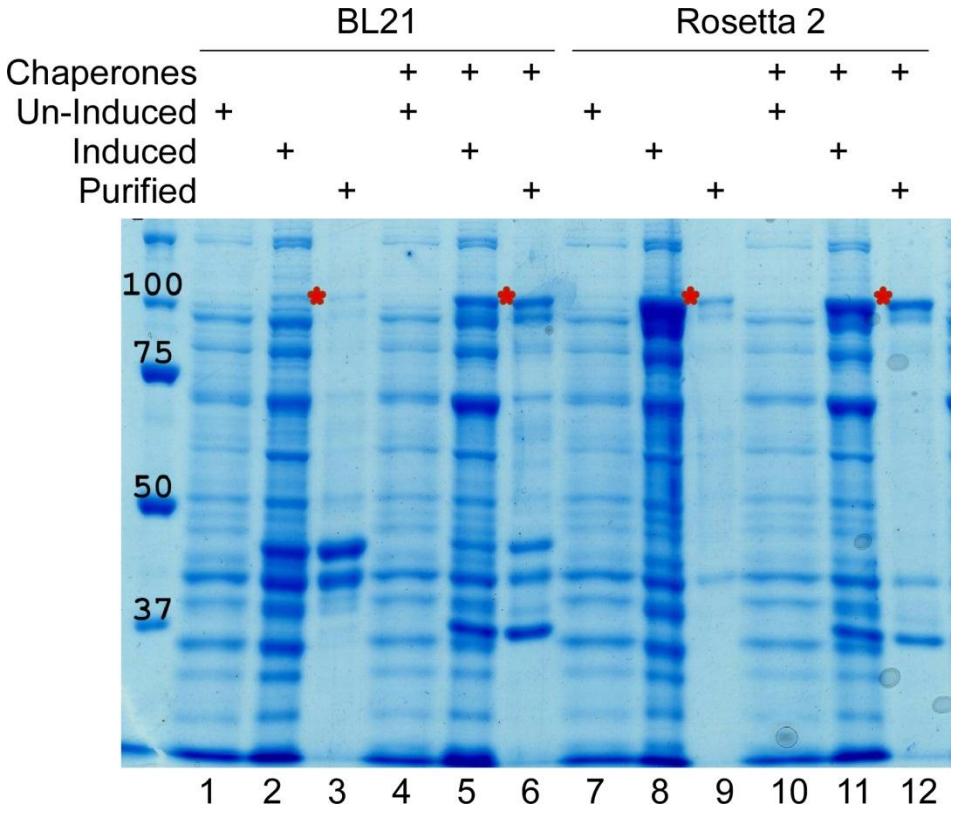


Figure 2.7- Effect of *E. coli* strain and chaperones on over expression of medaka TERT fragment

Medaka TERT protein fragment TRBD-RT was expressed in BL21 and Rossetta2 *E. coli* strains with or without co-expression of molecular chaperones dnaJ, dnaK and grpE. Un-induced and induced are whole cell lysate samples from + or – IPTG induction. Purified refers to partially purified protein by amylose affinity resin. Red star indicates MBP-TRBD-RT protein fragment.

2.4.4 Construction of a synthetic mdTERT gene

Every mdTERT fragment tested for over expression yielded an observable over expressed protein and expression in the presence of extra copies of rare tRNAs led to higher expression and purity. This led to the hypothesis that removing rare codons by synthesizing a gene encoding the mdTERT but using the codon bias of *E. coli* would result in higher expression and purity. Using a synthetic gene would also allow the use of host strains other than Rosetta 2 such as the Scarab Genomics multiple deletion series strains in which non-essential genes from the *E. coli* were removed reducing the genome size by 15%(Posfai et al., 2006).

The synthetic mdTERT gene was constructed in two parts, TRBD through CTE, and TEN to the start of TRBD. The two portions contained a small overlap allowing them to be combined by overlap extension PCR into one coding DNA sequence. The amino acid sequence of mdTERT was entered into the DNA works website (http://helixweb.nih.gov/dnaworks/) (Hoover and Lubkowski, 2002) which generates an initial DNA coding sequence that avoids low frequency rare codons. Based on the codon usage of 780 genes, three classes of *E. coli* genes have been defined (Medigue et al., 1991). Other than codon usage differences, two of the classes can be defined by expression levels. Class I genes are typically low expression level genes constitutively expressed. Class I genes are not limited to low expression as at times they can be unregulated when the correct conditions are met (Medigue et al., 1991). Class II genes are constitutively highly expressed genes. Since over expression was the goal, the

codon usage table for *E. coli* class II was used. A more recent study has come out since the mdTERT gene was synthesized indicating over expression is more highly correlated to a subset of 10 codons which are not the most commonly used for those amino acids(Welch et al., 2009). Another study found the secondary structure near the translation start site had a larger effect on expression than codon bias with more stable secondary structures in the –4 to +38 region leading to lower expression(Kudla et al., 2009). The sequences for restriction sites used during cloning (EcoR1, XhoI, NdeI) disallowed in the final sequence. In addition to codon optimizing, the algorithm used by the DNA works website attempts to remove any repetitive sequences that may lead to mis-priming during PCR or during gene construction (Hoover and Lubkowski, 2002).

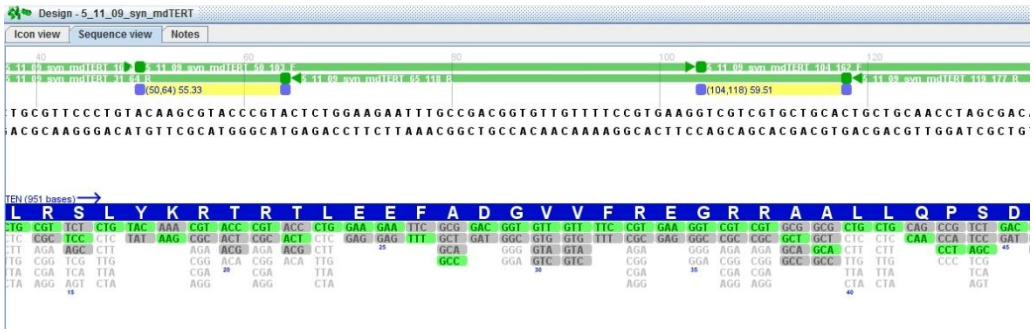


Figure 2.8- Synthetic gene oligo design with Gene Designer
A screen image showing Gene Designer program (Villalobos et al., 2006) used to design the oligonucleotides for synthetic mdTERT gene construction. Green lines indicate oligonucleotides with arrow heads indicating direction (forward or reverse, arrowhead at 3' end). Blue capped yellow bars indicate the Tm for the sequence they span, the overlap between oligonucleotide pairs. On the bottom the amino acid sequence is shown with all possible codons for each amino acid. Codons highlighted green are the codon in use at that position, grey codons indicate they are used greater than 10% of the time in class II *E. coli.* genes, uncolored are below the 10% threshold.

The DNA sequence generated was then used to design oligonucleotides, which when ligated together would result in the full gene. Using the program Gene Designer from DNA2.0 (Villalobos et al., 2006), oligos of approximately 55nt were designed to completely cover the gene in both the forward and reverse directions leaving no gaps. Oligos were paired (top/bottom strand) with each pair having about 30 base pairs and 5' overhangs of about 12 bases (Figure 2.8) at either end. The overhang sequences were individual optimized to have melting temperatures of 55-60^{o}C (Figure 2.8 Blue/yellow bar) as well as to avoid a sequence which could form hairpins. Codons in the overhang sequences were modified in order to adjust Tm or sequence. The two parts of the gene, TRBD-CTE and TEN, required 41 and 18 pairs of oligos respectively. Oligos were phosphorylated prior to ligation in sets of 5 with a single overlapping oligo per set (i.e. F5-F9, F9-F13, R5-R9, R9-R13…). The 5' terminal oligos (F1 and R41) were omitted from the phosphorylation reaction, added only after the kinase was heat denatured, to prevent mulitmers of the gene from forming during ligation. Since phosphorylation occurs before annealing of oligo pairs 5' or 3' overhangs can be used, if annealing is performed before phosphorylation 5' overhangs are preferred as 5' extensions are more readily phosphorylated by T4 PNK than recessed 5' ends. Forward and reverse sets of oligos were combined (i.e. F1-F5 combined with R1-R5) to be annealed by denaturing at 80^{o}C and slow cooling to room temperature in a heat block placed on a piece of insulating Styrofoam on the lab bench. Alternatively a PCR thermo cycler can be used to control cooling, decreasing one degree per min. Annealed oligos were diluted into ligation buffer

and ligated overnight with T4 DNA ligase. After an overnight ligation reaction samples were separated on a 3% agarose gel, multiple ligation products were observed creating a ladder composed of mono, di, tri up to the desired penta oligo-pair ligation products. After the initial ligation of oligos the rest of the gene construction was completed with PCR

Ligation products (Figure 2.9A/B) were gel extracted and used for overlap PCR amplification, a method of combining the two DNA fragments which contain a short stretch of overlap at their termini originally described for making point mutations (Higuchi et al., 1988). During the initial rounds of overlap PCR, the overlapping sequence of the two fragments act as super-primers annealing together and are extended to form the full length product. The full length product is then PCR amplified as normal with primers at either end. The overlap PCR products(figure 2.9C), which each contained the sequence of 9 of the original oligo pairs (i.e. 1-9 built by PCR with 1-5 and 5-9 ligation products) were gel extracted. The entire gene could be combined at this point by overlap PCR however; we chose to clone the gene fragments to a blunt end vector for sequencing at this point.

The products of overlap PCR were cloned to a blunt end vector and sequenced at this point as the average size of about 500 bp can easily be sequence by a single reaction in either direction. It was expected that some of the constructs would have mutations as the success during each round of oligonucleotide synthesis is approximately 99% at each step, since long oligonucleotides of about 55 bases were used it was expected that only about 58%

of each oligonucleotide would be full length (0.99^{54}=0.58). Many of these aberrant oligonucleotides result in oligo truncations. When a base is missed during synthesis, that molecule is capped disallowing further nucleotides to be added in subsequent rounds, resulting in a 5' truncated oligonucleotide. However, as with any chemical reaction, capping is not 100% and internal deletions can occur due to missed capping. For the TRBD-CTE portion of the gene, 3 plasmids for each of the 5 constructs (corresponding to oligonucleotide pairs 1-9, 9-17, 17-25, 25-33, 33-41) were sequenced. Of the 15 plasmids sequenced, only 4 had no insertions or deletions. In the constructs sequenced there were 15 instances of single nucleotide deletions and 2 instances of di-nucleotide insertions.

To complete the gene synthesis the insert of plasmids that contained the correct sequence, or a sequence that could be easily repaired by PCR, were amplified and used for a final overlap extension PCR in which all the individual pieces were combined. For example, with TRBD-CTE the 4 correct inserts and two halves of the final insert, for which no correct sequence was found, were individually amplified. The individual PCR products were then combined in a 6 way overlap PCR and amplified with oligonucleotides F1 and R41 as primers. The final gene construct was gel extracted, cloned to a blunt end vector and sequenced one final time.

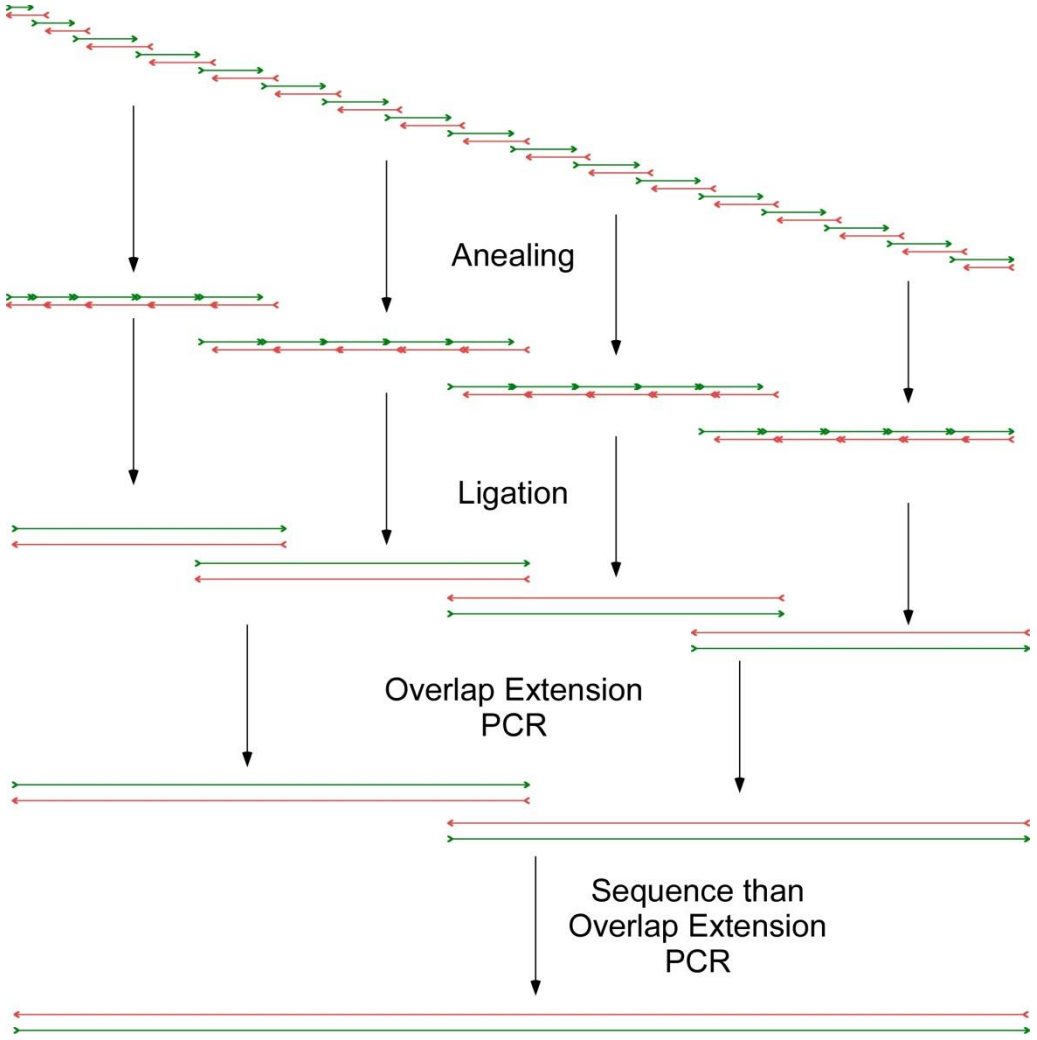
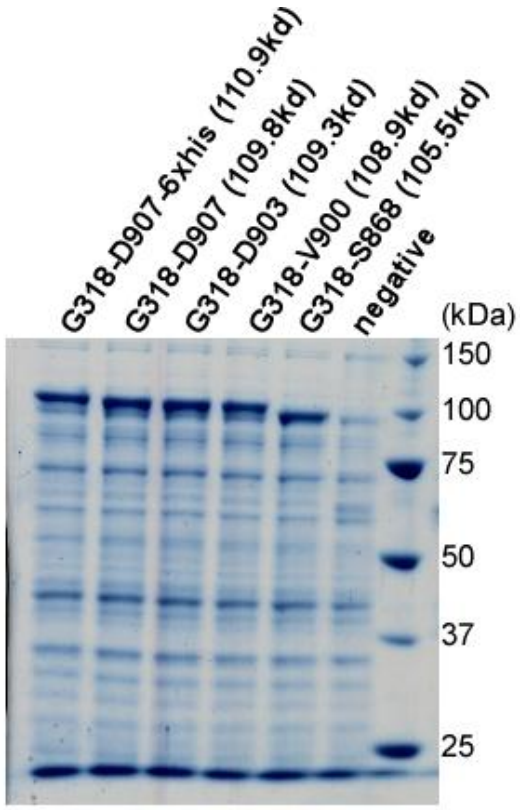


Figure 2.9- Overview of synthetic gene construction
An overview of the process used to create synthetic gene sequences. (A) Pairs of overlapping oligonucleotides of about 55bases each span the gene. (B) Oligos are combined in sets of 5 pairs and annealed by slow cooling. (C) Annealed oligos are ligated. Ligation products (B) are gel extracted and combined using PCR. PCR products (C) are cloned to a blunt end vector, sequenced, and combined by PCR to create the full synthetic gene. For simplicity only 17 oligo pairs are shown.

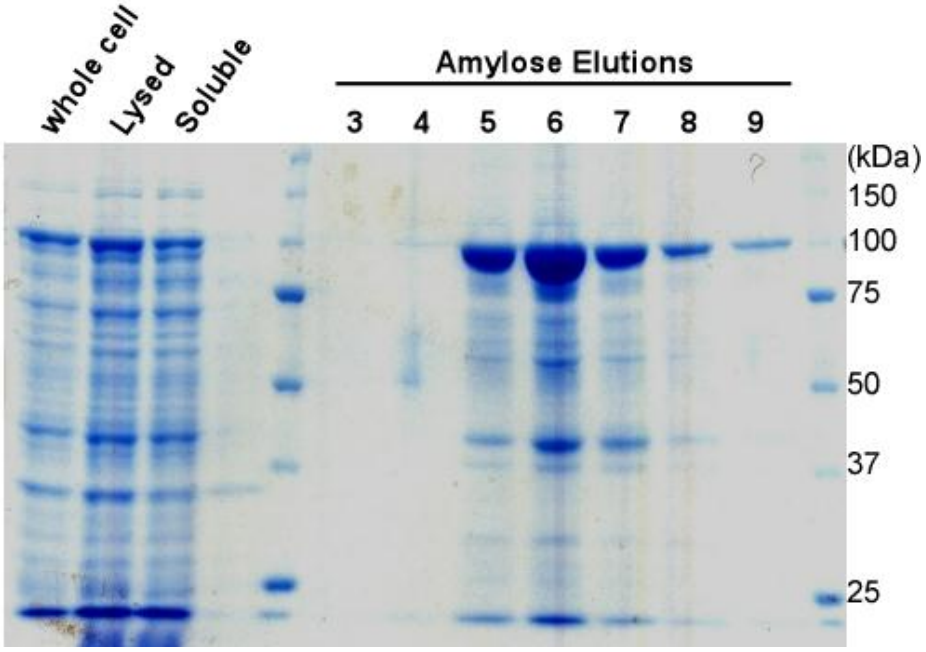
2.4.5 Expression/purification of synthetic mdTERT TRBD-RT fragments

Using the synthetic gene constructed above five constructs spanning from TRBD through the RT domain (TRBD-RT) were created. The constructs all encoded an N-terminal MBP tag and one included a C-terminal 6xHis tag as well. The five constructs encoded proteins from 105kDa to 111 kDa, all 5 exhibited over expression when whole cell lysates were analyzed (figure 2.10A) when compared with an uninduced control. Since all constructs expressed the largest (MBP-G318-D907) was chosen for further analysis. Upon lysis, the fusion protein remained soluble and bound to amylose resin. The protein eluted at a high concentration from the amylose resin but co-purified numerous smaller proteins (Figure 2.10B). When loaded to a gel filtration column the protein eluted in the void volume (figure 2.10C). The gel filtration used, sephadex S200, has a void size of about 2 mega Daltons. With an expected weight of 110.9kDa even a di-mer, tri-mer, or tetra-mer would resolve on the column rather than be found in the void volume indicating the construct formed a soluble aggregate.

A



B



(Legend on page 47)

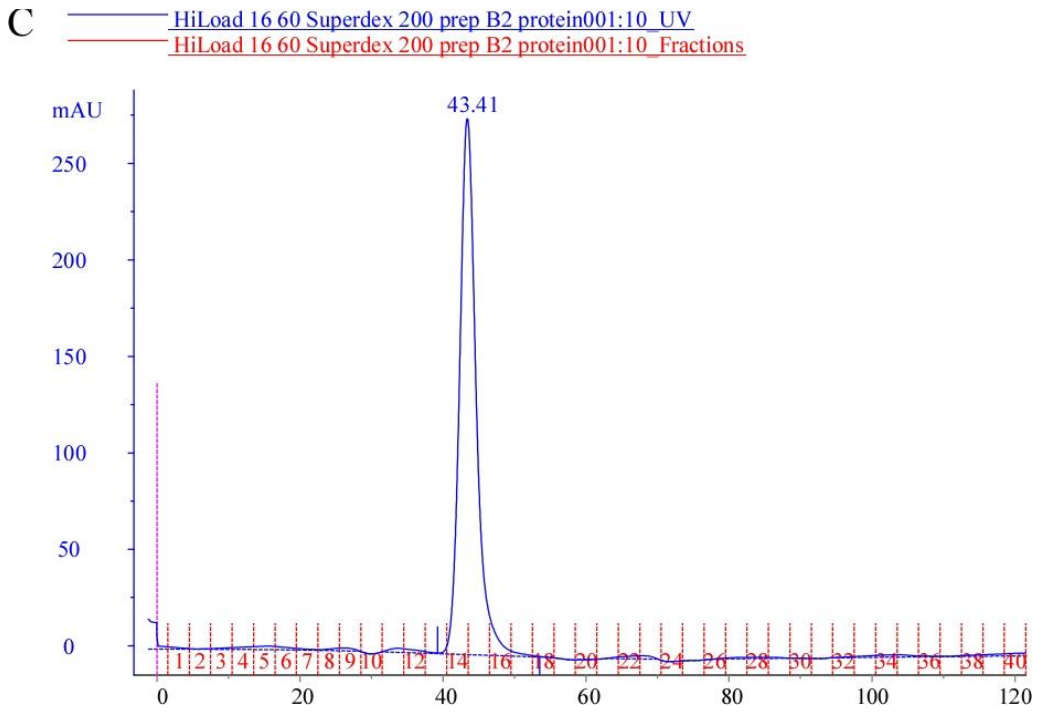


Figure 2.10C- Expression and purification of synthetic mdTERT TRBD-RT results in soluble aggregate protein.
(A)Whole cell samples from overexpression of 5 protein constructs of mdTERT TRBD-RT as MBP fusion proteins encoded by a synthetic mdTERT gene are compared to an un-induced control. Substantial bands near 100kDa are overexpressed MBP-TRBD-RT protein. (B) Samples from purification of MBP-TRBD-RT (G318-D907) are shown. Amylose elution samples were pooled and applied to (C) S200 gel filtration column. Blue line indicates 280nm UV absorbance. X axis is in milliliters. Elution of sample at 43.4ml corresponds to void volume elution.

2.5 DISCUSSION

The purpose of this set of experiments was to identify fragments of the TERT which yield soluble over expressed protein. Multiple expression systems were considered and ultimately the *E. coli* host system was chosen. Protein fragments from Human and medaka TERTs were tested for over expression and solubility.

While co-expression with molecular chaperones increased the amount of purified protein, most likely by increasing solubility, there are inherent problems to this method. It has been shown with a number of kinases that co-expression with the molecular chaperones dnaJ, dnaK grpE or GroEL and GroES can also increase the level of soluble aggregates (Haacke et al., 2009) and that co-purification of chaperones with the protein of interest can indicate misfolding. All of the human CTE fragments of hTERT that were expressed had a large amount of co-purifying proteins including possible molecular chaperones, the fragments were not ideal for further biochemical analysis due to low yield and a large portion of the protein in a likely misfolded state.

Of the 20 CTE fragments expressed during CTE walking, only 3 did not yield purified protein. Interestingly the fragment starting at N969, which falls in the middle of the non-expressing fragments, did yield purified protein showing how a relatively minor change in the starting amino acid can have a major impact on protein purification.

In addition to the 20 CTE constructs from hTERT, 8 additional constructs were created spanning different regions of the protein. Minor over expression

was observed for 5 of the constructs at 37^{o}C and for 7 constructs at room temperature. While observing any over expression is an improvement when expressing hTERT, no significant amounts of protein were purified from these constructs. Using TERT fragments from the teleost fish medaka, a higher level of over expression was seen. Even full length mdTERT was successfully over expressed and even partially purified, however protein yield and purity were low and protein eluted as a large complex in the void volume of gel filtration. Interestingly, TRBD from the synthetic human gene (fragment 1 figure 2.5) did not express while TRBD from medaka is highly over expressed and was ultimately characterized biochemically for the RNA-protein interactions (see chapter 4).

When faced with low expression or a high abundance of co-purifying truncation products, gene synthesis is an attractive technique to use. In the case of medaka TERT fragments over expression in Rosetta2 removed some impurities and greatly reduced the level of others when compared to expression in BL21 (Figure 2.7). However, the need for pRARE, which provides the rare tRNAs in Rosetta2, eliminated the possibility of using other modified *E. coli* hosts such as C41 and C433 from Lucigen which are reported to better express toxic and membrane proteins or other modified *E. coli* strains. For this reason we generated a synthetic coding sequence for mdTERT optimized for expression in *E. coli*. Nearly 10% of the codons in the native medaka coding sequence, 104 out of 1091 total codons, use the low abundance tRNAs supplied by Rossetta2 cells. In addition there are 6 repeats of 12 or more nucleotides in the native coding

sequence. The synthetic coding sequence synthesized contains no rare codons or repeats of 12 bases or longer.

During gene synthesis, high number of deletions and some insertions were observed. Because overlap PCR is used for combining the gene segments, restriction sites were not engineered to digest the individual pieces out of the cloning vector. Deletions in the terminal oligonucleotides can be ignored if they are near the 5' portion of the PCR primer annealing region as primer annealing during PCR would obligatorily correct these mutations. When no plasmids contain the correct sequence for a segment of the gene two options are available; either more plasmids can be sequenced or if one plasmid has the correct front half and a different plasmid has the correct rear half, they can be combined by overlap PCR. It was hypothesized that the deletions are by product of oligonucleotide synthesis. An alternate explanation is that deletions were introduced during PCR as the samples went through multiple PCR reactions before they were cloned and sequenced. However, internal deletions are the most common mistake in synthesized oligonucleotides (Hecker and Rill, 1998).

Using the synthetic mdTERT gene multiple constructs were created spanning TRBD only, TRBD-RT, TRBD-CTE as well as full length mdTERT as fusion proteins to MBP. Of all the constructs tested only TRBD was monomeric, all other constructs tested eluted in the gel filtration void volume (>2000kDa). It is for this reason that the TRBD was chosen for purification optimization and biochemical characterization.

REFERENCES

Agashe, V.R., Guha, S., Chang, H.C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U., and Barral, J.M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. Cell 117, 199-209.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol 22, 1399-1408.

Castanie, M.P., Berges, H., Oreglia, J., Prere, M.F., and Fayet, O. (1997). A set of pBR322-compatible plasmids allowing the testing of chaperone-assisted folding of proteins overexpressed in Escherichia coli. Anal Biochem 254, 150-152.

Cristofari, G., and Lingner, J. (2006). Telomere length homeostasis requires that telomerase levels are limiting. EMBO J 25, 565-574.

Dennis, P.P., and Bremer, H. (1974). Differential rate of ribosomal protein synthesis in Escherichia coli B/r. J Mol Biol 84, 407-422.

Dennis, P.P., and Nomura, M. (1974). Stringent Control of Ribosomal-Protein Gene-Expression in Escherichia-Coli. P Natl Acad Sci USA 71, 3819-3823.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400, 693-696.

Forsythe, H.L., Jarvis, J.L., Turner, J.W., Elmore, L.W., and Holt, S.E. (2001). Stable association of hsp90 and p23, but Not hsp70, with active human telomerase. J Biol Chem 276, 15571-15574.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405-413.

Haacke, A., Fendrich, G., Ramage, P., and Geiser, M. (2009). Chaperone over-expression in Escherichia coli: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. Protein Expr Purif 64, 185-193.

Hecker, K.H., and Rill, R.L. (1998). Error analysis of chemically synthesized polynucleotides. Biotechniques 24, 256-260.

Higuchi, R., Krummel, B., and Saiki, R.K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16, 7351-7367.

Holt, S.E., Aisner, D.L., Baur, J., Tesmer, V.M., Dy, M., Ouellette, M., Trager, J.B., Morin, G.B., Toft, D.O., Shay, J.W., et al. (1999). Functional requirement of p23 and Hsp90 in telomerase complexes. Genes Dev 13, 817-826.

Hoover, D.M., and Lubkowski, J. (2002). DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res 30, e43.

Jacobs, S.A., Podell, E.R., and Cech, T.R. (2006). Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol 13, 218-225.

Jacobs, S.A., Podell, E.R., Wuttke, D.S., and Cech, T.R. (2005). Soluble domains of telomerase reverse transcriptase identified by high-throughput screening. Protein Sci 14, 2051-2058.

Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol 6, 494-500.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science 324, 255-258.

Masutomi, K., Kaneko, S., Hayashi, N., Yamashita, T., Shirota, Y., Kobayashi, K., and Murakami, S. (2000). Telomerase activity reconstituted in vitro with purified human telomerase reverse transcriptase and human telomerase RNA component. J Biol Chem 275, 22568-22573.

Medigue, C., Rouxel, T., Vigier, P., Henaut, A., and Danchin, A. (1991). Evidence for horizontal gene transfer in Escherichia coli speciation. J Mol Biol 222, 851-856.

Pedersen, S. (1984). Escherichia coli ribosomes translate in vivo with variable rate. EMBO J 3, 2895-2898.

Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., de Arruda, M., et al. (2006). Emergent properties of reduced-genome Escherichia coli. Science 312, 1044-1046.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure 15, 1403-1412.

Straus, D., Walter, W., and Gross, C.A. (1990). DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32. Genes Dev 4, 2202-2209.

Villalobos, A., Ness, J.E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006). Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics 7, 285.

Welch, M., Govindarajan, S., Ness, J.E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009). Design parameters to control synthetic gene expression in Escherichia coli. PLoS One 4, e7002.

Wu, C.K., Gousset, K., and Hughes, S.H. (2007). Targeting to the endoplasmic reticulum improves the folding of recombinant human telomerase reverse transcriptase. Protein Expr Purif 56, 8-19.

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F., and Chen, J.J.-L. (2008). Structure and function of the smallest vertebrate telomerase RNA from teleost fish. J Biol Chem 283, 2049-2059.

Chapter 3

PURIFICATION OF TELOMERASE RNA BINDING DOMAIN

3.1 ABSTRACT

Telomerase is composed of two essential core components, Telomerase RNA and Telomerase Reverse Transcriptase protein. In order to further understand the interaction between telomerase protein and RNA, an RNA binding domain of the medaka TERT protein is purified. An optimized 3 step purification is determined using salting out, affinity, and size exclusion chromatography yielding milligram quantities of TRBD. Purified TRBD is greater than 95% pure as estimated by SDS PAGE and monomeric as determined by gel filtration. Also described is a purification of a TRBD-CR4/CR5 complex with generation of 3' homogenous RNA achieved through the use of glms-ribozyme. Only upon complex formation can the ionic strength of the protein buffer be reduced without precipitation of TRBD.

3.2 INTRODUCTION

Eukaryotic chromosomes are linear DNA capped by telomeres that are maintained by the ribonucleoprotein telomerase (Greider and Blackburn, 1985). The core components required for *in vitro* activity are Telomerase Reverse Transcriptase (TERT), the catalytic protein portion, and Telomerase RNA (TR). Two portions of the vertebrate TR are required for activity: the pseudoknot and CR4/CR5 (Mitchell and Collins, 2000). While the function of CR4/CR5 remains elusive it is known to interact with the Telomerase RNA Binding Domain (TRBD) of TERT(Moriarty et al., 2004) a protein site with high binding affinity to TR (Lai et al., 2001).

