Exploring the Telomeric Repeat Addition Processivity of Vertebrate Telomerase

by

Mingyi Xie

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

ARIZONA STATE UNIVERSITY

May 2010

Exploring the Telomeric Repeat Addition Processivity of Vertebrate Telomerase

by

Mingyi Xie

has been approved

April 2010

Graduate Supervisory Committee:

Julian J.L. Chen, Chair
Hao Yan
Rebekka M. Wachter

ACCEPTED BY THE GRADUATE COLLEGE

ABSTRACT

Telomerase is a special reverse transcriptase that extends the linear chromosome termini in eukaryotes. Telomerase is also a unique ribonucleoprotein complex which is composed of the protein component called Telomerase Reverse Transcriptase (TERT) and a telomerase RNA component (TR). The enzyme from most vertebrate species is able to utilize a short template sequence within TR to synthesize a long stretch of telomeric DNA, an ability termed "repeat addition processivity". By using human telomerase reconstituted both *in vitro* (Rabbit Reticulocyte Lysate) and *in vivo* (293FT cells), I have demonstrated that a conserved motif in the reverse transcriptase domain of the telomerase protein is crucial for telomerase repeat addition processivity and rate. Furthermore, I have designed a "template-free" telomerase to show that RNA/DNA duplex binding is a critical step for telomere repeat synthesis. In an attempt to expand the understanding of vertebrate telomerase, I have studied RNA-protein interactions of telomerase from teleost fish. The teleost fish telomerase RNA (TR) is by far the smallest vertebrate TR identified, providing a valuable model for structural research.

This thesis is dedicated to my most beloved ones:

To my parents who have always believed in me and loved me unconditionally, to my wife who is always supportive of whatever I do, and to my unborn child.

ACKNOWLEDGEMENT

I give my deepest thanks to all the people who have mentored and helped me through my graduate life.

Dr. Julian J.L. Chen, my PhD advisor, wholeheartedly mentored me to grow into a scientist. His words will accompany me in the journey of pursuing my scientific career.

Dr. Hao Yan generously offered me a lot of research opportunities and invaluable advice.

Dr. Rebekka Watcher, Dr. Giovanna Ghirlanda and Dr. George Wolf, in my graduate and oral committee, kindly helped me through the years.

Dr. Mary Armanios is very supportive of my research and career development.

I am very grateful to all the fellow graduate students in Chen Laboratory: Chris Bley, Lena Li, Sandhya Tadepalli, Xiaodong Qi, Tracy Niday, Xiaowei Liu, Josh Podlevsky and Andrew Brown. Their company made my life in lab enjoyable. I would also like to thank the undergrads who have assisted me in performing the experiments: Matt Bean, Heather Lim and Yang Xuan. Last but not least, my thanks go to all the friends in our department.

TABLE OF CONTENTS

Page

Page

APPENDIX

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

The Nobel Prize in Physiology or Medicine of 2009 was awarded to Elizabeth H. Blackburn, Carol W. Greider and Jack W. Szostak for their discovery of how chromosomes are protected by telomeres and the enzyme telomerase. The importance of telomere and telomerase study is apparent as they have strong implications in cancer and aging. Telomeres are protective capping structures located at the ends of linear chromosomes. Replication of chromosomes results in telomere shortening due to the end replication problem (1). The loss of telomeres leads to senescence and programed cell death. Thus, telomere shortening acts as a molecular clock for cellular aging. Telomerase is a specialized reverse transcriptase that can extend telomeres and counter balance the chromosomal termini shortening. In most multi-cellular organisms, telomerase is inactive in somatic cells and active in stem cells and germ line cells. On the other hand, about 90% of the human cancer have detectable telomerase activity, making it a promising target for anticancer therapies (2). The enzyme telomerase is a special ribonucleoprotein (RNP) containing two core components: Telomerase RNA (TR) and Telomerase Reverse Transcriptase (TERT). The unique RNA-protein association, repeat addition processivity (detailed in section 1.2) of telomerase reaction and its evolutional relation to all the other reverse transcriptases make this enzyme an invaluable model to study the biogenesis, evolution and functional interactions of RNP complexes.

1.1 History of telomere and telomerase research

The study of telomere and telomerase can be dated back to the 1930s, when Herman Muller and Barbara McClintock proposed that the natural termini of linear chromosomes have special sequence or structure which could be distinguished from DNA breaks (3, 4). After 40 years, the first telomeric sequence (repetitive T_2G_4) from the ciliate *Tetrahymena thermophila* was determined by Elizabeth Blackburn (5). Subsequently, repetitive G_4T_4 telomeric sequence was found in another ciliate, *Oxytricha,* (6) and another GT rich telomere was identified in *Saccharomyces cerevisiae* (7). These results indicated that the telomere is comprised of both double stranded region and 3' overhang ends that are repetitive and rich in G and Ts. The vertebrate telomere sequence was also discovered to be TG-rich, with regular T_2AG_3 repeats.

Although there were multiple hypotheses of the mechanisms by which telomere sequences are maintained or extended at that time, Carol Greider and Liz Blackburn correctly predicted that telomere extension is carried out by a polymerase. The enzyme telomerase which extends telomeres was first identified in post-mating *Tetrahymena* cell-free extract (8). They further demonstrated that the purified telomerase activity was ensitive to RNase digestion. And the enzyme uses a template sequence residing in the RNA moiety to carry out telomere extension (9, 10). Telomerase was thus proven to be a ribonucleoprotein complex enzyme, which consists of both essential RNA and protein components.

In the following decade, extensive efforts had been directed to determine the core components of the telomerase complex. The telomerase RNA was first identified in

Tetrahymena by Greider as mentioned previously (10). Like all other large RNA molecules in the RNPs, telomerase RNA ought to form a conserved secondary and tertiary structure correspondent to its function. In 1991, phylogenetic comparative analysis was rendered to determine the conserved secondary structure of ciliate TRs (11). Later, the first yeast and vertebrate TRs were cloned from *S. cerevisiae* and human in 1994 and 1995 (12, 13). Discovery of these TRs revealed a surprising divergence in primary sequence from different groups of species. Therefore, more TRs from the same group of species were needed to deduce the conserved secondary structure. In 2000, vertebrate TR structure was finally determined using phylogentic comparative analysis, after the cloning of TR from 35 species (14). The yeast TR structure, due to its extremely large size (>1500 nt), was even more difficult to determine. The completetion of yeast TR secondary structure determination in 2004 revealed a pseudoknot structure near the template is a conserved core structure among all three groups of species: ciliates, vertebrates and the yeast (15).

The TERT gene from *Euplotes*, *S. cerevisiae, S. pombe* and human were simultaneously identified in the year 1997 (16,17). Interestingly, TERTs from various species are quite conserved when compared to each other. With only a couple of exceptions, they all include a TEN (Telomerase Essential N-terminus) domain important for DNA binding, a TRBD (Telomerase RNA Binding Domain) that tightly associates with the TR, a central RT Domain (Reverse Transcriptase Domain) which is responsible for the catalytic activity of telomerase and a less conserved C terminus extension domain (CTE) (18).

Ever since its discovery, the enzyme telomerase has been closely implicated to cell immortality. In both *S. cerevisiea* and mouse systems, a telomerase knock-out strain is viable in the first few generations, but the gradual loss of telomere sequence eventually results in cellular death or animal infertality in the following generations (19, 20). Telomerase RNA is universally expressed in all human cells. However, it is discovered that the TERT gene is not expressed in human somatic cells, but present in the cells that requires proliferative abilities, such as stem cells (21). Interestingly, most tumor cells also require telomerase activity to maintain their telomere length and infinite growth (2, 22, 23). Inhibition of telomerase activity in cancer cells thus will stop cell proliferation (24). Conversely, cell life span could be extended by introducing TERT gene expression (25). Therefore, telomerase provides a potential target for anti-cancer drugs design (26). On the other hand, dysfunctional, deficient or deregulated telomerase holoenzyme in stem cells will lead to multiple diseases, such as Dyskeratosis Congenita (DKC), Idiopathic Pulmonary Fibrosis (IPF), and Aplastic Anemia (AA). Mutations in TERT, TR gene and the telomerase associated Dyskerin gene (DKC) have been linked to these diseases (27).

Comprehensive mutagenesis studies aiming at deciphering the critical functional motifs/domains/groups in both TR and TERT were carried out after the domain organization of both components became more clear and after the discovery of disease related mutations (18,27). From these studies, mutations affecting overall telomerase activity, repeat addition processivity, repeat extension rate, RNP assembly, nucleotide incorporation fidelity and RNA/protein biogenesis were identified.

In the last few years, unveiling the structure of telomerase enzyme become urgent as it would provide a final explanation of a lot of biochemical and functional data obtained previously. Solving the structure of the critical pseudoknot domain and three-way junction (CR4-CR5) domain of human TR received the highest priority in the RNA component. The NMR structure of these two fragments offered structural basis for interpreting possible roles of hTR contribution to telomerase activity (28, 29, 30). The structure of *Tetrahymena* TR stem IV suggested it might be analogous to hTR CR4-CR5 domain (31). Additionally, the structure of the template boundary of *Tetrahymena* TR was also determined (32).

The crystal structure of full length TERT protein was difficult to solve due to the obstacle of large-scale protein expression and protein solubility. Therefore, the first TERT protein crystal structure was a fragment obtained from the *Tetrahymena* TEN domain (33). Soon after, the TRBD from *Tetrahymena* was also successfully crystalized (34). Both structures revealed possible DNA and/or RNA binding surfaces on the molecule. The full length TERT crystal was finally determined from the flour beetle *Tribolium castaneum*. Although lacking the TEN domain, *Tribolium* TERT provided the first insight into the spatial domain organization of the whole TERT protein (35).

In summary, the discovery of telomerase has revealed a route to understanding diverse biological questions, e.g. the molecular bases of cancer, stem cell biology, as well as many human diseases. The intensive ongoing research aiming at deciphering the molecular mechanism of how telomerase extends telomere substrates will further uncover the fundamental basis of many telomerase-deficiency related diseases.

1.2 Telomerase specific repeat addition processivity

One unique property of telomerase is the "repeat addition processivity" or "type II processivity" (hereafter referred to as "processivity"), which occurs when the short template in TR is used to repetitively synthesize a long stretch of telomeric DNA without complete dissociation between telomerase and the DNA substrate. (36) The processive addition of telomere repeats by telomerase relies on a translocation event, in which the template RNA dissociates from the telomeric DNA strand and then realigns with it to provide a template for subsequent repeat synthesis. This translocation process can be divided into three basic steps: (I) the dissociation between the RNA template and the telomeric DNA, (II) the movement of RNA template in relation to the telomeric DNA, and (III) the base-pairing between the alignment region of the RNA template and the telomere DNA 3' terminus (Figure 1.1). During this translocation process, the TERT protein presumably undergoes step-wise or concerted conformational changes that might involve multiple TERT domains to accommodate the events of RNA/DNA duplex melting and reformation.

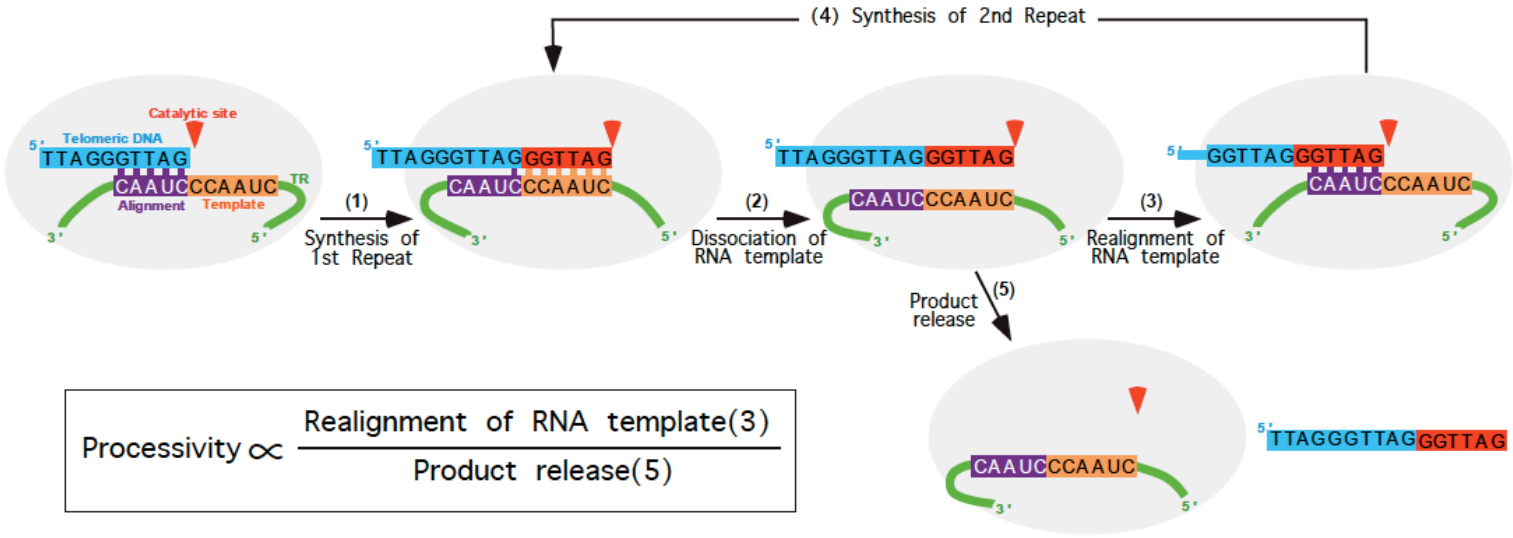


Fig. 1.1 A working model for repeat addition processivitySee next page for figure legend

Fig. 1.1 A working model for repeat addition processivity. Schematic of the basic steps for the translocation of the RNA template. Positioned within the active site (red arrowhead), the 3' end of the telomeric DNA primer (blue) base-pairs with the alignment region of telomerase RNA (violet) leaving the template region (orange) available for the synthesis of the first telomere repeat (step 1). As new nucleotides (red) are incorporated, both strands of the duplex move relatively away from the active site, but maintain a similar number of base-pairings. Once the end of the template is reached, the translocation process begins with the dissociation of the RNA template from the telomeric DNA (step 2). The dissociated RNA will then re-align with the telomeric DNA (step 3) and provide template for the synthesis of the second repeat (step 4). Alternatively, the telomeric DNA product can dissociate entirely from the enzyme (step 5). As shown in the inset, the repeat addition processivity of telomerase is proportional to the ratio of template realignment events over the product release events. Higher probability of template realignment event will lead to higher processivity of repeat addition.

Telomerase processivity varies to a great extent among different species. The telomerase from most ciliates and vertebrates are processive. The apparent low processive property of mouse and Chinese hamster telomerase is due to the lack of an alignment region in the TR template (37, 38). In contrast, most yeast telomerase can only add no more than two telomeric repeats onto the primer *in vitro* before dissociation (39, 40), with the exception of *S. castellii* (41).

Processivity can be affected by the reaction conditions such as temperature, ionic strength, primer concentration, dNTP substrate and certain nucleotide homologs. (42, 43, 44, 45, 46, 47). Moreover, accessory proteins in the telomerase holoenzyme, such as Euplotes p43 and human TPP1-POT1 complex, can also affect processivity (48, 49).

In spite of numerous processivity-related factors, the two essential core components alone can reconstitute the active and processive telomerase in Rabbit Reticulocyte lysate, indicating processivity as an intrinsic property. In most telomerase RNAs, the template sequence is usually one and a half repeats long and complementary to the telomeric DNA (38). The 3' region of TR template is referred to as the alignment region and contributes to the processivity by providing extra complementary sequence for the DNA to align during translocation (38, 50). Other regions of the RNA, such as the pseudoknot structure and stem IV in tetrahymena TR, also contribute to processivity (18).