Expression of recombinant TERT proteins in *E. coli* has met with limited success, especially with vertebrate TERTs. A soluble *Tetrahymena* TEN was identified via a high throughput GFP solubility screen (Jacobs et al., 2005). Following this the crystal structure of the TEN was solved (Jacobs et al., 2006). An inactive TRBD from *Tetrahymena* TERT was also successfully purified and crystallized (Rouda and Skordalakes, 2007). The full length TERT of the beetle *Tribolium castaneum* was also expressed and crystallized though the RNA component has yet to be identified (Gillis et al., 2008).

In order to purify medaka TRBD expressed using an *E. coli* host multiple expression strains were tested for optimal over expression. BL21 (DE3) is a common *E. coli* expression strain. As an *E. coli* B strain, BL21 lacks the *Lon* protease, a protease that degrades misfolded proteins. BL21 also has the *ompT* protease removed. *OmpT* is an endoprotease found on the outer membrane which

is problematic because after cell lysis *OmpT* degrades released proteins(Grodberg and Dunn, 1988). Scarab Genomics went beyond the removal of a couple of genes, engineering a K-12 *E. coli* strain with 15% of the genome removed (Posfai et al., 2006). They targeted genome deletions to remove non-essential DNA and genes of both unknown function or of known function but hypothesized to reduce growth, plasmid DNA amplification, and expression of recombinant proteins. Unlike other multiple deletion strains, MDS40 generated by Scarab genomics maintains a normal growth rate (Sharma et al., 2007).

Salting out of proteins by ammonium sulfate, used in MBP-TRBD purification, has long been a staple of protein purification (Foster et al., 1976). The technique was first observed in the late 1800s by Lewith and Hofmeister (Baldwin, 1996). Hofmeister determined that salts at low concentrations can have a stabilizing effect on proteins and can cause proteins to precipitate at high salt concentrations. Hofmeister ranked both anion and cations creating two series based on the stabilizing effects of salts on proteins. Precipitation by ammonium sulfate is generally considered a gentle precipitation and does not affect the protein folding as does aggregation induced precipitation.

Overexpression of TRBD as a maltose binding protein (MBP) fusion protein is described previously (see chapter 2). In addition to enhancing solubility MBP can also be used for affinity purification. MBP-fusion proteins can be purified with immobilized amylose a polysaccharide of glucose monomers. Elution from amylose resin is achieved with free maltose, a glucose dimer.

Amylose affinity purification is used as the second purification step of the optimized MBP-TRBD purification.

Gel filtration serves to separate molecules based on shape and size. Gel filtration resin contains porous beads of controlled sizes. Larger molecules, either overall size or more linear in shape, are less likely to enter the resin passing around the beads. Smaller more compact molecules enter into the resin taking a longer path through the column eluting later than large molecules. The column is unable to separate molecules that are too large ever enter the beads, these molecules all elute at the same volume corresponding to the smallest amount of liquid required to pass through the column. This volume is known as the void volume. Using resin of the correct size non-interacting molecules can be separated by gel filtration based on size.

3.3 MATERIALS AND METHOD

3.3.1 Cloning synthetic mdTERT TRBD-RT and TRBD-

DNA seqeunces encoding mdTERT fragments were PCR amplified using previously created synthetic mdTERT gene (see chapter 2). PCR amplification introduced 5' EcoR1 and 3' XhoI sites to PCR product. DNA was ligated to EcoR1-XhoI (NEB) digested pMAL-2CX (ASU88) or pMAL-TEV-2CX(ASU 538) as indicated. Ligation products were transformed to DH10B-T1R or Mach1 *E. coli* and sequenced to verify PCR introduced no mutations.

3.3.2 Small scale expression test of different colony-

One nanogram of mini-prep DNA was transformed to BL21 (DE3) (MBP-TRBD-RT) or ScarabXpress T7 lac (SGT7) (Scarab Genomics) (MBP-TRBD, MBP-TRBD-RT, and MBP only) and plated to carbenicillin containing LB plates. Three individual colonies of SGT7 cells for each MBP-TRBD-RT, MBP-TRBD (G318-S569) and MBP-TRBD (G318-G579) and a single colony of SGT& containing MBP only vector and a single colony f BL21 containing MBP-TRBD-RT were inoculated in 2ml LB+ 50μg/ml carbenicillin. Cultures were grown at 37^{o}C with shaking (200RPM) to OD_{600} 0.5 and induced with 1mM IPTG (GoldBio). Upon induction, cultures were transfered to a 20^{o}C refrigerated shaker incubator. After 4 hours a 200μl sample was removed, pelleted and resuspended in 50μl 4X SDS loading buffer (40% (v/v) glycerol, 20% (v/v) beta-mercaptoethanol, 250mM Tris-Cl pH 6.8, 8% (w/v) SDS 0.01% bromophenol blue), and then stored at -20^{o}C until used. After overnight incubation, a second sample was removed from culture as above. Samples were denatured for 10

minutes at 90°C and 5μl as loaded to an 8% acrylamide SDS gel. Gel was stained with BioSafe Coomassie stain (BioRad).

3.3.3 Ammonium sulfate titration

The pellet of a 250 ml culture of expressing MBP-TRBD (G318-S569) in SGT7 cells was resuspended in 10ml lysis buffer (50mM Tris-Cl pH 8.0, 1mM $MgCl_2$, 10% glycerol, 150mM NaCl, 5mM beta-mercaptoethanol, 1mM PMSF, 1X complete protease inhibitor cocktail (Roche)). The resuspended pellet was sonicated 4x for 30sec each time at power level 3, 50% duty cycle. Lysate was clarified at 18K RCF for 20 minutes. Clarified lysate was slowly brought to 20% ammonium sulfate saturation by the addition of 1.06g ammonium sulfate. The sample was centrifuged for 15min at 18K RCF and the pellet and supernatant separated. Ten milliliters of supernatant was raised to 35% ammonium sulfate saturation by the addition of 0.83g ammonium sulfate and separated as above. Procedure continued as above increasing to 50% ammonium sulfate and 65% ammonium sulfate in series by addition of 0.87g and 0.92g ammonium sulfate. An aliquot was removed from all soluble fractions and combined with 4X SDS loading buffer. Pellets were resuspended in 10ml lysis buffer less protease inhibitors and an aliquot taken and combined with 4X SDS loading buffer. SDS samples were denatured for 10 minutes at 90°C and separated by SDS PAGE.

3.3.4 MBP-TRBD generic purification (used unless otherwise indicated)

SGT7 cells harboring pMAL-tev-TRBD (G318-G579) were grown in LB + 50μg/ml carbenicillin + 2% glucose to OD_{600} 0.3, induced with 1mM IPTG and incubated for 4 hours at 20°C with shaking. The cell pellet was resuspended in

10ml TMN-500 buffer (50mM Tris-Cl pH 7.5, 500mM NaCl, 1mM TCEP, 1mM $MgCl_2$, 10% glycerol) per gram of pellet and supplemented with 1mM PMSF and 1X complete protease inhibitor (Roche). Cells were sonicated 4x for 30 seconds each time at power level 4, 50% duty cycle. Lysate was centrifuged at 18K RCF at 4^{o}C and the resulting supernatant was brought to 55% ammonium sulfate saturation and stirred for 30 minutes. A pellet of precipitated protein was collected by centrifugation at 18K RCF for 15 minutes, resuspended in TMN-500 buffer, then applied to 20ml of amylose resin (NEB). The resin was washed with >150ml column buffer, before protein was eluted into 1.5ml fractions by TMN-500 buffer supplemented with 10mM maltose. Peak fractions, as detected by UV absorbance, were pooled and applied to a HiLoad 16/60 Superdex 200pg gel filtration column. UV spectra of gel filtration fractions were measured on a Nanodrop ND-1000 using TMN-500 buffer as a blank. Concentration of protein was obtained by multiplying correcting Nanodrop measurement by 0.7 to correct extinction coefficient of MBP-TRBD as determined by Expasy.

3.3.5 Gel filtration calibration-

HiLoad 16/60 Superdex 200pg gel filtration column was calibrated using 7 standards of known molecular weight (Sigma MWGF1000). Standards were applied to column in three sets loading 1ml of protein mix per run to avoid overlapping peaks. Gel flirtation (GF) buffer (20mM Tris-Cl pH 7.4, 150mM NaCl, and 0.1mM DTT) was used for calibration runs. Blue Dextran (~2,000kDa) was loaded separately from all other standards at 2mg/ml in GF buffer. 3mg carbonic anhydrase (29kDa), 5mg alcohol dehydrogenase (150 kDa)

and 10 apoferritin were dissolved in 1ml GF buffer, and applied to column together. 10mg albumin (66 kDa), 4mg beta-amylase (200 kDa) and 8mg thyroglobulin (669 kDa) were dissolved in 1ml GF buffer and applied to column together. Elution volume of each marker was measured with blue dextran eluting in the void volume. Elution volume (V_e), void volume (V_o) and molecular weight of the standards was used to create a standard curve by plotting molecular weight against V_e/V_o on a semi log scale.

3.3.6 TEV digestion-

Protein concentrations of each sample were determined by Bradford assay (BioRad) using gamma-globulin as the standard. Due to different 260/280 ratios UV absorbance could not be used to determine protein concentration. 19μg of each sample was digested with 10 units acTEV (Invitrogen) in a 100μl digestion in 1X column buffer supplemented with 1mM fresh DTT. Samples were digested at 16^{o}C and room temperature (~25^{o}C) overnight. After digestion 20μl of sample was combined with 7μl 4X SDS loading buffer, monomeric protein digests were centrifuged for 15 minutes at 18K RCF and a sample of the supernatant was taken. Samples were heat denatured (90^{o}C 10 min) and 20μl was separated on a 12% acrylamide SDS gel and visualized with BioSafe Coomassie dye.

3.3.7 Sensitivity of MBP-TRBD to ionic strength-

2ml of pooled MBP-TRBD amylose elution (purified as described above in generic purification) was dialyzed 3x against low salt buffer (column buffer with 200mM NaCl) in 10,000 MWCO Snake Skin Dialysis Tubing (Pierce) before being applied to gel filtration column. Dialyzed sample was centrifuged at

18K RCF for 25 minutes to remove any precipitate and applied to gel filtration column in column buffer.

3.3.8 Sensitivity of MBP-TRBD-CR4/CR5 complex to ionic strength-

CR4/CR5 RNA for complex formation was transcribed with a 3' glms ribozyme and prepared as described in 3.3.9. MBP-TRBD (G_{318}-G_{579}) was purified as described in generic purification in 3.3.4. 0.84mg of MBP-TRBD, determined by UV absorbance, was combined with 0.23mg of CR4/CR5 in 10ml of column buffer (TMN-500) resulting in a 1:1.25 molar ratio. RNA and protein were incubated for 15mins at room temperature before concentrating to 7.5ml using a Millipore stir cell equipped with a Millipore YM-10 membrane. 2.5ml aliquots were placed in dialysis tubing (10kDa cut off, Pierce) and placed in 200ml TMN-500, TMN-300, of TMN-150 in which only the sodium chloride concentration differed. Samples were dialyzed overnight with one buffer change. 2ml of each sample was applied to gel filtration column (HiLoad 16/60 Superdex 200pg) equilibrated in appropriate buffer.

3.3.9 In vitro transcription and purification of glms-CR4/CR5

Medaka CR4-CR5 with an upstream T7 promoter was cloned into KpnI site of a glms vector. Template for transcription was created by PCR using primers specific to T7 promoter and the 3' end of the glms. CR4/CR5-glms RNA was transcribed in a 2x 1ml reactions in 1X HEPES buffer (80mM HEPES-KOH pH 7.5, 24mM MgCl2 2mM spermidine, 40mM DTT) containing 5mM each NTP, 20μg T7 RNA polymerase (purified in house), 7U pyrophosphatase and 10μg template. Transcription reaction was incubated at 37^{o}C for 4 hours,

centrifuged twice at 18K RCF retaining the supernatant each time and 2ml was applied to HiLoad 16/60 Superdex 200pg gel filtration column equilibrated in RNA column buffer (50mM HEPES-KOH pH 7.5, 150mM NaCl and 10mM MgCl2). Peak fractions containing full length CR4/CR5-glms (centered on 64ml elution volume) were pooled and sodium concentration brought to 0.3M with 3M sodium acetate. RNA was precipitated by adding an equal volume isopropanol and stored at -20^{o}C until use. RNA was pelleted in microcentrifuge tubes and resuspended in 1.8ml RNA column buffer; 0.2ml of 10mM glucosamine-6-phosphate was added and RNA was incubated at room temperature for 15 minutes to allow cleavage. RNA was then applied to HiLoad 16/60 Superdex 200pg equilibrated in TRBC column buffer. Peak CR4/CR5 fractions, centered around a 90ml elution volume, were pooled.

3.3.10 MBP-TRBD-CR4/CR5 complex separated on gel filtration-

MBP-TRBD was purified using the generic purification described in 3.3.4, and RNA transcribed and purified as above were combined in a 1:2 molar ratio of protein to RNA in a 400µl volume and incubated for 30 minutes at room temperature. Sample was loaded to gel filtration column (HiLoad 16/60 Superdex 200pg) and elution profile was compared to a mock binding in which buffer was substituted for RNA.

3.4 RESULTS

3.4.1 Cell type and expression time affects expression levels

The synthetic gene constructed in Chapter 2 was used to create expression constructs encoding fragments of mdTERT. The construct encoding TRBD-RT was transformed in both BL21 and SGT7 *E. coli* strains. When expressed from BL21, a commonly used overexpression strain, little to no overexpression was observed. In contrast, when expressed in SGT7 strain the overexpression band was the dominant band when the proteins were separated on an SDS gel (Figure 3.1 compare lane 2 and 4-6). The background bands of the two cell lines were highly similar. In SGT7 cells the mdTERT fragments were expressed at room temperature for 4 hours as well as overnight (Figure 3.1A vs. B). When expressed overnight, the overexpression band was more intense for the MBP-TRBD-RT fragment; however it did not markedly increase for the MBP-TRBD only fragment. Further analysis showed that the level of soluble protein was the same in the two cultures, despite more overexpression seen in the overnight culture.

3.4.2 Separate but not equal colonies

Plasmids encoding MBP-TRBD-RT, and two MBP-TRBD constructs were transformed to SGT7 cells. Three colonies of each construct were tested in a small scale expression culture. The three colonies appeared equivalent on the petri dish in shape and size, yet the resulting cultures were not equal. In two of the TRBD (G318-G569) cultures an extra overexpression band was observed. The extra band, of approximately 40kDa, was observed in both the 4 hour and

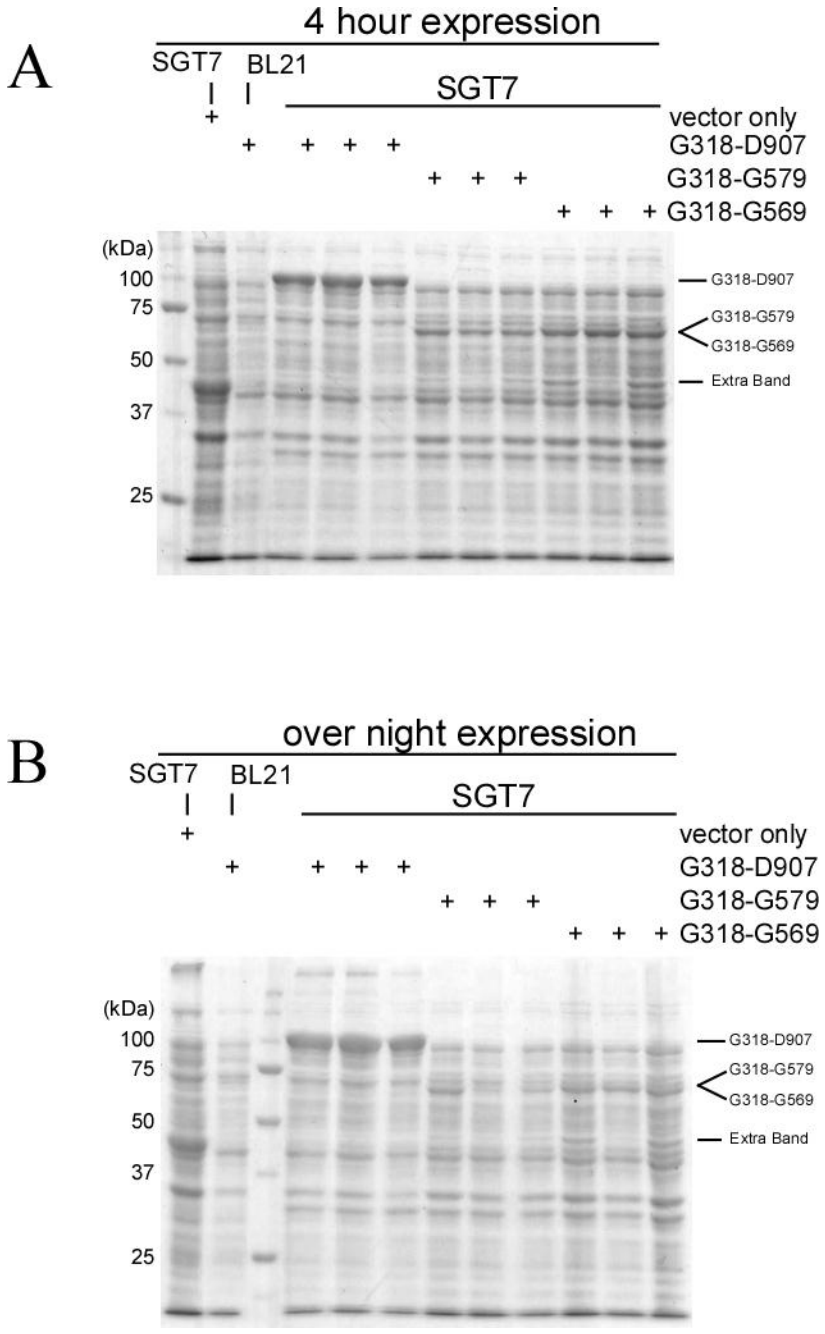


Figure 3.1- Separate but not equal expression cultures
Three colonies of each MBP-TRBD-RT (G318-D907), MBP-TRBD(G318-G579), and MBP-TRBD(G318-G569) in SGT7 cells were used to start expression cultures. After expression for 4 hours (A) or overnight (B) whole cell lysates were loaded to 8% acrylamide SDS gels. Bands corresponding to over expressed proteins are indicated as well as an "extra band" that was not similar between multiple cultures containing the same plasmid.

overnight expression tests. In the other TRBD fragment expressed (G318-G579), only one of the three cultures had an overexpressed band of appropriate size. The two non-expressing cultures grew at the same rate as the rest of the cultures while in the presence of antibiotic, indicating that the vector was present providing carbenicillin resistance yet there was not overexpression of the fusion tag only or a different fragment to indicate contamination. For further analysis, the cultures that did not exhibit extra bands but did exhibit the desired overexpression of were used. For all future *E. coli* expression tests when a new construct was transformed, multiple colonies were tested for expression after transformation. Only those cultures that did not show extra protein bands were used to create glycerol stocks.

3.4.2 Purification optimization: Ammonium Sulfate precipitation

Ammonium sulfate precipitation was tested as one step in the purification of MBP-TRBD. A clarified cell lysate was treated with a range of ammonium sulfate concentrations from 20-65% saturating ammonium sulfate. After mixing for half an hour to allow equilibration the sample was centrifuged and the pellet and soluble fractions were separated and analyzed on an SDS gel. Differential precipitation of separate proteins was not observed. In this test, all proteins appeared to start precipitating between 35% and 50% saturation and were nearly completely precipitated by 65% saturation with ammonium sulfate (Figure 3.2A). Further analysis showed the majority of MBP-TRBD precipitated with a single 55% saturation treatment with ammonium sulfate (Figure 3.2B). The ammonium sulfate precipitation did not remove a large number of proteins as most co-

precipitated with MBP-TRBD. A single 55% ammonium sulfate cut served to remove one contaminating band of approximately 35kDa that previously co-purified with MBP-TRBD (figure 3.2B).

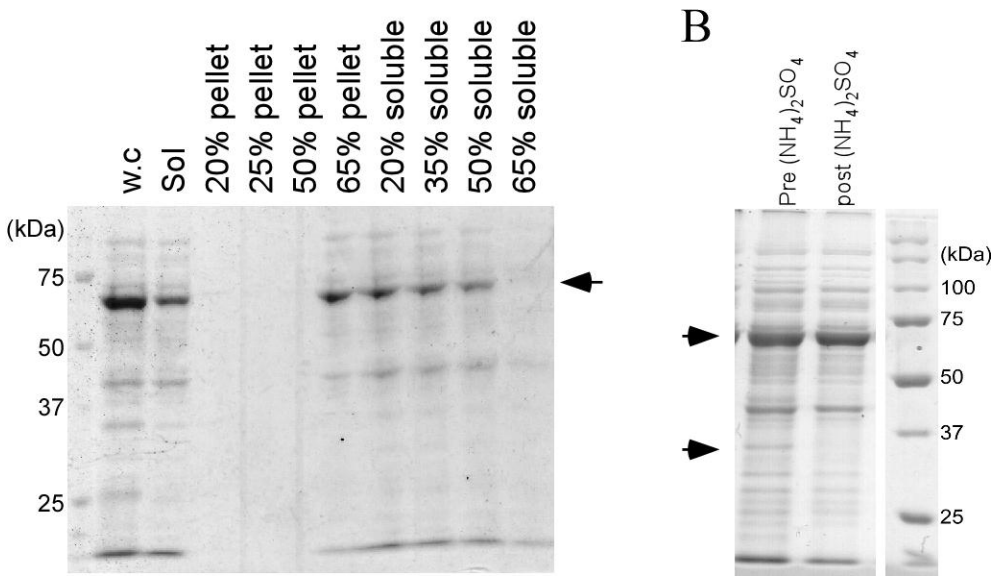


Figure 3.2- Ammonium sulfate precipitation of MBP-TRBD
(A) A titration of ammonium sulfate was performed from 20-65% saturation to precipitate MBP-TRBD (upper arrow both panels). The pellet fractions were free of protein as visualized by coomassie blue staining up to 50% ammonium sulfate. At 65% ammonium sulfate the supernatant is mostly depleted of proteins which resolve in the pellet fraction. (B) A 55% ammonium sulfate cut was taken and the pellet resuspened in lysis buffer, this procedure removed a single protein indicated by the lower arrow which otherwise co-purifies over amylose affinity column.

3.4.3 Purification optimization: Gel filtration separates monomer from soluble aggregate

When MBP-TRBD was separated on gel filtration two peaks were observed (Figure 3.3 A). One corresponding to the void volume, and the other matching the size of a monomer of MBP-TRBD compared to standard proteins

run over the column (Figure 3.4). When the UV spectra of the load and two elution peaks are compared, a noticeable difference is observed in the 260/280 ratio (Figure 3.3B), which is a measure of protein versus nucleic acid content. The load sample has a 260/280 of 1.04 while the void volume peak ratio is 1.24 and the monomer peak ratio is 0.55. A larger 260/280 ratio indicates a significant amount of contamination with nucleic acids, the peak absorbance of which is 254nm. The peak absorbance of proteins is dependent upon the amino acids it is composed of. Aromatic amino acids tend to absorb at 280nm. Pure protein, not contaminated by nucleic acids, will have a low 260/280 ratio as seen with the monomer peak (Figure 3.3B).

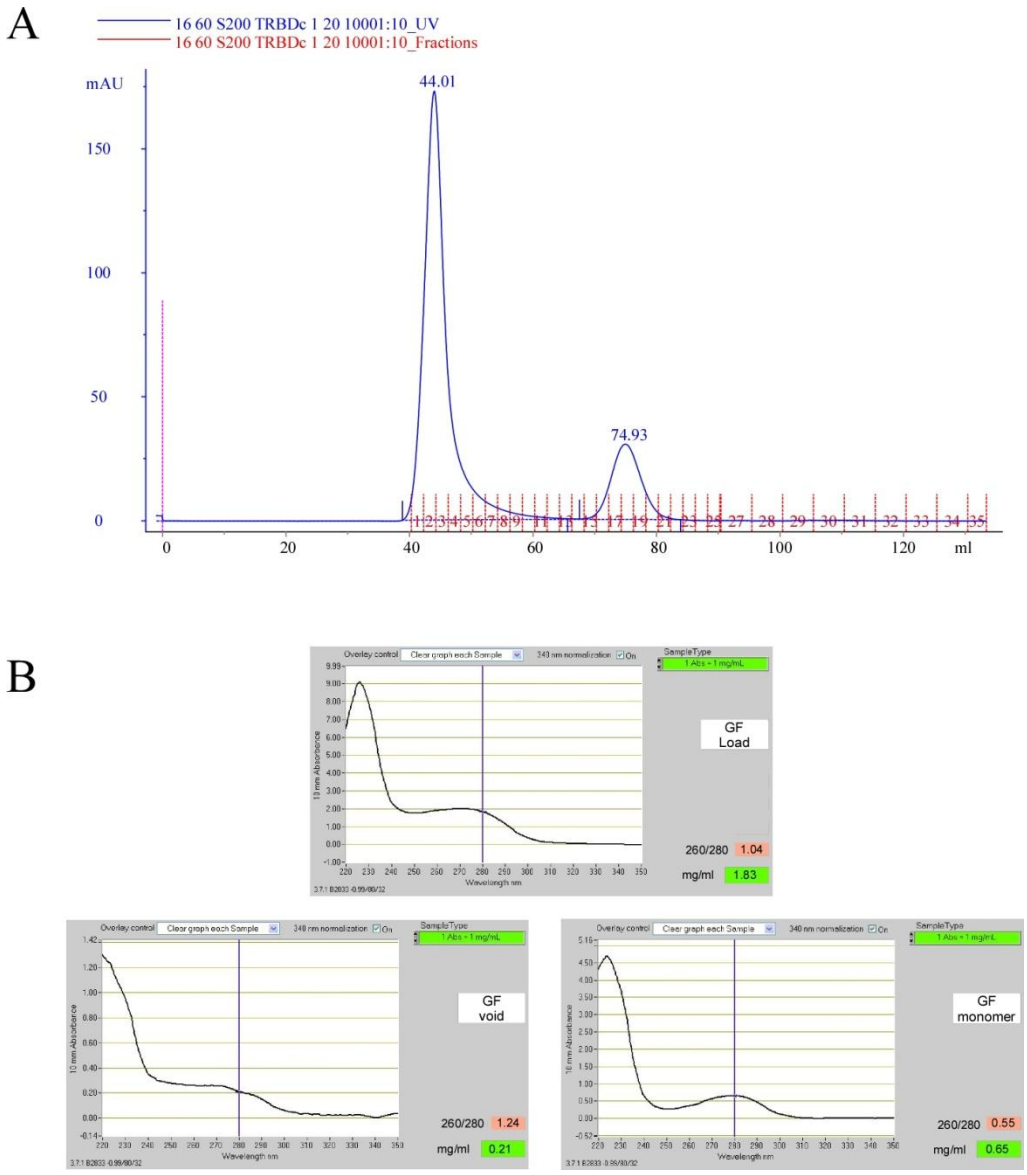


Figure 3.3-Gel filtration of MBP-TRBD

MBP-TRBD purified by ammonium sulfate precipitation followed by amylose affinity purification was applied to a 16/60 S200 gel filtration column. (A)Two peaks corresponding to the void volume (~44ml) and monomeric MBP-TRBD (~75ml) were observed. (B) Absorbance spectrum of the load sample and two elution peaks (post concentration) were measured on a nanodrop. The void volume had a high 260/280 ratio (1.24) and the monomeric fraction a low 260/280 ratio indicative of nucleic acid free protein (0.65).

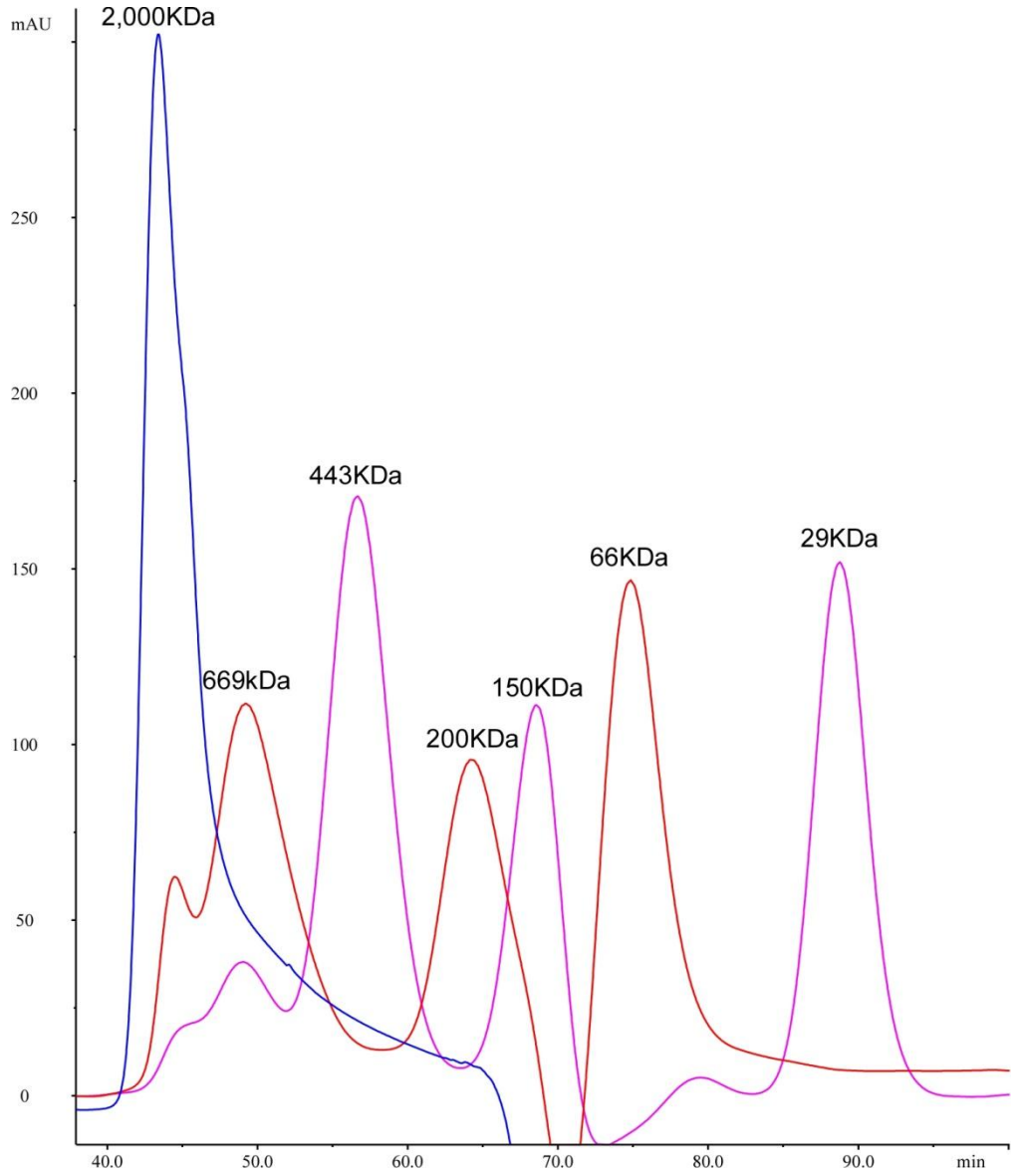


Figure 3.4- MW calibration of 16/60 S200 gel filtration column
A molecular weight standard (Sigma) was applied to 16/60 S200 column used for purification of MBP-TRBD. Blue dextran (~2MDa) eluted in the void volume of 43.3ml, thyroglobulin (669kDa) at 49.2ml, aproferritin (443 kDa) at 56.6 ml, beta-amylase (200kDa) at 64.2ml, alcohol dehydrogenase (150kDa) at 68.5, albumin (66kDa) at 74.8 ml, and carbonic anhydrase (29kDa) eluted at 88.7ml.