As a specialized ribonucleoprotein complex, the TERT has probably co-evolved with the TR component for the optimal processivity for each species. Comparing to other RTs, the unique domain(s) in TERT might harbor the critical element for processivity. Because the translocation requires complete dissociation between the RNA template and

the DNA product, it has been proposed that a portion(s) of TERT serves as an "anchor site" to keep telomeric DNA bound to the telomerase, thus facilitating the translocation. (18) Direct interaction between the telomere primer and TERT has been shown by crosslinking experiments in *E. aediculatus*, Tetrahymena and *S. cerevisiea* TERTs in the presence of TR. In tetrahymena and yeast TERTs, the crosslinking site was mapped in the TEN domain (33, 51, 52, 53). The crystal structure of the TEN domain from tetrahymena TERT revealed a positively charged patch on the surface. Mutating certain amino acids on this patch severely reduces primer binding ability and telomerase activity, interestingly, without apparent decrease of processivity (33, 52). However, functional assays performed with yeast TERT bearing mutations of the homologous TEN residues indeed reduce the second repeat products (54). Furthermore, several human TERT TEN domain deletions and single amino acid mutants decrease the processivity. Amino acids on these mutation sites are proposed to form a "proximal anchor site" for the primers (55, 56, 57). A recent study of Tetrahymena TERT suggests that the Leu14 in the TEN domain acts as a switch to control processivity without affecting telomerase activity and telomere binding. Thus, the TEN domain might have two independent processivity related functions (58). On the other hand, the TEN domain is not the only portion of TERT interacting with the telomere substrate. Primer binding assays with different fragments of TERT have shown that the RT domain and the C-terminal extension also contribute to DNA binding, probably in a cooperative manner (57, 59). Consistent with interaction data, mutations in the RT domain and C-terminal extension affect processivity as well. Human TERT C-terminus deletion mutants decrease processivity (60). Point

mutations in *Tetrahymena* TERT Motif C promotes processivity by increasing affinity between TERT and DNA (61). Lastly, the long insertion in between motif A and B in the *S. cerevisiea* TERT RT domain promotes processivity by stabilizing the short RNA-DNA duplex formed in the template region (62).

1.3 Projects

Although the mechanism of telomerase reaction has been vigorously studied, the detailed procedure during template translocation and the functional motifs in TR/TERT are unclear. I have chosen the vertebrate telomerases, especially human telomerase, to study telomerase processivity. As aforementioned, vertebrate telomerases synthesize regular TTAGGG repeats, and the human telomerase reconstituted *in vitro* is very processive, and is able to synthesize more than 100 repeats, providing an excellent model to study processivity. Another advantage is that understanding of human telomerase function might provide direct information to reveal the molecular basis of several disease-related mutations. Besides, telomerase RNAs and TERT proteins in the vertebrate lineage are diverse enough to provide information to help understand RNP evolution. The telomerase from teleost fish were missing from the original study of telomerase RNA identification. One of my tasks, therefore, was to fill this gap in the vertebrate telomerase evolution map. The research projects described in each chapter are outlined below.

In Chapter 2, focus was on specific motif(s) evolved in TERT that is responsible for processivity. As a unique function specifically evolved in telomerase, a specialized structural domain in TERT or TR should be responsible for processivity. Indeed, we have identified a motif, termed motif 3, in the catalytic domain of TERT protein that is

specifically evolved in telomerases with high processivity. Mutations in this motif reduce telomerase processivity presumably by affecting RNA/DNA duplex binding during translocation. Repeat addition rate was also altered by some of the motif 3 mutants. This was not surprising because repeat addition rate is also an outcome affected by template translocation.

In Chapters 3 and 4, as a continuation of motif 3 study, I designed a template free telomerase system to directly test RNA/DNA duplex binding by the telomerase catalytic site. The data suggested that telomerase still retains the ability to recognize duplex substrate like conventional RTs. The studies in Chapter 3 provide a comprehensive understanding of how a variety of different substrates are preferred by the telomerase active site, and what is the difference of duplex extension activity between telomerase and conventional RTs. In Chapter 4, the template free telomerase was applied to test duplex binding abiliy of different processivity mutants. It was evident that less-processive mutants have lower affinity towards the telomeric duplexes, suggesting duplex binding is one of the steps during template translocation. More surprisingly, telomerase active site shows sequence specificity toward different circular permuted duplexes. This suggested a possible novel mechanism used by telomerase to determine a complete telomeric sequence and initiate translocation at the end of a complete repeat.

In Chapter 5, vertebrate telomerase research is extended to a previously missing area, the teleost fish. The divergent nature of telomerase RNA sequence hampered the identification of TR genes in the fast evolving teleost fish. Using a novel bioinformatics method supplied by our collaborator, Dr. Peter Stadler in University of Leipzig, five

teleost fish TRs were identified from completed fish genome databases. *In vitro* reconstituted telomerase were assayed for activity. The data proved that teleost fish telomerases are also processive and the key determinants for processivity in TR are similar to human TR.

1.4 References

1. Watson, G., and Paigen, K. (1972) *Nat New Biol* 239, 120-122

2. 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) *Science* 266, 2011-2015

3. Muller, HJ. (1938) *Collecting Net* 13, 181-198

4. McClintock, B. (1939) *Proc Natl Acad Sci U S A* 25, 405-416

5. Blackburn, E. H., and Gall, J. G. (1978) *J Mol Biol* 120, 33-53

6. Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981) *Proc Natl Acad Sci U S A* 78, 3015-3019

7. Walmsley, R. W., Chan, C. S., Tye, B. K., and Petes, T. D. (1984) *Nature* 310, 157-160

8. Greider, C. W., and Blackburn, E. H. (1985) *Cell* 43, 405-413

9. Greider, C. W., and Blackburn, E. H. (1987) *Cell* 51, 887-898

10. Greider, C. W., and Blackburn, E. H. (1989) *Nature* 337, 331-337

11. Romero, D. P., and Blackburn, E. H. (1991) *Cell* 67, 343-353

12. Singer, M. S., and Gottschling, D. E. (1994) *Science* 266, 404-409

13. 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., and et al. (1995) *Science* 269, 1236-1241

14. Chen, J. L., Blasco, M. A., and Greider, C. W. (2000) *Cell* 100, 503-514

15. Dandjinou, A. T., Levesque, N., Larose, S., Lucier, J. F., Abou Elela, S., and Wellinger, R. J. (2004) *Curr Biol* 14, 1148-1158

16. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T. R. (1997) *Science* 276, 561-567

17. 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) *Science* 277, 955-959

18. Autexier, C., and Lue, N. F. (2006) *Annu Rev Biochem* 75, 493-517

19. Lundblad, V., and Szostak, J. W. (1989) *Cell* 57, 633-643

20. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) *Cell* 91, 25-34

21. Shay, J. W., and Wright, W. E. (2005) *Carcinogenesis* 26, 867-874

22. Hayflick, L. (1998) *Keio J Med* 47, 174-182

23. Collins, K., and Mitchell, J. R. (2002) *Oncogene* 21, 564-579

24. Li, S., Crothers, J., Haqq, C. M., and Blackburn, E. H. (2005) *J Biol Chem* 280, 23709-23717

25. 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) *Science* 279, 349-352

26. Blasco, M. A. (2005) *Nat Rev Genet* 6, 611-622

27. Armanios, M. (2009) *Annu Rev Genomics Hum Genet* 10, 45-61

28. Leeper, T., Leulliot, N., and Varani, G. (2003) *Nucleic Acids Res* 31, 2614-2621

29. Leeper, T. C., and Varani, G. (2005) *RNA* 11, 394-403

30. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) *Mol Cell* 17, 671-682

31. Chen, Y., Fender, J., Legassie, J. D., Jarstfer, M. B., Bryan, T. M., and Varani, G. (2006) *EMBO J* 25, 3156-3166

32. Richards, R. J., Theimer, C. A., Finger, L. D., and Feigon, J. (2006) *Nucleic Acids Res* 34, 816-825

33. Jacobs, S. A., Podell, E. R., and Cech, T. R. (2006) *Nat Struct Mol Biol* 13, 218-225

34. Rouda, S., and Skordalakes, E. (2007) *Structure* 15, 1403-1412

35. Gillis, A. J., Schuller, A. P., and Skordalakes, E. (2008) *Nature* 455, 633-637

36. Greider, C. W. (1991) *Mol Cell Biol* 11, 4572-4580

37. Prowse, K. R., Avilion, A. A., and Greider, C. W. (1993) *Proc Natl Acad Sci U S A* 90, 1493-1497

38. Chen, J. L., and Greider, C. W. (2003) *EMBO J* 22, 304-314

39. Lue, N. F., and Peng, Y. (1997) *Nucleic Acids Res* 25, 4331-4337

40. Fulton, T. B., and Blackburn, E. H. (1998) *Mol Cell Biol* 18, 4961-4970

41. Cohn, M., and Blackburn, E. H. (1995) *Science* 269, 396-400

42. Hammond, P. W., and Cech, T. R. (1997) *Nucleic Acids Res* 25, 3698-3704

43. Hammond, P. W., and Cech, T. R. (1998) *Biochemistry* 37, 5162-5172

44. Sun, D., Lopez-Guajardo, C. C., Quada, J., Hurley, L. H., and Von Hoff, D. D. (1999) *Biochemistry* 38, 4037-4044

45. Maine, I. P., Chen, S. F., and Windle, B. (1999) *Biochemistry* 38, 15325-15332

46. Hardy, C. D., Schultz, C. S., and Collins, K. (2001) *J Biol Chem* 276, 4863-4871

47. Jarstfer, M. B., and Cech, T. R. (2002) *Biochemistry* 41, 151-161

48. Aigner, S., and Cech, T. R. (2004) *RNA* 10, 1108-1118

49. Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., and Lei, M. (2007) *Nature* 445, 506-510

50. Autexier, C., and Greider, C. W. (1995) *Genes Dev* 9, 2227-2239

51. Hammond, P. W., Lively, T. N., and Cech, T. R. (1997) *Mol Cell Biol* 17, 296-308

52. Lue, N. F. (2005) *J Biol Chem* 280, 26586-26591

53. Romi, E., Baran, N., Gantman, M., Shmoish, M., Min, B., Collins, K., and Manor, H. (2007) *Proc Natl Acad Sci U S A* 104, 8791-8796

54. Lue, N. F., and Li, Z. (2007) *Nucleic Acids Res* 35, 5213-5222

55. Moriarty, T. J., Marie-Egyptienne, D. T., and Autexier, C. (2004) *Mol Cell Biol* 24, 3720-3733

56. Moriarty, T. J., Ward, R. J., Taboski, M. A., and Autexier, C. (2005) *Mol Biol Cell* 16, 3152-3161

57. Wyatt, H. D., Lobb, D. A., and Beattie, T. L. (2007) *Mol Cell Biol* 27, 3226-3240

58. Zaug, A. J., Podell, E. R., and Cech, T. R. (2008) *Nat Struct Mol Biol* 15, 870-872

59. Finger, S. N., and Bryan, T. M. (2008) *Nucleic Acids Res* 36, 1260-1272

60. Huard, S., Moriarty, T. J., and Autexier, C. (2003) *Nucleic Acids Res* 31, 4059-4070

61. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2000) *J Biol Chem* 275, 24199-24207

62. Lue, N. F., Lin, Y. C., and Mian, I. S. (2003) *Mol Cell Biol* 23, 8440-8449

CHAPTER 2

A NOVEL MOTIF IN TELOMERASE REVERSE TRANSCRIPTASE REGULATES TELOMERE REPEAT ADDITION RATE AND PROCESSIVITY

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

Author contributions: M. Xie designed and performed the mutagenesis and telomerase kinetics experiments. J.D. Podlevsky performed multiple sequence alignment and molecular modeling. X. Qi reconstituted telomerase in 293 FT cells and performed Northern blotting and Western blotting. C.J. Bley predicted protein secondary structure. J.J. Chen conceived the project.

2.1 Abstract

Telomerase is a specialized reverse transcriptase that adds telomeric DNA repeats onto chromosome termini. Here, we characterize a new telomerase-specific motif, called motif 3, in the catalytic domain of telomerase reverse transcriptase, that is crucial for telomerase function and evolutionally conserved between vertebrates and ciliates. Comprehensive mutagenesis of motif 3 identified mutations that remarkably increase the rate or alter the processivity of telomere repeat addition. More importantly, the rate and processivity of repeat addition are affected independently by separate motif 3 mutations. The processive telomerase action relies upon a template translocation mechanism whereby the RNA template and the telomeric DNA strand separate and realign between each repeat synthesis. By analyzing the mutant telomerases reconstituted *in vitro* and in cells, we show that the hyperactive mutants exhibit higher repeat addition rates and faster enzyme turnovers, suggesting higher rates of strand-separation during template translocation. In addition, the strong correlation between the processivity of the motif 3 mutants and their ability to use an 8 nt DNA primer, suggests that motif 3 facilitates realignment between the telomeric DNA and the template RNA following strand-separation. These findings support motif 3 as a key determinant for telomerase activity and processivity.

2.2 Introduction

Telomerase is a specialized reverse transcriptase (RT) responsible for adding telomeric DNA repeats onto the 3' ends of chromosomes. Telomere elongation counter-balances the natural shortening of linear chromosomes due to the end-replication

problem, preventing senescence, apoptosis and genome instability (1). The deficiency in telomerase function leads to limited renewal capacity in highly proliferative cells, and is associated with human diseases including dyskeratosis congenita, aplastic anemia and idiopathic pulmonary fibrosis (2,3).

Reconstitution of catalytically active telomerase *in vitro* requires two core components: the telomerase RNA (TR) and the telomerase reverse transcriptase (TERT) protein. The TR contains a short template region for the synthesis of telomeric DNA repeats, and conserved structural domains essential to *in vivo* biogenesis and assembly with the TERT protein. The TERT subunit is a multi-domain protein comprised of an N-terminal extension (NTE), a central catalytic RT domain and a C-terminal extension (CTE) (4). In most eukaryotes, the NTE consists of a telomerase-essential N-terminal (TEN) domain (5) that binds telomeric DNA and a TR binding domain. However, the TEN domain is dispensable in certain species, such as insects (6). The catalytic RT domain encompasses seven essential motifs (1, 2, A, B, C, D and E) that are universally conserved among RTs (7).

Telomerase has the unique ability to add multiple telomeric repeats to a given primer before complete dissociation from the DNA, called “repeat addition processivity” (abbreviated to “processivity”). Unlike conventional RTs which can utilize a variety of single-stranded RNA templates, telomerase uses only a short sequence from its intrinsic RNA component as template. During telomere DNA synthesis, the 3’ end of telomeric DNA base pairs with the RNA template forming an RNA/DNA duplex which is positioned within the catalytic site of TERT protein for nucleotide addition. When

telomerase completes the synthesis of one telomeric DNA repeat, a "template translocation" must occur whereby the RNA template dissociates from the DNA strand, translocates and realigns relative to the 3' end of the DNA, providing an unoccupied template for the next round of repeat synthesis. This translocation process is the rate-limiting step in a processive telomerase reaction, as indicated by the strong pause after each round of repeat synthesis, giving rise to the characteristic ladder banding pattern of telomere products (8,9). While the repeat addition rate is determined by the rate-limiting translocation step, the processivity of the reaction is determined by the probability, or efficiency, of RNA/DNA realignment over complete product release during translocation.

The extent of telomerase processivity varies dramatically among species. Telomerase from ciliates and most vertebrates are highly processive (8-10). In contrast, telomerase from certain rodents and yeasts have little to no processivity (11-14). Reaction conditions and accessory proteins appear to contribute to the disparity in processivity observed *in vitro* (15-17). However, the intrinsic properties of TERT and TR components are by and large the major determinants for the varying degrees of processivity observed among species.

Mutations in both TERT and TR components have been found to affect the rate and processivity of the telomerase reaction. Several elements within TERT were shown to be crucial for telomerase processivity. The TEN domain contains an "anchor site" that binds single-stranded telomeric DNA, preventing complete product release from the enzyme during template translocation (4,18). A recent study has shown that a mutation at Leu14 in the *Tetrahymena* TEN domain abolished processivity while leaving activity or DNA

binding affinity intact. The Leu14 residue was proposed to function as an intra-molecular switch for processivity (19). In *S. cerevisiae,* a motif called IFD (insertion in finger domain), located in the RT domain between motifs A and B, contains four conserved residues shown to be important for second repeat synthesis (20). Also within the RT domain, a point mutation in motif C of *Tetrahymena* TERT increases processivity by increasing protein-DNA primer affinity (21). Beyond the RT domain, a mutation in the CTE, a putative homologue to the HIV RT thumb, was shown to reduce repeat addition processivity (22). Within the TR, the template length and conserved structural elements also contribute to telomerase processivity, through affecting telomeric DNA/template RNA base-pairing interactions during template translocation (23,24) or the RNA/TERT protein interactions (25,26). A previous study by Drosopoulos *et al.* showed that varying the template sequence can alter the rate of telomere repeat addition, possibly through modulating interactions between the template RNA, DNA product and TERT protein (27). Although the TERT protein was shown to contribute to the processivity of telomerase activity, its involvement in the regulation of telomere repeat addition rate is unclear.

In this study, we carried out a comprehensive alanine-substitution screening in a novel motif of the TERT protein and discovered mutations that surprisingly increased the rate or altered the processivity of telomere repeat addition. Characterization of the *in vitro* reconstituted telomerase enzymes containing these unusual hyperactive mutations indicates a higher rate of enzyme turnover or product dissociation. In addition, mutations that alter processivity alter the ability of the enzymes to use the short 8 nt primer as

substrate, the use of which resembles the realignment of the 3'-end of DNA with the template RNA - the second step of the template translocation. We conclude that this novel TERT motif is an important determinant for telomerase activity and processivity, regulating both strand-separation and realignment of telomeric DNA and template RNA during template translocation.

2.3 Materials and Methods

2.3.1 Sequence alignment analysis

The sequence alignment of the RT domain for TERT and other RTs was performed within the program BioEdit using the ClustalW algorithm, and further refined manually using the highly conserved RT motifs as anchor points. The alignment was carried out initially for individual groups of closely related species, then expanded to include sequences from more divergent species. The complete sequence alignment is available at the telomerase database (http://telomerase.asu.edu) (28).

2.3.2 Plasmid construction and mutagenesis

Specific mutations in the human TERT (hTERT) genes were introduced into pNFLAG-hTERT (a generous gift from Dr. Vinayaka Prasad) by site-directed mutagenesis using an overlapping PCR strategy (29). For *in vivo* expression of TERT, gene fragments that contain specific mutations were sub-cloned from pNFLAG-hTERT into a modified pcDNA-hTERT vector (generous gifts from Dr. Joachim Lingner) via two SacII sites within the TERT gene. Intended mutations were confirmed by sequencing.

2.3.3 Reconstitution of telomerase *in vitro* and in cells

Human telomerase was reconstituted *in vitro* using the TNT Quick Coupled rabbit reticulocyte lysate system (Promega) as described previously with minor modifications (30). To assemble telomerase, 1 μM of human TR (hTR) pseudoknot (nt 32-195) and CR4-CR5 (nt 239-328) RNA fragments were added to the hTERT synthesis reactions, and incubated at 30ºC for 30 min.

To purify sufficient amount of mutant telomerase from cells for telomerase direct assay, we used the telomerase reconstitution system developed by Lingner's group with minor modifications (31,32). Recombinant telomerase enzyme was reconstituted by over-expressing the hTERT and hTR genes in 293FT cells (Invitrogen) using pcDNA-hTERT and pBS-U1-hTR (generous gifts from Dr. Joachim Lingner) as described previously. Cells were grown in DMEM media supplemented with 10% FBS at 37ºC with 5% CO_2. After reaching 80-90% confluency, cells were transfected with 1 μg of plasmid DNA (200 ng of pcDNA-hTERT and 800 ng of pBS-U1-hTR diluted in 50 μl of FBS-free DMEM media) and 4 μl of Fugene HD transfection reagent (Roche) in 12-well plates, following the manufacturer's instruction. Two days post transfection, cells were harvested and lysed (31).

2.3.4 Telomerase activity assay

Telomerase reconstituted *in vitro* or in cells was assayed using the conventional direct primer-extension assay as previously described (30), with the exception of that 0.165 μM of [α-^{32}P]dGTP (3000 Ci/mmole, Perkin Elmer) was used in a 10 μl reaction. Telomerase processivity was determined by measuring the intensity of each major band, normalized by the numbers of ^{32}P-dGTP incorporated and plotted versus the repeat

numbers as previously described (15). Processivity was calculated using the equation: Processivity=-ln2/(2.303k), where k is the slope. The processivity of mutant telomerase was presented relative to the wild-type enzyme.

The pulse-chase time course experiments were carried out using *in vitro* reconstituted telomerase and the conventional direct assay as described above. The pulse-chase assay tracks the progressive extension of telomere products that were labeled (or pulsed) during the pulse reaction and extended by the processive telomerase enzymes during the chase reaction. During the pulse reaction, the enzymes add telomere repeats to the DNA primer with the incorporation of radioactive [α-^{32}P]dGTP. In brief, telomerase was incubated with 4 μM $(TTAGGG)_3$ primer in the presence of 1X PE buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$ and 1 mM spermidine) at 30ºC for 5 min to allow sufficient primer-enzyme complex formation prior to the pulse reaction. The pulse reaction was initiated by the addition of 1 mM dATP, 1 mM dTTP, 2 μM dGTP and 0.33 μM α-^{32}P-dGTP (3000 Ci/mmole, Perkin Elmer) and incubated at 30ºC for 5 min. The chase reaction was then carried out by adding non-radioactive dGTP to 100 μM in the reaction and incubated at 30ºC for various amount of time as indicated. Upon addition of 50-fold excess non-radioactive dGTP, the initially labeled telomere products continue to be extended in the chase reaction by the same enzyme. During the chase reaction, the primers extended by the enzymes dissociated from the initially pulsed telomere products will not be seen. The chase reaction was terminated by ethanol precipitation and analyzed by gel-electrophoresis. For each chase reaction, 10 bands with highest intensity above initial pulse product bands were used to deduce a "modal band"

and calculate the extension rate as previously described (27). Because of the short extension time, the processivity cannot be accurately measured during the pulse reaction and the chase reactions at the early time points. Instead, telomerase processivity in the pulse-chase assay was determined from the first ten major bands (1-10 repeats added) in the chase reaction at the last time point (10 min).

For the short-primer telomerase assay, different telomere primers tel8, tel10, tel12, tel15, tel18 were used as indicated. For *Km* measurements, the primers were supplied at varying concentrations and the reaction mixture was incubated for 5-60 min, which falls in the linear range of product formation. A ^{32}P end-labeled oligonucleotide (50 nt) was used as the recovery and loading control. The product intensity of each reaction was quantitated, normalized with loading control and expressed as a relative activity compared to the reaction with highest primer concentration. The relative activities were plotted against primer concentration and the Michaelis-Menten equation, Y=Vmax*X/(*Km*+X), was used to fit the nonlinear curve to determine the *Km* (Prism 5, Graphpad Software).

To measure enzyme turnover rate, the *in vitro* reconstituted telomerase was pre-incubated with 10 μM tel7 primer (5'-AGGGTTA-3') in the presence of 1X PE buffer at 30°C for 10 min. Telomerase reaction was initiated by addition of 2 μM dGTP and 0.33 μM [α-^{32}P]dGTP (3000 Ci/mmole, Perkin Elmer), and aliquots were removed from the reaction mixture at different time points. The intensity of product was first adjusted by the TERT protein level, and normalized by the intensity of loading controls. The intensity of each band was normalized by that of the wild-type reaction at the last time point (10

min). The relative product intensities were then plotted against the amount of time. The slope of linear trend line represents the enzyme turnover rate. The enzyme turnover rates of mutant telomerases were indicated as relative values to the wild-type enzyme.

Tetrahymena telomerase was reconstituted *in vitro* using pCITE-5XT7-tTERT as human telomerase reaction. Telomerase activity assay: a 10 μL reaction was carried out with 2.5 μL of *in vitro* reconstituted telomerase in the presence of 1X tPE buffer (50mM Tris-HCl, pH 8.3, 5mM KCl, 1.25 mM MgCl2 and 5 mM DTT), 200 μM dTTP, 9 μM dGTP, 1 μM $(TTGGGG)_3$ telomere primer and 0.165 μM (α-^{32}P dGTP (3000 Ci/mmole, 10m Ci/ml, Perkin Elmer) at 30°C for 1 h. The subsequent steps were the same as human telomerase activity assay, except processivity was quantitated by the ration of the intensity of the second repeat to the first of the primer extension products.

2.3.5 Western blot analysis

Ten micrograms of total protein of 293FT cell lysate was heated at 95°C for 5 min in 1X Laemmli buffer (0.125M Tris-HCl, pH6.8, 2% SDS, 10% glycerol, 5% 2-mercaptoethanol and 0.0025% bromophenol blue), fractionated on a 6% or 8% SDS-PAGE gel, and electro-transferred onto the PVDF membrane. Blocking (overnight at 4ºC) and incubation with antibodies (1 hour at room temperature) were carried out in 5% nonfat milk/1X TTBS (20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.05% Tween 20). Anti-hTERT goat polyclonal antibody L-20 (Santa Cruz Biotechnology) and anti-GAPDH mouse monoclonal antibody 6C5 (Ambion) were used as the primary antibodies. After incubation with the HRP-conjugated secondary antibody (Santa Cruz Biotechnology), the blots were developed using the Immobilon Western

Chemiluminescent HRP substrate (Millipore), and the blot images were acquired and analyzed using a Gel Logic440 system (Kodak).

2.3.6 Northern blot analysis

Total RNA was extracted from transfected cells using Tri-reagent (Molecular Research Center, Inc.) following manufacturer's instruction. Three micrograms of total RNA was resolved on a 4% polyacrylamide/8M urea denaturing gel and electro-transferred to the Hybond-XL membrane (GE Healthcare). Preparation of the riboprobes and hybridization of the blot were carried out as described previously (10).

2.3.7 Homology modeling

The RNA-DNA duplex was modeled into the *Tribolium* TERT structure (3DU6) by superimposition with the HIV1 p66 structure (1HYS) containing an RNA/DNA duplex. The two pdb files were superimposed in the CCP4 program using the following seven conserved resides within the RT domain as anchor points: three Asp residues in motifs A and C; Arg in motif 2; Glu and Gly in motif B; and Gly in motif E. These seven residues are highly conserved between TERTs and retroviral RTs.

2.3.8 Direct DNA/RNA Binding Assay with TERT RT Fragments

Oligonucleotides were synthesized and gel-purified by IDT (Integrated DNA technologies). All oligonucleotides were prepared to a final concentration of 50 μM in 1X annealing buffer (100 mM Tris-HCl, pH 7.5, 500 mM NaCl, 50 mM EDTA). For preparing the RNA/DNA duplex, DNA oligo $(GTTAGG)_2$ was added in 10% excess of molar concentration to the biotin-labeled RNA oligo. The mixtures were heated at 80°C for 3 min and cooled down slowly to room temperature. The TERT-oligo binding assay

was carried out by mixing 5 µL of TNT expressed hTERT 601-939, 1 µL of MBP-hTERT 601-927 (internal control), 50 µM oligo-dT18 in a 10 µL reaction in 1X PE buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$ and 1 mM spermidine) and the biotin-labeled oligonucleotides at the final concentration of 10 µM and incubated at 30°C for 1 h. Before use, 50 µL Dynabeads M-280 Streptavidin (Invitrogen) was washed twice with 250 µL Hypobuffer (90.7 mM HEPES, pH7.3, 7 mM KCl, 2.3 mM $MgCl_2$, 0.05 mM BME), surface blocked with 50 µL blocking buffer (hypobuffer with 0.75 mg/mL BSA, 0.15 mg/mL glycogen, 0.015 mg/mL yeast total RNA, 0.5% NP40 and 1 µM oligo-dT18) for 30 min at 4ºC with agitation and resuspended in 100 µL blocking buffer. A hundred µL pre-treated beads were then added with 10 µL TERT/biotin-oligo binding reaction and mixed at 4 ºC for 15 min. The beads were washed three times with 300 µL wash buffer (hypobuffer with 0.5% TritonX 100) and resuspended in 25 µL 2X SDS loading buffer (125 mM Tris-HCl pH6.8, 10% β-mercapto-ethanol, 4% SDS, 0.01% Bromophenol blue, 20% glycerol), boiled for 5 min, and 20 µL of the supernatant was loaded onto a 12% SDS-PAGE gel. After electrophoresis, the gel was dried, exposed to a film and analyzed using a Bio-Rad FX Pro Imager.

2.4 Results

2.4.1 Sequence conservation of motif 3

The novel motif 3, located in the catalytic RT domain of TERT between motifs 2 and A (Fig. 2.1a), was previously found conserved in vertebrate and invertebrate chordates (33). To determine the extent of conservation of this motif in species beyond

chordates, we extended the sequence alignment analysis to include additional groups of eukaryotes (detailed in Materials and Methods). Remarkably, the sequence of motif 3 is conserved also in non-yeast fungi, plants and ciliates (Fig. 2.1b). The conservation of motif 3 sequences over a large evolutionary distance, from vertebrates to ciliates, implies necessity for telomerase function. Secondary structure prediction of the motif 3 sequences from a large number of species suggested a putative helix-coil-helix fold (Fig. S2.1a). Interestingly, the recent crystal structure of TERT from an insect, *Tribolium castaneum*, contains a helix-coil-helix structure between motif 2 and A, supporting the secondary structure prediction (34). Despite the apparent similarity in predicted secondary structure, TERTs from insects, nematodes and yeasts did not show the same degree of sequence conservation, particularly in the central region of motif 3 (Fig. S2.2). Previously in the RTs closely related to the TERT, including the Penelope-like retrotransposons, non-LTR retrotransposons and group II introns, a conserved motif between motif 2 and motif A was identified as motif 2a (35-37), which however shares no common sequence to the TERT motif 3 (Fig. S2.2). Since motif 3 appears conserved specifically in vertebrates and ciliates whose telomerase is highly processive, but not in yeasts whose telomerase is not processive, we speculated that motif 3 might be important for telomerase processivity.

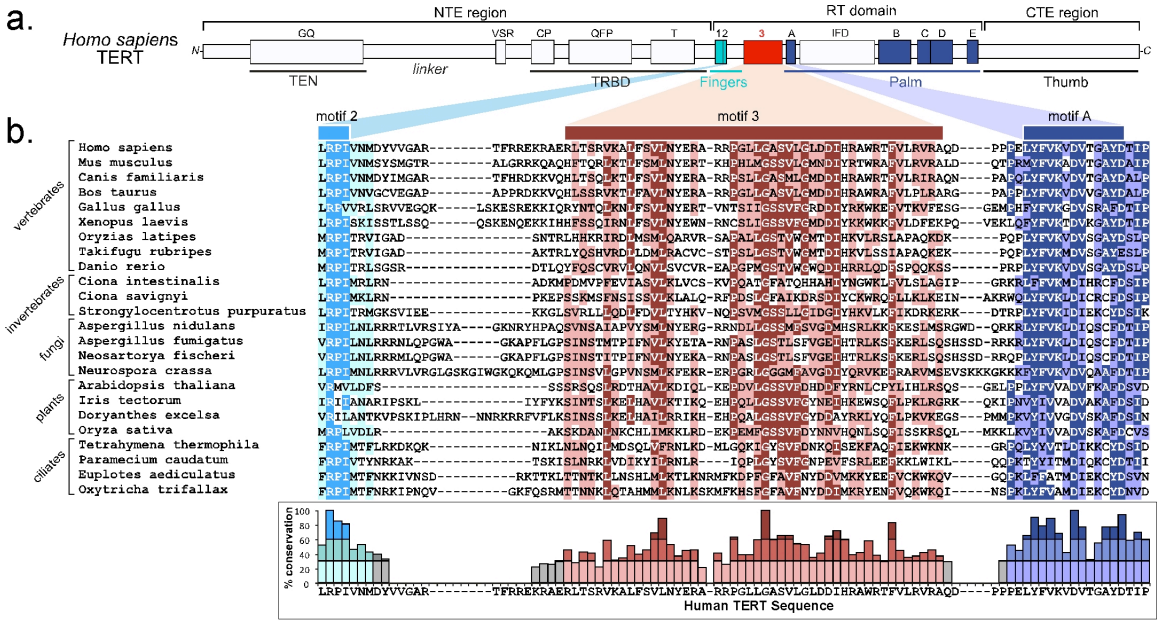


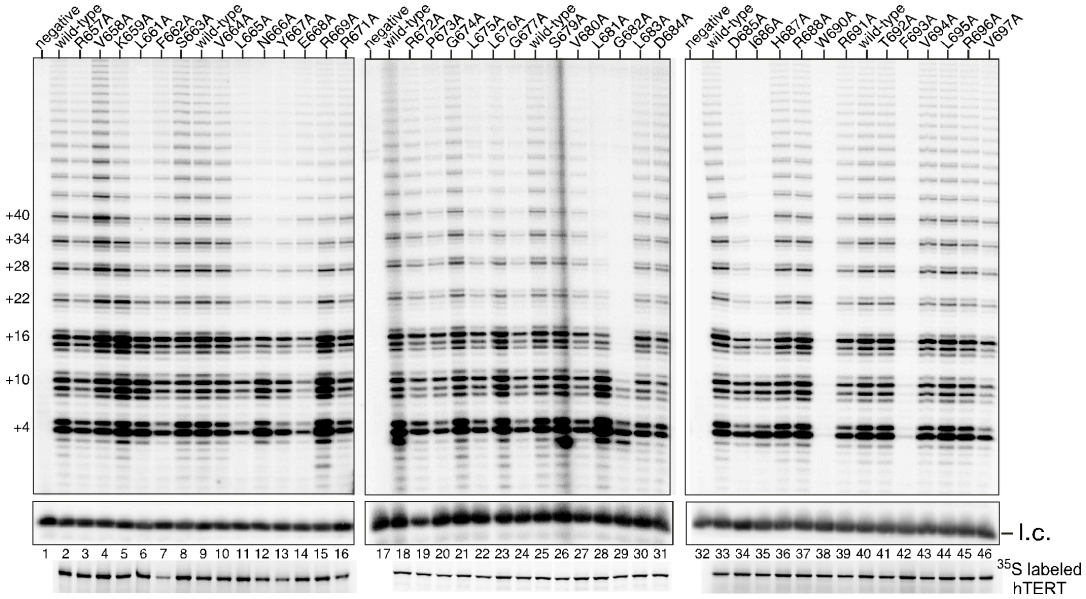
Fig. 2.1. Multiple sequence alignment of TERT motif 3. (A) Schematic of domain and motif organization of human TERT protein. Motif 3 (red) and the conserve RT motifs 1, 2 and A-E (cyan and blue) are colored. (B). Sequence alignment of TERT motif 3 from vertebrates, invertebrates, fungi, plants and ciliates. Shading indicates a minimum of 55% identity (dark cyan, red and blue) and 55% similarity (light cyan, red and blue) conservation. The degree of identity conservation with the human sequence at each residue is shown below the sequence alignment. No conservation determined where more than two sequences had gaps present. Darker shading indicates greater identity conservation with the human sequence (<30% light, 30-60% medium and >60% dark).