3.4.4 Monomeric MBP-TRBD is cleaved by TEV better than void volume MBP-TRBD

The MBP-TRBD, found in both the void volume fractions and the monomer fractions, remained soluble when spun at 18K RCF for extended periods of time. To see if the two samples of MBP-TRBD differed in more than just 260/280 ratio, the void and monomeric MBT-TRBD was digested with acTEV protease. A TEV site had previously been engineered between the MBP fusion tag and the medaka TRBD domain to allow release of TRBD from the MBP tag.

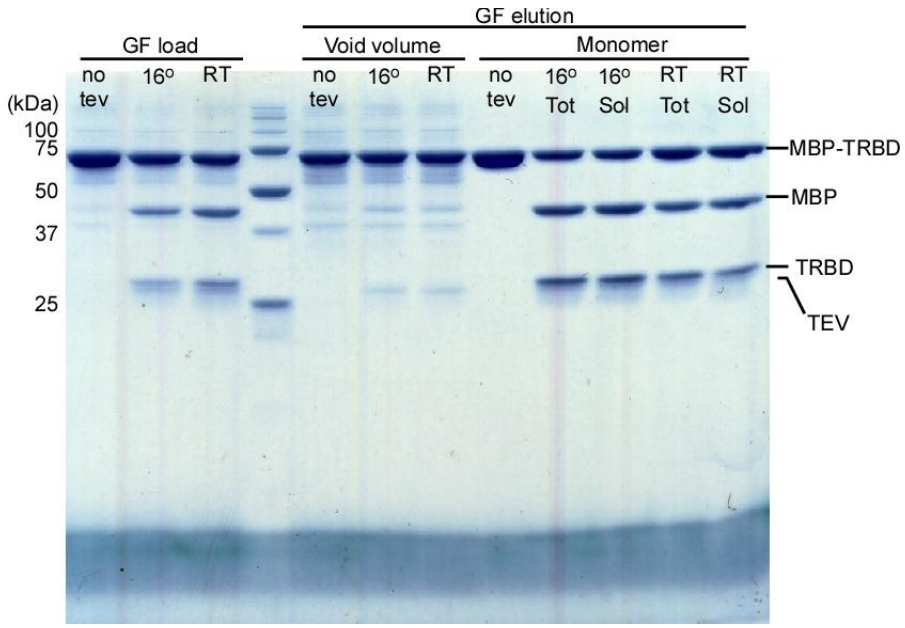


Figure 3.5- TEV digest monomer vs soluble aggregate
acTEV digestion of MBP-TRBD (seperates the MBP tag from TRBD) was performed on the sample loaded to the gel filtration (GF) column as well as samples from the void volume and monomer elution peaks. Digest was performed at 16°C and room temp, a noTEV sample was removed prior to acTEV addition and incubation. MBP-TRBD (74.2kDa), MBP fusion tag (43.8kDa), and free TRBD (30.4kDa) are indicated. TEV protease (27kDa) resolves directly below free TRBD.

MBP-TRBD samples from gel filtration load, void volume peak, and the monomer peak were digested overnight with acTEV at 16^{o}C and room temperature (Figure 3.5). Cleavage products from the monomer peak were tested for post cleavage solubility. The cleavage reactions were compared to mock digests for each protein fraction to identify the cleavage products. Intermediate cleavage was observed for the pre-gel filtration sample with a higher level of cleavage at room temperature. The protein found in the void volume had only a minor amount of cleavage, made most apparent by comparing the MBP band intensity of no TEV with the two cleavage lanes (Figure 3.5 lanes 5-7). The protein found in the monomer peak is cleaved significantly better than that in the load and void volume. A marked decrease in band intensity of the full length uncleaved protein (MBP-TRBD) and corresponding increase in cleavage products (TRBD and MBP) was observed for the monomer fraction (Figure 3.5 lanes 9-14). The monomeric protein exhibited more cleavage at 16^{o}C than at room temperature. It is possible that the protein is not stable at room temperature and begins to aggregate or misfold leading to lower cleavage.

The monomeric protein was tested for solubility post cleavage. At 16^{o}C the band intensity of TRBD is the same for total protein and soluble protein indicating the cleaved TRBD remained soluble. At room temperature however, there is a decrease in soluble TRBD but not cleaved MBP. A decrease in only the TRBD band indicates that a loading error is not responsible for the decrease of TRBD, but rather a solubility difference.

3.4.5 Monomeric MBP-TRBD remains monomer when re-applied to gel filtration column.

It is possible that the MBP-TRBD protein is in equilibrium between monomer and multimer species leading to the two peaks observed from gel filtration. To determine if the monomer would convert to multimer, either soluble aggregate or a more controlled multimeric state, sample from the monomer peak was incubated at 4^{o}C for one week and re-applied to the gel filtration column. If the protein maintains equilibrium between monomer and multimer, removing the multimer by gel filtration should drive conversion from the monomer to multimer to maintain equilibrium. However, the protein remained monomer (Figure 3.6) after a one week incubation at 4^{o}C, indicating that the monomer was stable, not in equilibrium with the multimer, nor prone to aggregation when stored for one week at 4^{o}C.

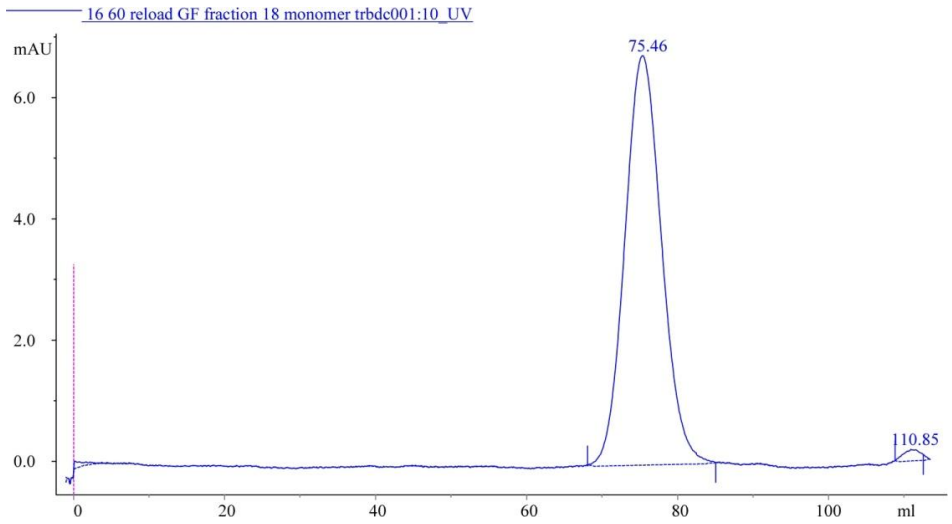


Figure 3.6- Gel filtration monomer remains monomer
MBP-TRBD monomer fraction from gel filtration was incubated at 4^{o}C for one week then reapplied to gel filtration column. Monomeric protein remained monomeric with no aggregation seen eluting in void volume.

3.4.6 Purification optimization: High salt is required for purification of MBP-TRBD monomer

During purification of MBP-TRBD, the buffers contained 500mM sodium chloride. This is hypertonic when compared to the intracellular sodium equivalent which is 170-240mM(Kao-Huang et al., 1977). Since higher ionic strength is known to interrupt some RNA-protein interactions, high ionic strength buffers would be non-ideal for studying the RNA interactions of TRBD. To determine if the sodium chloride level could be decreased, the post amylose column protein was dialyzed to 200mM sodium chloride buffer, separated by gel filtration, and compared to sample which was dialyzed against the original column. Dialysis to 200mM sodium chloride eliminated the monomer peak completely (Figure 3.7). For further studies of MBP-TRBD alone the monovalent salt concentration was maintained at 500mM in order to maintain the monomeric protein.

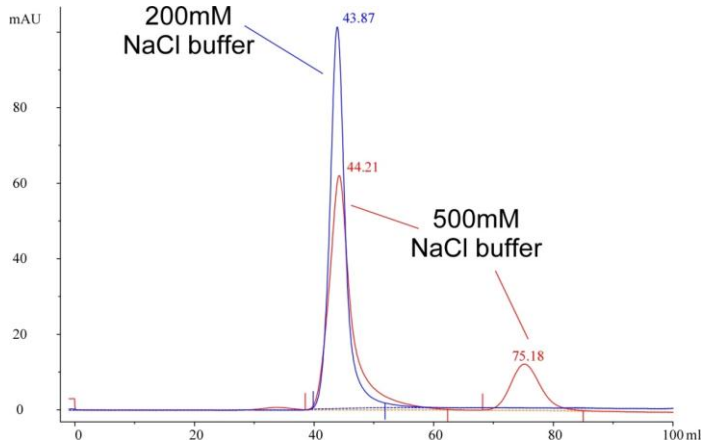


Figure 3.7-MBP-TRBD: High Salt is required to maintain monomer
MBP-TRBD purified by ammonium sulfate precipitation and amylose affinity chromatography was dialysed against low salt (200mM NaCl) or standard column buffer (500mM NaCl) and applied to gel filtration column. Removal of salt caused a shift of all protein to the void volume peak.

3.4.7 Purification optimization: High salt is not required after RNA protein complex formation

Though purified MBP-TRBD does not remain as a monomer when the monovalent salt is reduced from 500mM to 200mM, it is possible that the high level of negative charges found on the RNA could shield the protein during salt removal. A MBP-TRBD-CR4/CR5 complex was incubated to allow RNA-protein binding and then dialyzed against lower salt buffer. Three salt buffers were used: TMN-500, TMN-300, and TMN-150 differing only in the concentration of sodium chloride (500mM-150mM). The dialyzed samples were applied to a gel filtration column to look for aggregation as seen by an increase in sample eluting in the void volume (Figure 3.8). Sample in each of the three buffers eluted in two peaks, free RNA and as MBP-TRBD complex. Free RNA eluted around 90ml (87.2-89.7). The other elution peak observed, around 70ml, is composed of RNA-protein complex and any free protein remaining. The location of the peaks shift to eluting earlier as salt is removed, for example the complex peak was found at 73.8, 71.7, and 69.7ml in the three buffers TMN-500, TMN-300 and TMN-150 respectively (Figure 3.8). In the TMN150 sample, a minor void volume peak was observed. However, unlike RNA-free TRBD, upon dialysis to low salt the majority of the protein is found in the RNA-protein complex.

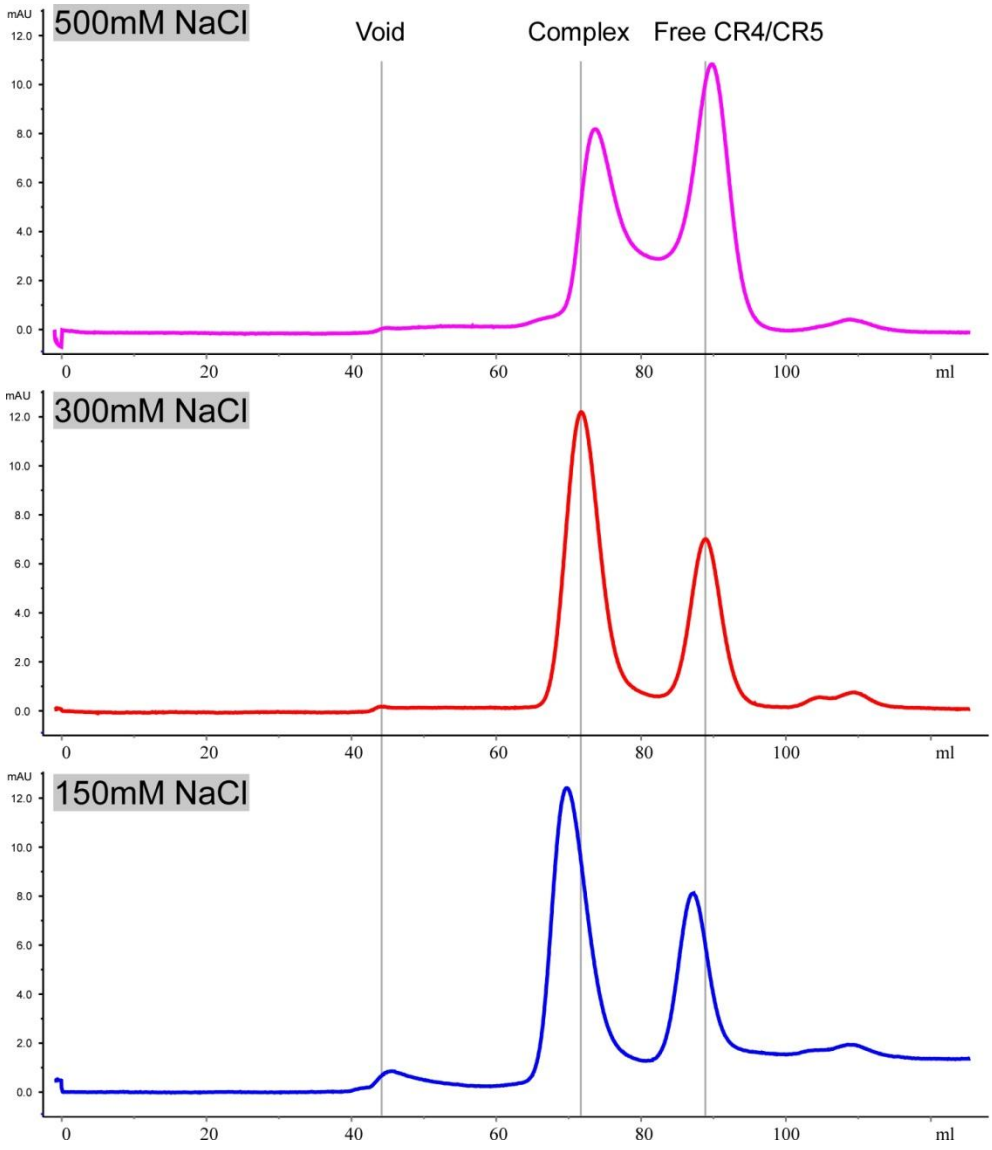


Figure 3.8-TRBD-CR4/CR5 High salt is not required to maintain monomer
Fully purified monomeric MBP-TRBD was bound to CR4/CR5 RNA concentrated, and dialyzed against TMN-500, TMN-300, and TMN-150 with reducing concentrations of sodium chloride. Dialyzed samples were applied to gel filtration column (16/60 S200). After addition of RNA only a minor portion of the protein shifted to the void volume compared to RNA-free protein (see figure 3.7). Two major elution peaks observed correspond to free CR4/CR5 and RNA-protein complex. A salt concentration dependent shift in all peaks is observed. Grey lines used as reference centered on TMN-300 peaks.

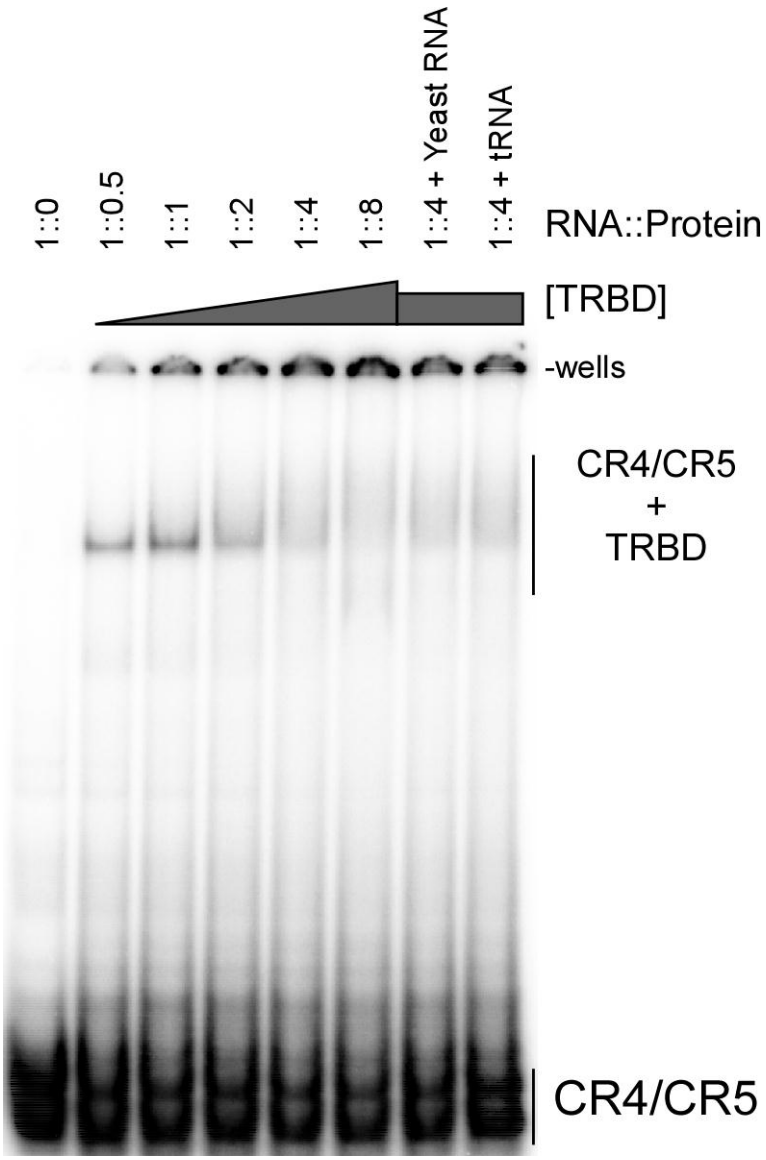


Figure 3.9-Excess of RNA required in MBP-TRBD-CR4/CR5 gel shift assay
MBP-TRBD and P32 labeled CR4/CR5 were bound and loaded to a native polyacrylamide gel. RNA level was held constant and bound by increasing amounts of protein. Radio labeled RNA was used to visualize products. Free CR4/CR5 is found at the bottom of the gel. CR4/CR5 complex formation retards the mobility of the bound RNA resulting in a shifted band. Complex is indicated. A significant amount of RNA was found in the wells of the gel and complex band began to smear when an excess of protein over RNA was used. Neither an excess of yeast total RNA nor tRNA resolved band smearing with excess protein.

MBP-TRBD-CR4/CR5 complex was analyzed by the electrophoretic mobility shift assay (Figure 3.9). An increasing amount of MBP-TRBD protein was bound to a constant concentration of P-32 labeled CR4/CR5 RNA. When CR4/CR5 was equimolar or in excess of MBP-TRBD, the complex resolved as a single band when run on a native gel. As the RNA became limiting, and protein excess, a smearing of the complex band appeared along with a significant amount of RNA retained in the wells of the gel (Figure 3.9). When RNA was run with buffer only, no RNA was retained in the wells. The 0.5X TBE running buffer used during gel shift is a low salt buffer compared to TMN-500. When protein is applied to the gel in absence of excess RNA, it appears to aggregate as seen in protein only gel filtration samples with reduced salt.

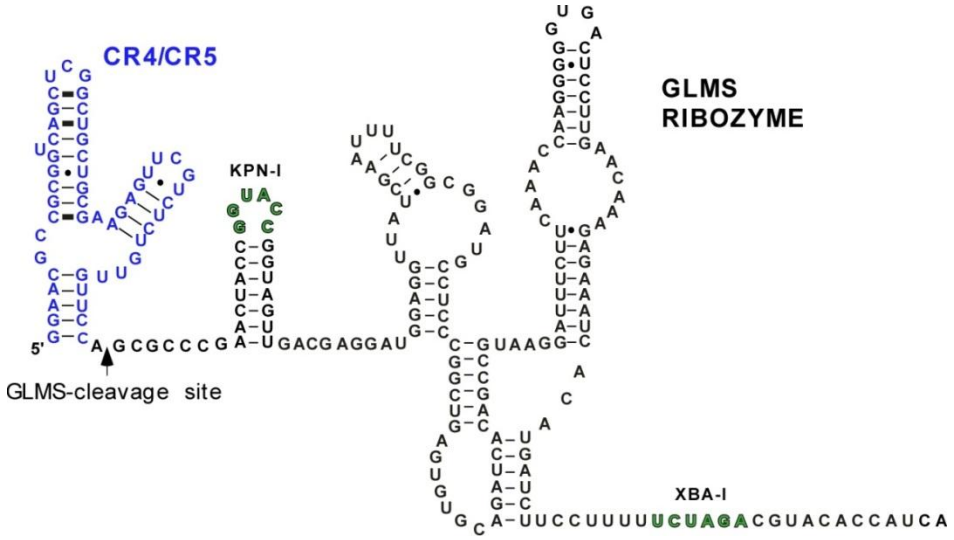


Figure 3.10- glms ribozyme secondary structure
Run off transcription with T7 RNA polymerase adds non template nucleotides to the 3' end of the transcript. CR4/CR5(blue) was transcribed *in cis* to GLMS ribozyme, a self-cleaving riboswitch in order to generate CR4/CR5 with a homogeneous 3' end. Hetoerogenity caused by T7 RNA polymerase occurs after the glms ribozyme, which is cleaved by the addition of glucosamine 6-phosphate. KpnI and XbaI sites indicated are used for cloning purposes. Cleavage occurs at the arrow leaving a single extra adenine at the 3' end of CR4/CR5 RNA.

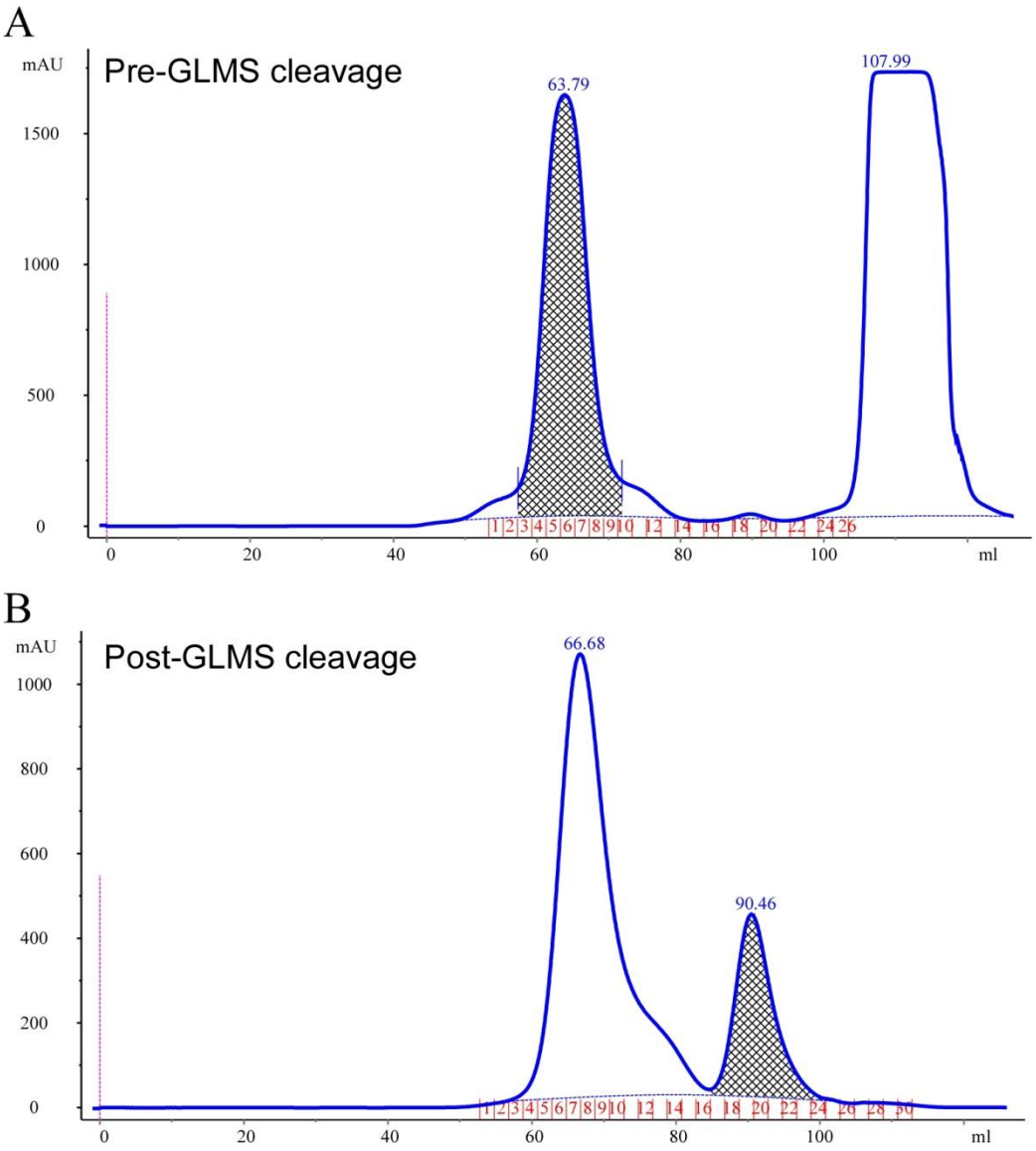


Figure 3.11-Purification of glms cleaved CR4/CR5

CR4/CR5 transcribed with a 3' glms ribozyme was purified by gel filtration. (A) Uncleaved RNA was purified by gel filtration in HEPES buffer. Fractions 3-9 containing uncleaved RNA were pooled, ethanol precipitated to concentrate, and resuspended in glucosamine-6-phospphate containing buffer. (B) Cleaved CR4/CR5 was separated from free glms and uncleaved CR4/CR5-glms by gel filtration in TRBD column buffer. Fractions 17-24 were pooled and used for binding to MBP-TRBD.

3.4.8 Purification optimization: Purification of MBP-TRBD-CR4/CR5 complex

One difficulty when attempting to crystalize RNA is the 3' heterogeneity of RNA as a result of *in vitro* transcription using T7 RNA polymerase (Price et al., 1995). Run off transcription using T7 RNA polymerase often results in the addition of non-template nucleotides at the 3' run off site (Milligan et al., 1987). In addition, many aborted transcription products are found containing only the first 2 to 6 nucleotides (Milligan et al., 1987). One method to remove the smaller abortive transcription products, and with limited success the larger products, is gel extraction of the newly transcribed RNA. Larger RNA fragments differing by only a couple nucleotides will not be removed by gel extraction due to the lower resolution of preparative gels. In addition, elution from polyacrylamide gels is relatively low yield, greatly hampering the recovery of RNA. Gel filtration followed directly by isopropanol precipitation, separates the small abortive RNA products with a higher yield than gel extraction by eliminating the need for passive elution from gel slices. As with gel extraction, gel filtration does not have the resolution to remove the 3' heterogeneic RNA which differ by only a couple nucleotides. To remove the 3' heterogeneity from CR4/CR5 RNA, a GLMS ribozyme was used *in cis* at the 3' end of CR4/CR5 RNA. The GLMS ribozyme is a self-cleaving ribozyme (Figure 3.10); cleavage is induced in the presence of glucosamine-6-phopshate. The only requirement of GLMS 5' of the cleavage site is a single adenine residue, any RNA can be placed 5' to this position without affecting the activity of the ribozyme. GLMS cleavage to produce 3' homogenous RNA has previously been described (Kieft and Batey, 2004). The

CR4-CR5-GLMS RNA was transcribed using T7 RNA polymerase and purified by gel filtration (Figure 3.11A), and the full length product was precipitated with isopropanol. The RNA was resuspended in the presence of glucosamine-6-phopsphate to induce cleavage and again separated via gel filtration (Figure 3.11B). The fractions containing cleaved CR4/CR5 were pooled and used for RNA-protein binding. Tris based buffers have been shown to induce cleavage by GLMS ribozyme necessitating the use of HEPES based buffers prior to induced cleavage (Kieft and Batey, 2004).

CR4/CR5-TRBD complex was concentrated and purified by gel filtration to remove free RNA. Due to the resolution of the column, unbound TRBD could not be removed. For this reason excess RNA was added during the binding reaction. The RNA-protein complex eluted from the gel filtration column earlier than protein only, as expected due to the increased size. Samples from the peak suspected to contain complex were analyzed by SDS PAGE and urea denaturing PAGE to verify co-purification of protein and RNA. Both the RNA and protein levels peaked between fractions 16 and 17 compared to gel filtration of protein only, which peaked at fraction 18 (Figure 3.8B).

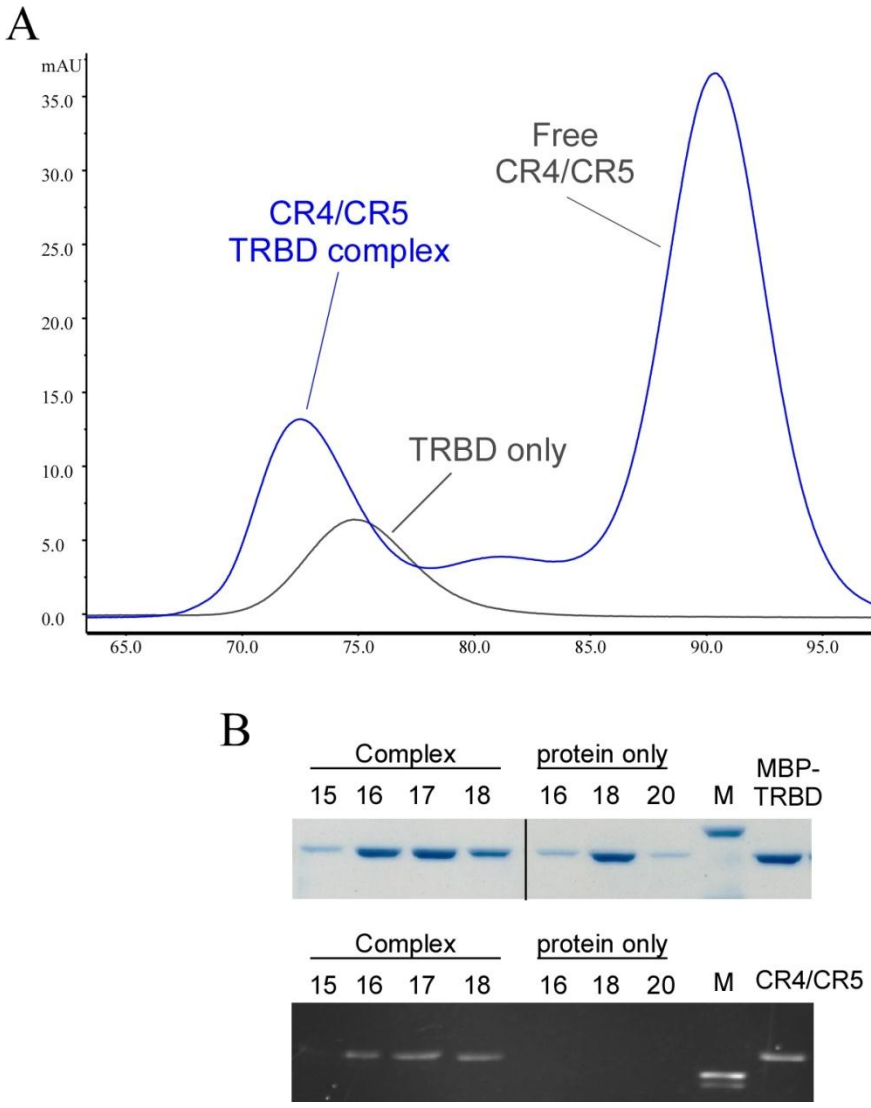


Figure 3.12 Gel filtration purification of MBP-TRBD-CR4/CR5 complex
MBP-TRBD was incubated with excess CR4/CR5 generated by glms-cleavage. (A) Protein-RNA complex (Blue) was analyzed by gel filtration and compared to UV absorbance of protein only (Grey). (B) Fraction of the first gel filtration peak of each run (~72.5 and 75ml) were analyzed by SDS-PAGE and urea denaturing PAGE to separately visualize protein and RNA content of each fraction. M indicates molecular weight marker, MBP-TRBD and CR4/CR5 indicate load of free protein or RNA as a marker.