2.4.2 Mutations in motif 3 affect telomerase activity and processivity independently

To experimentally determine the function of motif 3, we conducted comprehensive alanine substitution mutagenesis on human TERT and analyzed the telomerase mutants reconstituted in rabbit reticulocyte lysate (RRL) for activity and processivity (see Materials and Methods). Certain residues in motif 3 appear to be critical for telomerase catalysis, as the alanine substitutions W690A and F693A nearly abolished telomerase activity (Fig. 2.2a, lanes 38 and 42). This is not unexpected for the F693 residue as it is one of the three most conserved residues in motif 3 (Fig. 2.1b). The W690 residue, although not as highly conserved as F693, is naturally substituted with hydrophobic leucine or valine in most species (Fig. 2.1b).

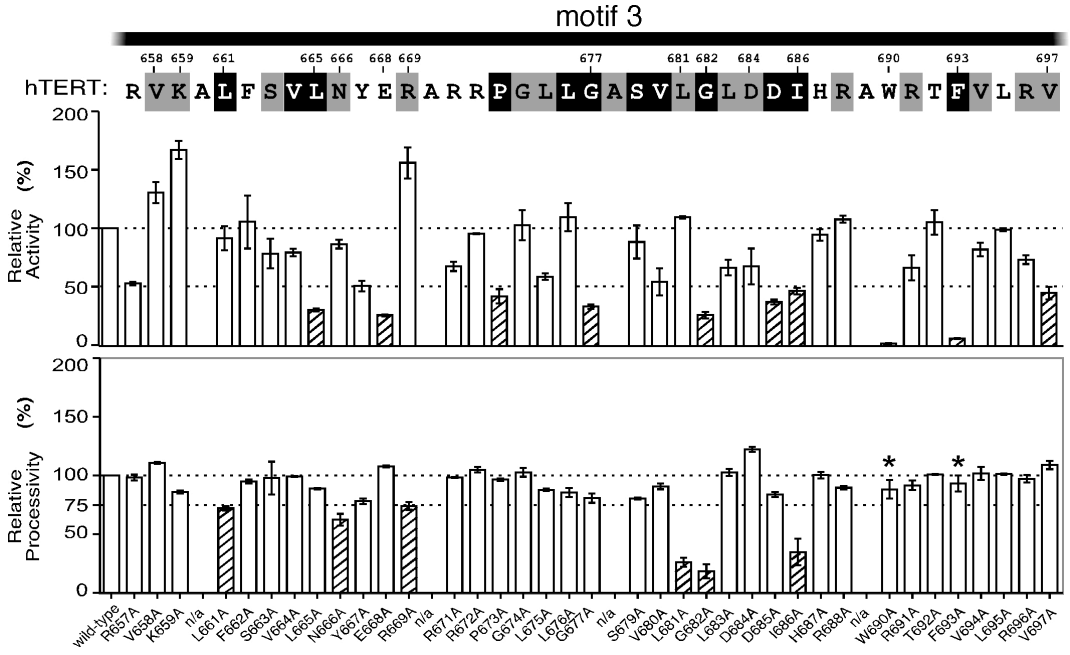
One intriguing finding from analyzing these motif 3 mutants was that telomerase activity and processivity can be independently altered (Fig. 2.2). Three mutations (V658A, K659A and R669A) in the N-terminal region of motif 3 increased telomerase activity up to 1.7 fold, but had different effects on processivity (Fig. 2.2b). For example, mutant V658A is hyperactive and hyper-processive, compared to wild-type enzyme (Fig. 2.2a, lane 4). In contrast, hyperactive mutants K659A and R669A had reduced processivity (Fig. 2.2a, lanes 5 and 15). Unlike V658A, mutants E668A, D684A and V697A showed greater processivity, but lower activity (Fig. 2.2b). Mutations L661A, N666A, R669A, L681A, G682A and I686A, while all reduced processivity, altered activity differently (Fig. 2.2b). In summary, the analysis of motif 3 mutants showed no correlation between changes in activity and changes in processivity, suggesting that telomerase activity and processivity are independent and possibly regulated through

separate mechanisms.



a. Activity assays of the motif 3 mutants

Fig. 2.2. Alanine substitution screening of motif 3. (a) Activity assays of the motif 3 mutants. Human telomerases with alanine substitutions in motif 3 were reconstituted *in vitro* and assayed for activity. Numbers on the left (+4, +10, +16 etc.) of the gel indicate the number of nucleotides added to the primer in each major band. l.c.: loading control, a ^{32}P end-labeled 15 nt DNA oligonucleotide, shown with the contrast adjusted. Below the gel, the [^{35}S] methionine labeled TERTs analyzed by SDS-PAGE for quantitation are shown.



b. Quantitation of activity and processivity of motif 3 mutants.

Fig. 2.2. continued (b) Quantitation of activity and processivity of motif 3 mutants. The residues in the human motif 3 sequence are shaded according to their identity and similarity as shown in Fig. 1B. Below the sequence, the bar graph shows the activity and processivity of each mutant relative to wild-type. The dash line across the graph indicates the wild-type level of activity and processivity. The shaded bars indicate a relative activity lower than 50% or a processivity lower than 75% of the wild-type level. Alanine residues in the wild-type motif 3 sequence were omitted from the analysis and labeled as n/a (not available). Error bars indicate the standard deviation derived from 2-4 independent experiments. *: For the mutants with an extremely low activity, a special assay (3-fold more enzyme and reagents, and longer exposure time of the gel) was performed to determine processivity.

To determine if the telomerase mutants assembled in vitro and in cells behave similarly, we assayed two hypo-processive mutants (G682A and I686A) and three hyper-processive mutants (E668A, D684A and V697A). The mutant telomerases were reconstituted by over-expressing the full-length hTR and mutant hTERT genes in the 293FT cells to generate a high telomerase activity sufficient for direct telomerase assay, a system developed by Cristofari and Lingner (31,32). The endogenous telomerase activity from 293FT cells transfected with the empty vector was undetectable by the direct primer-extension assay (Fig. 2.3, lane 1). The telomerase mutants reconstituted in RRL and in human cells exhibited similar levels of activity and processivity (Fig. 2.3, lanes 7, 8, 16, 17 and 18), confirming the results from the in vitro reconstituted enzymes. Additionally, we rescued two nearly inactive alanine-substituted mutants, L665A and F693A, with substitutions of conservative amino acids L665I and F693Y (Fig. 2.3, compare lanes 10 to 11, and lanes 12 to 13), suggesting the bulky hydrophobic side-chains of these two residues are required for telomerase function.

We also tested two disease-associating motif 3 mutations, G682D and V694M, previously identified in aplastic anemia patients (38,39). While both mutations significantly reduced telomerase activity (Fig. 2.3, lanes 9 and 20), the G682D mutation caused also a significant reduction in processivity (Fig. 2.3, lane 20). The reduced telomerase activity and processivtiy of the disease-associating motif 3 mutants, together with the shortened telomere length in the patients harboring the mutations, support the importance of motif 3 for telomerase function and telomere maintenance *in vivo*.

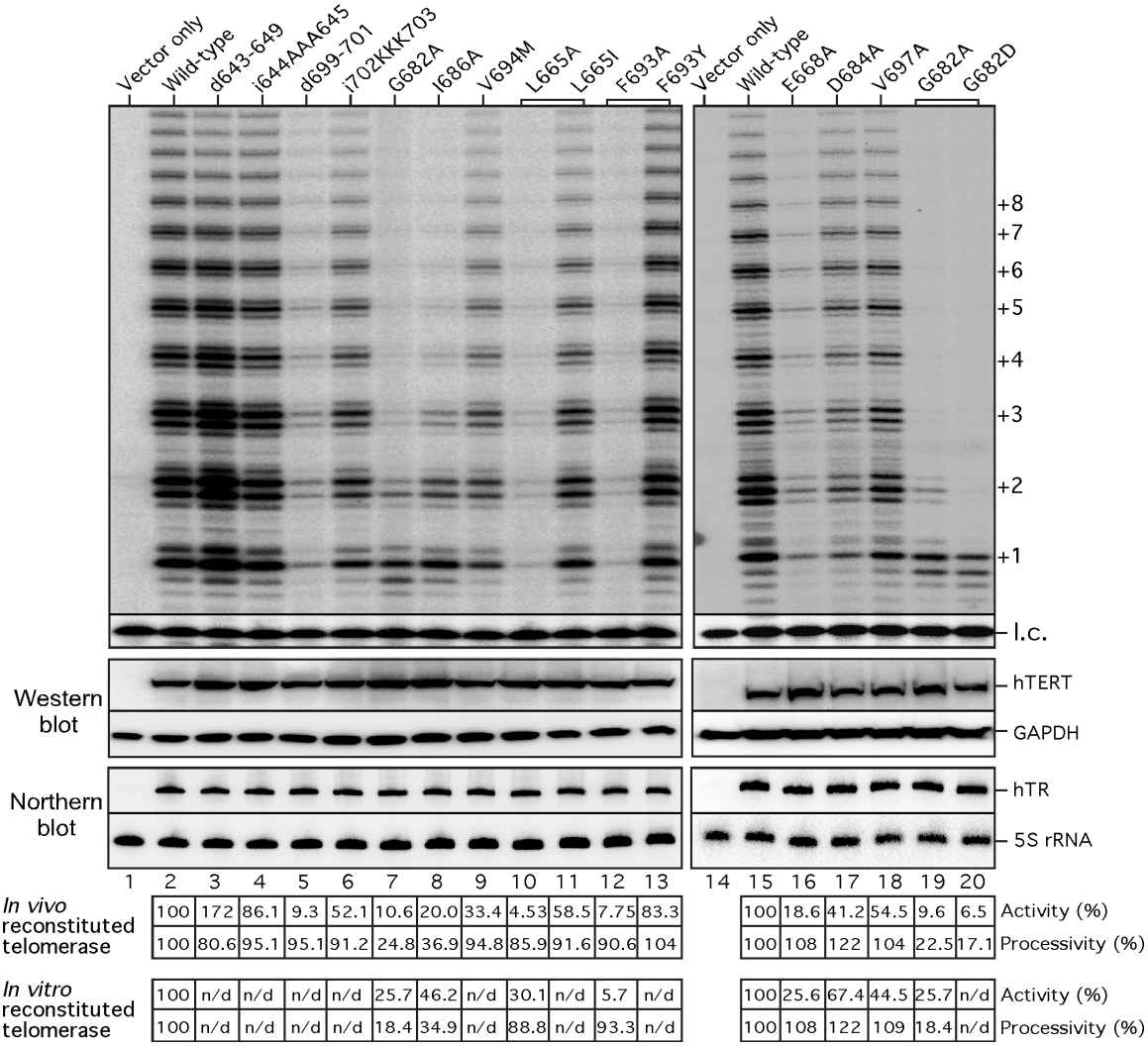


Fig. 2.3 Activity assay of telomerase mutants reconstituted in cells. (Top panel) Mutant telomerases were reconstituted in 293FT cells and analyzed for activity. The TERT mutants d643-649 and d699-701 contain deletions of 7 and 3 residues, respectively. The TERT mutants i644AAA645 and i702KKK703 contain insertions of three alanine residues and three lysine residues between 644-645 and 702-703, respectively. The TERT mutants that contain different amino acid substitutions at the same residue are indicated by brackets. Numbers on the right (+1, +2, +3 etc.) indicate the number of repeats added to the telomeric primer. A ^{32}P end-labeled 15-mer DNA oligonucleotide is used as a loading control (l.c.).

Fig.2.3. continued (Middle panel) Expression level of hTERT protein in the transfected cells was analyzed by western blots of hTERT and GAPDH using anti-hTERT L-20 and anti-GAPDH antibodies. The level of GAPDH was used as a loading control. (Bottom panel) Expression level of hTR in the transfected cells was analyzed by northern blots of hTR and 5S ribosomal RNA (rRNA) using riboprobes against hTR or 5S rRNA. The level of 5S rRNA was used as a loading contrnnnnnol. The endogenous hTR is not visible in the vector-only sample (lane 1) due to the short exposure time. (*In vivo* reconstituted) Quantitation of telomerase activity and processivity of telomerase reconstituted in cells in relation to the wild-type TERT are shown below the gel. (*In vitro* reconstituted) Activity and processivity of telomerase mutants reconstituted *in vitro* analyzed in Fig. 2.2 is shown at bottom for comparison. n/d: not determined.

2.4.3 The length of N-terminal linker of motif 3 affects telomerase activity

Based on the sequence alignment, the upstream linker connecting motif 2 to motif 3 is more variable in length than the downstream linker connecting motif 3 to motif A (Fig. 2.1b). To assess the functional importance of the motif 3 flanking linkers, we generated TERT mutants with insertions (i644AAA645 and i702KKK703) or deletions (d643-649 and d699-701) in the linker regions and assayed the *in vivo* reconstituted enzymes for telomerase activity.

Both linker regions are more sensitive to deletions than insertions. Insertions i644AAA645 (N-terminal linker) and i702KKK703 (C-terminal linker) did not significantly decrease telomerase activity or processivity (Fig. 2.3, lanes 4 and 6). In contrast, deletions in the linker regions caused dramatic alterations in telomerase activity. The 3-residue deletion d699-701 in the C-terminal linker 3/A nearly abolished activity (Fig. 2.3, lane 5). Surprisingly, a 7-residue deletion d643-649 in the N-terminal linker increased activity by nearly two-fold without significant changes in processivity (Fig. 2.3, lane 3). The N-terminal linker appeared more flexible than of the C-terminus, as a 12-residue deletion in the N-terminal linker did not affect telomerase activity (Fig. S2.4,

compare lane 1 and 4). Notably, the hyperactive alanine-substitution mutations, V658A, K659A, R669A, are located within the N-terminal portion of motif 3 near the upstream linker, implicating a similar role for the N-terminal linker and the N-terminal portion of motif 3 in regulating telomerase activity.

2.4.4 Motif 3 is functionally conserved in *Tetrahymena* TERT

The sequence of motif 3 is well conserved between vertebrates and ciliates (Figure 2.1b). To determine if motif 3 is required for telomerase function in ciliates, we reconstituted eight *Tetrahymena* telomerase motif 3 mutants *in vitro* (L565A, R573A, G584A, F588A, D589A, I593A, F597A and F600A) with alanine substitutions at positions homologous to the residues critical for activity or processivity in human telomerase (Figure S2.6a). All eight mutants exhibited 3-20 fold reductions in activity (Figure S2.6b), indicating that motif 3 is critical for telomerase activity in *Tetrahymena*. Similar to the human telomerase mutants, W690A and F693A, the *Tetrahymena* mutants, F597A and F600A, impaired only telomerase activity, not processivity (Figure S2.6b, lanes 8 and 9). Furthermore, four *Tetrahymena* mutants, R573A, G584, F588A and I593A, showed 10-20 fold reductions in processivity (Figure S2.6b, lanes 3-5 and 7). We concluded that motif 3 is essential and evolutionarily conserved for telomerase processivity.

2.4.5 Hyperactive motif 3 mutants exhibit faster rates of repeat addition

To determine if the greater activity observed within the hyperactive mutants was due to faster repeat addition, we carried out a pulse-chase time-course assay to measure repeat addition rate. Our results demonstrate that, regardless of their differences in

processivity, the hyperactive d643-649, V658A, K659A and R669A mutants add telomere repeats at a faster rate of 5-6 repeats/min, higher than the wild-type enzyme, 3-4 repeats/min (Fig. 2.4, lanes 7-32). Conversely, the hypoactive mutants E668A, D684A and V697A present slower repeat addition rates than the wild-type enzyme (Fig. 2.4, lanes 38-52). These results suggest a critical role for motif 3 in regulating repeat addition rate of telomerase enzyme.

The increase or decrease in repeat addition rate is independent of the processivity level of the mutants. In the pulse-chase assay, the extent of telomerase processivity was measured at the last time point where the processive enzyme-product complexes have already moved up to the top of the gel and separated from the products dissociated from the enzyme during the time course. While the hyperactive V658A mutant was more processive, the other two hyperactive mutants, K659A and R669A, were less processive (Fig. 2.4, compare lanes 17, 22, 27 and 32). Moreover, by combining two motif 3 mutations (d643-649 and V658A) and an hTR-57C template mutation that increases processivity (23), we generated a telomerase mutant that is super-active and super-processive (Fig. S2.3, lane 6), demonstrating an additive effect for these mutations in telomerase activity and processivity.

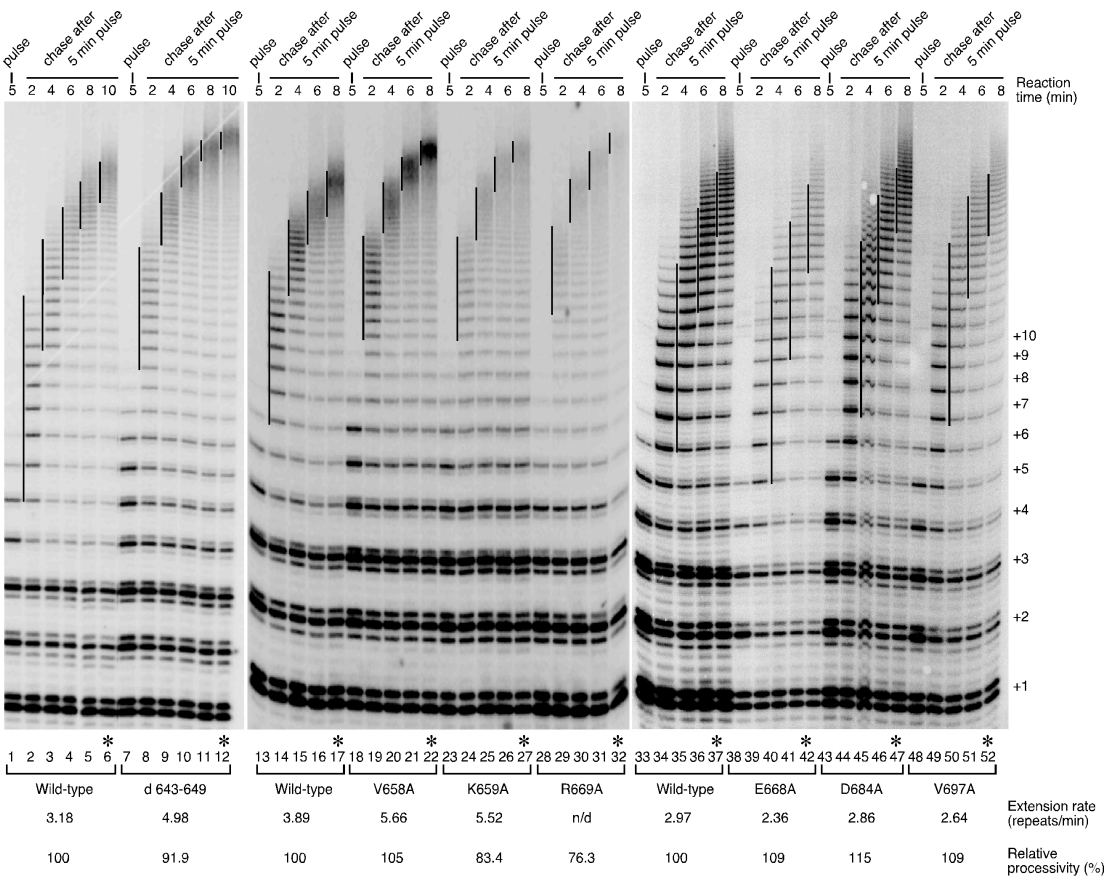


Fig. 2.4. Pulse-chase time course analysis to measure repeat addition rates of the motif 3 mutants. *In vitro* reconstituted wild-type, hyperactive (d643-649, V658A, K659A and R669A) or hypoactive (E668A, D684A and V697A) telomerase mutants were incubated with $(TTAGGG)_3$ primer in the pulse reaction for 5 min in which the [α-^{32}P]dGTP is incorporated to the newly synthesized telomere repeats. After 5 min of pulse reaction, non-radioactive dGTP was added to 100 μM to initiate the chase reactions and the reactions were terminated at different time points (2-10 min). The vertical lines on the gel denote the major bands of telomere products synthesized and labeled in the initial 5-min pulse reaction, and extended in the following chase reactions. Numbers on the right (+1, +2, +3 etc.) indicate the number of repeats added to the telomeric primer. Repeat-extension rate, expressed as repeats per minute, of each enzyme were calculated (see Materials and Methods) and indicated below the gel. Asterisk (*) denotes the 10 min chase reaction from which the processivity of each mutant was measured based on the first 10 major bands (repeats 1 to 10) (see Materials and Methods).

2.4.6 Low-processivity motif 3 mutants are defective in utilizing short DNA primers

The motif 3 mutations that altered processivity presumably affected either the template realignment or product release step of translocation, as telomerase processivity correlates to the probability of template realignment over product release. To determine if the low-processivity mutations affect the template realignment step, we designed a short primer assay to analyze the ability of telomerase to use short primers that, when base-pair with the RNA template, leave no single-stranded overhang for the TEN domain anchor site to bind (Fig. 2.5b). This assay thus discounts the effect of the TEN domain on substrate binding, as it binds to the upstream single-stranded region of a longer telomeric DNA primer (40). The base-pairing between the short tel8 DNA primer and the template RNA resembles the realignment of the 3'-end of telomeric DNA with the RNA template during translocation. By using a short primer, i.e. the 8-nt tel8 primer, in the telomerase assay, we can then determine solely the contribution of motif 3 in facilitating formation, or recognition, of the RNA/DNA duplex inside the active site. Thus, a low-processivity mutant with an inability to complete the realignment step in a translocation cycle would be predicted unable to use a short primer as substrate.

Using the short-primer assay, we tested six motif 3 mutants (V658A, N666A, L681A, G682A, D684A and I686A) with either increased or decreased processivity. In addition to motif 3, we also analyzed low-processivity mutations that are located in other parts of TERT, or in the TR component. Mutations N95A (TEN), L980A (CTE), 790-VVIE-793-4A (IFD) and hTR-A55G (RNA template) have been previously shown to reduce telomerase processivity (20,22,23,41). The wild-type telomerase can utilize all

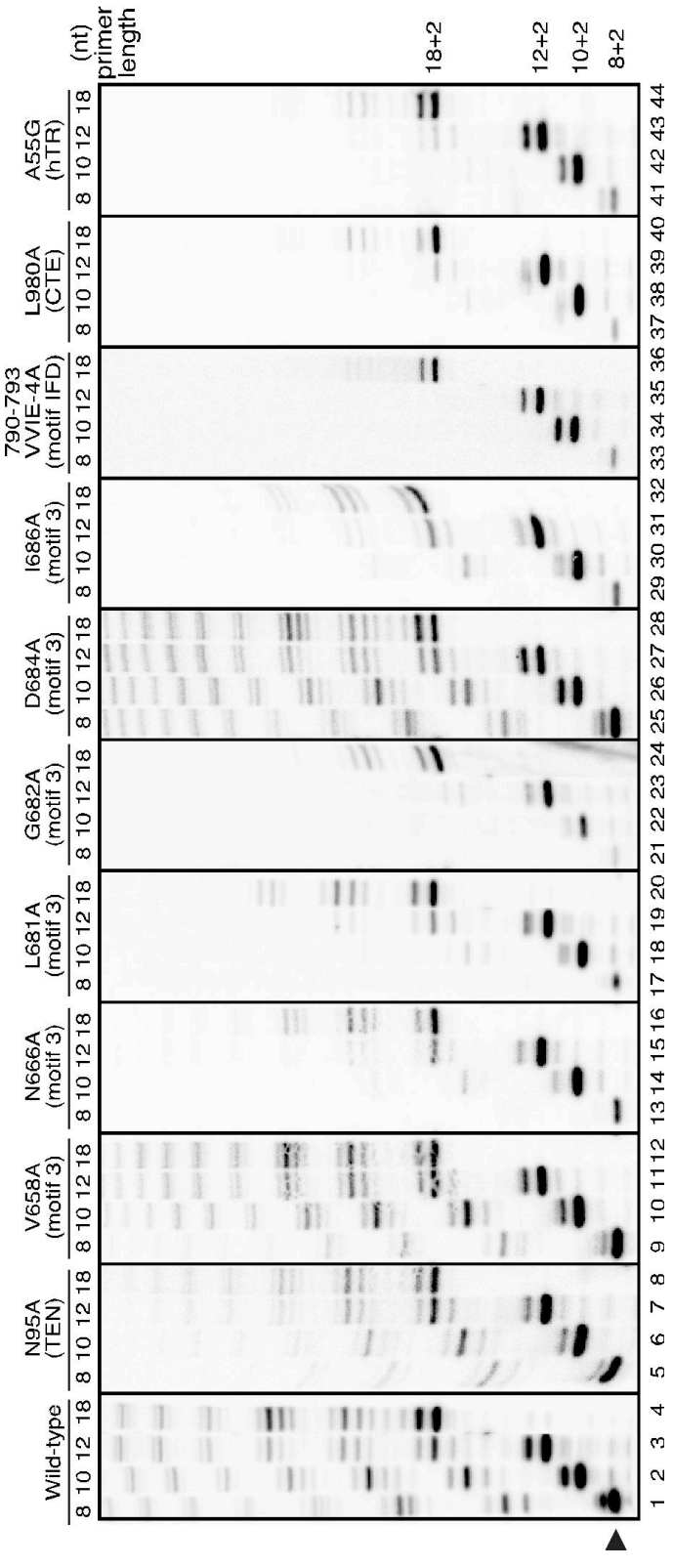
primers (8, 10, 12 or 18 nt) tested with similar activity (Fig. 2.5A, lanes 1-4). Remarkably, the low-processivity mutants, N666A, L681A, G682A, I686A, 790-VVIE-793-4A, L980A and hTR-A55G, that can extend longer primers normally, had little to no activity when using the short tel8 primer (Fig. 2.5a, lanes 13, 17, 21, 29, 33, 37 and 41). This suggests these mutations compromised the ability of TERT in promoting RNA/DNA duplex formation or positioning the duplex into the active site for the first repeat synthesis. As would be expected, the hyper-processive mutants V658A and D684A utilized the short tel8 and the longer primers with equal efficiency (Fig. 2.5a, lanes 9 and 25). The TEN domain N95A mutant, while having a low processivity, can however extend the short tel8 primers efficiently (Fig. 2.5a, lanes 1 and 5). The TEN domain thus does not appear to play a role in facilitating primer/template realignment, rather preventing product release during template translocation through DNA binding. When using tel10 and tel12 primers, all enzymes gave rise to stronger first bands (10+2 and 12+2) than the subsequent bands, indicating a lower efficiency for the first translocation event (Fig. 2.5a). This phenomenon is also consistent with the notion that TEN domain binding to the longer DNA primer facilitates template translocation.

To quantitatively determine the ability of these processivity mutants to utilize short primers as substrate, we measured the *Km* of the wild-type and mutant enzymes to the tel8 DNA primer. All mutants that failed to extend the tel8 primer had higher *Km* values ranging from 2.76 to 5.31 μM, 5-10 fold higher than the 0.56 μM of the wild-type enzyme (Fig. 2.5c). In contrast, mutants that retained the ability to extend the tel8 primer had *Km* values similar to or lower than the wild-type enzyme. For example, the most

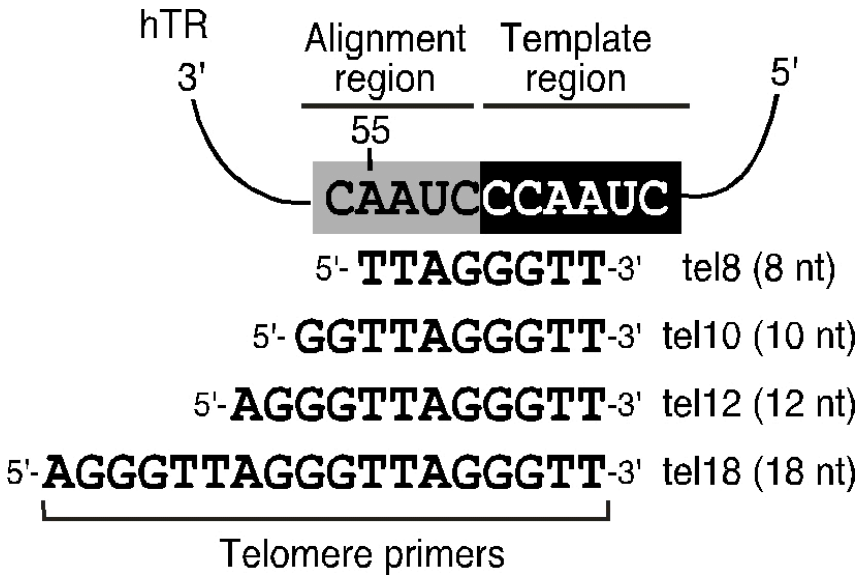
processive mutant D684A has a *Km* of 0.33 μM, significantly lower than that of the wild-type (Fig. 2.5c). Compared to tel8, primers tel10 and tel12 can still be used efficiently by the wild-type and mutant enzymes, with *Km* ranging from 80 to 680 nM (Fig. 2.5a). When the longer tel18 primer (18 nt) was used, the wild-type enzyme and all mutants, with the exception of the TEN N95A mutant, had similar *Km* values around 100 nM (Fig 2.5c).

The hTR-A55G template mutation resulted in a mismatch between the RNA template and the DNA substrate (Fig. 2.5b), leading to a high *Km* for the tel8 primer (Fig. 2.5c) and a lower processivity (Fig. 2.5a, lanes 41-44). The IFD motif, found first in yeast, also contributes to the repeat addition processivity (20). Based on the sequence alignment, we divided the IFD into three regions, termed IFD-a, -b and -c (see Figs. S2.1b and S2.4). Our human IFD-b mutant (790-VVIE-793-4A) causes a phenotype similar to motif 3, CTE and the hTR template mutants, confirming that IFD-b is indeed required for processivity (Fig. 2.5a, lanes 33-36).

The TEN N95A mutant had a higher *Km* value of 0.16 μM to the 18 nt tel18 primer, presumably due to a reduced binding affinity to the 5'-end single-strand region of the longer DNA primer (Fig. 2.5c). This supports that the TEN domain is a major contributor for the overall substrate affinity as proposed previously. The strong correlation between the processivity and the *Km* to the tel8 primer of the motif 3, IFD and CTE mutants ascertain that these elements contribute to the formation or positioning of an extendable RNA/DNA duplex substrate in the active site.



a. Telomerases with specific mutations in TERT or TR reaction with primers of various lengths



b. The four different primers are aligned with hTR template sequence.

Fig. 2.5. Activity assay of telomerase mutants using primers with various lengths.

Mutation (Domain)	Km^{app} to telomere primers (μM)			
	tel8	tel10	tel12	tel18
Wild-type	0.56 (± 0.10)	0.20 (± 0.06)	0.09 (± 0.02)	0.10 (± 0.01)
N95A (TEN)	0.45 (± 0.14)	0.23 (± 0.08)	0.13 (± 0.02)	0.16 (± 0.02)
V658A (motif 3a)	0.58 (± 0.11)	0.25 (± 0.07)	0.10 (± 0.02)	0.12 (± 0.02)
N666A (motif 3a)	3.60 (± 0.07)	n/d	n/d	n/d
L681A (motif 3b)	2.96 (± 0.66)	0.47 (± 0.15)	0.14 (± 0.01)	0.11 (± 0.01)
G682A (motif 3b)	5.31 (± 0.75)	0.68 (± 0.05)	0.15 (± 0.02)	0.11 (± 0.01)
D684A (motif 3b)	0.33 (± 0.06)	0.16 (± 0.01)	0.08 (± 0.03)	0.12 (± 0.02)
I686A (motif 3b)	3.63 (± 0.20)	n/d	n/d	n/d
L980A (CTE)	2.76 (± 0.96)	n/d	n/d	n/d
790VVIE793-4A (IFD)	4.92 (± 0.70)	n/d	n/d	n/d
A55G (hTR template)	3.43 (± 1.21)	n/d	n/d	n/d

c. The Km^{app} values for telomere primers of various lengths.