3.5 DISCUSSION

Creation of a synthetic gene, to remove rare codons expanded the available host strains beyond Rosetta 2. SGT7 was engineered by Scarab Genomic to remove excess regions of the genome including non-essential genes. 14.7% of the genome was removed including 704 genes(Posfai et al., 2006). Use of the SGT7 cell line greatly increased the expression level of MBP-TRBD-RT fusion protein indicating that when faced with low expression, a simple test is to sample various expression strains. It is most likely that the TRBD-RT fragment is somewhat toxic to the *E. coli* leading to the low expression level in BL21 as SGT7 tolerates toxic proteins better(Posfai et al., 2006).

Expression cultures grown from different *E. coli* colonies containing the same plasmid were found to show different expression profiles. When testing three cultures apparently equivalent in all ways (transformed at the same time, picked from the same plate, expressed in the same LB and antibiotics with the same culture size temperature and shaking), the resulting overexpression was different. In one case, some of the cultures an extra band while in another no overexpression was observed. It is possible this discrepancy arises from dirty test tubes used for the cultures. The test tubes used could contain a small amount of residual detergent from washing which could stress the cultures. This stress could result in expression of stress induced proteins such as heat shock proteins. While this explanation cannot be ruled out, it fails to explain the lack of expression exhibited by the second set of TRBD cultures. Since a single colony is used to initiate overexpression cultures, and that colony arose from a single bacteria being

transformed with plasmid, a genomic bottleneck event occurred. This genomic bottleneck could lead to different background mutations in the *E. coli* of different colonies. Different *E. coli* background mutations would lead to varied expression of *E. coli* genes resulting in different banding patters. It is also possible that for one reason or another there is more protease activity in some cultures than in others, and the extra band is due to protease cleavage of the target protein. The underlying cause of the observed expression differences was not determined; the ultimate goal was overexpression of the mdTERT fragment and purification for biochemical analysis. To this end, it is beneficial to start from the culture with the fewest extra protein bands and highest expression level. For this reason expression of multiple transformation colonies has been adopted as standard practice for future constructs. During further expression using these constructs the extra bands did not appear, indicating that an individual culture with optimal expression can be kept for further use without having to re run the test each time the glycerol stock is streaked out.

One common method used during protein purification is salting out with ammonium sulfate. The typical method is to take a cut below the precipitation level of the protein of interest, then performing a cut that precipitates the protein of interest. The dual precipitation method serves to remove those proteins with lower solubility in the pellet of the initial cut, and those with a much higher solubility with the supernatant of the second cut. In the case of MBP-TRBD, most proteins precipitated at about 55% saturation of ammonium sulfate along with the target protein. Performing a cut below 55% served to remove a protein

that consistently co-purified with MBP-TRBD. It is to remove this ~35 kDa protein that the ammonium sulfate precipitation continues to be a part of the purification scheme of MBP-TRBD. It is likely that in a different buffer with lower sodium chloride levels, differential precipitation of background bands would be seen.

When separated on gel filtration, the MBP-TRBD protein sample segregates into two peaks which match the void volume and MBP-TRBD monomer. The void volume peak represents a larger portion of the UV absorbance. When the area under the peaks is integrated, the monomer peak only accounts for 10-20% (depending on the purification run) of the total UV absorbance seen. However, the monomer peak does not contain the smaller fraction of the protein (Figure 3.13 compare lanes S200 Void and S200 monomer). The void volume contains a significant amount of nucleic acids, either DNA or RNA, which contribute to the void volume peak. As TRBD is the RNA binding domain of telomerase it is not surprising to see co-purification of nonspecific nucleic acids. It is not known if the protein found in the void peak is a soluble aggregate that has RNA or DNA trapped or if monomeric TRBD non-specifically bound to DNA or RNA is shifted to the void volume due to the large size of nucleic acids. The gel filtration step of purifying MBP-TRBD removes multiple contaminants which co-segregate to the void volume. Multiple smaller proteins are visible in the lanes from the gel filtration load and void peak on SDS gel, but not found in the monomer peak (Figure 3.5 compare no TEV lanes). Simple dilution beyond the sensitivity of coomassie stain cannot explain the

removal of these proteins as they are visible in the void volume peak despite a lower intensity of the MBP-TRBD band. When the protein is concentrated and run on a gel these bands are still absent indicating they are removed by gel filtration.

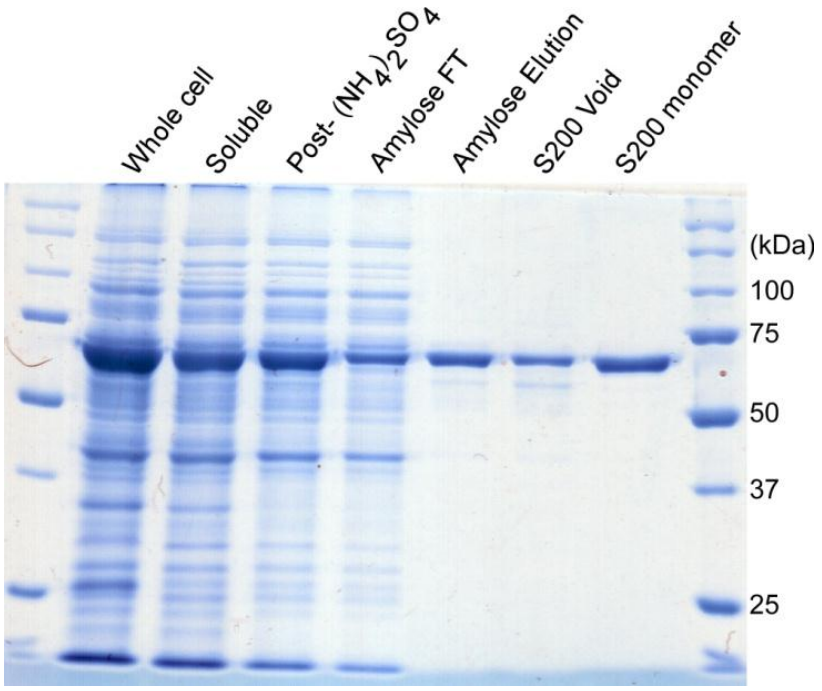


Figure 3.13-Complete purification of MBP-TRBD yeilds >95% purity
SDS-gel showing protein at each step of final optimized purification. Whole cell is post lysis, pre-clarification. Soluble indicates protein after centrifugation. Post ammonium sulfate is protein resuspended from 55% saturation ammonium sulfate pellet which is loaded to an amylose column. Amylose FT indicates protein which does not bind amylose resin, elution is pooled protein to be applied to gel filtration column (S200). Gel filtration results in two peaks, one at void volume the other corresponding to monomeric MBP-TRBD.

Not only does the gel filtration column segregate the protein in monomer and large complex corresponding to nucleic acid free and nucleic acid containing fractions but also segregates the protein into TEV sensitive and TEV resistant species. The resistance of the protein in the void volume toward TEV is indicative of misfolded or aggregated protein in which the TEV site is not accessible. The MBP fusion tag is a solubility enhancer of carried proteins (Kapust and Waugh, 1999). One concern in using a solubility enhancer is that it can keep an otherwise misfolded protein soluble. Retention of cleaved TRBD is solution after removal of the MBP tag indicates that TRBD is not a misfolded protein kept in solution only by fusion to the solubility enhancer. It is possible that the MBP tag kept the TRBD from aggregating long enough for TRBD to adopt a soluble folding confirmation. Once in the soluble state, it appears to be stable. The decreased solubility of TRBD after room temperature TEV cleavage indicates the freed TRBD protein can unfold during extended room temperature incubation. In future crystallization trials with MBP-TRBD it is important to know that TRBD has decreased solubility at room temperature but more stable at 16^{o}C.

MBP-TRBD bound to CR4/CR5 RNA appears to be more stable than free protein. Pre-incubation with RNA shields the MBP-TRBD protein from forming a soluble aggregate upon removal of salts. A similar phenomenon is seen when performing the electrophoretic mobility shift assay with MBP-TRBD and CR4/CR5. A single discreet band for bound protein is only visible when CR4/CR5 is in molar excess to TRBD. When CR4/CR5 is limiting, the shifted

band containing the complex begins to smear. Also, the amount of RNA in the wells increases, indicating the RNA is trapped in the wells by precipitated protein. An alternate explanation to protein precipitation is that multiple proteins bind a single RNA. A mixture of RNA-protein complexes of different ratios would results in multiple bands which could appear as a smear. However, this is an unlikely explanation as a large amount of free RNA remains unbound. Given a large amount of free RNA, the excess free protein is more likely to encounter unbound RNA rather than protein bound RNA.

As the salt concentration is decreased, the gel filtration elution volume of the MBP-TRBD-CR4/CR5 complex decreases (elutes earlier). It is possible this shift is due to less of the protein being found in the RNA-free state. Binding of CR4/CR5 to MBP-TRBD does not shift the elution volume of the complex far enough from the free protein for the peaks to not overlap. Any free protein will increase UV absorption in the trailing end of the complex peak, giving rise to an apparent shift in peak elution volume. A change in the ratio of RNA-bound and RNA-free protein would not explain the concurrent shift in free CR4/CR5 elution volume. It is more likely that the dynamics of the gel filtration column are shifted slightly by the decrease in monovalent salt.

The purification steps of MBP-TRBD were investigated and an optimal purification scheme for MBP-TRBD and MBP-TRBD-CR4/CR5 complex were determined. The final purification of protein requires high salt, 500mM sodium chloride, in order to maintain monomeric protein after gel filtration. A single ammonium sulfate cut is used to remove a ~35kDa protein that otherwise co-

purifies with MBP-TRBD. Amylose affinity purification enriches the MBP-TRBD but the protein eluted is a mixture of free protein and protein either bound to large polynucleotides or aggregated. Gel filtration separates monomeric MBP-TRBD from all other significant contaminants, including nucleic acid or protein. Using the protein purification scheme devised, greater than 95% purity was achieved using a three step purification yielding approximately 5mg MBP-TRBD per liter of culture.

REFERENCES

Baldwin, R.L. (1996). How Hofmeister ion interactions affect protein stability. Biophys J 71, 2056-2063.

Foster, P.R., Dunnill, P., and Lilly, M.D. (1976). The kinetics of protein salting-out: precipitation of yeast enzymes by ammonium sulfate. Biotechnol Bioeng 18, 545-580.

Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the Tribolium castaneum telomerase catalytic subunit TERT. Nature 455, 633-637.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405-413.

Grodberg, J., and Dunn, J.J. (1988). ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. J Bacteriol 170, 1245-1253.

Jacobs, S.A., Podell, E.R., and Cech, T.R. (2006). Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol 13, 218-225.

Jacobs, S.A., Podell, E.R., Wuttke, D.S., and Cech, T.R. (2005). Soluble domains of telomerase reverse transcriptase identified by high-throughput screening. Protein Sci 14, 2051-2058.

Kao-Huang, Y., Revzin, A., Butler, A.P., O'Conner, P., Noble, D.W., and von Hippel, P.H. (1977). Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. Proc Natl Acad Sci U S A 74, 4228-4232.

Kapust, R.B., and Waugh, D.S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8, 1668-1674.

Kieft, J.S., and Batey, R.T. (2004). A general method for rapid and nondenaturing purification of RNAs. RNA 10, 988-995.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol Cell Biol 21, 990-1000.

Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15, 8783-8798.

Mitchell, J.R., and Collins, K. (2000). Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. Mol Cell 6, 361-371.

Moriarty, T.J., Marie-Egyptienne, D.T., and Autexier, C. (2004). Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. Mol Cell Biol 24, 3720-3733.

Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., de Arruda, M., et al. (2006). Emergent properties of reduced-genome Escherichia coli. Science 312, 1044-1046.

Price, S.R., Ito, N., Oubridge, C., Avis, J.M., and Nagai, K. (1995). Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. J Mol Biol 249, 398-408.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure 15, 1403-1412.

Sharma, S.S., Blattner, F.R., and Harcum, S.W. (2007). Recombinant protein production in an Escherichia coli reduced genome strain. Metab Eng 9, 133-141.

Chapter 4

IDENTIFICATION OF RNA- PROTEIN INTERACTION INTERFACE

4.1 ABSTRACT

Telomerase is a specialized reverse transcriptase composed of two essential components: a template containing Telomerase RNA and the catalytic Telomerase Reverse Transcriptase protein. Two essential pieces of Telomerase RNA, the pseudoknot and CR4/CR5, can act *in trans* and are sufficient when combined with TERT to reconstitute activity *in vitro*. The active telomerase complex requires correct assembly of TERT and TR, identifying the interaction interface on both the protein and RNA is essential to understanding complex assembly. To facilitate identification of interaction interface, photo cross-linking was used to map RNA and protein residues in close proximity within the interaction interfaces of TRBD and CR4/CR5. Alkaline hydrolysis was used to generate RNA cleavage fragments and map the RNA cross-linking sites. A combination of MALDI-TOF and MALDI-TOF/TOF combined with RNase digestions and alkaline hydrolysis allowed the amino acids involved in cross-linking to be identified. This data provides an insight in understanding the function of essential telomerase RNA components and the methods developed could be applied to study other RNA-protein interactions.

4.2 INTRODUCTION

Telomerase is a ribonucleoprotein that regenerates telomeres, which serve to cap linear chromosomes (Greider and Blackburn, 1985). Without active telomerase, telomeres shorten with each replication cycle. Telomere shortening serves as a natural molecular clock limiting the number of times a cells genomic material can be copied, thereby limiting the replicative capacity of telomerase negative cells. Ninety percent of tumors have telomerase activity allowing the tumor to bypass this molecular clock (Kim et al., 1994).

Telomerase protein and RNA combine in a number of intricate interactions in order to form the active enzyme. Two regions of the telomerase RNA are required for activity, the pseudoknot and CR4/CR5 region. These two RNA regions each bind to separate regions of the TERT protein (Moriarty et al., 2002). CR4/CR5 is a highly studied, yet poorly understood RNA domain composed of p5, p6 and p6.1. Disruption of p6.1 base pairing inhibits binding of CR4/CR5 to TERT, while extending the p6.1 stem and point mutations in the p6.1 loop interrupt activity without affecting binding (Chen et al., 2002b). In the absence of protein the p6.1 loop was found to cross-link to the 5' end of the template in pseudoknot indicating an interaction between the two essential RNA domains (Ueda and Roberts, 2004).

Two separate RNA Interaction Domains (RIDs) within TERT have been identified. RID1 is defined in human TERT as amino acids 30-159 (Moriarty et al., 2002). RID1 has binding affinity to the telomerase pseudoknot (Moriarty et al., 2004). The second RNA interaction domain, also known as Telomerase RNA

Binding Domain (TRBD), is comprised of a discontinuous regions of amino acids 250-299 and 481-547 in human TERT (Moriarty et al., 2002). TRBD is responsible for high affinity binding of TERT to TR (Lai et al., 2001). TRBD has been shown to interact with the enzyme activating CR4/CR5 RNA (Moriarty et al., 2004) (see Figure 4.1). TRBD has also been implicated in binding to the pseudoknot, though with lesser affinity.

The photo-reactive-cross-linking reagents 5-Iodo-Uridine (5IU), 5-Iodocytosine (5IC), and 4-Thiouridine (4SU) were used to randomly label CR4/CR5 by inclusion of 10% photo-cross-linker in the T7 transcription reaction. Photo-cross-linker labeled RNA was used for cross-linking from RNA to protein. As opposed to shortwave length UV induced cross-linking, photo-reactive cross-linking reagents are excited at longer wave lengths (>300nm vs 254nm). Peak absorbance of 4SU is at 330nm while 5IU and 5IC peak absorbance is around 290nm though excitation can be achieved even up to 325nm (Meisenheimer and Koch, 1997). At longer wavelengths less UV induced damage is endured by the protein and RNA due to the lack of chromophores absorbing in this range (Willis et al., 1993). The photo cross-linkers used do not interfere with base pairing, and limited substitution does not affect RNA folding or affinity toward proteins. However, multiple substitutions with 5IU or 5IC can decrease interaction of labeled RNA toward protein (Meisenheimer and Koch, 1997). Cross-linking often occurs with aromatic amino acids thought to stack with the labeled nucleotide.

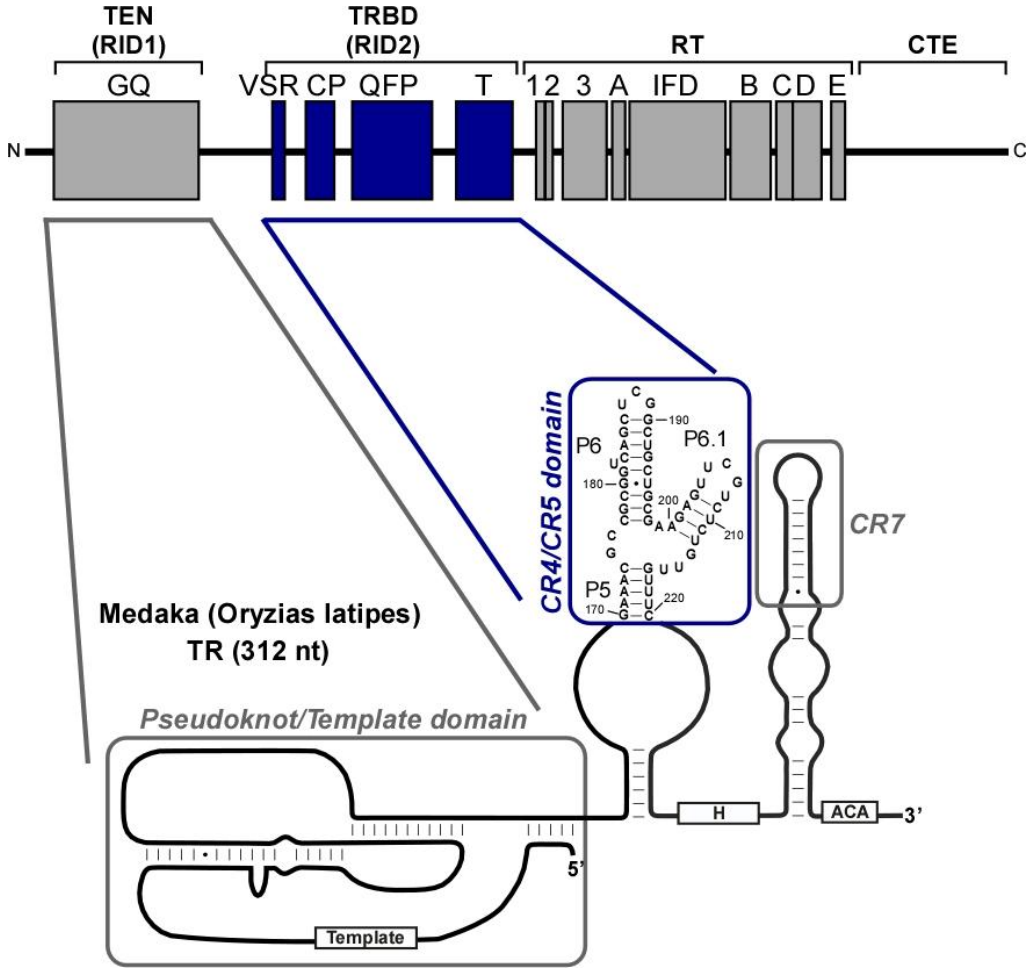


Figure 4.1- Domain structure of medaka TERT and TR
Top: Domains and motifs of medaka TERT are shown, Telomerase essential N-terminal domain is thought to bind the pseudoknot region of the TR while TRBD interacts with the CR4/CR5 of TR. Motifs 1-E compose the catalytic reverse transcriptase domain followed by the C-terminal thumb domain. Bottom: Secondary structure of medaka TR. The sequence of CR4/CR5 which is used in this study is shown.

4.3 MATERIALS AND METHODS

4.3.1 Protein expression and purification- A synthetic codon optimized TRBD was cloned to pMAL-2CX vector modified to insert a TEV cleavage site between MBP and TRBD and result in N-terminal MBP and C-terminal 6xhis tags. Protein was expressed in ScarabXpress® T7lac cells. Cells were lysed by sonication in 50mM Tris-Cl pH 7.5, 500mM NaCl, 1mM MgCl2 10% glycerol and 1mM TCEP 1mM PMSF 1X complete protease inhibitor (Roche). Lysate was brought to 55% ammonium sulfate saturation and pelleted at 18K RCF for 15 minutes. The pellet was resuspended in column buffer (lysis buffer less protease inhibitors) and applied to amylose resin (New England Bio labs). Protein was eluted in column buffer supplemented with 10mM maltose. Peak fractions were pool and applied to gel filtration column (sephadex? 200, GE). Protein eluted as a monomer was used for further studies.

4.3.2 Gel shift assay CR4/CR5- 0.8μM 5-P32 labeled CR4/CR5 was incubated with and without 0.8μM MBP-TRBD (G318-G579) for 30 minutes at room temperature in 20μl of TMN-500 buffer. CR4/CR5 was heated to 70^{o} for 3 minutes and snap cooled on ice 2 minutes before use to promote hairpin folding. Ten microliters of each binding reaction was loaded to a 5% polyacrylamide native gel (5%-polyacrylamide, 0.5X TBE, 5% glycerol, 1mM $MgCl_2$, prerun for 30 minutes). Gel was run at $4^{o}C$ for 2 hours at 10V/cm. Gel was dried and exposed to a phosphor screen for imaging.

4.3.3 Gel shift assay Pseudoknot- End labeled pseudoknot RNA was prepared by *in vitro* transcription and 5' P32 phosphorylated using standard reactions. 5 ng of

pseudoknot RNA (nt 1-97, 33-97, or 1-149) were incubated with 0.8μM MBP-TRBD in TMN-500 buffer for 30 minutes at room temperature in the presence of 1μg tRNA. For top strand RNA experiments 5' ACAAAAAAAGCCAGAAAA-3' RNA was added at 10μM final concentration. Ten microliters of each sample was loaded to a native gel and run as above except for 2.5hours instead of 2 hours.

4.3.4 Alkaline hydrolysis- RNA was in vitro transcribed in the presence of 10% 5-Iodo-Uridine triphosphate (Trilink Biotechnologies) and 90% UTP. RNA was gel extracted and 5' P32 labeled by standard kinasing reaction or 3' PcP labeled (described below). RNA (2ng) was combined with protein (~23μg) in 50μl of column buffer (50mM Tris-Cl pH 7.5, 500mM NaCl, 1mM $MgCl_2$, 1mM TCEP and 10% glycerol) and incubated 30 minutes. Sample was pipetted into a petri dish which was placed on an aluminum block in ice. Sample was irradiated with UV light using a 302nm handheld light through a petri dish lid for 2 hours. Sample was run on an 8% SDS gel. The gel was exposed to X-ray film and the gel slice with the desired band was excised. Gel slice was washed 3 times for 5 minutes with 1ml water. Gel slice was placed in 100μl of 100mM sodium carbonate pH 9.5, left at room temperature for 5 minutes with occasional mixing then incubated at 90^{o}C for 10 minutes. Sample was returned to room temperature and 200μl 100mM Tris-Cl pH7.5 and 300μl phenol chloroform was added. Sample was allowed to elute overnight and ethanol precipitated. Sample was resuspended in 20μl 2X formamide loading buffer and separated on a 10% sequencing gel along with RNaseT1 and alkaline hydrolyzed free RNA to serve as ladders.

4.3.5 P-32 labeling RNA at 3' end for alkaline hydrolysis-RNA was 3' labeled through pCp labeling reaction. P-32 pCp was synthesized by combining 2µl 20mM 3'cytosine monophosphate, 6µl γ-^{32}P ATP, and 10U t4 polynucleotide kinase (PNK) in a 10µl reaction with 1X PNK buffer (NEB). Sample was incubated for 5 hours at 37^{o} before heat inactivation of kinase at 80^{o}C. P-32 pCp was stored for future use at -20^{o}C. pCp labeling reaction was set up with 2.5nmol of RNA, 4 nmol ^{32}pCp, 10% DMSO, and 10U T4 RNA ligase were combined in 15µl 1x T4 RNA ligase buffer and incubated at 4^{o} overnight. Reaction was brought to 50µl and free ^{32}pCp removed using an oligo quick spin column (Roche).

4.3.6 MALDI-TOF MS- RNA and protein (MBP-TRBD-6xhis), each ~3.8µM, were combined in 1-2 ml of column buffer. Sample was irradiated for 2 hours with UV light as above. Sample was denatured with 3CV denaturing buffer A (100mM sodium phosphate pH 8.0, 10mM Tris, 250mM NaCl, 0.1% Triton X-100, 8M urea) and combined with 400µl pre-washed Ni-NTA beads (Qiagen). Sample was allowed to bind overnight with mixing then washed 3x with denaturing buffer A. Ni-NTA resin was washed twice with water and resuspended in trypsin buffer (25mM ammonium bicarbonate 1M urea). Twenty micrograms of sequencing grade trypsin (Promega) was added and sample was incubated overnight at 37^{o}C with mixing. Ni-NTA was pelleted at 1K RCF and the supernatant was collected. Resin was washed 2X with 0.5ml trypsin buffer collecting the supernatant each time. The supernatants were combined with 2X formamide loading dye and loaded to a 4% acrylamide 1X TBE gel, free RNA

was also loaded on the gel to be used as a negative control during MALDI-TOF. RNA was visualized by UV shadowing and band was excised and placed in 2.5M ammonium acetate pH 5.2 and allowed to elute overnight. Sample was ethanol precipitated and resuspended in 25mM ammonium bicarbonate. Samples were treated with XU RNase A or RNaseT1 for 1 hour and cleaned with micro C18 ZipTip according to manufactures protocol for oligonucleotides. One microliter of sample was combined with 2µl saturated 3-hydroxy picolinic acid matrix and 1µl was spotted for analysis. MALDI-TOF was performed on Voyager DE STR MALDI-TOF in positive ion reflectron mode.

4.3.7 MALDI-TOF/TOF MS- sample was prepared as for MALDI-TOF except in the place of RNase treatment the sample was hydrolyzed overnight in 5% ammonia solution at 90^{o}C. Ammonia was removed from sample by evaporation in a speedvac, dried sample was resuspended in 30µl water, dried again and resuspended in 30µl 0.1M triethylammonium acetate pH 8.0. 1U of alkaline phophatase (NEB) was added to sample and incubated at 50^{o}C for 30 minutes. Sample was cleaned by µC18 ZipTip and 1µl was combined with 2µl saturated alpha-cyano-4-hydroxycinnamic acid. Three microliters were spotted on a plate and analyzed on a Bruker Ultraflex III MALDI TOF/TOF.

4.3.8 Filter binding assay and Kd determination- 0-5µm MBP-TRBD was combined with 1ng (~1pM) 5'-P32 labeled RNA in column buffer. Samples were incubated for 10 minutes at room temperature and passed over a 25mM nitrocellulose membrane (Whatman) on a Millipore filter binding apparatus. Sample was washed with 1ml column buffer. Membranes were dried and

exposed to a phosphor imager screen. Dots were quantitated and percent RNA bound calculated by comparing to free RNA spotted on a membrane. All data points were run in triplicate. Kd was determined using Prizm software fitting data to one-site specific binding.

4.3.9 Homology modeling using SWISS MODEL Workspace-

The medaka TERT sequence was threaded through *Tetrahymena* TRBD (2R4G) and duplex bound *Tribolium* TERT (3KYL). Threading was performed using SWISS-MODEL Workspace in automated mode following the directions for submission.

4.4 RESULTS

4.4.1 Recombinant medaka TRBD binds CR4/CR5 RNA

In order to map the binding surface of TRBD involved in the CR4/CR5-TRBD interaction the purified protein must bind specifically to CR4/CR5. After obtaining highly pure protein (figure 4.2A), we tested if the recombinant protein could interact specifically with CR4/CR5. A gel shift experiment showed a single band upon addition of both CR4/CR5 and TRBD (Figure 4.2B). Using the filter binding assay the Kd of binding was determined to be 0.59 +/- 0.07 µM, while a mutant deleting the P6 U-bulge increased the Kd to greater than 4µM and saturated binding was not reached. Human CR7, a negative control for non-specific RNA binding, did not bind significantly using the filter binding assay (Figure 4.3).

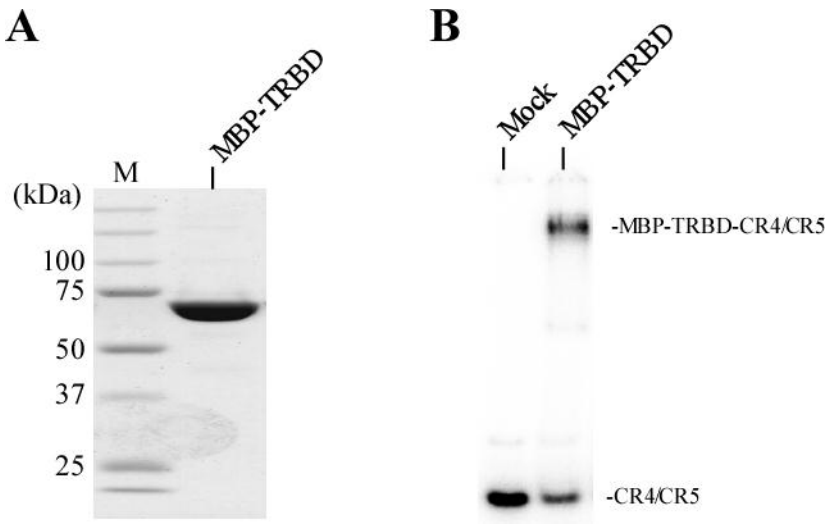


Figure 4.2-Purified MBP-TRBD is pure and binds CR4/CR5
(A) SDS gel showing the final purified MBP-TRBD protein after gel filtration, sample was concentrated to show purity, minor impurities can be seen around 100kDa. (B) Electrophoretic mobility shift assay shows interaction of purified MBP-TRBD with radio labeled CR4/CR5 RNA. Mock lane contains binding buffer in place of MBP-TRBD.

4.4.2 Recombinant medaka TRBD binds pseudoknot

Previous reports have indicated TRBD may have low affinity binding to the pseudoknot RNA. To determine if recombinant TRBD has affinity to the pseudoknot RNA, three pseudoknot RNA truncations were created. Full length pseudoknot spanning residues 1-149 including an intact P1 did interact with TRBD as indicated by a mobility shift of radio labeled RNA only upon addition of TRBD protein (Figure 4.4 lanes 3 and 4). A pseudoknot truncation removing a large single stranded loop within the pseudoknot (J2a/3), as well as paired region P3 greatly reduced the interaction, nearly eliminating signal from the shifter RNA band (Figure 4.4 lanes 1 and 2).