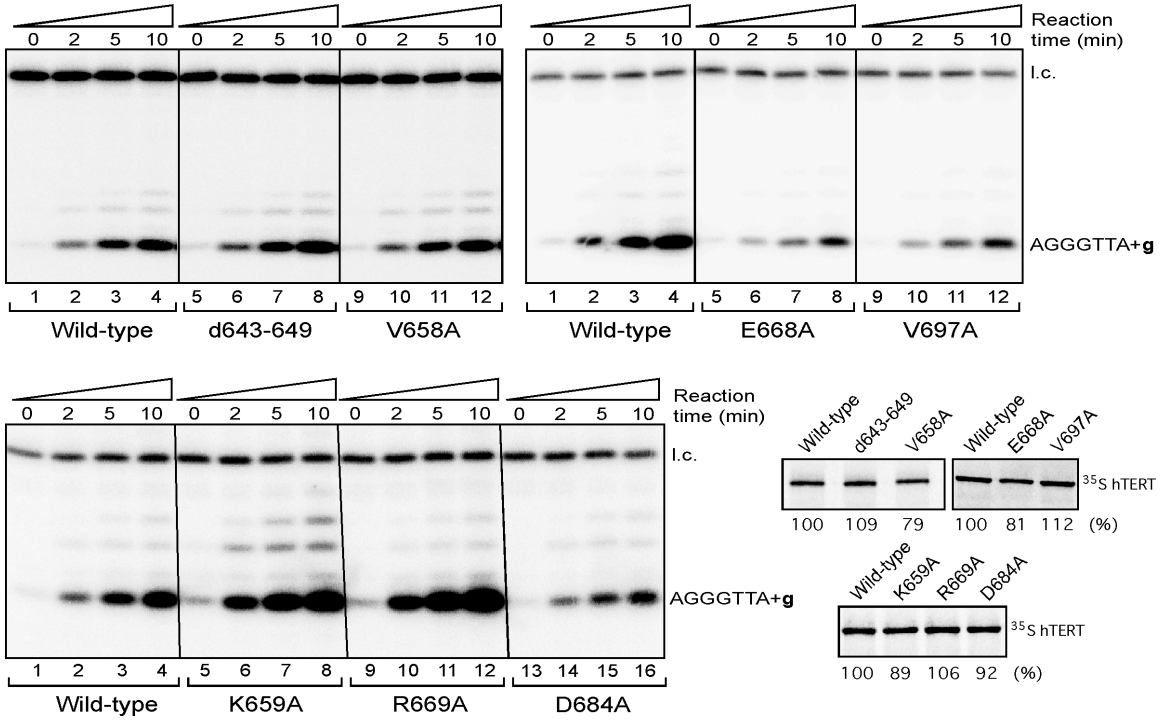
Fig.2.5. continued. Activity assay of telomerase mutants using primers with various lengths. (a) Telomerases with specific mutations in TERT or TR indicated were assayed for activity using telomere primers, tel8, tel10, tel12 or tel18, with length ranging from 8 to 18 nt. Due to the difference in overall activity between mutants, the gel image of each mutant was adjusted to have similar intensity for better comparison of the products of different primers. The numbers (8+2, 10+2, 12+2 etc.) labeled on the right of the gel indicate the length of the primer plus the number of nucleotides added. The black triangle on the left of the gel indicates the first repeat product extended from the tel8 primer. (b) The four different primers are aligned with hTR template sequence. The alignment (nt 52-56) and template (nt 46-51) regions are shaded in grey and black, respectively. (c) The Km^{app} values for telomere primers of various lengths. The apparent *Km* values are determined from experiments using tel8, tel10, tel12 or tel18 primers at various concentrations and by fitting the data to the Michaelis-Menten equation (see Materials and Methods). Standard deviations (n=3-4) are given in parentheses. n/d: not determined.

2.4.7 Hyperactive telomerase mutants have higher enzyme turnover rates

Since template translocation is the rate-limiting step in a processive telomerase reaction, the increased repeat addition rate observed with the hyperactive mutants should result from a greater translocation rate. Strand-separation between the telomeric DNA and the template RNA is a crucial step of template translocation. Taking advantage of the short primer assay, we asked if the hyperactive mutants have a faster dissociation rate for extended telomere product from the template, which resembles the strand-separation step of template translocation. Since the tel8 primer still gave rise to multiple repeats products, indicating successful translocation events, we thus used an even shorter tel7 DNA primer, 5'-AGGGTTA-3' (7-nt) and only dGTP in the reaction to prevent any possible realignment of the telomeric DNA product with the RNA template after one round of repeat synthesis. Without the interference from TEN domain or other DNA binding sites, this time-course assay focused primarily on the rate of product dissociation from the active site.

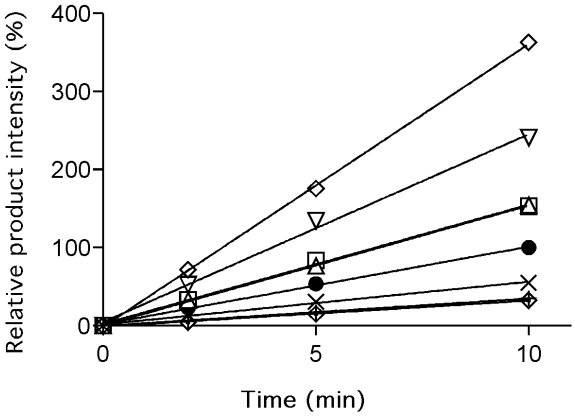
Our results from this time-course analysis indicated that all hyperactive mutants d643-649, V658A, K659A and R669A have higher enzyme turnover rates than the wild-type enzyme (Fig. 2.6). The higher turnover rates of these mutants were not due to higher substrate binding affinity, as these hyperactive mutants had a *Km* higher or similar to that of the wild-type enzyme (Fig. 2.6b) and the reactions were performed at a saturated substrate concentration of 10 μM (see Materials and Methods). The high *Km* values of the hyperactive mutants K659A and R669A were consistent with their low processivity as shown above (Fig. 2.4 and 2.6b). Conversely, the hypoactive mutants E668A, D684A

and V697A that showed lower repeat addition and template translocation rates (Fig 2.4) had lower enzyme turnover rates (Fig. 2.6b). Together, these results suggest that the increased template translocation rates of the hyperactive motif 3 mutants are likely due to the higher dissociation rates of products from the active site after each round of repeat synthesis.



a. Telomerase activity time course analysis of the hyperactive (d643-649, V658A, K659A and R669A) and hypoactive (E668A, D684A and V697A) mutant telomerases.

Fig. 2.6. Enzyme turnover rates of the hyperactive and hypoactive motif 3 mutants. (a) Telomerase activity time course analysis of the hyperactive (d643-649, V658A, K659A and R669A) and hypoactive (E668A, D684A and V697A) mutant telomerases. The reactions were performed using a 7 nt telomere primer 5'-AGGGTTA-3' and incubated for various amounts of time (0, 2, 5 and 10 min) as indicated. The reactions were carried out in the presence of only ^{32}P-dGTP nucleotide to prevent processive reactions. A ^{32}P end-labeled 15 nt DNA oligonucleotide is used as the loading control (l.c.). The *in vitro* synthesized TERT proteins (wild-type, d643-649, V658A, K659A, R669A, D684A, E668A and V697A) were labeled with [^{35}S]-methionine and quantitated after SDS-PAGE.



Mutation*	Relative enzyme turnover rate (%)	Km^{app} (μM)**
◇ R669A (↑)	334 ± 25.0	3.55 ± 0.97
▽ K659A (↑)	240 ± 18.8	2.13 ± 0.08
△ V658A (↑)	146 ± 7.1	1.96 ± 0.11
□ d643-649 (↑)	151 ± 15.1	1.36 ± 0.10
● Wild-type	100	1.31 ± 0.04
× D684A (↓)	57.0 ± 2.0	1.16 ± 0.24
◈ E668A (↓)	27.9 ± 4.6	1.27 ± 0.06
+ V697A (↓)	33.1 ± 1.8	1.31 ± 0.02

*: (↑) = hyperactive; (↓) = hypoactive
**: measured using the tel7 primer (5'-AGGGTTA-3')

b. Quantitation of enzyme turnover rates of the hyperactive and hypoactive telomerase mutants.

Fig. 2.6. continued. Enzyme turnover rates of the hyperactive and hypoactive motif 3 mutants. (b) Quantitation of enzyme turnover rates of the hyperactive and hypoactive telomerase mutants. For each telomerase enzyme, the intensity of products was adjusted with protein amount and normalized by the intensity of loading control. For each set of reactions, the product intensities are further normalized to that of the wild-type reaction at the 10 min time point. The relative product intensities were then plotted against the amount of time. Wild-type (filled circle); d643-649 mutant (square); V658A (triangle); K659A (reverse triangle); R669A (diamond); D684A (cross); E668A (dotted diamond); V697A (plus). The relative enzyme turnover rates were determined from slopes of the linear trend lines in relation to that of the wild-type enzyme. The apparent *Km* values of different telomerase mutants toward the tel7 primer were determined by fitting the data to Michaelis-Menten equation. The standard deviation was derived from 3 independent experiments.

2.4.8 Mutations in motif 3b do not affect the affinity of the RT domain to ssDNA, ssRNA or RNA/DNA duplex

To determine if the higher *Km* values of motif 3 mutants are due to a lower DNA-binding affinity, we carried out *in vitro* direct binding assay in which biotin-labeled telomeric DNA oligonucleotides and streptavidin beads were used to pull down the in vitro synthesized TERT protein fragment. In addition to ssDNA, we also tested ssRNA and RNA/DNA duplex for their affinity to the TERT protein fragment (Fig. S2.7a). The TERT fragment (a.a. 601-939) tested here includes the whole RT domain and exhibits weak affinity to DNA as shown previously. Our results did not reveal any significant differences in binding affinity to either the ssDNA, ssRNA or RNA/DNA duplex between the motif 3 mutant and wildtype protein fragments (Fig. S2.7b). We thus conclude that *in vitro* direct binding assay is not sensitive enough to detect the affinity difference between wild-type and processivity mutant telomerase toward RNA/DNA duplexes.

2.5 Discussion

The repeat addition processivity of telomerase relies on a unique template translocation mechanism that presumably requires novel structural elements within the TERT protein. The telomerase-specific motif 3 we characterized in this study has been overlooked in the past, in part due to the low degree of sequence conservation among eukaryotic lineages, the presence of variable linkers and the inefficiency of alignment algorithms (Fig. 2.1 and S2.2). In this study, through a comprehensive mutagenesis analysis within motif 3 (Fig. 2.2), we identified mutations that unusually increased the rate or altered the processivity of telomere repeat addition (Fig. 2.3 and 2.4). By using a novel short-primer assay to determine the binding affinity of the mutants to short DNA primers (Fig. 2.5) and the time-course analysis to measure enzyme turnover rates (Fig. 2.6), we showed that motif 3 mutations affect repeat addition rate and processivity, suggesting a crucial role for motif 3 in strand-separation and realignment during template translocation.

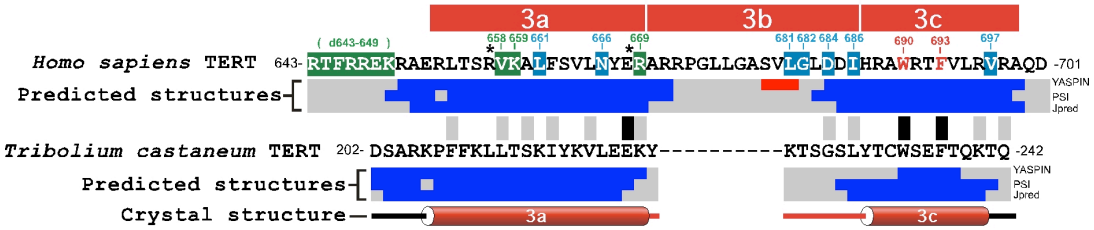
Our sequence alignment analysis and secondary structure prediction on motif 3 provide useful insights into the function and evolution of this motif. The secondary structures of motif 3 predicted from different TERT homologs are surprisingly conserved (Fig. S2.1A) and consistent with the crystal structure of *Tribolium castaneum* TERT (34). The secondary structure prediction of TERT sequences from all available species suggests that the motif 3 region consists of two α-helices separated by a conserved linker (Fig. S2.1A), consistent with the *Tribolium* crystal structure. To display the physical location of motif 3 in relation to other TERT domains and the RNA/DNA duplex, we

mapped the sequence of the two putative α-helices of human motif 3 onto the crystal structure *Tribolium* TERT based on the structural and sequence homology. To specifically denote the different structural features, we divided motif 3 into three sub-motifs, 3a, 3b and 3c, where 3a and 3c designate the two separate α-helices and 3b designates the spanning linker (Fig. 2.7A). While the sequences of helices 3a and 3c are well conserved in most organisms, the sequence of linker 3b is conserved most specifically within vertebrates, non-yeast fungi, plants and ciliates (Fig. 2.1B and 2.7A). This group-specific sequence conservation of linker 3b suggests a role important for telomerase function in most species, yet dispensable and lost in species including nematodes, insects and yeasts (Fig. S2.2).

Our comprehensive mutagenesis surveyed the functional effects of alanine-substitution at individual residues of motif 3. Although alanine-substitution at most of the highly conserved residues resulted in significant changes in telomerase activity or processivity, mutations at the residues V664A, L676A and S679A showed no dramatic changes (Fig. 2.2B). It was however expected that alanine substitutions do not always give the same degree of effects for all conserved residues, due to unique structural and chemical properties within various amino acids. Substitutions to amino acids other than alanine will presumably produce different phenotypes.

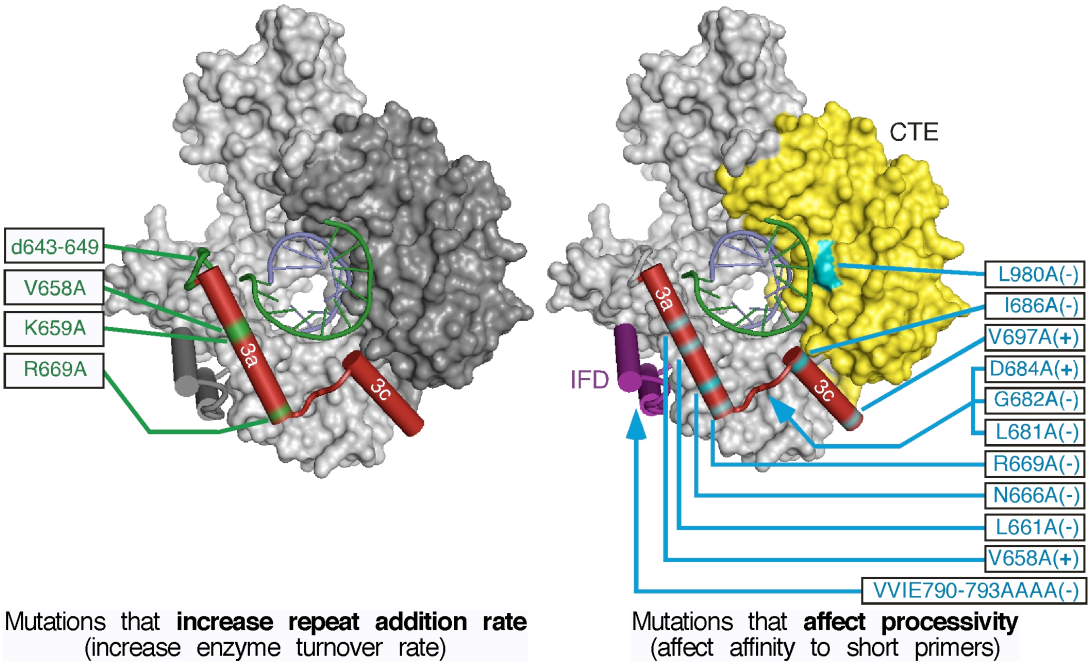
The mutations (L681A, G682A and I686A) that severely impair processivity are located in linker 3b (Fig. 2.2B and 2.7B, right), suggesting a primary role for this linker in regulating the template translocation efficiency and repeat addition processivity. In comparison, mutations (d643-649, V658A, K659A and R669A) that significantly

increase telomerase activity are located in the helix 3a and its N-terminal linker (Fig. 2.7B, left), suggesting that helix 3a is more important in regulating the rate of template translocation and repeat addition. However, helix 3a might play an additional role in telomerase processivity as several mutations (V658A, L661A, N666A, E668A and R669A) in motif 3a also substantially altered processivity.



a. Sequence alignment and predicted secondary structures of human and *Tribolium* motif 3.

Fig. 2.7. Homologous locations of human TERT mutations on the *Tribolium* TERT structure. (a) Residues critical for repeat addition rate are shaded in green, while critical residues for processivity are shaded in blue; asterisk (*) indicates both. The residues that abolish activity when mutated are colored red. The secondary structure (α-helix shown as a cylinder) based on the crystal structure of *Tribolium* TERT is shown below the predicted secondary structures. The predicted secondary structures (α-helix shown in blue and β-sheet shown in red) based on three algorithms, YASPIN, PSI, and JPred (see Supplementary Fig. S1A). Black/grey boxes located between human and *Tribolium* sequences indicate identity/similarity. The alignment is based on optimal positioning within the predicted helix sequence.



b. Mutations in human TERT that affect repeat addition rate and processivity are mapped onto the crystal structure of *Tribolium* TERT.