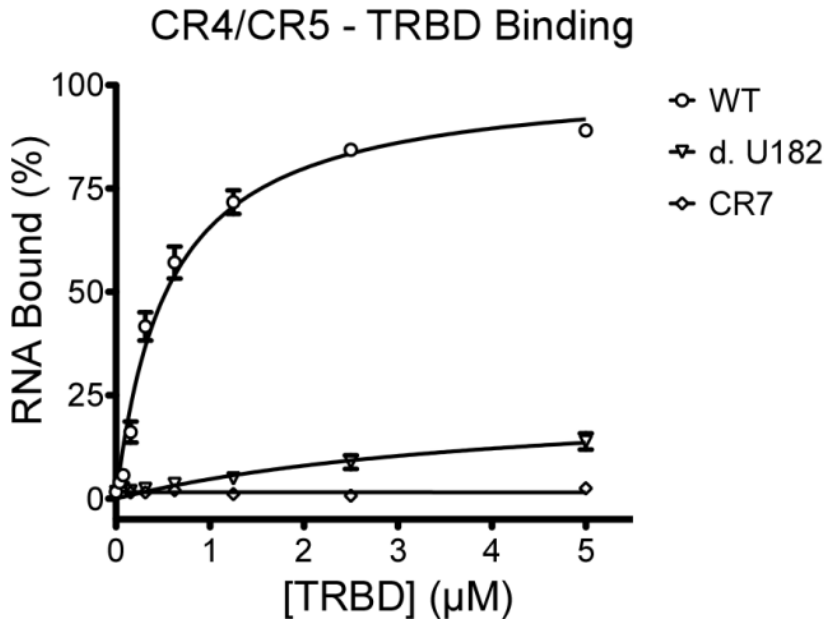


Figure 4.3- CR4/CR5 – MBP-TRBD binding curve
Using the filter binding assay (discussed in appendix 1) the dissociation constant (Kd) of MBP-TRBD toward medaka CR4/CR5 was determined by titrating protein concentration and measuring the percent of RNA bound. All points were run in triplicate; error bars indicate standard error (SEM). CR7 served as a negative control; CR7 RNA retained on the protein binding membrane was independent of protein concentration. A mutant CR4/CR5 deleting U_{182} has severely decreased binding; Kd was shifted beyond the highest protein concentration of 5μM.

To determine if binding was occurring at the core of the pseudoknot, a third truncation containing only the central stem of the pseudoknot (nucleotides 33-97) was tested for TRBD binding. Pseudoknot 33-97 did not exhibit binding within the sensitivity of the gel shift assay (Figure 4.4).

The solution structure of a human pseudoknot fragment revealed the presence of a 5 base triple helix between p2 and J2b/3 (Theimer et al., 2005). To determine if the pseudoknot TRBD interaction requires the triple helix the top strand of the triple helix was *in trans* to the pseudoknots 1-97 and 33-97 (Figure 4.5B) which previously exhibited low and no binding respectively. A minor increase in binding was observed when the top strand was added to pseudoknot 1-97 (figure 4.5A lanes 1-4), yet binding was not returned to the level of full length pseudoknot (Figure 4.5A compare lanes 4 and 10). Addition of the top strand did not restore binding of TRBD to pseudoknot 33-97 (Figure 4.5A lanes 5-8). To verify that the top strand RNA did not impair binding to pseudoknot, the top strand was added to the full length pseudoknot and no difference was observed in the presence or absence of top strand RNA (Figure 4.5A lanes 9-12).

To determine if TRBD can concurrently bind pseudoknot and CR4/CR5 RNA, unlabeled CR4/CR5 was added to a binding reaction of TRBD to radiolabeled full length pseudoknot. Addition of cold CR4/CR5 completely abolished binding to pseudoknot (Figure 4.5A lane 14).

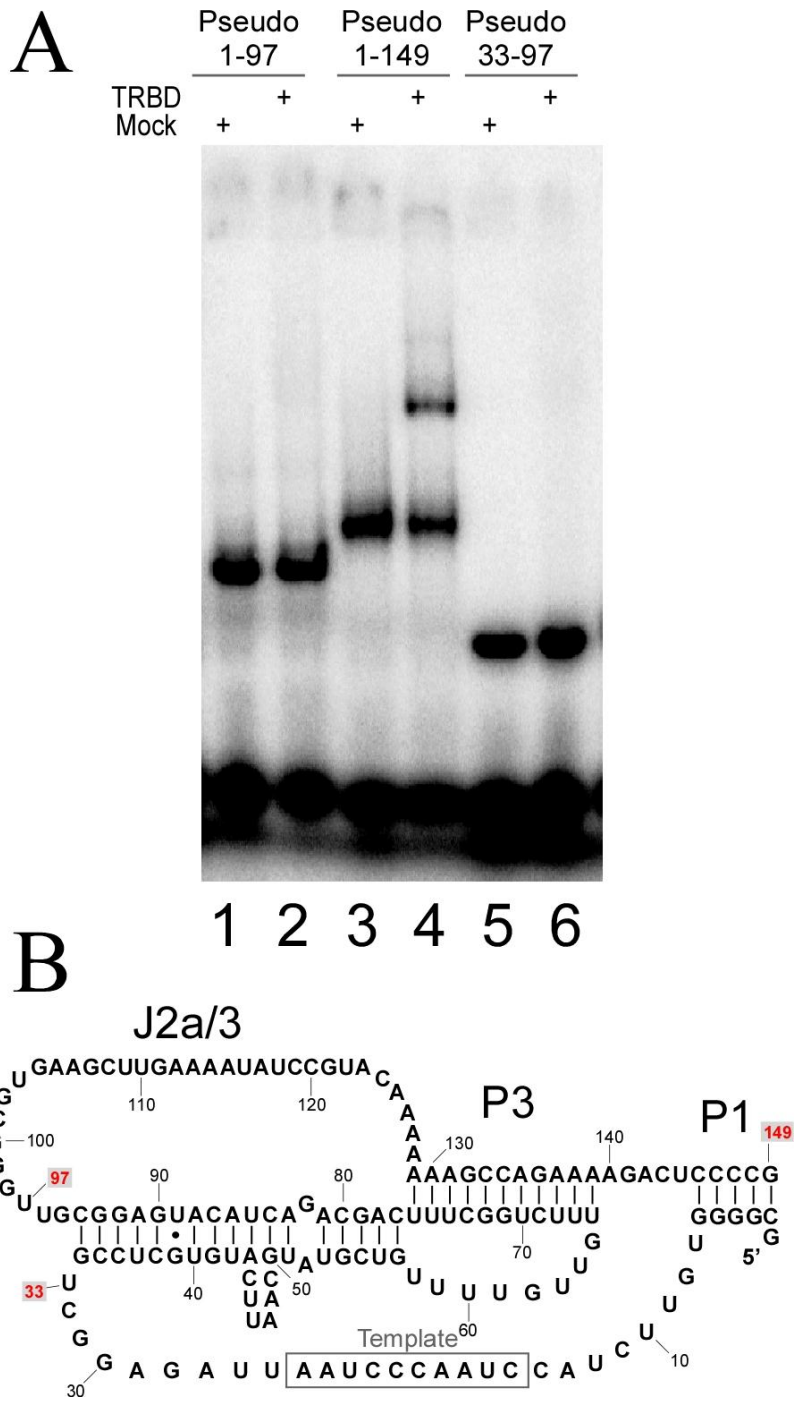
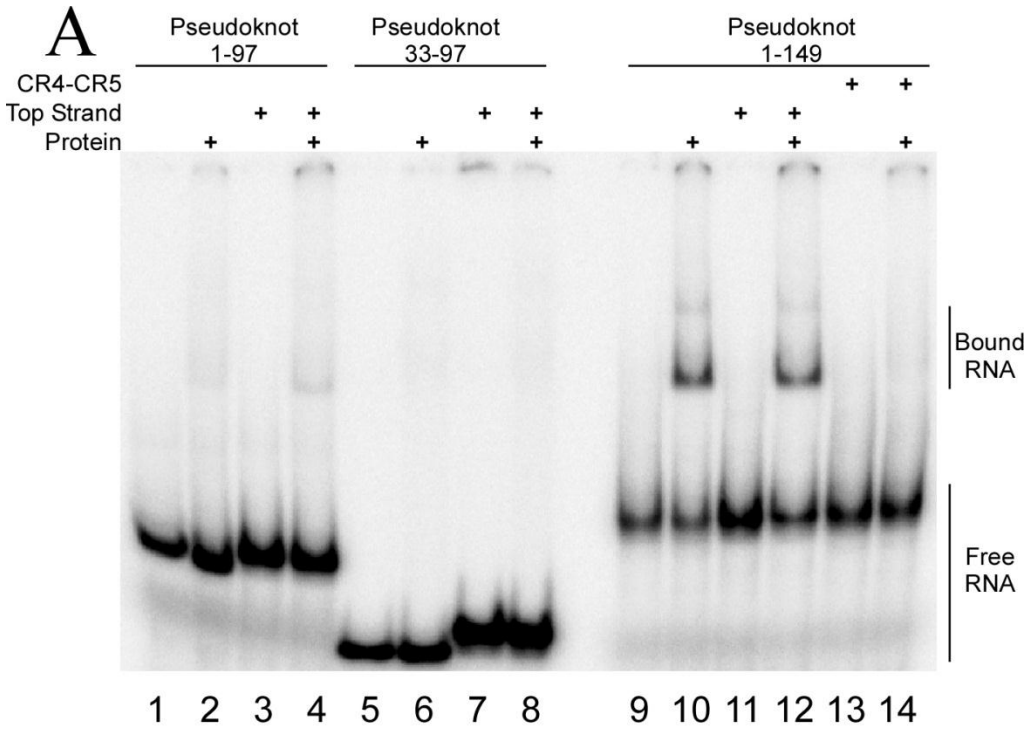


Figure 4.4- TRBD Binds pseudoknot in gel shift assay
(A) Electrophoretic mobility shift assay was used to visual binding of MBP-TRBD to medaka pseudoknot RNA. Three RNA constructs spanning nucleotide 1-97, 1-149, and 33-97 were tested for binding to MBP-TRBD. Mock reactions contain buffer only in the place of protein. (B) Secondary structure of medaka pseudoknot, numbers in red indicate truncation boundaries used.



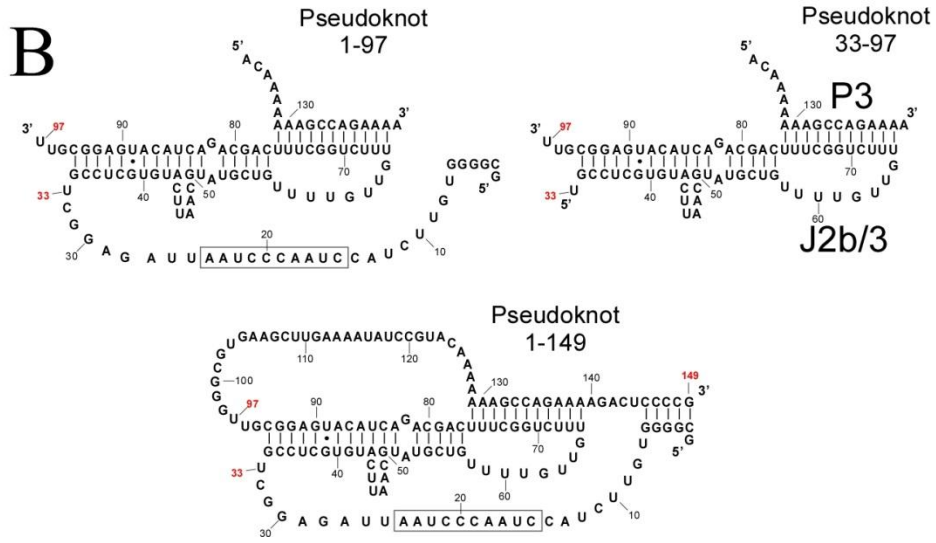


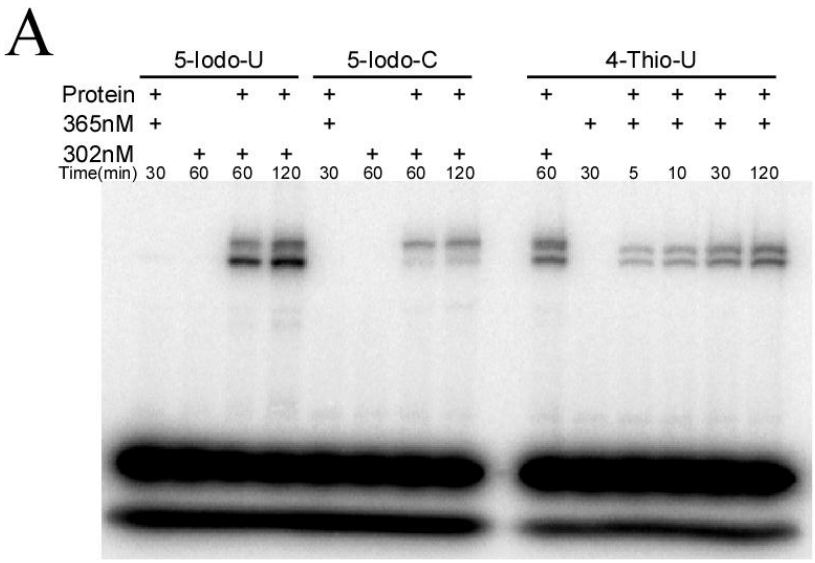
Figure 4.5- Triple helix not determinant of Pseudoknot-TRBD interaction
Two pseudoknot truncations and full length pseudoknot were tested for binding to MBP-TRBD. Triple helix was formed by adding top strand RNA *in trans*. Unlabeled CR4/CR5 was added as indicated in competition with pseudoknot for binding to TRBD. (B) Secondary structure of pseudoknot constructs used, P3 and J2b/3 which form the triple helix region are indicated on pseudoknot 33-97.

4.4.3 Comparison of photo-cross-linkers 5IU, 5IC and 4SU

Medaka CR4/CR5 was randomly labeled by including 10% photo cross-linker labeled UTP and 90% normal UTP in the transcription reactions. In order to compare efficiency of the available cross-linkers, RNA was labeled individually with 5-Iodo-Uridine (5IU), 5-Iodo Cytodine (5IC), and 4-Thio-Uridine (4SU) (structures shown in 4.6B). A cross-linking time course was performed from 30 minutes to 2 hours using 302nm UV light for 5IU and 5IC and 365nm light for 4SU with P32 labeled RNA. The cross-linking products were visualized by SDS-PAGE. No cross-linking was observed for 5IU or 5IC when treated with non-optimal 365nm UV light; however 4SU did exhibit cross-linking when irradiated with 302nm light (Figure 4.6A). 5IU labeled RNA had the greatest yield of cross-linking products resolved as two bands. 5IC RNA predominantly resulted in a single cross-linking band of similar mobility to the less intense 5IU band. 4SU resulted in an equal mixture of two cross-linking bands. 5IU was used for subsequent experiments as it yielded the highest level of cross-linking product (Figure 4.6).

4.4.4 UV induced 5-Iodo-Uridine mediated cross-linking CR4/CR5 to TRBD

During the cross-linking reaction a petri dish lid is used to minimize sub300nm UV light which can damage protein and RNA as well as induce shortwave length cross-linking. The UV filter used is not expected to completely block sub 300nm UV light so cross-linking was perform with 5IU labeled and non-5IU labeled CR4/CR5 RNA to identify 5IU mediated cross-linking products.



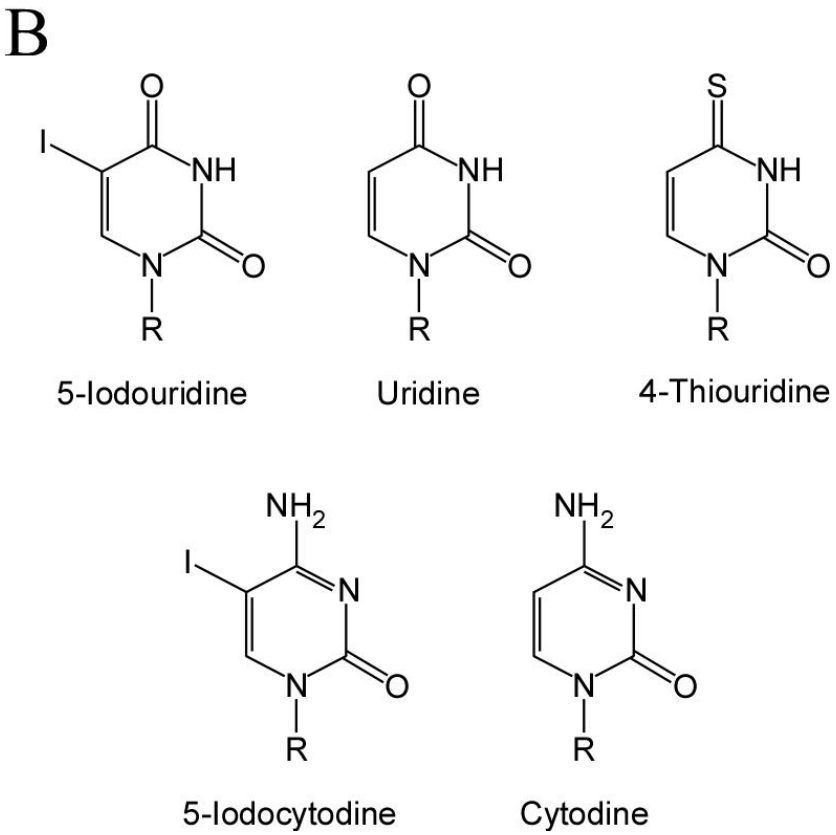


Figure 4.6 Comparison of photo-cross-linking reagents.
(A) A time course of UV irradiation was performed to cross-link CR4/CR5 individually labeled with one of three photo-cross-linking reagents 5IU, 5IC or 4SU. Wavelength of lamp used in irradiation (302 or 365nm) is indicated above each lane. (B) Structure or 5IU, 4SU and 5IC with unlabeled uridine and cytodine for comparison.

Protein free samples did not result in any RNA only cross-link bands (Figure 4.7A). When treated with trypsin after cross-linking, the cross-linking products disappear indicating that the RNA is cross linked to protein rather than a protein dependent RNA-RNA cross-link. One major band (Figure 4.7A indicated by triangle) is present in the 5IU cross-linking reaction but is only a minor product in the non-5IU cross-linking reaction (Figure 4.7A). The efficiency of cross-linking for the major band is about 5%, and the major band was used in the subsequent RNA mapping experiments.

To determine if the cross-linking observed was due to a specific interaction, as opposed to random RNA-protein interaction allowed by the extended cross-linking time, 10% IU labeled human CR7 was used as a negative control (Figure 4.7B). MBP-TRBD protein was incubated with CR4/CR5 or CR7 in the presence or absence of excess tRNA competitor prior to cross-linking. Competitor tRNA did not change the cross-linking pattern of CR4/CR5 RNA. Human CR7 exhibited no cross-linking to MBP-TRBD (Figure 4.7B).

4.4.5 Cross-linking sites identified on RNA

In order to map which nucleotides were involved in cross-linking we modified an approach previously described for mapping nucleotide-nucleotide interactions (Hiley et al., 2002). After UV-irradiation RNA-peptide hetero-conjugates were gel purified on SDS-PAGE, removing non-cross-linked RNA. The hetero-conjugates were treated with limited in gel alkaline hydrolysis resulting in a ladder of all possible RNA sizes. When separated on a high resolution sequencing gel, only free RNA is resolved, as any RNA still covalently

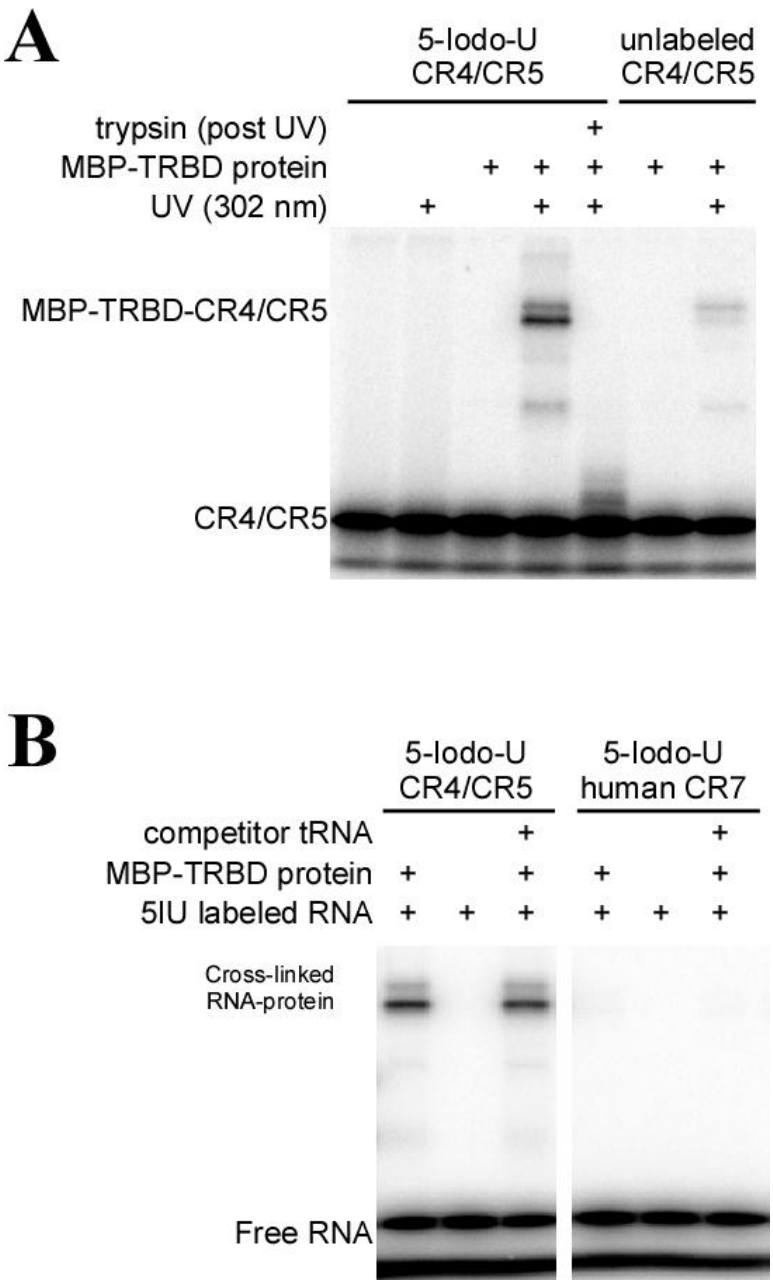


Figure 4.7- 5-Iodo-Uridine dependent cross-linking
(A) SDS gel showing cross-linking of CR4/CR5 RNA to MBP-TRBD, triangle indicates the major 5IU dependent cross-linking band. 5IU labeled CR4/CR5 RNA and non-5IU RNA were irradiated for 2 hours with 302nm UV light in the presence or absence of MBP-TRBD. No UV controls were used to indicate UV dependence of bands observed. + Trypsin reaction was UV irradiated before trypsin digestion. (B) 5IU labeled medaka CR4/CR5 and human CR7 were incubated with MBP-TRBD and UV irradiated to induce cross-linking. Cross-linking was not observed between MBP-TRBD and human CR7.

bonded to the protein would have a greatly reduced mobility. Only RNA fragments ending at residues between the crosslinking site and the 5' end of the RNA, the site of P32 labeling, will resolve. In this manner all sized RNAs up to the -1 of crosslinking site will be visualized on the gel (Figure 4.8). The cross-linking nucleotide is one nucleotide larger than the last resolved nucleotide.

The major 5IU crosslinking band was excised from an SDS gel for mapping. As expected, a complete RNA ladder stop was observed in the RNA hydrolysis experiment. However, a second potential cross-linking site was observed corresponding to a second dramatic decrease of intensity (Figure 4.9). A decrease of intensity would be observed if the cross-linking resulted in a mixture of two species of hetero-conjugates cross-linked at either of two sites. The complete ladder stop was observed at U_{205} corresponding to a highly conserved nucleotide in P6.1 loop. A mutation of this nucleotide was previously characterized using mouse telomerase to interrupt telomerase activity but not the interaction of CR4/CR5 with TERT when reconstituted *in vitro* (Chen et al., 2002b). The second RNA cross-linking site mapped to U_{182} which forms a U-bulge in p6. While the U is not conserved, a bulge in p6 is conserved in species that contain an RNA three way junction.

To verify the two crosslinking sites especially the U-bulge crosslinking which resulted in a decrease in band intensity when using 5' P-32 labeled RNA, 3' P-32 labeled RNA was used in a cross-linking experiment. RNA 3' heterogeneity, a result of run-off transcription (Milligan et al., 1987), resulted in doublet bands observed on the sequencing gel (Figure 4.10). Though the doublet

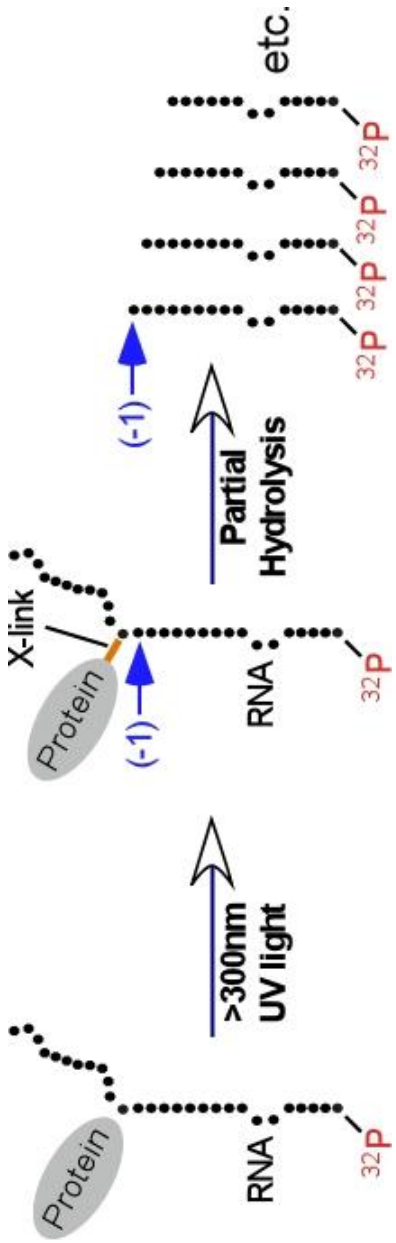


Figure 4.8- Schematic of mapping RNA crosslink site
To map the crosslinking site limited alkaline hydrolysis is used. First the protein and RNA are incubated and allowed to reach equilibrium binding. UV induced crosslinking forms a covalent bond (orange line) from one nucleotide in the RNA to an amino acid of the protein. The -1 site indicates nucleotide one reside 5' to crosslinking site. The crosslinking product is separated from free RNA by SDS-page and treated by partial alkaline hydrolysis. Hydrolysis products are separated by high resolution PAGE resulting in single nucleotide resolution. All RNA fragments up to the -1 site resolve on the gel as a nucleotide ladder. Cross-linking is mapped as one nucleotide beyond the -1 site.

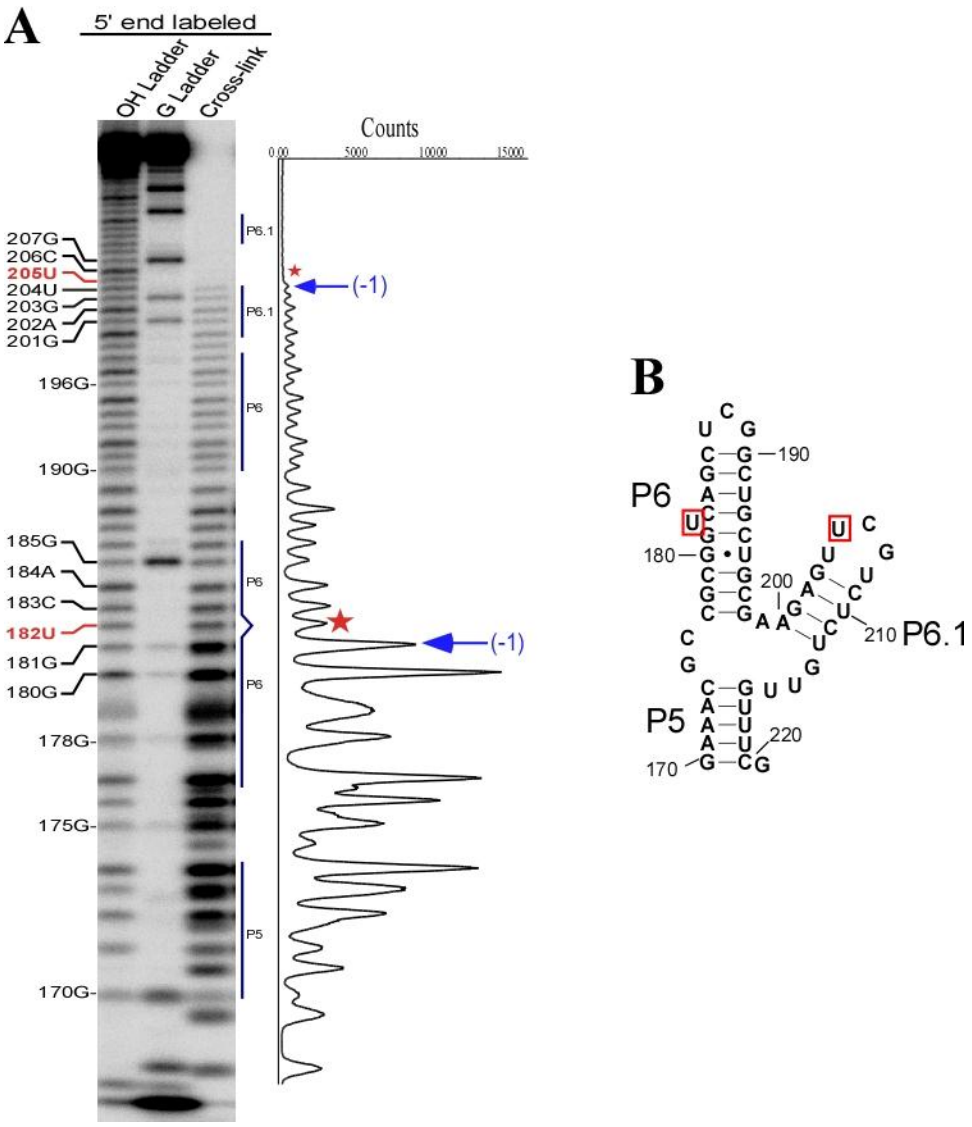


Figure 4.9- Mapping cross-linking site on 5' P32 labeled CR4/CR5
(A) Sequencing gel resulting from partial alkaline hydrolysis of cross-linking product composed of medaka CR4/CR5 and MBP-TRBD. Alkaline hydrolysis of free RNA was used to generate a single nucleotide ladder for comparison (OH ladder). RNase T1 digested free RNA results in ladder of G residues (G-ladder) allowing assignment of bands to RNA sequence. Trace of band intensity for Cross-linking lane is shown to right of gel. A decrease in band intensity is observed corresponding to crosslinking at U182. A complete ladder stop indicates crosslinking at U205. (B) Secondary structure of medaka CR4/CR5. Nucleotides mapped as cross-linking sites are boxed in red.

bands convolute mapping, the experiment provides high enough resolution to verify the previous assignment of a crosslinking site at U_{182}. A decrease in intensity corresponding to U_{205} is not observed with 3' labeled RNA though this is not entirely unexpected as the decrease at U_{182} (with 5' labeled RNA) was more intense than the ladder continuing to U_{205} indicating a stronger cross-linking at U_{182}.

Two mutant RNAs, $U_{182}C$ and $U_{205}A/G_{207}C$, were created to remove the two cross-linking sites individually to isolate the individual cross-linking products and simplify protein mapping. The p6.1 mutant $U_{205}A/G_{207}C$ was used because it had previously been described for mouse TR to not inhibit TERT binding (Chen et al., 2002b). Mutant $U_{205}A/G_{207}C$, which cross-links predominately from the P6 U-bulge, has higher cross-linking efficiency than $U_{182}C$, consistent with the larger decrease in band intensity of the sequencing gel. To verify that the mutants did not a have a major impact on binding, the Kd of each RNA was determined using the filter binding assay (Figure 4.11). Mutant $U_{205}A/G_{207}C$ has a similar Kd to wild type, 0.51 +/- 0.07μM versus 0.55 +/- 0.06μM. Mutant $U_{182}C$ exhibited a slightly decreased binding, shifting the Kd to 1.4 +/- 0.1μM. In contrast a mutant removing U182 greatly shifted the Kd, and saturated binding was not observed even up to the 5μM highest protein concentration. No significant binding was observed with human CR7 RNA, the negative control (Figure 4.11).

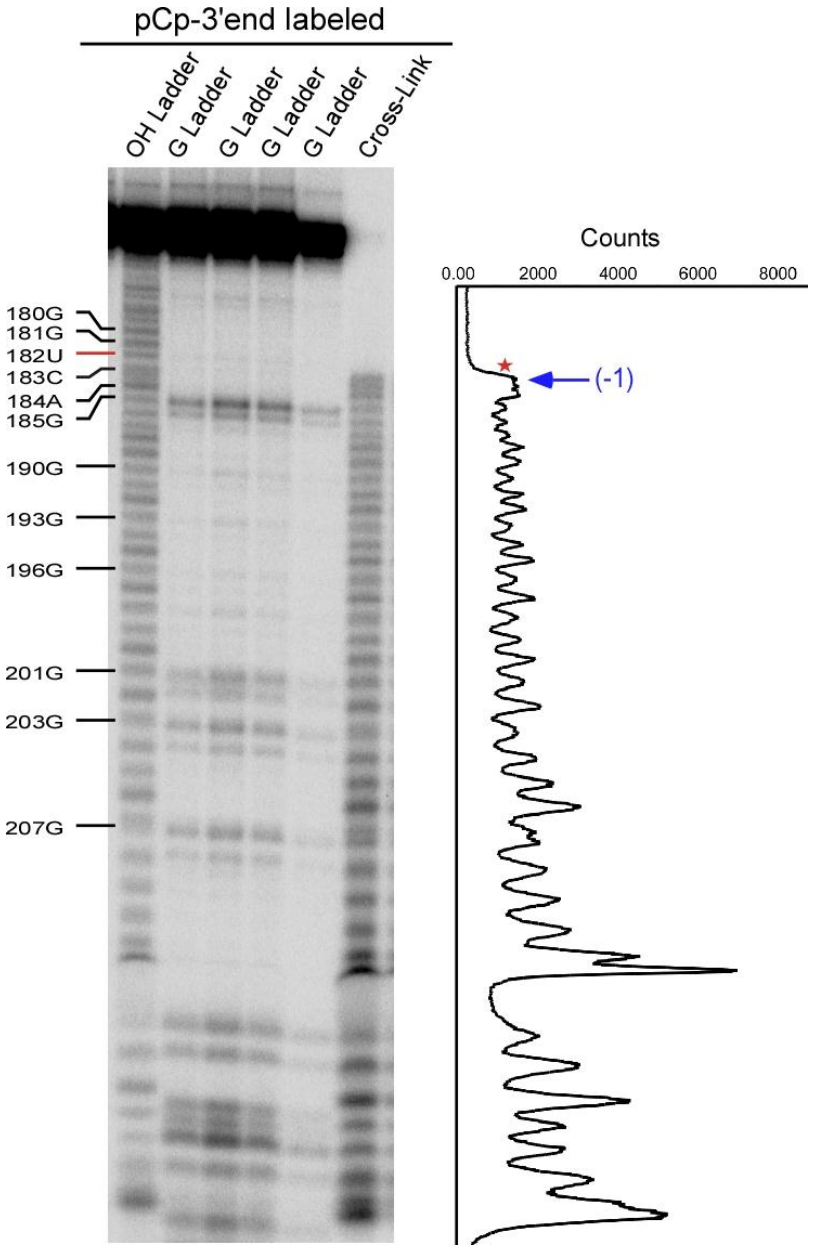


Figure 4.10- Mapping cross-linking site on 3' P32 labeled CR4/CR5
To verify the two crosslinking sites observed with 5' P32 labeled RNA, 3' labeled RNA was cross-linked to MBP-TRBD and mapped by partial alkaline hydrolysis. Due to 3' heterogeneity doublet bands are observed. Crosslinking stops match previously determined assignments.

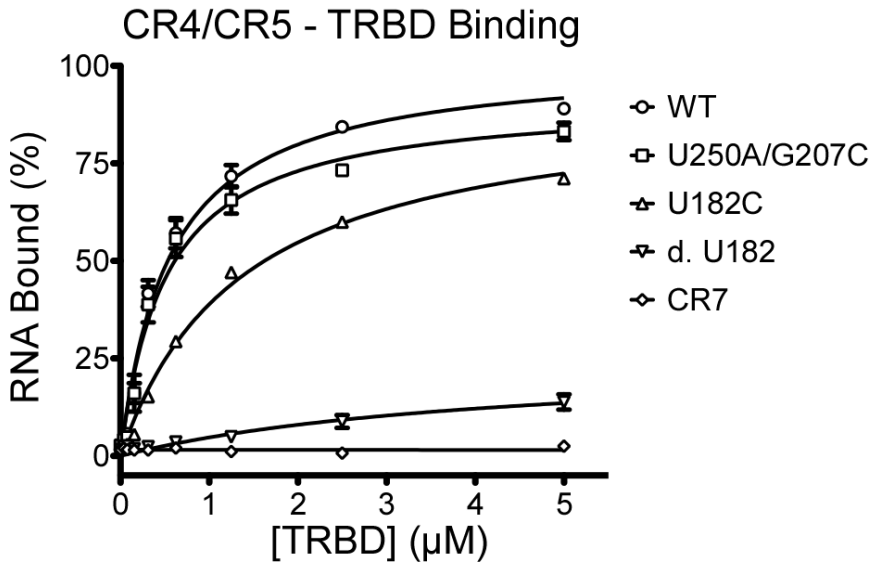


Figure 4.11- Binding of U182C and U205A/G207C to MBP-TRBD
Filter binding assay was used to determine Kd of RNA mutants U182C and U205A/G207C. Kd of U205A/G207C (squares) was determined to be 0.51+/- 0.07μM, similar to wt (0.55 +/- 0.6μM). U182C exhibited impaired binding with a Kd 1.4 +/- 0.1μM.

4.4.6 Cross-Linking of U_{182} and U_{205} mapped to two different peptides

In order to obtain a clean sample suitable for mass spec analysis the RNA-peptide, the heteroconjugate must be purified away from competing ions. Post cross-linking, the protein was denatured in urea and purified through a C-terminal 6xhis tag allowing the removal of non-cross-linked RNA. The sample was then protease digested and gel purified. Since the heteroconjugate still contained intact RNA the sample easily separated from the smaller trypsin fragments on a 4% polyacrilamide urea denaturing gel (Figure 4.12). The sample free of non-cross-linked peptides or full RNAs was eluted and RNase treated. The sample was then cleaned with a μC18 ZipTip and eluted into 50% acetonitrile. As there was no further purification to remove extra RNA fragments after the RNase treatment, a free RNA control was used to eliminate mass spec peaks resulting from non-cross-linked RNAs.

Using both RNase A and RNase T1 treated samples no significant peaks were observed during MALDI-TOF when using the peptide matrix alpha-cyano-4-hydroxycinnamic acid. Use of 3-hydroxy picolinic acid matrix, typically used with nucleic acids, resulted in observable peaks from all samples. Using a combinatorial matrix the observed peaks were compared with all RNase-trypsin fragments calculated *in silico* allowing up to 1 missed RNase and trypsin cleavage. When calculating the mass of the RNA, one hydrogen was substituted for iodine accounting for the labeling of Iodo-Uridine. During cross-linking the iodine is lost from the RNA and a single hydrogen is lost from the target aromatic

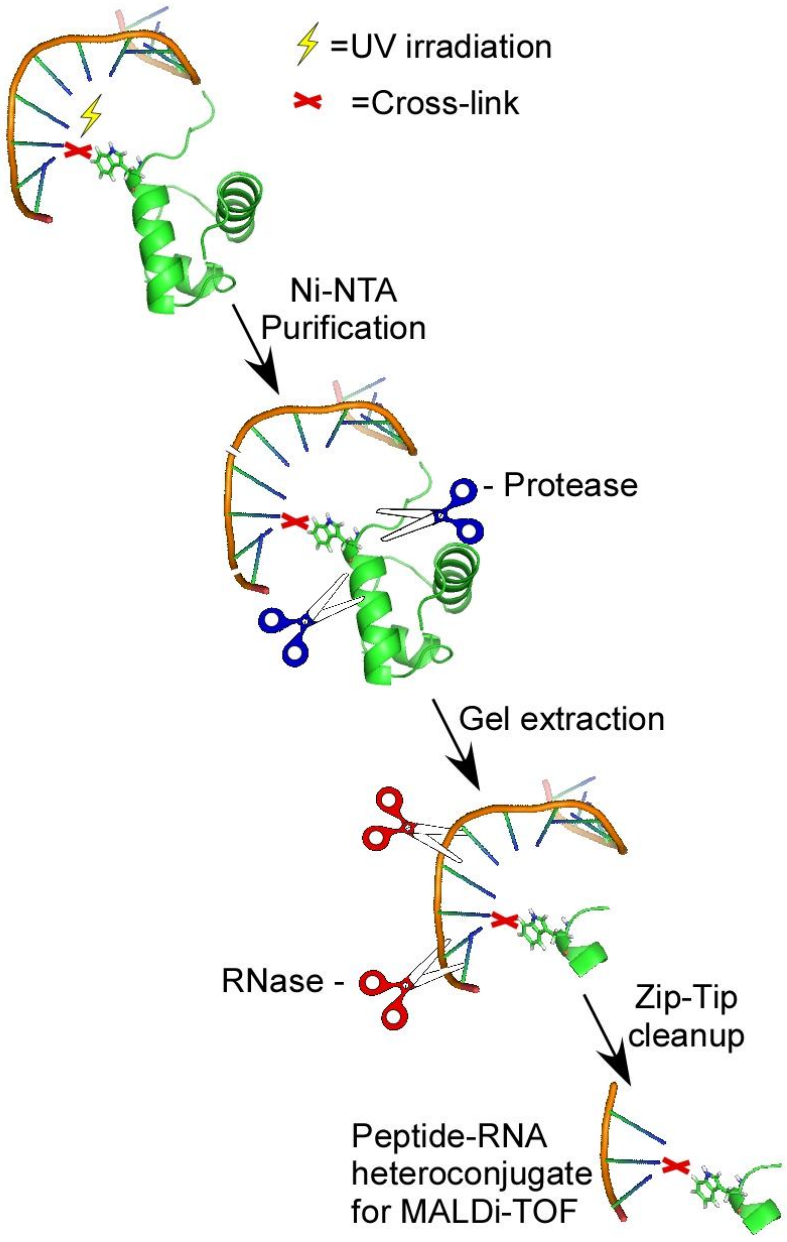
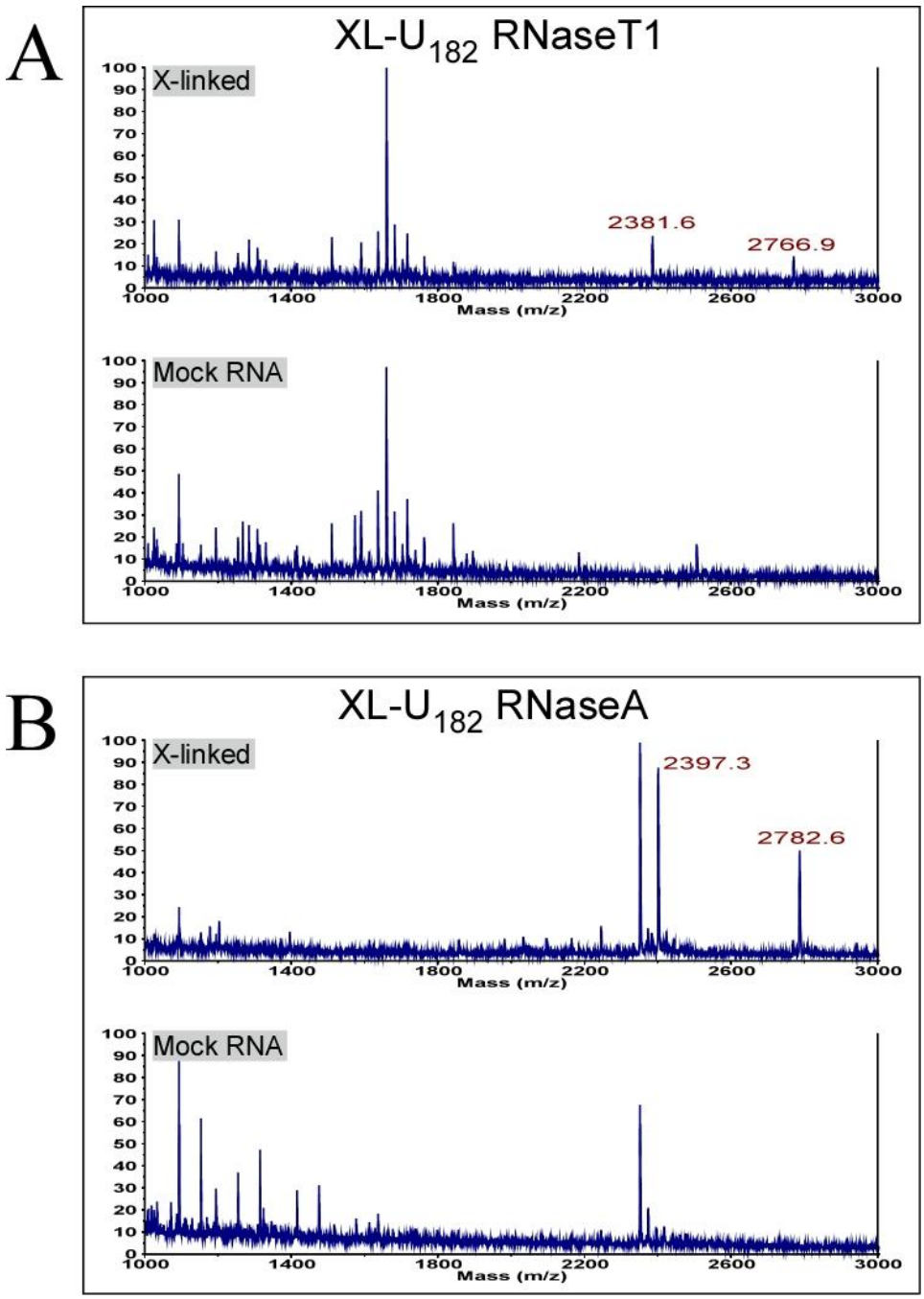


Figure 4.12-Schematic of mapping protein cross-link site
RNA and protein are incubated reaching equilibrium binding and cross-linked creating a covalent bond between RNA and protein. Crosslink product is purified from free RNA then protease digested. Sample is gel extracted from a urea denaturing polyacrylamide gel removing free peptides. Crosslink is RNase treated and cleaned up by μC18-ZipTip. Cleaned sample is used for MALDI-TOF.



(Legend on next page)

C

RNase	Peptide	RNA	Loss during X-link	[M+H] calc	[M+H] exp
T1	FPPSELAYR 1078.6	(IU)CAG 1429.1	(I + H) 127.9	2380.8	2381.6
T1	TAGRFPPSELAYR 1463.6	(IU)CAG 1429.1	(I + H) 127.9	2765.8	2766.9
A	FPPSELAYR 1078.6	GG(IU)C 1445.1	(I + H) 127.9	2396.8	2397.3
A	TAGRFPPSELAYR 1463.8	GG(IU)C 1445.1	(I + H) 127.9	2782.0	2782.6
	a	b	c	=a+b-c+H	

D

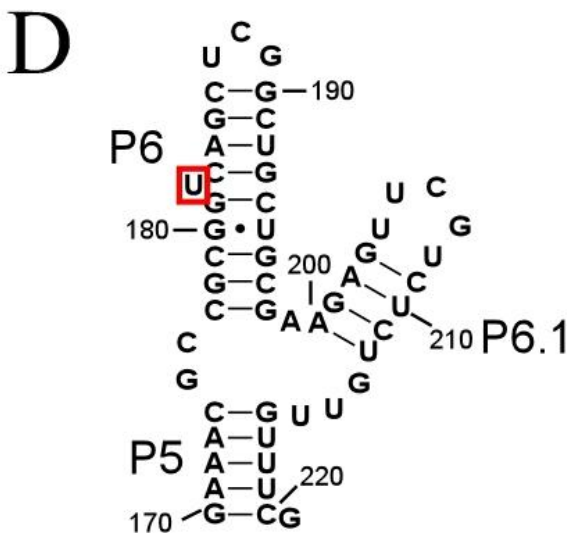
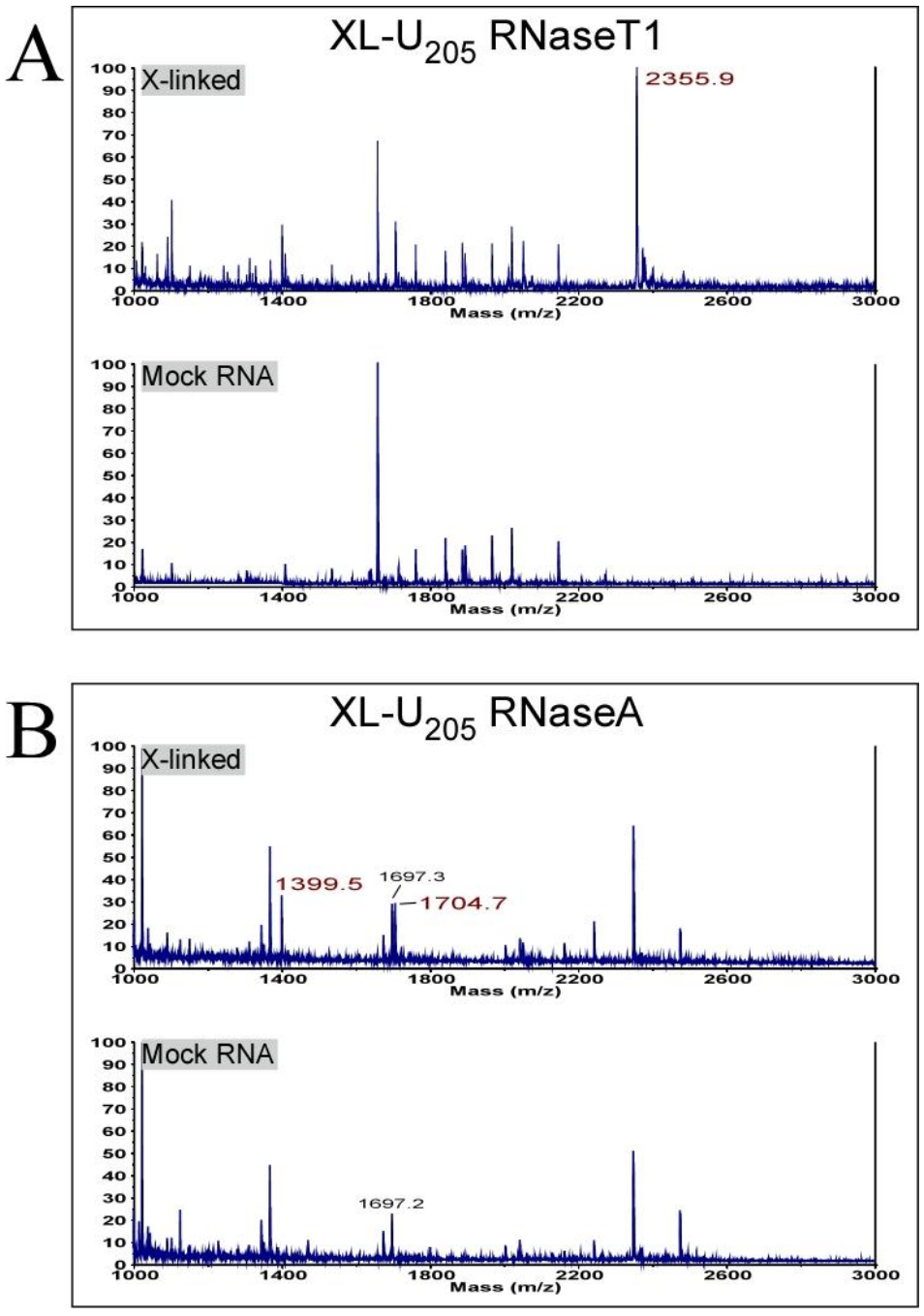


Figure 4.13- MALDI-TOF identifies peptide cross-linked to U182
CR4/CR5 RNA mutant U205A/G207C which cross-links predominantly at U182 was prepared for MALDI-TOF with either (A)RNaseT1 or (B)RNaseA treatment as indicated. Mock RNA indicates digestion and purification of free RNA to serve as a control to eliminate peaks due to RNA only. Masses in red indicate peaks in cross-link sample not present in mock RNA sample. (C) Summary of peptide-RNA combinations matching observed peaks. Monoisoptic uncharged masses of peptide and RNA are listed (a and b), mass lost during cross-linking(c) is due to loss of iodine and hydrogen. Calculated mass (a+b-c+H) includes mass of peptide and RNA less mass lost during crosslinking as well as mass of hydrogen for as the single charged +1 is measured on MALDI. Experimental mass is listed in the last column. (D) Secondary structure of CR4/CR5 with crosslinking RNA residue boxed in red.

amino acid to maintain the aromatic structure of the amino acid (Norris et al., 1996; Steen et al., 2001). A mass tolerance of 1da was allowed during the initial data analysis due to lower peak resolution and external calibration using a different matrix and anylate type. Cross-linking of mutant $U_{205}A/G_{207}C$ resulted in two MALDI-TOF peaks not observed in the free RNA sample. In both the RNaseA and RNase T1 sample the peaks were separated by 385.3 Daltons. The lower mass peak of the RNaseT1 matches the mass of UCAG-cross-linked to $_{496}$FPPSELAYR$_{504}$ (Figure 4.13A). The larger mass peak matched the same RNA, UCAG, cross-linked to the same peptide extended by a single missed trypsin cut $_{492}$TAGRFPPSELAYR$_{504}$ (Figure 4.13B). The RNaseA treated sample matches the same two peptides cross-linked to RNA G_2UC. RNaseA cleaves after pyrimidine residues, it is hypothesized that the cross-linked peptide interfered with recognition of the U, leading to the missed RNase cleavage. It was expected for this mutant to cross-link to U_{182}; however since the entire cross-linking reaction was used rather than isolating the more intense lower band it is possible that the cross-linking occurred from a second RNA site. However, the sole 4nt RNA segment composed of one each of U,C,A and G within CR4/CR5 occurs at the P6 U-bulge (Figure 4.13).

The same combinatorial matrix approach was used with cross-linked $U_{182}C$ RNA. With RNase T1 treatment a single peak, unique to the cross-linked species, was found to match the mass of U_2CG bonded to the peptide $_{470}$ISVAELMWK$_{478}$ (Figure 4.14A). When the $U_{182}C$ mutant was treated with

RNaseA two peaks were observed corresponding to UC-ISVAELMWK and U-ISVAELMWK (Figure 4.14B).



(Legend on next page)

C

RNase	Peptide	RNA	Loss during X-link	[M+H] calc	[M+H] exp
T1	ISVAELMWK 1075.6	U(IU)CG 1406.7	(I + H) 127.9	2355.4	2355.9
A	ISVAELMWK 1075.6	(IU) 449.9	(I + H) 127.9	1398.6	1399.5
A	ISVAELMWK 1075.6	(IU)C 755.0	(I + H) 127.9	1703.7	1704.7
	a	b	c	=a+b-c+H	

D

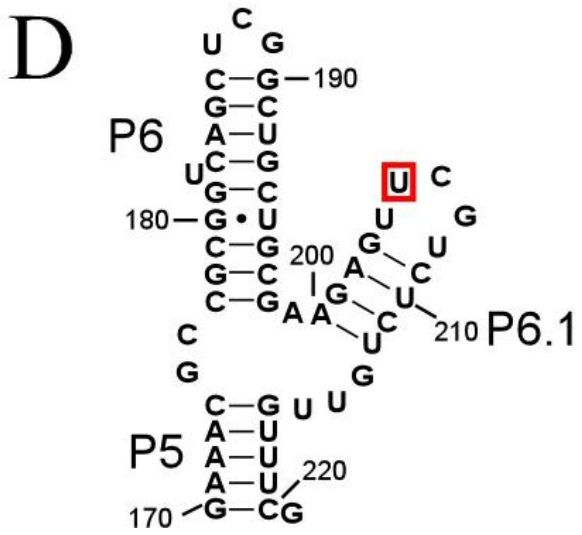


Figure 4.14- MALDI-TOF identifies peptide cross-linked to U205
CR4/CR5 RNA mutant U182C crosslinking at U205 was prepared for MALDI-TOF with either (A)RNaseT1 or (B)RNaseA treatment as indicated. Mock RNA indicates digestion and purification of free RNA to serve as a control to eliminate peaks due to RNA only. Masses in red indicate peaks in cross-link sample not present in mock RNA sample. (C) Summary of peptide-RNA combinations matching observed peaks. Monoisoptic uncharged masses of peptide and RNA are listed (a and b), mass lost during cross-linking(c) is due to loss of iodine and hydrogen. Calculated mass (a+b-c+H) includes mass of peptide and RNA less mass lost during crosslinking as well as mass of hydrogen for as the single charged +1 is measured on MALDI. Experimental mass is listed in the last column. (D) Secondary structure of CR4/CR5 with crosslinking RNA residue (U205) boxed in red.

4.4.7 Y_{503} and W_{477}, two amino acids on the protein-RNA interaction interface

Iodo-modified bases typically, though not exclusively, cross-link with aromatic amino acids. It was predicted that crosslinking occurred to one of the aromatic amino acids in each peptide; however, to determine the actual amino acid involved in cross-linking MALDI-TOF/TOF was used. Originally the same RNaseA and RNaseT1 treated samples were used during MALDI-TOF/TOF however only fragmentation of the RNA was observed. In order to obtain peptide fragmentation, we used alkaline hydrolysis followed by phosphatase treatment to remove all extra nucleotides and the last remaining phosphate, thus leaving only a single uridine nucleoside attached to the peptide. The modified sample preparation allowed the use of alpha-cyano-4-hydroxycinnamic acid which resulted in better resolution of monoisotopic mass peaks. Experimental masses matched the expected masses to within 0.1da (Figures 4.16A and 4.17A).

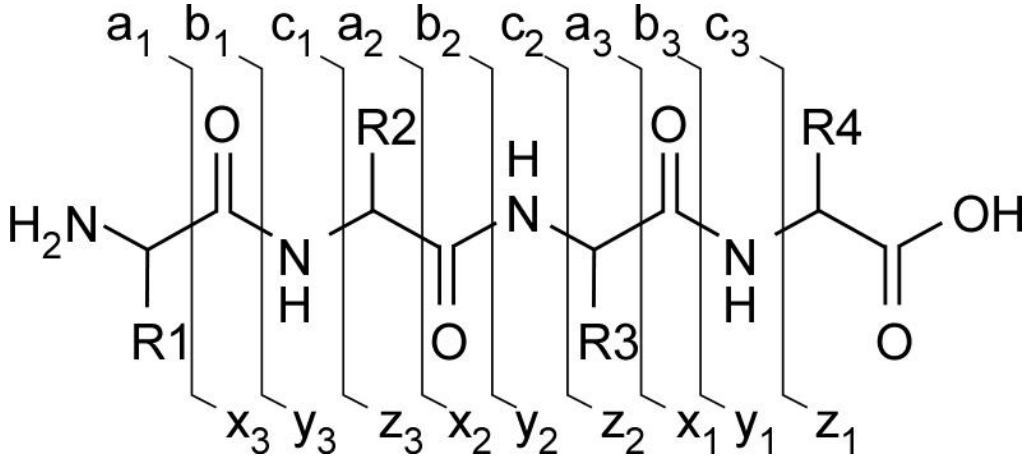


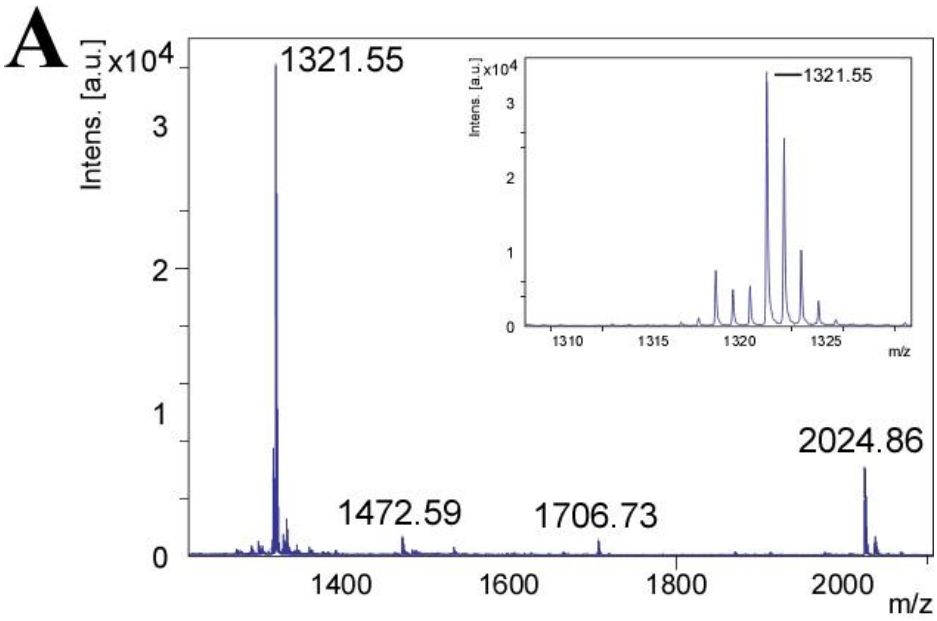
Figure 4.15- MALDI-TOF TOF fragmentation pattern nomenclature
Roepstorff–Fohlmann–Biemann nomenclature for fragmentation of peptides is indicated. Ions a, b, c represent the fragment to the left of the breakage point, while x, y, z represent the ions to the right. The most common ions observed are b and y ions from breakage at the peptide bond.

Fragmentation during MALDI-TOF//TOF occurs along the peptide backbone. Fragmentation can occur at any bond along the backbone and standard Roepstorff–Fohlmann–Biemann nomenclature has become the standard used in the field (Steen and Mann, 2004). After fragmentation, one half of the molecule will carry the +1 charge the other will not. In the MALDI-TOF/LIFT/TOF used, the ions are reaccelerated after fragmentation (Suckau et al., 2003). Only the peptide fragment carrying the +1 charge will be seen from that individual parent molecule. After fragmentation the ions on the N-terminal side of the breakage are termed a, b, and c based on fragmentation position. The fragments C-terminal to the breakage are termed x, y, and z (Figure 4.15) (Steen and Mann, 2004). Fragmentation between the alpha carbon and carbonyl results in a and x ions. Cleavage of the amide bond, resulting in b and y ions, results from the cleavage of the lowest energy pathway and as a result are the most abundant ions in a typical MALDI-TOF/TOF spectrum (Steen and Mann, 2004). The final cleave between the nitrogen of the peptide bond and the alpha carbon results in c and z ions.