Fig. 2.7. continued (b) Mutations in human TERT that affect repeat addition rate and processivity are mapped onto the crystal structure of *Tribolium* TERT. (Left panel) Mutations that increase repeat addition rate are located in helix 3a and its N-terminal linker that connects motif 2 to helix 3a. (Right panel) Mutations that affect processivity and RNA/DNA duplex formation are dispersed in the IFD, motif 3 and the CTE. The blue arrowheads indicate putative locations for the human TERT sequences (IFD-b and motif 3b) absent from the *Tribolium* TERT. The (+) and (-) following the mutations denote an increase or decrease in processivity. Superimposed hetero-duplex of RNA strand (green) and DNA strand (purple).

2.5.1 The role of motif 3 on repeat addition rate

The three mutations in motif 3a, V658A, K659A and R669A, and the deletion d643-649 in N-terminal linker of motif 3a remarkably increased the repeat addition rate (Fig. 2.4), presumably due to an increase in template translocation rate. This increase in repeat addition rate is independent of the processivity, as some hyperactive mutants have decreased processivity (Fig. 2.4). The combination of a high addition-rate mutation with a low-processivity mutation did not reduce the repeat addition rate of the enzyme (Fig. S2.5, compare lane 5-8 and 9-12), supporting that the rate and processivity of telomerase are regulated separately, as previously proposed (27). Likewise, low repeat addition rate mutant combining with high processivity mutation did not increase enzyme extension rate (Fig. S2.5, compare lane 25-28 and 29-32).

The higher enzyme turnover rates measured in the short primer assay suggest faster product dissociation (or strand-separation) rates for the motif 3 hyperactive mutants (Fig. 2.6), assuming the product dissociation is the rate-limiting step in the assay. This is consistent with the faster template translocation rates of the hyperactive mutants observed in the processive pulse-chase time-course analysis, in where nucleotide polymerization is not rate limiting (Fig. 2.4). We propose that the putative helix 3a and its N-terminal linker regulate the strand-separation step of template translocation and thus modulate the rate of repeat addition. It remains unclear if the strand-separation of the RNA/DNA hybrid involves a conformational change to helix 3a or its N-terminal linker. We hypothesize that the long N-terminal linker could function in allowing helix 3a to swing away from the active site, permitting the RNA/DNA duplex to dissociate from the active site and the

two strands to separate from each other (Fig. 2.7B). A source of energy for such a conformational change could originate from the movement and distortion of DNA/RNA duplex during repeat synthesis as previously proposed (42).

2.5.2 The role of motif 3 on repeat addition processivity

Mutations at conserved residues within motif 3, the CTE and the IFD of TERT as well as the template region of TR affected telomerase processivity and the ability to use short telomere primers (Fig. 2.5 and 2.6). The retained ability of the wild-type enzyme to extend the short tel7 and tel8 primers indicates that the catalytic core of TERT protein alone is capable of promoting the formation of, or recognizing, the RNA/DNA duplex substrate inside the active site independent of the TEN domain. The ability of telomerase to use the short primer correlates to the processivity of repeat addition as it resembles the second step of the translocation event, where the RNA and DNA realign to form the hetero-duplex inside the active site for the next round of repeat synthesis. The fact that the TEN N95A mutant can efficiently elongate the tel8 primer is consistent with the notion that this TEN mutation impairs the binding to the upstream single-stranded region of telomeric DNA primer, representing a different mechanism to affect telomerase processivity.

It was however unexpected that the wild-type enzyme would be capable of adding more than one telomere repeats to the 8 nt DNA primer in the absence of upstream single-stranded sequence for TEN binding (Fig. 2.5A, lane 1). The synthesis of multiple repeats indicates the occurrence of template translocation after the synthesis of the first repeat. Although the tel8 primer does not initially leave a single-stranded overhang when

base-paired with the RNA template, it would potentially have the 5'-end unpaired from the template during cycles of nucleotide addition, while maintaining a constant number of base-pairings between the telomeric DNA and the template RNA, as previously proposed (43). Since the TEN domain requires a longer single-stranded DNA overhang for binding, a more adjacent DNA binding site (the template-proximal anchor site) in the RT domain would thus seem responsible for binding the partially unpaired 5'-end of the short tel8 DNA primer (41,44).

Our results suggest that motif 3, CTE and IFD contribute to the realignment of telomeric DNA and the RNA template, i.e. the reformation of RNA/DNA duplex. Interestingly, based on the crystal structure of *Tribolium* TERT, these three motifs are located adjacent to the RNA/DNA duplex, forming a horseshoe shaped structure encircling the duplex (Fig. 2.7B). The majority of the motif 3 mutations that severely impaired the enzyme's processivity and short primer usage are located in linker 3b (Fig. 2.7). We envision the conserved motif 3b would act as a molecular hinge, positioning helix 3a and the CTE to facilitate the RNA/DNA duplex formation or positioning the duplex within the active site (Fig. 2.7B). Since motif 3b is not conserved in insects, a crystal structure of a vertebrate or ciliate TERT would be necessary to elucidate the structural and functional purpose of motif 3b in template translocation.

Our phylogenetic and biochemical studies of motif 3 shed light upon the molecular mechanism of the translocation process for the processive telomerase reaction. The implication of our data provides testable hypotheses and elicits critical questions for future studies of telomerase action. Moreover, our hyperactive and hyper-processive

motif 3 mutants demonstrate the feasibility of enhancing telomerase enzymatic activity through motif 3 targeting. Altering telomerase function can possibly affect the proliferative capacity of adult stem cells. Drugs that augment enzymatic activity and processivity of telomerase, similar to our mutants, might provide treatments for patients suffering from telomerase-insufficiency diseases. Additionally, the abated telomerase motif 3 mutants provide potential drug target locals for anti-cancer therapies.

2.6 References

1. Shawi, M. and Autexier, C. (2008) 129, 3-10.

2. Armanios, M.Y., Chen, J.J.L., Cogan, J.D., Alder, J.K., Ingersoll, R.G., Markin, C., Lawson, W.E., Xie, M., Vulto, I., Phillips, J.A., 3rd *et al.* (2007) *N. Engl. J. Med.*, 356, 1317-1326.

3. Walne, A.J. and Dokal, I. (2008) *Mechanisms of ageing and development*, 129, 48-59.

4. Autexier, C. and Lue, N.F. (2006) *Annu. Rev. Biochem.*, 75, 493-517.

5. Jacobs, S.A., Podell, E.R. and Cech, T.R. (2006) *Nat. Struct. Mol. Biol.*, 13, 218-225.

6. Osanai, M., Kojima, K.K., Futahashi, R., Yaguchi, S. and Fujiwara, H. (2006) *Gene*, 376, 281-289.

7. Xiong, Y. and Eickbush, T.H. (1990) *EMBO J.*, 9, 3353-3362.

8. Greider, C.W. (1991) *Mol. Cell. Biol.*, 11, 4572-4580.

9. Morin, G.B. (1989) *Cell*, 59, 521-529.

10. Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F. and Chen, J.J.L. (2008) *J. Biol. Chem.*, 283, 2049-2059.

11. Cohn, M. and Blackburn, E.H. (1995) *Science (New York, N.Y*, 269, 396-400.

12. Prowse, K.R., Avilion, A.A. and Greider, C.W. (1993) *Proc. Natl. Acad. Sci. USA*, 90, 1493-1497.

13. Lue, N.F. and Peng, Y. (1997) *Nucleic Acids Res.*, 25, 4331-4337.

14. Bosoy, D. and Lue, N.F. (2004) *Nucleic Acids Res.*, 32, 93-101.

15. Wang, F., Podell, E.R., Zaug, A.J., Yang, Y., Baciu, P., Cech, T.R. and Lei, M. (2007) *Nature*, 445, 506-510.

16. Aigner, S. and Cech, T.R. (2004) *RNA*, 10, 1108-1118.

17. Sun, D., Lopez-Guajardo, C.C., Quada, J., Hurley, L.H. and Von Hoff, D.D. (1999) *Biochemistry*, 38, 4037-4044.

18. Harrington, L.A. and Greider, C.W. (1991) *Nature*, 353, 451-454.

19. Zaug, A.J., Podell, E.R. and Cech, T.R. (2008) *Nat. Struct. Mol. Biol.*, 15, 870-872.

20. Lue, N.F., Lin, Y.C. and Mian, I.S. (2003) *Mol. Cell. Biol.*, 23, 8440-8449.

21. Bryan, T.M., Goodrich, K.J. and Cech, T.R. (2000) *J. Biol. Chem.*, 275, 24199-24207.

22. Huard, S., Moriarty, T.J. and Autexier, C. (2003) *Nucleic Acids Res.*, 31, 4059-4070.

23. Chen, J.-L. and Greider, C.W. (2003) *EMBO J.*, 22, 304-314.

24. Gavory, G., Farrow, M. and Balasubramanian, S. (2002) *Nucleic Acids Res*, 30, 4470-4480.

25. Lai, C.K., Miller, M.C. and Collins, K. (2003) *Mol. Cell*, 11, 1673-1683.

26. Moriarty, T.J., Marie-Egyptienne, D.T. and Autexier, C. (2004) *Mol. Cell. Biol.*, 24, 3720-3733.

27. Drosopoulos, W.C., Direnzo, R. and Prasad, V.R. (2005) *J. Biol. Chem.*, 280, 32801-32810.

28. Podlevsky, J.D., Bley, C.J., Omana, R.V., Qi, X. and Chen, J.J.L. (2008) *Nucleic Acids Res*, 36, D339-343.

29. Ge, L. and Rudolph, P. (1997) *Biotechniques*, 22, 28-30.

30. Alder, J.K., Chen, J.J.L., Lancaster, L., Danoff, S., Su, S.C., Cogan, J.D., Vulto, I., Xie, M., Qi, X., Tuder, R.M. *et al.* (2008) *Proc. Natl. Acad. Sci. USA*, 105, 13051-13056.

31. Cristofari, G., Adolf, E., Reichenbach, P., Sikora, K., Terns, R.M., Terns, M.P. and Lingner, J. (2007) *Mol. Cell*, 27, 882-889.

32. Cristofari, G. and Lingner, J. (2006) *EMBO J.*, 25, 565-574.

33. Li, Y., Yates, J.A. and Chen, J.J.L. (2007) *Gene*, 400, 16-24.

34. Gillis, A.J., Schuller, A.P. and Skordalakes, E. (2008) *Nature*, 455, 633-637.

35. Malik, H.S. and Eickbush, T.H. (1998) *Mol Biol Evol*, 15, 1123-1134.

36. Zimmerly, S., Hausner, G. and Xc, W. (2001) *Nucleic Acids Res*, 29, 1238-1250.

37. Arkhipova, I.R. (2006) *Syst Biol*, 55, 875-885.

38. Liang, J., Yagasaki, H., Kamachi, Y., Hama, A., Matsumoto, K., Kato, K., Kudo, K. and Kojima, S. (2006) *Haematologica*, 91, 656-658.

39. Yamaguchi, H., Calado, R.T., Ly, H., Kajigaya, S., Baerlocher, G.M., Chanock, S.J., Lansdorp, P.M. and Young, N.S. (2005) *N. Engl. J. Med.*, 352, 1413-1424.

40. Moriarty, T.J., Ward, R.J., Taboski, M.A. and Autexier, C. (2005) *Mol. Biol. Cell*, 16, 3152-3161.

41. Wyatt, H.D., Lobb, D.A. and Beattie, T.L. (2007) *Mol. Cell. Biol.*, 27, 3226-3240.

42. Greider, C.W. (1995) In Blackburn, E. H. and Greider, C. W. (eds.), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 35-68.

43. Forstemann, K. and Lingner, J. (2005). *EMBO reports*, 6, 361-366.

44. Finger, S.N. and Bryan, T.M. (2008) *Nucleic Acids Res.*, 36, 1260-1272.

CHAPTER 3

TEMPLATE FREE TELOMERASE UTILIZES RNA/DNA DUPLEX LIKE CONVENTIONAL RTS

3.1 Abstract

Telomerase is a specialized reverse transcriptase (RT) that recognizes single stranded DNA substrates and synthesizes telomere repeats onto the chromosome termini using an intrinsic RNA template. Here, we showed that the template free human telomerase reconstituted in either TNT lysate or 293FT cells is capable of extending RNA/DNA heteroduplex like other conventional RTs. The template free telomerase, which is composed of human telomerase reverse transcriptase (hTERT) and essential human telomerase RNA (TR) fragments lacking the template, can reverse transcribe up to 8 nucleotides along the RNA template with the RNA/DNA duplex ranging from 5 to 16 base pairs long. Supplying 5' DNA overhang to the duplex greatly enhanced the substrate's affinity to the enzyme. A comprehensive study on various RNA and DNA hybrid substrates suggested RNA/DNA duplex is the preferred substrate of telomerase. Ribonucleotide residues can prime the telomerase catalyzed DNA polymerization on the RNA/DNA duplex. No telomerase specific RNA polymerase activity was detected using this system.

3.2 Introduction

Telomerase is a unique reverse transcriptase (RT) that synthesizes telomeric DNA onto the linear chromosome termini according to the template sequence within its intrinsic RNA component. Telomerase activity is important for cell mortality as the germ line cells and 90% of the tumor cells require telomerase to maintain telomere length (1). Meanwhile, somatic cells undergo telomere shortening and only allow for limited cell proliferation capacity due to the lack of telomerase activity (2). As a ribonucleoprotein complex, telomerase holoenzyme contains two core components: a telomerase RNA (TR) carrying the template sequence and the telomerase reverse transcriptase (TERT).

The TR component not only contains the RNA template sequence, but also has conserved secondary structure among different groups of species. To date, the secondary structure is only determined for TRs from ciliates, vertebrates and yeast (3). The pseudoknot structure located to the 3' of the template appears to be a universal element essential for telomerase activity (3,4). In vertebrates, a CR4-CR5 domain interacts with TERT protein independent from the pseudoknot domain and is also important for telomerase enzymatic activity. The scaRNA domain, located in the 3' half of the vertebrate TRs, is critical for RNA biogenesis (5). Human telomerase activity can be rconstituted from only the pseudoknot and CR4-CR5 domains (6). The solution structure of hTR pseudoknot triple helix and a critical stem loop within CR4-CR5 domain are available (7, 8). However, the detailed mechanism of how TR domains contribute to telomerase function is still unclear. It has also been shown that telomerase can be over-expressed in 293 FT cells, and the resulting "super telomerase" cell extract allows for

direct analysis of telomerase activity (9).

The TERT protein is composed of an N terminal extension (NTE), a central RT domain and a C terminal extension (CTE) (10). The NTE and CTE are specific for telomerase RT. The NTE contains the Telomerase essential N terminus (TEN) domain for single stranded DNA binding and a TR binding domain (TRBD) (11, 12, 13). The CTE is not well conserved among different species and possibly contributes to both nucleotide and repeat addition processivity (14, 15). The central RT domain, which possesses seven signature reverse transcriptase motifs, is responsible for telomerase catalytic reaction (16). The *Tribolium* TERT structure demonstrated that the domain organization of TERT RT domain resembles the HIV RT domain (17). Although lacking the TEN domain, the *Tribolium* TERT has the RNA binding domain making extensive contacts with the CTE to form a ring-shape structure. The catalytic site falls in the center of the ring structure and can potentially fit 7 base pairs of RNA template/DNA duplex (17). Interestingly, the estimation of duplex length within *S. cerevisiea* active site also reveals a constant 7 base pairs between template and telomeric DNA (18).

The telomerase rection differs from the reactions catalyzed by conventional RTs in several aspects. Telomerase is the only RT specialized to recognize single stranded DNA. ssDNA specific binding motifs, especially the TEN domain, have evolved within the TERT protein from different species (11, 19). The RNA template helps position the 3' end of telomeric DNA into the catalytic site using a complementary sequence. The one and a half repeat of RNA template sequence enables telomerase to evolve a template translocation event in order to synthesize a long stretch of telomeric repeats (20). On the

other hand, conventional RTs recognize the RNA/DNA duplex as substrate to initiate reverse transcription. The conventional RTs also synthesize an extensive strand of DNA along very long RNA templates with limited dissociation occasions during this process (21). For example, the HIV RT has to complete the replication of 9 kb viral RNA into DNA sequence. The retroviral RTs not only conducts reverse transcription but also exhibits DNA dependent DNA polymerase activity. Moreover, during the initiation of DNA synthesis, the retroviral RTs use an RNA/RNA duplex formed by the viral RNA and host tRNA (22). Telomerase RT and the conventional RTs presumably are homologous and sharing the common ancestor during evolution. Therefore, telomerase might also have the ability to utilize diverse substrates to carry out various polymerization reactions. It has been recently reported that telomerase has limited ability to function as RNA dependent RNA Polymerase (RdRP) and DNA dependent DNA polymerase (23, 24). It has also been suggested that telomerase RT can synthesize the full length anti sense RNase MRP RNA in specific conditions (25). This RdRP reaction is presumably initiated by a self loop back hairpin at the 3' end of the RMRP RNA.