Using the mutant $U_{205}A/G_{207}C$ RNA two peptides were sequenced, (Figure 4.16B) confirming the peptide assignments made previously. The MS/MS data indicated U_{182} cross-linking occurred to Tyr_{503}. The Y_1 ion contains only the C-terminal arginine, Y_2 contains the arginine as well as tyrosine cross-linked to uridine. All Y-ions beyond Y_2 also contain the mass of the uridine. In contrast the b-ions observed up to b7 match only the mass of the peptide sequence. Ion b8 is shifted by the summed masses of tyrosine and uridine. The shift in mass only

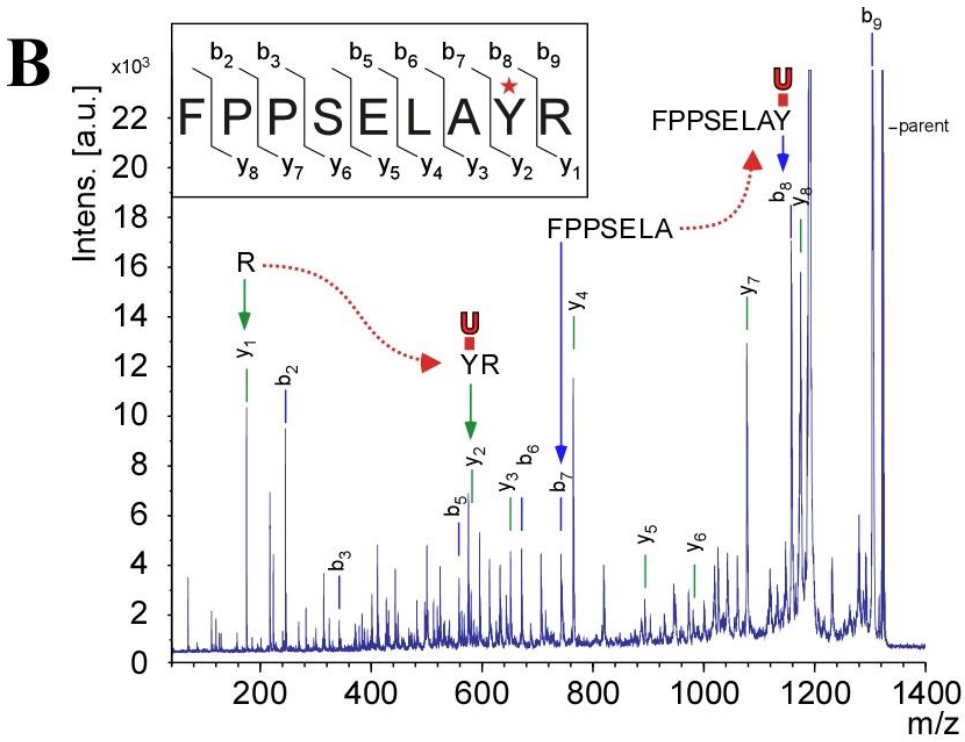
when a peptide includes the tyrosine, indicates cross-linking occurred at this amino acid (Figure 4.16B).

Using $U_{182}C$ RNA, cross-linking of p6.1 to peptide ISVAELMWK was confirmed. The entire y-ion series can be observed in the TOF/TOF spectrum including a large mass shift from y_1 to y_2 matching the summed masses of uridine and tryptophan (Figure 4.17) indicating cross-linking to Trp_{477}. A second MALDI-TOF peak shifter upward 16 m/z (Figure 4.17A) was sequenced and confirmed to correspond to the same peptide with an oxidized methionine residue.



Peptide [M]	RNA [M]	Loss during X-link [M]	[M+H] calc	[M+H] exp
FPPSELAYR 1078.55	IU 369.97	(I + H) 127.90	1321.62	1321.55
TAGRFPPSELAYR 1463.76	IU 369.97	(I + H) 127.90	1706.83	1706.73
A	B	C	=A+B-C+1	

(Legend on next page)



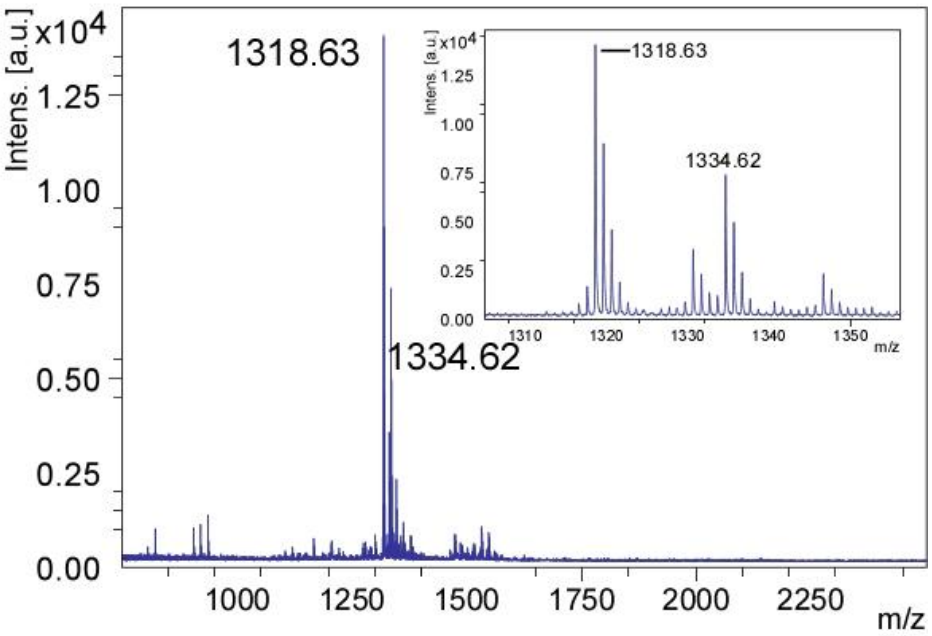
C

Error (ΔDa)

	1	2	3	4	5	6	7	8	9
	Phe	Pro	Pro	Ser	Glu	Leu	Ala	Tyr-U	Arg
a	-0.2	0.05	0.04	-0.02	---	0.09	0.09	---	---
b	---	0.06	0.09	---	0.03	-0.06	0.05	-0.09	-0.1
	Arg	Tyr-U	Ala	Leu	Glu	Ser	Pro	Pro	Phe
y	-0.01	-0.003	0.02	-0.03	0.02	0.02	-0.1	0.02	-0.002

Figure 4.16- MALDI TOF/TOF analysis of cross-link at U182
(A) Post alkaline hydrolysis and Zip-Tip clean up sample was spotted in alpha alpha-cyano-4-hydroxycinnamic acid for MALDI-TOF analysis to choose parent ion for MALDI-TOF/TOF. Inset shows a narrower window of peak at 1321.55 amu. The expected mass is calculated as in previous MALDI-TOF analysis. (B) MALDI-TOF/TOF spectrum of 1321.55 peak. Expected sequence is shown in inset with cross-link site indicated by a red star. Observed y and b ions are indicated in inset and indicated in main spectrum. Large mass shift of b7-b8 and y1-y2 is indicated matching the expected shift of Tyr-Uridine. (C) The mass difference between the expected mass and observed mass of fragmentation peaks is indicated in Daltons.

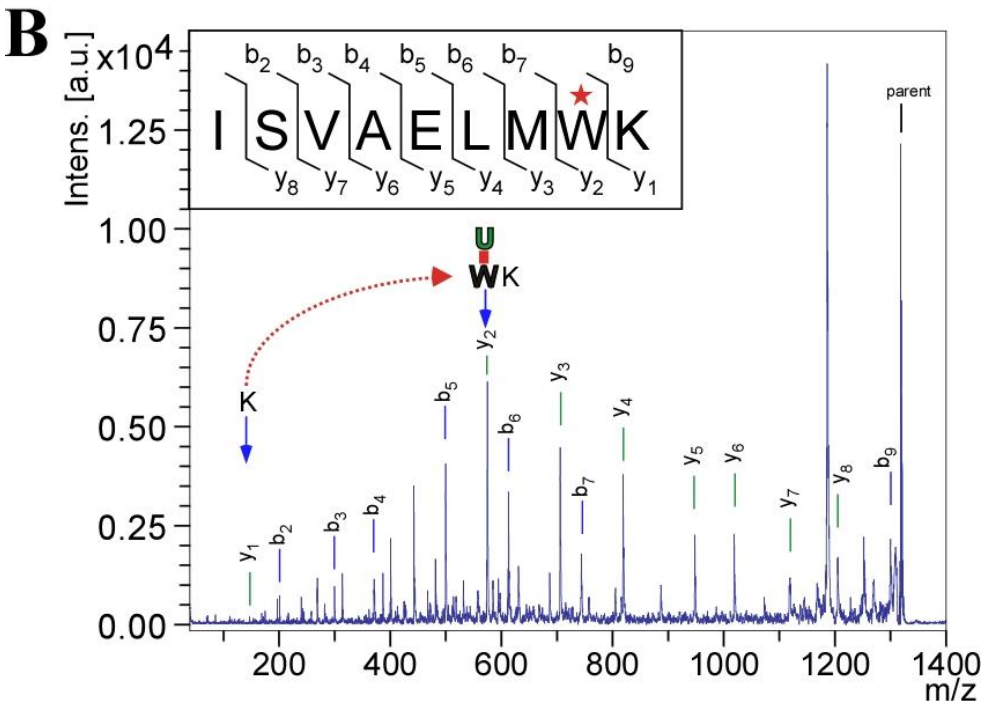
A



Peptide [M]	RNA [M]	Loss during X-link [M]	[M+H] calc	[M+H] exp
ISVAELMWK 1075.58	IU 369.97	(I + H) 127.90	1318.65	1318.63
ISVAELM*WK 1091.58	IU 369.97	(I + H) 127.90	1334.65	1334.62
A	B	C	=A+B-C+1	

* = oxidized methionine

(Legend on next page)



C

Error (ΔDa)

	1	2	3	4	5	6	7	8	9
	Ile	Ser	Val	Ala	Glu	Leu	Met	Trp-U	Lys
b	---	-0.3	-0.4	-0.4	-0.4	-0.4	-0.7	---	-0.3
	Lys	Trp-U	Met	Leu	Glu	Ala	Val	Ser	Ile
y	-0.1	-0.3	-0.4	-0.4	-0.4	-0.3	-0.5	-0.2	0.03

Figure 4.17- MALDI TOF/TOF analysis of cross-link at U205
(A) MALDI-TOF spectrum of alkaline hydrolyzed cross-link product using U182C RNA. Inset shows a narrower window of peak at 1318 and 1334 amu corresponding to predicted peptide and peptide with oxidized methionine. Calculated and expected masses are indicated. (B) MALDI-TOF/TOF spectrum of 1318.63 peak. Expected sequence is shown in inset with cross-link site indicated in by red star. Observed y and b ions are indicated in inset and indicated in main spectrum. Large mass shift of y1-y2 is indicated matching the expected shift of Trp-uridine. (C) Mass difference between the expected mass and observed mass of each fragmentation peak is indicated in Daltons.

4.5 DISCUSSION

4.5.1 MBP-TRBD –RNA interacts with both pseudoknot and CR45/CR5-

Purified MBP-TRBD protein binds to both CR4/CR5 and pseudoknot RNAs. Binding to pseudoknot RNA does not appear to involve the conserved triple helix. The greatest level of binding is observed only when an intact P1 is included. It is not known if binding is directly to P1 or the template containing single stranded region also included in the full length pseudoknot construct used. While pseudoknot binding was observed, it was competed away by inclusion of CR4/CR5 in the binding reaction.

4.5.2 RNA-cross-linking sites-

While cross-linking indicates two residues are in close proximity to each other, it does not indicate actual physical contact. Cross-linking with 5-Iodo nucleotides typically occurs with aromatic amino acids that can be stacked with nucleotides. The U-bulge identified does appear to be important for binding as deletion of this bulge leads to a dramatic decrease in binding as seen by the Kd shift of 0.55μM to ~4μM. It is possible that the function of the U-bulge is to allow the RNA helix to allow it to wrap around the protein similar to the required bending of *Tetrahymena* Stem IV by p65(Stone et al., 2007). Mutating the uridine in the bulge to a cytosine leads to a modest Kd shift, an increase of about 3 fold, indicating that the U may be involved in a protein interaction rather than just bending the RNA. It is possible that the increase in Kd is due to a perturbed secondary structure. U_{182} forms the bulge and is followed by a cytodine. By mutating U_{182} to cytodine, it is possible that the RNA now forms two different

base pairings: one with a bulge at residue 182, the other with a bulge at 183. This change in secondary structure would lead to a disturbed tertiary structure, changing the folding angle slightly thereby decreasing the protein-RNA interaction.

It has recently been shown that the P6.1 U_{205} equivalent in human TR, U_{307}, may be pseudouridylated *in vivo* (Kim et al., 2010*)*. The pseudouridylated p6.1 had a significantly different loop structure; however, the conserved loop U is still solvent exposed, though slightly less accessible for stacking or interacting with the protein. It is possible that pseudouridylation could affect how the CR4/CR5 interacts with TRBD. It would be interesting to determine if pseudouridylation of the p6.1 U_{205} would affect the Kd measured for CR4/CR5 to recombinant TRBD.

Mutation of $U_{205}A/G_{207}C$ did not affect the Kd concurring with previously published data for mouse TR (Chen et al., 2002a). Mutation of $U_{205}A$ would not necessarily eliminate a stacking interaction between the base and an aromatic amino acid. Mutation of the corresponding amino acid W_{477} to another aromatic amino acid did not affect binding. Mutation to a hydrophobic non-aromatic amino acid did affect binding, shifting the Kd about 4 fold. This change in binding appears to indicate that an aromatic amino acid is important at this site to maintain a stacking interaction, though it is possible that the leucine residue alters protein structure more than phenylalanine does. It is also possible that $U_{205}A/G_{207}C$ did not affect the Kd, because at position 204 is a second U which could also interact with W_{477}. This is supported by the presence of a minor peak

located at 1318 amu, the mass of cross-link at U_{205}, seen even when $U_{205}A/G_{207}C$ mutation is used (See figure 4.16A inset). This minor peak is located close to the 1321.55 peak, which was sequenced, yet does not match the isotopic mass pattern expected, indicating it is a separate peak rather than an isotope difference of the sequenced peak. With U_{205} mutated, this peak should disappear completely since U_{205} is no longer available for crosslinking. Continued presence of this peak could be due to cross-linking at U_{204}.

4.5.3 Mapping cross-linking site onto known structures to identify interaction interface

Mapping of the two identified cross-linking sites onto the TRBD of *Tribolium* TERT (Gillis et al., 2008) and *Tetrahymena* TRBD (Rouda and Skordalakes, 2007) indicates the two peptides are in close proximity, on helix 9 and 11 of *Tetrahymena* TRBD and helix 6 and 8 in the *Tribolium* TERT (Figures 4.18A/B and 4.19). These helices do not correspond to the RNA-binding pocket previously proposed for *Tetrahymena* TRBD (Rouda and Skordalakes, 2007). The RNA binding pocket of *Tetrahymena* TRBD was proposed to interact with the template boundary element, so it is not entirely surprising that CR4/CR5 interaction interface does not co-locate to the template boundary element RNA binding pocket. The TRBD-CR4/CR5 interaction interface is a predominantly positively charged patch on both the *Tetrahymena* and *Tribloium* structures. When the medaka TERT sequence is threaded through either structure, using SWISS MODEL Worskspace (Arnold et al., 2006), the surface remains a

positively charged patch as would be expected of an RNA binding interface. (Figure 4.18 C/D and 4.20)

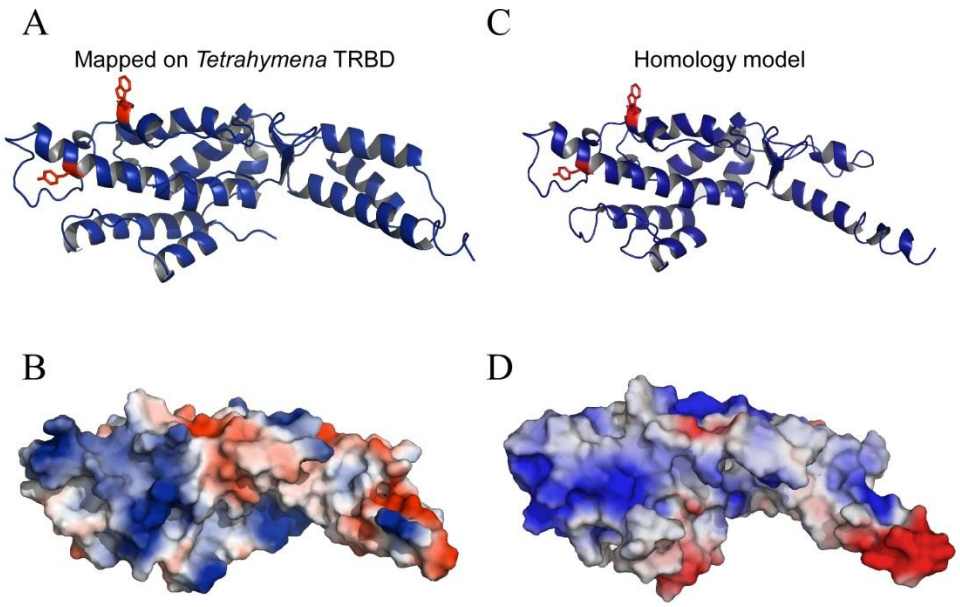


Figure 4.18-Mapping and modeling with Tetrahymena TRBD-
(A) Cross-linking amino acids (red) of medaka were mapped to *Tetrahymena* TRBD structure. (B) Surface rendering of structure in A with coloring by partial charge with blue being positive and red negative. (C) Medaka TRBD sequence was threaded through the *Tetrahymena* TRBD structure using the SWISS-MODEL workspace. Trp_{477} and Tyr_{503} are colored blue. (D) Surface rendering of threading model colored by partial charge.

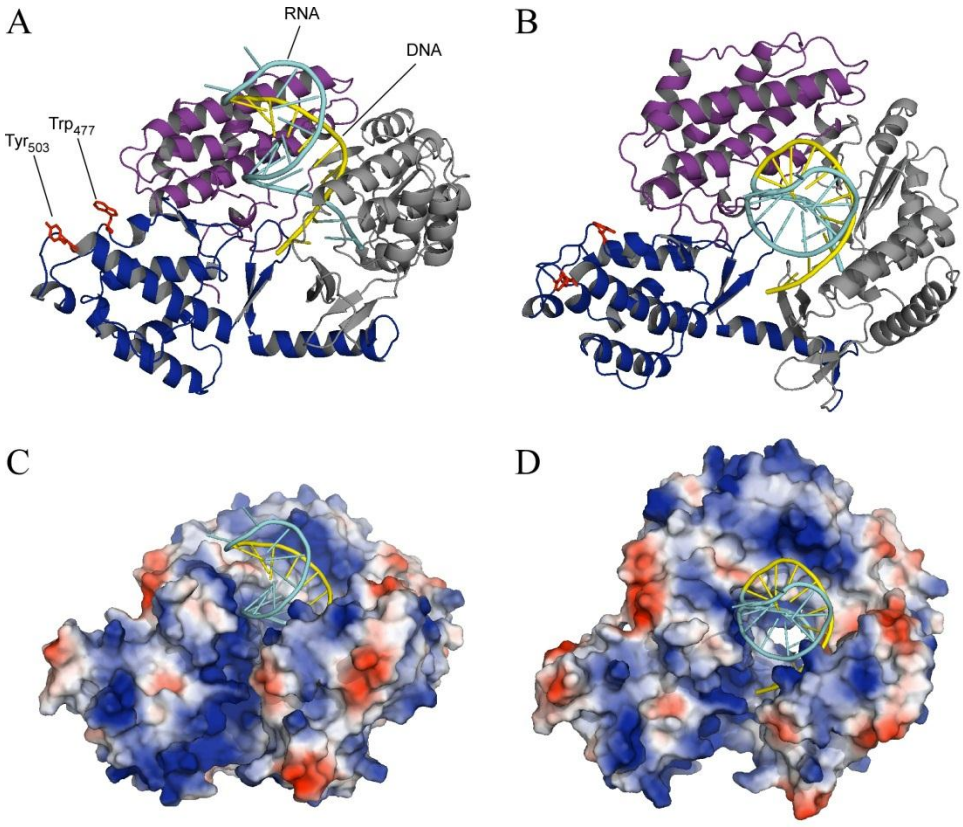


Figure 4.19- Trp_{477} and Tyr_{503} mapped onto *Tribolium* TERT structure
Cross-linking residues identified were mapped to the corresponding positions of *Tribolium* TERT structure. (A) Corresponding Amino acids were mutated to Trp and Tyr and colored red. TRBD is colored in blue while RT is grey and CTE is purple. Yellow is the DNA strand of a heteroduplex while cyan is RNA. (B) Rotated view of structure in A. (C) The surface of the structure is colored by partial charge with blue being positive and red negative. (D) Rotated view of C.

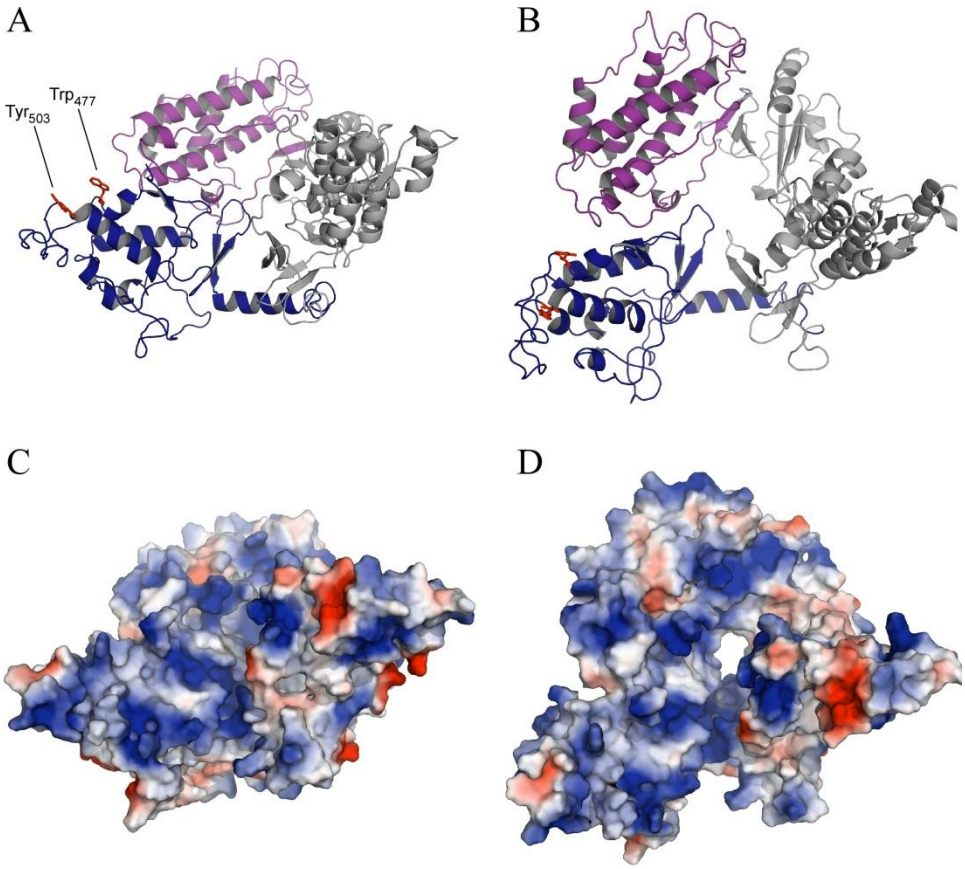


Figure 4.20- Homology model of medaka TERT
The medaka sequence was threaded through the *Tribolium* TERT structure using the SWISS-MODEL Workspace. (A) Cartoon of resulting structure with Trp_{477} and Tyr_{503} indicated in red, TRBD in blue, RT in grey, and CTE in purple. (B) Rotated view of structure in A. (C) The surface of the structure is colored by partial charge with blue being positive and red negative. (D) Rotated view of C.

4.5.4 Implication of cross-linking on CR4/CR5 tertiary structure-

The spacing of the two amino acids provides insight into the possible tertiary structure of CR4/CR5. When mapped onto the existing *Tetrahymena* TRBD crystal structure, or the crystal structure used for homology modeling, the two cross-linking residues are separated by about 20 angstroms. Mapped or threaded with the *Tribolium* structure, the residues are separated by about 15 angstroms. When the NMR structures for the human p6 and p6.1 (Leeper et al.,

2003; Leeper and Varani, 2005) are coaxial stacked, the nucleotides homologues to U_{182} and U_{205} are separated by 45 angstroms (Figure 4.21). The cross-linking data eliminates coaxial stacking as a potential configuration of p6 and p6.1. It is more likely that the two helices are closer to a parallel configuration as this puts the two nucleotides approximately 20 angstroms apart. The bulged nucleotide, a cytosine in human, is buried in the major groove of the RNA when lacking the interacting protein. Protein binding could flip the bulged nucleotide farther out of the helix allowing a larger angle, than parallel positioning of the helices, while maintaining the separation of about 20 angstroms.

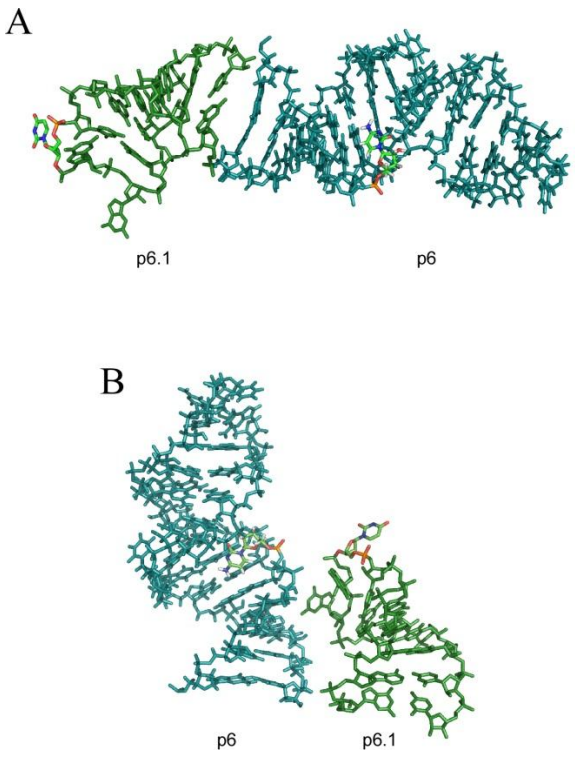


Figure 4.21- Human p6 and p6.1 RNA
NMR structures of human p6 (teal) (Leeper and Varani, 2005) and p6.1 (green) (Leeper et al., 2003) are shown. Nucleotides homologoues to U_{182} and U_{205} of medaka TR are colored by element. (A) Coaxial stacking of p6 and p6.1 (B) Parallel positioning of p6 and p6.1.

It has been observed that the U_{205} and G_{207} of p6.1 are important for activity. When mapped onto duplex bound *Tribolium* crystal structure tryptophan 477, it is located approximately 45 angstroms from the 3' end of the telomere. Cross-linking of U_{205} to tryptophan 477 places the p6.1 loop residues too far from the 3' end of the DNA to be directly involved in catalysis. P6.1, however, could interact with the 3' end of the template as template threads into and out of the active site during extension and translocation.

The methods developed for this study are simple method to map the RNA-protein interaction of any RNA binding protein. The methods could be built upon through the use of RNAs from closely related species. For instance medaka TERT can be reconstituted with CR4/CR5 from fugu and zebrafish (Xie et al., 2008) allowing the use of two additional CR4/CR5 with differently positioned uredines to map positions within the binding interface. Additionally, it would be interesting to apply these methods to *Tetrahymena* TRBD determine which portion of the RNA interacts with the TRBD. Since Stem IV of *Tetrahymena* has been proposed to be similar to CR4/CR5 (Robart et al., 2010), it would be interesting to see if this stem cross-links to the same interface of TRBD.

REFERENCES

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195-201.

Chen, J.-L., Opperman, K.K., and Greider, C.W. (2002a). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res 30, 592-597.

Chen, J.L., Opperman, K.K., and Greider, C.W. (2002b). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res 30, 592-597.

Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the Tribolium castaneum telomerase catalytic subunit TERT. Nature 455, 633-637.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405-413.

Hiley, S.L., Sood, V.D., Fan, J., and Collins, R.A. (2002). 4-thio-U cross-linking identifies the active site of the VS ribozyme. EMBO J 21, 4691-4698.

Kim, N.K., Theimer, C.A., Mitchell, J.R., Collins, K., and Feigon, J. (2010). Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. Nucleic Acids Res 38, 6746-6756.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science 266, 2011-2015.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol Cell Biol 21, 990-1000.

Leeper, T., Leulliot, N., and Varani, G. (2003). The solution structure of an essential stem-loop of human telomerase RNA. Nucleic Acids Res 31, 2614-2621.

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

Meisenheimer, K.M., and Koch, T.H. (1997). Photocross-linking of nucleic acids to associated proteins. Crit Rev Biochem Mol Biol 32, 101-140.

Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15, 8783-8798.

Moriarty, T.J., Huard, S., Dupuis, S., and Autexier, C. (2002). Functional multimerization of human telomerase requires an RNA interaction domain in the N terminus of the catalytic subunit. Mol Cell Biol 22, 1253-1265.

Moriarty, T.J., Marie-Egyptienne, D.T., and Autexier, C. (2004). Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. Mol Cell Biol 24, 3720-3733.

Norris, C.L., Meisenheimer, P.L., and Koch, T.H. (1996). Mechanistic studies of the 5-iodouracil chromophore relevant to its use in nucleoprotein photo-cross-linking. J Am Chem Soc 118, 5796-5803.

Robart, A.R., O'Connor, C.M., and Collins, K. (2010). Ciliate telomerase RNA loop IV nucleotides promote hierarchical RNP assembly and holoenzyme stability. RNA 16, 563-571.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure 15, 1403-1412.

Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 5, 699-711.

Steen, H., Petersen, J., Mann, M., and Jensen, O.N. (2001). Mass spectrometric analysis of a UV-cross-linked protein-DNA complex: tryptophans 54 and 88 of E. coli SSB cross-link to DNA. Protein Sci 10, 1989-2001.

Stone, M.D., Mihalusova, M., O'Connor, C.M., Prathapam, R., Collins, K., and Zhuang, X. (2007). Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. Nature 446, 458-461.

Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P., Franzen, J., and Holle, A. (2003). A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal Bioanal Chem 376, 952-965.

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

Ueda, C.T., and Roberts, R.W. (2004). Analysis of a long-range interaction between conserved domains of human telomerase RNA. RNA 10, 139-147.

Willis, M.C., Hicke, B.J., Uhlenbeck, O.C., Cech, T.R., and Koch, T.H. (1993). Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. Science 262, 1255-1257.

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F., and Chen, J.J.-L. (2008). Structure and function of the smallest vertebrate telomerase RNA from teleost fish. J Biol Chem 283, 2049-2059.

Chapter 5

CONCLUSION

To begin to understand the many roles of telomerase RNA and how it assembles with TERT protein, soluble protein fragments of TERT protein were identified. Creation of a synthetic, codon optimized, gene was a key to generating over expressed, soluble protein of high purity. Multiple soluble TERT protein fragments were identified, including TRBD, and full length mdTERT. Only TRBD was a non-aggregate, suitable for biochemical study under the conditions tested. Purification conditions of TRBD protein only, and TRBD bound to CR4/CR5 were optimized to give high yield and purity of monomeric products. The optimized purification schemes devised will facilitate future structural studies of medaka Telomerase.

The RNA binding characteristics of medaka TRBD were analyzed. TRBD was found to interact with both pseudoknot, and CR4/CR5 RNA. Using the photo-reactive crosslinking reagent 5-Iodo-Uridine to randomly label CR4/CR5 RNA, cross-linking from RNA to protein was performed. Two cross-linking sites on the RNA were identified: U_{182} and U_{205}. Peptide sequencing by mass spectrometry identified the amino acid partners, Trp_{477} and Tyr_{503} of mdTERT. Through mapping of the cross-link sites onto know TERT structures, the CR4/CR5 RNA interaction interface of TRBD was determined. The cross-linking sites placed p6.1 too distant from the 3' end of the DNA to be directly involved in activity at the active site. Additionally, the spatial separation of the two cross-

linking sites would prevent coaxial stacking of p6 and p6.1, a proposed conformation of the CR4/CR5 RNA.

The methods developed could readily be adapted to most nucleic acid-protein interactions, in which the interaction interface is unknown. The methods can also be used to study the complex of distantly related telomerase to identify homologous regions of the RNA.

REFERENCES

Chapter 1

Allsopp, R.C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W., and Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc Natl Acad Sci U S A *89*, 10114-10118.

Armanios, M. (2009). Syndromes of telomere shortening. Annu Rev Genomics Hum Genet *10*, 45-61.

Avilion, A.A., Piatyszek, M.A., Gupta, J., Shay, J.W., Bacchetti, S., and Greider, C.W. (1996). Human telomerase RNA and telomerase activity in immortal cell lines and tumor tissues. Cancer Res *56*, 645-650.

Blackburn, E.H. (2010). Telomeres and telomerase: the means to the end (Nobel lecture). Angew Chem Int Ed Engl *49*, 7405-7421.

Blackburn, E.H., and Gall, J.G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol *120*, 33-53.

Blasco, M.A., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A., and Greider, C.W. (1997). Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. Cell *91*, 25-34.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. Science *279*, 349-352.

Chen, J.L., Blasco, M.A., and Greider, C.W. (2000). Secondary structure of vertebrate telomerase RNA. Cell *100*, 503-514.

Chen, J.L., Opperman, K.K., and Greider, C.W. (2002). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res *30*, 592-597.

Dandjinou, A.T., Lévesque, N., Larose, S., Lucier, J.-F., Abou Elela, S., and Wellinger, R.J. (2004). A phylogenetically based secondary structure for the yeast telomerase RNA. Curr Biol *14*, 1148-1158.

Feng, J., Funk, W.D., Wang, S.S., Weinrich, S.L., Avilion, A.A., Chiu, C.P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., *et al.* (1995). The RNA component of human telomerase. Science *269*, 1236-1241.

Folini, M., Brambilla, C., Villa, R., Gandellini, P., Vignati, S., Paduano, F., Daidone, M.G., and Zaffaroni, N. (2005). Antisense oligonucleotide-mediated inhibition of hTERT, but not hTERC, induces rapid cell growth decline and apoptosis in the absence of telomere shortening in human prostate cancer cells. Eur J Cancer *41*, 624-634.

Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the Tribolium castaneum telomerase catalytic subunit TERT. Nature *455*, 633-637.

Greider, C.W. (2010). Telomerase discovery: the excitement of putting together pieces of the puzzle (Nobel lecture). Angew Chem Int Ed Engl *49*, 7422-7439.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell *43*, 405-413.

Greider, C.W., and Blackburn, E.H. (1987). The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell *51*, 887-898.

Greider, C.W., and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature *337*, 331-337.

Greider, C.W., and Blackburn, E.H. (2004). Tracking telomerase. Cell *116*, S83-86, 81 p following S86.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet *19*, 32-38.

Jacobs, S.A., Podell, E.R., and Cech, T.R. (2006). Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol *13*, 218-225.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science *266*, 2011-2015.

Klobutcher, L.A., Swanton, M.T., Donini, P., and Prescott, D.M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. Proc Natl Acad Sci U S A *78*, 3015-3019.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol Cell Biol *21*, 990-1000.

Leeper, T., Leulliot, N., and Varani, G. (2003). The solution structure of an essential stem-loop of human telomerase RNA. Nucleic Acids Res *31*, 2614-2621.

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

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. Science *276*, 561-567.

Masutomi, K., Possemato, R., Wong, J.M.Y., Currier, J.L., Tothova, Z., Manola, J.B., Ganesan, S., Lansdorp, P.M., Collins, K., and Hahn, W.C. (2005). The telomerase reverse transcriptase regulates chromatin state and DNA damage responses. Proc Natl Acad Sci USA *102*, 8222-8227.

McClintock, B. (1939). The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis. P Natl Acad Sci USA *25*, 405-416.

Meyne, J., Ratliff, R.L., and Moyzis, R.K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci U S A *86*, 7049-7053.

Mitchell, J.R., Cheng, J., and Collins, K. (1999a). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. Mol Cell Biol *19*, 567-576.

Mitchell, J.R., Wood, E., and Collins, K. (1999b). A telomerase component is defective in the human disease dyskeratosis congenita. Nature *402*, 551-555.

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

Moriarty, T.J., Marie-Egyptienne, D.T., and Autexier, C. (2004). Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. Mol Cell Biol *24*, 3720-3733.

Morin, G.B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. Cell *59*, 521-529.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., and Wu, J.R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A *85*, 6622-6626.

Muller, H. (1938). The remaking of chromosomes. The Collecting Net *13*, 181-198.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. (1997). Telomerase catalytic subunit homologs from fission yeast and human. Science *277*, 955-959.

Pace, N.R., Smith, D.K., Olsen, G.J., and James, B.D. (1989). Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA- a review. Gene *82*, 65-75.

Ramirez, R.D., Morales, C.P., Herbert, B.S., Rohde, J.M., Passons, C., Shay, J.W., and Wright, W.E. (2001). Putative telomere-independent mechanisms of replicative aging reflect inadequate growth conditions. Genes Dev *15*, 398-403.

Romero, D.P., and Blackburn, E.H. (1991). A conserved secondary structure for telomerase RNA. Cell *67*, 343-353.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure *15*, 1403-1412.

Saretzki, G., Ludwig, A., von Zglinicki, T., and Runnebaum, I.B. (2001). Ribozyme-mediated telomerase inhibition induces immediate cell loss but not telomere shortening in ovarian cancer cells. Cancer Gene Ther *8*, 827-834.

Shampay, J., Szostak, J.W., and Blackburn, E.H. (1984). DNA sequences of telomeres maintained in yeast. Nature *310*, 154-157.

Shay, J.W., and Bacchetti, S. (1997). A survey of telomerase activity in human cancer. Eur J Cancer *33*, 787-791.

Shay, J.W., and Wright, W.E. (2002). Telomerase: a target for cancer therapeutics. Cancer Cell *2*, 257-265.

Szostak, J.W. (2010). DNA ends: just the beginning (Nobel lecture). Angew Chem Int Ed Engl *49*, 7386-7404.

Szostak, J.W., and Blackburn, E.H. (1982). Cloning yeast telomeres on linear plasmid vectors. Cell *29*, 245-255.

Vidal-Cardenas, S.L., and Greider, C.W. (2010). Comparing effects of mTR and mTERT deletion on gene expression and DNA damage response: a critical examination of telomere length maintenance-independent roles of telomerase. Nucleic Acids Res *38*, 60-71.

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J., and Dokal, I. (2001a). The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature *413*, 432-435.

Vulliamy, T.J., Knight, S.W., Mason, P.J., and Dokal, I. (2001b). Very short telomeres in the peripheral blood of patients with X-linked and autosomal dyskeratosis congenita. Blood Cells Mol Dis *27*, 353-357.

Vulliamy, T.J., Marrone, A., Knight, S.W., Walne, A., Mason, P.J., and Dokal, I. (2006). Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation. Blood *107*, 2680-2685.

Walne, A.J., and Dokal, I. (2009). Advances in the understanding of dyskeratosis congenita. British Journal of Haematology *145*, 164-172.

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F., and Chen, J.J.-L. (2008). Structure and function of the smallest vertebrate telomerase RNA from teleost fish. J Biol Chem *283*, 2049-2059.

Xie, M., Podlevsky, J.D., Qi, X., Bley, C.J., and Chen, J.J. (2010). A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. Nucleic Acids Res *38*, 1982-1996.

Zhong, F., Savage, S.A., Shkreli, M., Giri, N., Jessop, L., Myers, T., Chen, R., Alter, B.P., and Artandi, S.E. (2011). Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. Genes Dev *25*, 11-16.

Chapter 2

Agashe, V.R., Guha, S., Chang, H.C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.U., and Barral, J.M. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. Cell 117, 199-209.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol 22, 1399-1408.

Castanie, M.P., Berges, H., Oreglia, J., Prere, M.F., and Fayet, O. (1997). A set of pBR322-compatible plasmids allowing the testing of chaperone-assisted folding of proteins overexpressed in Escherichia coli. Anal Biochem 254, 150-152.

Cristofari, G., and Lingner, J. (2006). Telomere length homeostasis requires that telomerase levels are limiting. EMBO J 25, 565-574.

Dennis, P.P., and Bremer, H. (1974). Differential rate of ribosomal protein synthesis in Escherichia coli B/r. J Mol Biol 84, 407-422.

Dennis, P.P., and Nomura, M. (1974). Stringent Control of Ribosomal-Protein Gene-Expression in Escherichia-Coli. P Natl Acad Sci USA 71, 3819-3823.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400, 693-696.

Forsythe, H.L., Jarvis, J.L., Turner, J.W., Elmore, L.W., and Holt, S.E. (2001). Stable association of hsp90 and p23, but Not hsp70, with active human telomerase. J Biol Chem 276, 15571-15574.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405-413.

Haacke, A., Fendrich, G., Ramage, P., and Geiser, M. (2009). Chaperone over-expression in Escherichia coli: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. Protein Expr Purif 64, 185-193.

Hecker, K.H., and Rill, R.L. (1998). Error analysis of chemically synthesized polynucleotides. Biotechniques 24, 256-260.

Higuchi, R., Krummel, B., and Saiki, R.K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16, 7351-7367.

Holt, S.E., Aisner, D.L., Baur, J., Tesmer, V.M., Dy, M., Ouellette, M., Trager, J.B., Morin, G.B., Toft, D.O., Shay, J.W., et al. (1999). Functional requirement of p23 and Hsp90 in telomerase complexes. Genes Dev 13, 817-826.

Hoover, D.M., and Lubkowski, J. (2002). DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res 30, e43.

Jacobs, S.A., Podell, E.R., and Cech, T.R. (2006). Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol 13, 218-225.

Jacobs, S.A., Podell, E.R., Wuttke, D.S., and Cech, T.R. (2005). Soluble domains of telomerase reverse transcriptase identified by high-throughput screening. Protein Sci 14, 2051-2058.

Kane, J.F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol 6, 494-500.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science 324, 255-258.

Masutomi, K., Kaneko, S., Hayashi, N., Yamashita, T., Shirota, Y., Kobayashi, K., and Murakami, S. (2000). Telomerase activity reconstituted in vitro with purified human telomerase reverse transcriptase and human telomerase RNA component. J Biol Chem 275, 22568-22573.

Medigue, C., Rouxel, T., Vigier, P., Henaut, A., and Danchin, A. (1991). Evidence for horizontal gene transfer in Escherichia coli speciation. J Mol Biol 222, 851-856.

Pedersen, S. (1984). Escherichia coli ribosomes translate in vivo with variable rate. EMBO J 3, 2895-2898.

Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., de Arruda, M., et al. (2006). Emergent properties of reduced-genome Escherichia coli. Science 312, 1044-1046.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure 15, 1403-1412.

Straus, D., Walter, W., and Gross, C.A. (1990). DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32. Genes Dev 4, 2202-2209.

Villalobos, A., Ness, J.E., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006). Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics 7, 285.

Welch, M., Govindarajan, S., Ness, J.E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009). Design parameters to control synthetic gene expression in Escherichia coli. PLoS One 4, e7002.

Wu, C.K., Gousset, K., and Hughes, S.H. (2007). Targeting to the endoplasmic reticulum improves the folding of recombinant human telomerase reverse transcriptase. Protein Expr Purif 56, 8-19.

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F., and Chen, J.J.-L. (2008). Structure and function of the smallest vertebrate telomerase RNA from teleost fish. J Biol Chem 283, 2049-2059.

Chapter 3

Baldwin, R.L. (1996). How Hofmeister ion interactions affect protein stability. Biophys J 71, 2056-2063.

Foster, P.R., Dunnill, P., and Lilly, M.D. (1976). The kinetics of protein salting-out: precipitation of yeast enzymes by ammonium sulfate. Biotechnol Bioeng 18, 545-580.

Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the Tribolium castaneum telomerase catalytic subunit TERT. Nature 455, 633-637.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405-413.

Grodberg, J., and Dunn, J.J. (1988). ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. J Bacteriol 170, 1245-1253.

Jacobs, S.A., Podell, E.R., and Cech, T.R. (2006). Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol 13, 218-225.

Jacobs, S.A., Podell, E.R., Wuttke, D.S., and Cech, T.R. (2005). Soluble domains of telomerase reverse transcriptase identified by high-throughput screening. Protein Sci 14, 2051-2058.

Kao-Huang, Y., Revzin, A., Butler, A.P., O'Conner, P., Noble, D.W., and von Hippel, P.H. (1977). Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. Proc Natl Acad Sci U S A 74, 4228-4232.

Kapust, R.B., and Waugh, D.S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci 8, 1668-1674.

Kieft, J.S., and Batey, R.T. (2004). A general method for rapid and nondenaturing purification of RNAs. RNA 10, 988-995.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol Cell Biol 21, 990-1000.

Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15, 8783-8798.

Mitchell, J.R., and Collins, K. (2000). Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. Mol Cell 6, 361-371.

Moriarty, T.J., Marie-Egyptienne, D.T., and Autexier, C. (2004). Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. Mol Cell Biol 24, 3720-3733.

Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., de Arruda, M., et al. (2006). Emergent properties of reduced-genome Escherichia coli. Science 312, 1044-1046.

Price, S.R., Ito, N., Oubridge, C., Avis, J.M., and Nagai, K. (1995). Crystallization of RNA-protein complexes. I. Methods for the large-scale

preparation of RNA suitable for crystallographic studies. J Mol Biol 249, 398-408.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure 15, 1403-1412.

Sharma, S.S., Blattner, F.R., and Harcum, S.W. (2007). Recombinant protein production in an Escherichia coli reduced genome strain. Metab Eng 9, 133-141.

Chapter 4

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195-201.

Chen, J.-L., Opperman, K.K., and Greider, C.W. (2002a). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res 30, 592-597.

Chen, J.L., Opperman, K.K., and Greider, C.W. (2002b). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res 30, 592-597.

Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the Tribolium castaneum telomerase catalytic subunit TERT. Nature 455, 633-637.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43, 405-413.

Hiley, S.L., Sood, V.D., Fan, J., and Collins, R.A. (2002). 4-thio-U cross-linking identifies the active site of the VS ribozyme. EMBO J 21, 4691-4698.

Kim, N.K., Theimer, C.A., Mitchell, J.R., Collins, K., and Feigon, J. (2010). Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. Nucleic Acids Res 38, 6746-6756.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science 266, 2011-2015.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol Cell Biol 21, 990-1000.

Leeper, T., Leulliot, N., and Varani, G. (2003). The solution structure of an essential stem-loop of human telomerase RNA. Nucleic Acids Res 31, 2614-2621.

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

Meisenheimer, K.M., and Koch, T.H. (1997). Photocross-linking of nucleic acids to associated proteins. Crit Rev Biochem Mol Biol 32, 101-140.

Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15, 8783-8798.

Moriarty, T.J., Huard, S., Dupuis, S., and Autexier, C. (2002). Functional multimerization of human telomerase requires an RNA interaction domain in the N terminus of the catalytic subunit. Mol Cell Biol 22, 1253-1265.

Moriarty, T.J., Marie-Egyptienne, D.T., and Autexier, C. (2004). Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. Mol Cell Biol 24, 3720-3733.

Norris, C.L., Meisenheimer, P.L., and Koch, T.H. (1996). Mechanistic studies of the 5-iodouracil chromophore relevant to its use in nucleoprotein photo-cross-linking. J Am Chem Soc 118, 5796-5803.

Robart, A.R., O'Connor, C.M., and Collins, K. (2010). Ciliate telomerase RNA loop IV nucleotides promote hierarchical RNP assembly and holoenzyme stability. RNA 16, 563-571.

Rouda, S., and Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure 15, 1403-1412.

Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 5, 699-711.

Steen, H., Petersen, J., Mann, M., and Jensen, O.N. (2001). Mass spectrometric analysis of a UV-cross-linked protein-DNA complex: tryptophans 54 and 88 of E. coli SSB cross-link to DNA. Protein Sci 10, 1989-2001.

Stone, M.D., Mihalusova, M., O'Connor, C.M., Prathapam, R., Collins, K., and Zhuang, X. (2007). Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. Nature 446, 458-461.

Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P., Franzen, J., and Holle, A. (2003). A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal Bioanal Chem 376, 952-965.

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

Ueda, C.T., and Roberts, R.W. (2004). Analysis of a long-range interaction between conserved domains of human telomerase RNA. RNA 10, 139-147.

Willis, M.C., Hicke, B.J., Uhlenbeck, O.C., Cech, T.R., and Koch, T.H. (1993). Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. Science 262, 1255-1257.

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F., and Chen, J.J.-L. (2008). Structure and function of the smallest vertebrate telomerase RNA from teleost fish. J Biol Chem 283, 2049-2059.

Appendix A

Nirenberg, M., and Leder, P. (1964). Rna Codewords and Protein Synthesis. The Effect of Trinucleotides Upon the Binding of Srna to Ribosomes. Science 145, 1399-1407.

Smith, C.W.J. (1998). RNA:protein interactions : a practical approach (Oxford ; New York, Oxford University Press).

Wong, I., and Lohman, T.M. (1993). A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. Proc Natl Acad Sci U S A 90, 5428-5432.

APPENDIX A

FILTER BINDING ASSAY OPTIMIZATION

The filter binding assay used in chapter four was optimized for maximal signal of specific binding while minimizing non-specific binding. Variables optimized included, number and size of washes used, number and type of filters used, and incubation time prior to filtering. The filter binding assay was first described for measuring the interaction between tRNA and the ribosome(Nirenberg and Leder, 1964). The basic principle behind the filter binding assay is to separate free RNA from RNA bound to protein and quantitate the protein bound RNA. By keeping the RNA concentration constant at least 100-fold lower than the lowest protein concentration one can determine the dissociation constant by titrating the protein. The concentration of protein at 50% binding corresponds to Kd (Smith, 1998).

The filter binding assay uses membranes to separate free and protein bound RNA. The membrane is placed on an apparatus to which a vacume is applied (Figure A1.1), as the sample passes through the membrane sample that interacts with the membrane is retained. The primary membrane, often nitrocellulose, should not interact with free RNA, allowing it to pass through the membrane. Protein should be held up by the membrane without major structural changes that would cause any protein bound RNA to be released. In this way only RNA that is bound by protein is retained on the primary membrane. The filter binding assay methods have remained relatively unchanged since the first implementation of the experiment. One improvement that has been implemented in some cases is to add a second membrane below the primary membrane (Figure A1.1). Wong et al. used a DEAE membrane as the secondary membrane(Wong

and Lohman, 1993). The role of the secondary membrane is to capture all RNA that passes through the primary membrane. By measuring the radioactive RNA on each membrane, the percentage of RNA bound by protein in each sample is easily calculated. One advantage of the two membrane filter binding system is to correct for small errors in RNA level between samples that may arise from pipetting errors.

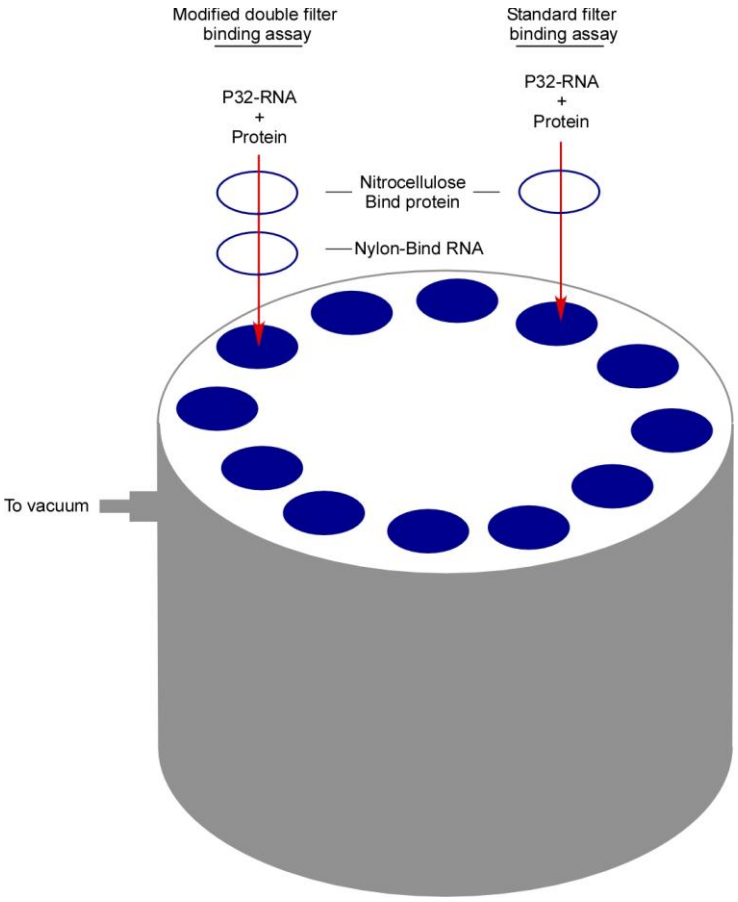


Figure A1.1- Filter binding apparatus
Twelve spot filter binding apparatus is shown comparing the standard and double and binding methods. In the double membrane method a nitrocellulose membrane captures protein bound RNA while the nylon membrane below it captures protein free RNA. The RNA captured on the membranes is used to calculate percent bound. In the standard method only a nitrocellulose membrane is used. Percent RNA bound is calculated by comparing to a separate membrane onto which total RNA is spotted but not washed.

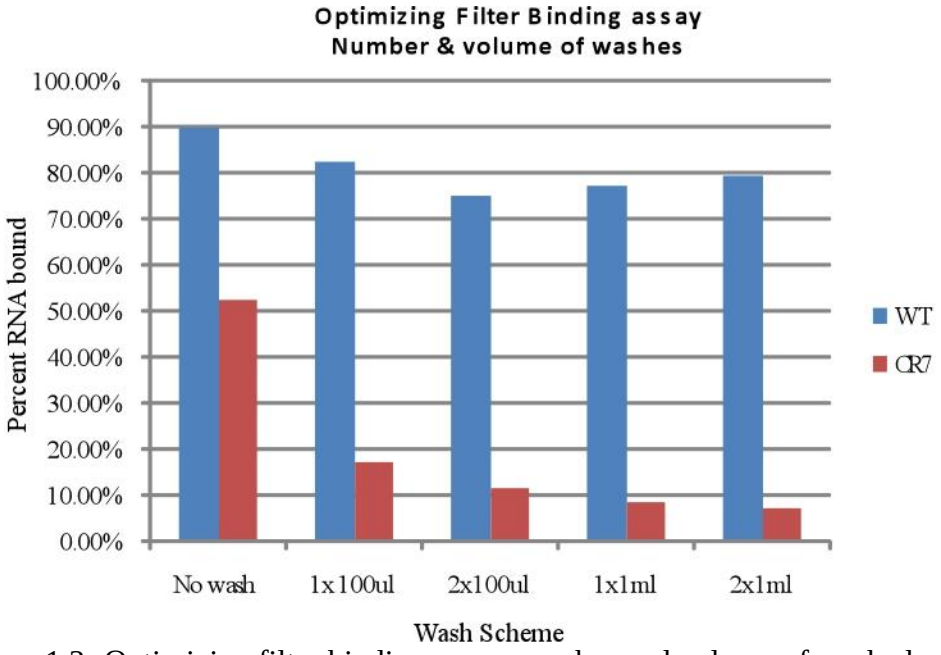


Figure 1.2- Optimizing filter binding assay, number and volume of washed
The number and size of washes after passing the binding reaction over membrane were tested. Signal of WT- CR4/CR5 did not change dramatically with the number or volume of washes. Using non-binding CR7 it was found a single 1 ml wash is sufficient to remove unbound RNA from the nitrocellulose membrane.

Since many variables of the filter binding assay, such as time of incubation before applying sample to the membrane, are protein specific, we set out to optimize the assay for TRBD-CR4/CR5 interaction. After passing the sample over the nitrocellulose membrane, the membrane is washed with binding buffer. We compared no washing with one or two washes of either one hundred microliters or one milliliter using CR4/CR5 as a positive sample and CR7 as a negative sample to control for non-specific background (Figure A1.2). No washing had significant background binding of CR7. With all washes, similar amount of specific RNA binding was observed (74-79% bound). Both 1 milliliter washes had slightly less nonspecific binding than the 100µl washes (9 and 7% vs

17 and 11%). For further studies we chose to use 1x1ml wash as optimal for both washing away non-specific binding as well as ease of use.

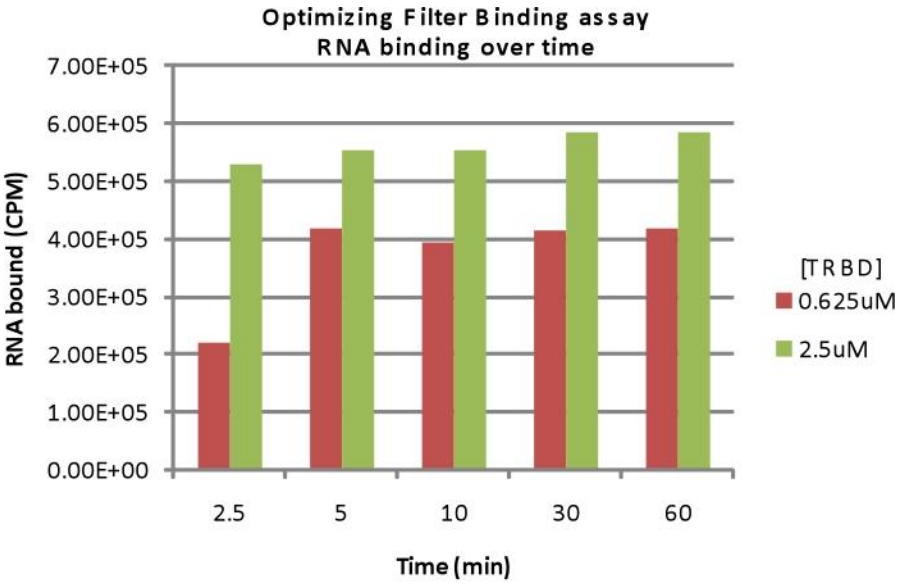


Figure A1.3- Optimizing filter binding assay, RNA binding over time
Incubation time of CR4/CR5 binding to MBP-TRBD before applying sample to filter binding apparatus was tested. Samples were allowed to bind from 60 minutes down to 2.5minutes. Two concentration of protein were used 0.625µM (red) and 2.5µM (green).

The optimal binding time before applying sample to membrane was also determined. Binding was compared for samples from 2 ½ minutes up to 1 hour at two different concentration of protein 0.625µM and 2.5µM, one above the suspected Kd and one near Kd (Figure A1.3). When using 2.5µM the amount of bound RNA did not significantly increase with time. When CR4/CR5 RNA is mixed with protein at a concentration close to the Kd, there was almost half as much binding at only 2.5 minutes compared to 5 minutes. Beyond 5 minutes there was no shift in the amount of RNA retained with the protein on the

nitrocellulose membrane. As there were no ill effects from binding for an extended period, a 10 minute binding step was chosen for further assays, as it allows time to set up the binding apparatus while binding is occurring.

The dual membrane version of the filter binding assay is reported to give more accurate results however we noticed an increase in total RNA reading when less protein was used. We hypothesized that some RNA was passing through the RNA binding membrane. CR4/CR5 is 50 nucleotides long, the same as the low end of RNA tolerance of the Hybond membrane being used, which may lead to decreased binding to the membrane. We tested the assay using one, two and three RNA capture membranes below the nitrocellulose protein capturing membrane. Though each successive RNA capture membrane had less bound RNA than the membrane above it, each had a significant amount of RNA bound (Figure A1.4). The total RNA was effectively increased with each additional RNA capture membrane used. Since the amount of RNA passing through the membranes was dependant on protein concentration we selected to compare to a single membrane spotted with the starting RNA level to indicate total RNA when determining the percent of RNA bound while determining Kd.

The conditions of the filter binding assay were compared and optimized for CR4/CR5 binding to medaka TRBD. Final conditions are to use the original single membrane assay due to the small size of the RNA being analyzed, a 10 minute incubation time, and a single 1ml wash.

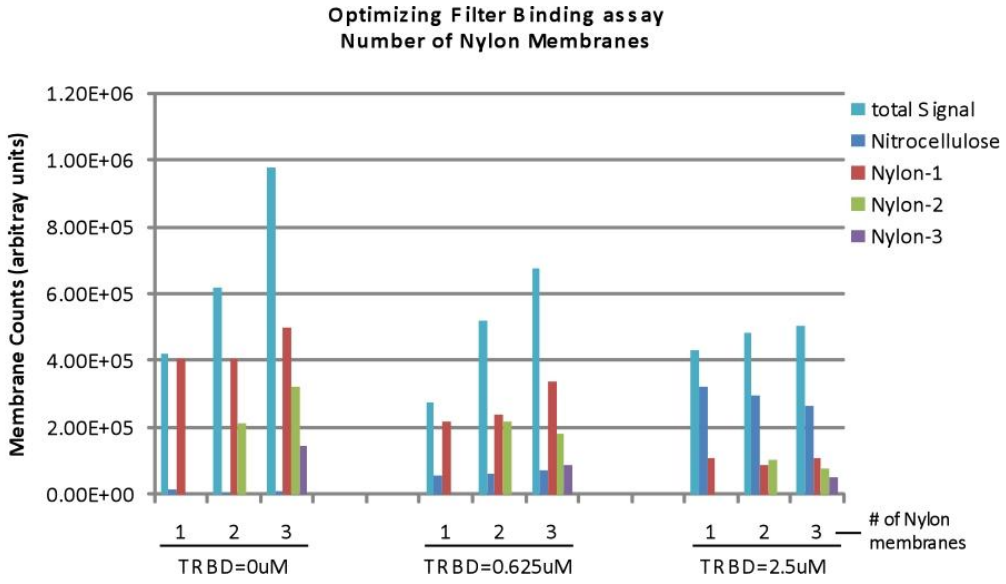


Figure A1.4- Optimizing filter binding assay, number of nylon membranes
The filter binding assay was performed with 0, 1, 2, and 3 nylon membranes below the nitrocellulose membrane. Two concentration of MBP-TRBD and buffer only series were tested. Signal from P-32 labeled CR4/CR5 was quantitated for each membrane and summed for each well (light blue). With each successive nylon membrane, total signal increased.

REFERENCES

Nirenberg, M., and Leder, P. (1964). Rna Codewords and Protein Synthesis. The Effect of Trinucleotides Upon the Binding of Srna to Ribosomes. Science 145, 1399-1407.

Smith, C.W.J. (1998). RNA:protein interactions : a practical approach (Oxford ; New York, Oxford University Press).

Wong, I., and Lohman, T.M. (1993). A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. Proc Natl Acad Sci U S A 90, 5428-5432.