To investigate whether the ssDNA specific telomerase still retains the ability of uitilizing RNA/DNA duplex substrate and how well telomerase recognizes other hybrids, such as RNA/RNA duplex, we designed a template-free human telomerase missing the intrinsic RNA template sequence. A smililar "template-free" system has been previously tested using *Tetrahymena* telomerase but without the supply of duplex substrates (26). Here, we show that our template free telomerase can use pre-annealed RNA/DNA as substrate and such activity depends on both TERT and TR elements. Various RNA/DNA

duplexes were tested. Duplex length didn't affect telomerase activity while 5' DNA overhang greatly enhance the activity and nucleotide addition processivity. Unlike conventional RTs, telomerase can only synthesize a short strand of DNA with limited nucleotide addition processivity. Using this template free telomerase, we also directly show that telomerase preferred its native RNA/DNA duplex over all other hybrids. This study is the first to reveal telomerase's ability of using RNA/DNA duplex without single stranded DNA as the substrate. It also provides a comprehensive understanding of how telomerase utilizes different duplex substrates.

3.3 Materials and Methods

3.3.1 Oligos and duplex substrates

Oligos are purchased from Integrated DNA technology Inc (IDT). Various oligos used to assemble duplex substrates in particular experiments are shown in the figures. To assemble the duplex, RNA template and DNA oligos are mixed to a final concentration of 100 μM in 100 mM Tris-HCl (pH7.5), 500 mM NaCl, 50 mM EDTA (1X annealing buffer). The mixture was heated at 80°C for 3 min and slowly cool down to room temperature. For fair comparison, single stranded RNA or DNA oligo substrates used in reactions were prepared in the same way. In Fig. 3.3b, all reactions were conducted in annealing buffer with 100 μM extra KCl, and duplex substrate was added to 100 μM to assure 5 bp duplex formation.

3.3.2 Tm measurement of the duplexes

To measure the melting of each duplex, both oligos for the duplex were mixed at equal molar in a 150μL volume at 10 μM final concentration in telomerase reaction

buffer (1X PE buffer, described below). The OD-260 of the sample was measured along a temperature range from 10 to 60°C at 0.5°C intervals in a Cary 300 thermal control Spectrophotometer (Varian tech.). The data was collected using the Cary 300 UV-vis spectrophotometer and calculated for annealing temperature value, Tm. Alternatively, data points were fit into sigmoid curve (variable slope) and calculate for Tm using Prism 5 (Graphpad).

3.3.3 *In vitro* reconstitution of human telomerase

Human telomerase was reconstituted in Rabbit Reticulocyte Lysate (Promega) as previously described (Chapter 2). Briefly, hTERT was first synthesized in RRL according to the manufacturer's instruction. *In vitro* transcribed and PAGE-purified hTR fragments (hTR nt 32-195/ hTR nt 44-184/ hTR nt 64-184 and hTR nt 239-328) were added to the final concentration of 1μM to assemble the telomerase enzyme. For the turnover rate assay, hTERTs were synthesized in the presence of ^{35}S Met (>1000 Ci/mmol, 10.2 mCi/mL, PerkinElmer) and analyzed by SDS-PAGE. The same amount of protein were added for different mutants to carry out the reaction.

3.3.4 Telomerase activity assays

3.3.4.1 Conventional and template free telomerase activity assay

Activity assay was carried out as previously described (Chapter 2). Basically, 3 μL of *in vitro* reconstituted telomerase was assayed in a 10 μL reaction with 50 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$, 1 mM Spermidine, 1 mM dTTP, 1 mM dATP, 2 μM dGTP, 0.165 μM α^{32}P-dGTP (3000 Ci/mmole, 10 mCi/mL, PerkinElmer), and 1 μM $(TTAGGG)_3$ telomeric primer or 10 μM duplex substrate. The

reaction was carried out for 60 min at 30°C or room temperature and terminated by phenol/chloroform extraction, followed by ethanol precipitation. The product was analyzed on 10% or 18% polyacrylamide gel, and the gel was analyzed by a Phosphor Imager. To measure the *Km* of different duplexes to the template free telomerase, reactions were carried out with different concentrations of duplex substrate from 0 to 125 μM as indicated in the figure. In the temperature gradient assay, reaction mixing were carried out and held on ice before proceeding to various temperatures to avoid background activity from room temperature condition. In the reactions that contain 5 bp duplex substrate, temperature was kept at 4°C for 2 hours.

3.3.4.2 Turnover rate assay

Telomerase reconstituted in Rabbit Reticulocyte Lysate (RRL) was incubated with 200 μM duplex substrate (indicated in the figure) in 1X telomerase reaction buffer (50 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$, 1 mM spermidine) in a 5 μL volume at 30°C for 10 min. 5 μL solution including 3.3 μM α-^{32}P-dGTP (3000 Ci/mmole, 10 mCi/mL, PerkinElmer) in 1X reaction buffer was added to the mixture to initiate the reaction. The reaction was carried out at 30°C and terminated at different time points by phenol/chloroform extraction. The product was analyzed as stated in conventional telomerase activity assay.

3.3.4.3 Exonuclease I digestion

Template free telomerase was reacted with 40 μM of a 7 bp duplex with 9 nt overhang (detailed in the figure) in a 20 μL reaction following the condition described in conventional activity assay for 1 hour at 30°C. The DNA oligo of the duplex was added

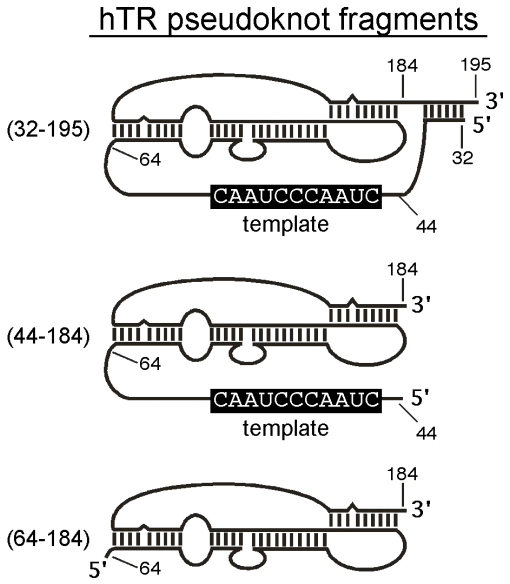
to a 10 μM excess in the reaction to prevent any re-annealing of the released single stranded DNA to the RNA template. After telomerase reaction, the solution was split equally and incubated with or without 20U Exonuclease I (20U/μL, New England Biolabs), 2.5 μM 12 bp duplex, 17 nt overhang, 20 nM 24 nt overhang DNA oligo for 30 min at 30°C. Reactions were terminated by phenol/chloroform extraction, followed by ethanol precipitation and the products were analyzed as described above.

3.4 Result

3.4.1 Template free telomerase reacts on the RNA/DNA duplex

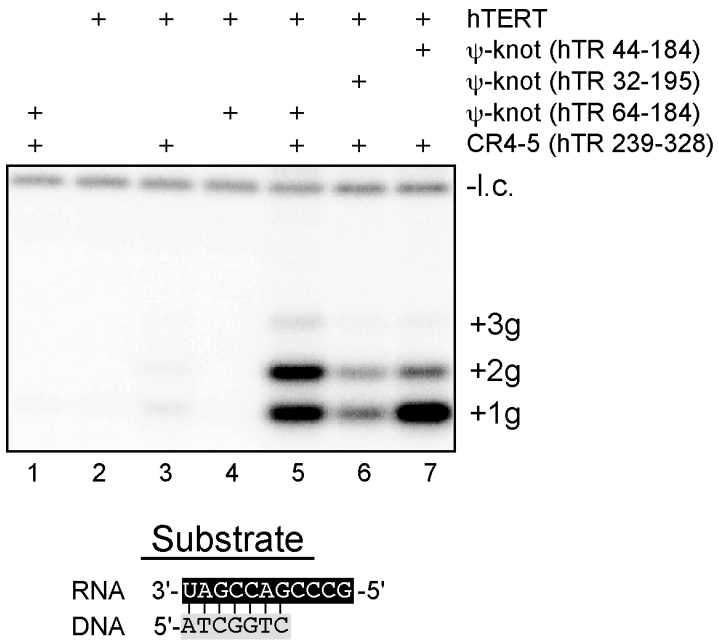
To reconstitute a human template-free telomerase, we assembled hTERT with essential hTR fragments omitting the template sequence, hTR pseudoknot nt 64-184 (Fig. 3.1a) and CR4-CR5 239-328, in Rabbit Reticulocyte Lysate (RRL). As expected, the template free telomerase failed to utilize single stranded telomeric DNA primer, while the telomerase with template show characteristic 6 nucleotides ladder pattern (Fig. S3.1). We then tried to test the ability of telomerase with or without intrinsic template sequence to use RNA/DNA duplex substrates (Fig. 3.1a. compare hTR 32-195, 44-184 and 64-184). In order to prevent potential base pairing between the intrinsic template and the DNA oligo in the duplex, we designed a non-native DNA sequence that would not be recognized by wild-type telomerase (Fig. 3.1b). The duplex was set to be 7 base pairs long according to the estimation from recently published structural and biochemical data that telomerase catalytic core can accommodate 7 base pairs between the RNA template and the DNA primer (17,18). When supplied with RNA template pre-annealed to the DNA primer, template free telomerase extended the primer by incorporating three dG

residues (Fig. 3.1b, lane 5). The observed activity is telomerase dependent, since neither hTR nor hTERT alone within RRL reacts on the RNA/DNA duplex (Fig. 3.1b, lane 1 and 2). Both essential hTR fragments are required for this activity. Assembling either pseudoknot or CR4-CR5 fragment with hTERT could not yield comparable level of activity as the template free telomerase (Fig 3.1b, lane 3 and 4). Compared to template free telomerase, the telomerase with template tethered by both flanking linkers (hTR 32-195, Fig. 3.1a) yielded 50% activity (Fig. 3.1b, lane 6). It is possible that the wild-type template held in place by the linkers reduced the accessibility of the activity site to the RNA/DNA duplex. In support with this notion, when the 5' linker was removed (hTR 44-195, Fig. 3.1a), the telomerase could utilize the duplex as efficiently as template-free telomerase (Fig. 3.1b, lane 7).



a. Schematic of hTR pseudoknot fragments

Fig 3.1 Template free human telomerase reacts on RNA/DNA duplex. a, Secondary structure of hTR pseudoknot fragments, hTR nt 32-195, 44-184 and 64-184, are presented. The RNA template sequences are highlighted with black background.

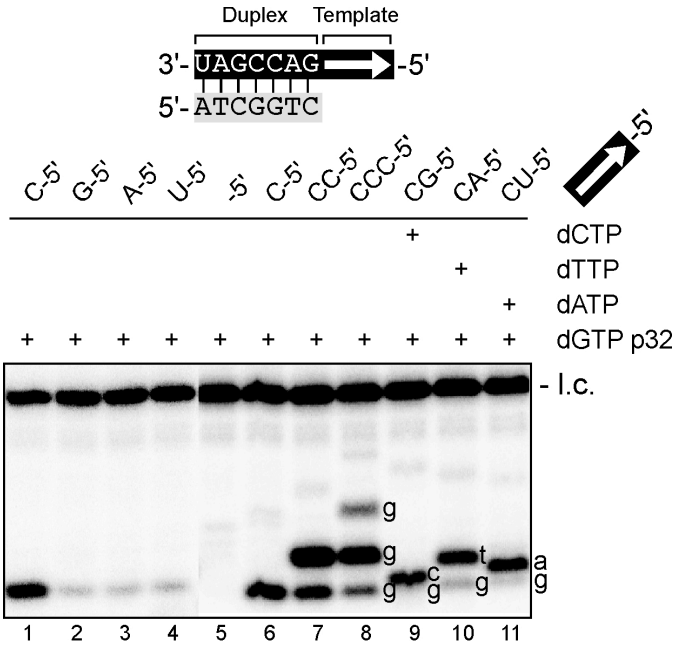


b. Template free human telomerase reaction

Fig 3.1 The template free human telomerase reacts on RNA/DNA duplex. b, Three different hTR pseudoknot fragments shown in Fig 3.1a are assembled with the CR4-CR5 fragment (hTR 239-328) and assayed for activity with pre-annealed RNA/DNA duplex substrate. The substrate (7 bp duplex) sequence is shown in the lower panel. The number of dG residues added is labeled on the right of the gel.

It has been shown that telomerase possesses a terminal transferase activity under specific conditions (27). We asked if the observed activity from template free telomerase is a terminal transferase activity. RNA/DNA duplex substrates with various template sequences and lengths attached to the same duplex stem were tested with template free telomerase. The α ^{32}P is incorporated at a much lower efficiency when the template is rG, rA or rU (Fig. 3.1c, compare lane 1 to lane 2, 3, 4). There is no nontemplated incorporation detected when RNA template protruding is not available on the duplex (Fig. 3.1c, compare lane 5 and 6). Furthermore, the extension of DNA oligo strictly followed the sequence presented in the RNA template. For example, the templates range

from one rC to three rCs directed the synthesis of one dG to three dGs. (Fig. 3.1c, lane 6 to 8) The rG, rA and rU templates lead to the incorporation of dC, dT and dA respectively (Fig. 3.1c, lane 9 to 11). The oligo products with same length did not show the same mobilities in the high percentage PAGE gel due to different molecular weight and electric properties of the residues added (compare lane 7, 9, 10 and 11). Collectively, these data suggested that template free telomerase conducts reverse transcription upon the RNA/DNA template rather than terminal transferring ^{32}P dGMPs.

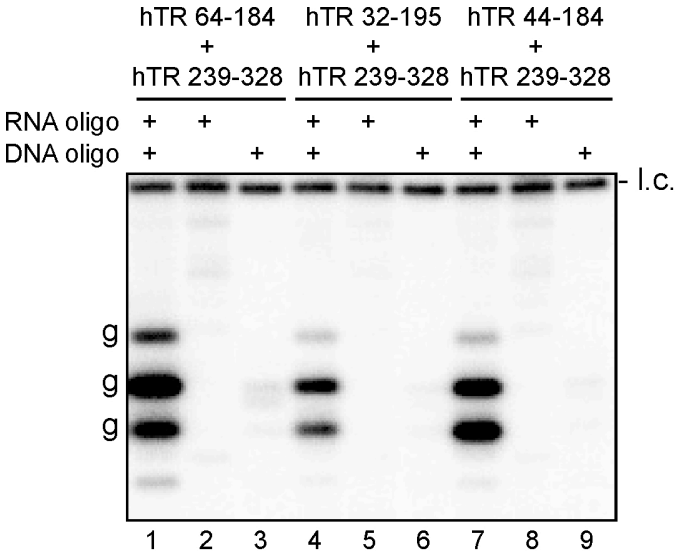


c. Template free telomerase reaction according to various template sequences.

Fig. 3.1 The template free human telomerase reacts on RNA/DNA duplex. c, Template free telomerase was assayed with duplexes having the same stem but different template sequences, ranging from 0 to 3 residues. The template residues are shown on the top of the gel. Different nucleotide substrate combinations were used as indicated. The ^{32}P dGTP is the only radioactive substrate. l.c.: ^{32}P end labeled 15nt DNA

3.4.2 Template free telomerase recognizes duplex substrate and releases duplex product

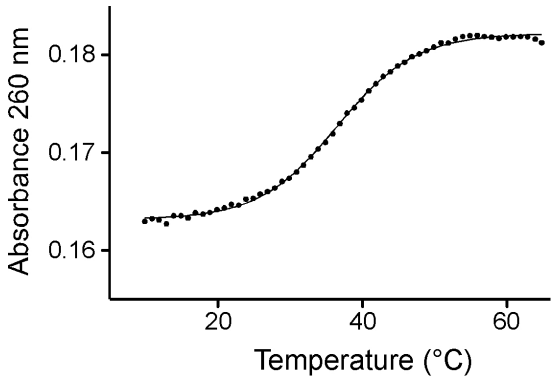
Even though the supplied substrates are presumably pre-annealed RNA/DNA duplex under our reaction condition, we tested whether the template free telomerase recognizes traces of single stranded RNA and DNA respectively and generate RNA/DNA duplex substrate within its active site. The following experiments were carried out to address this possibility. First of all, all three telomerase used showed activity only when supplied with both RNA and DNA oligos, but not with either one of the oligos (Fig. 3.2a, compare lane 1, 4 and 7 to lane 2, 3, 5, 6, 8 and 9).



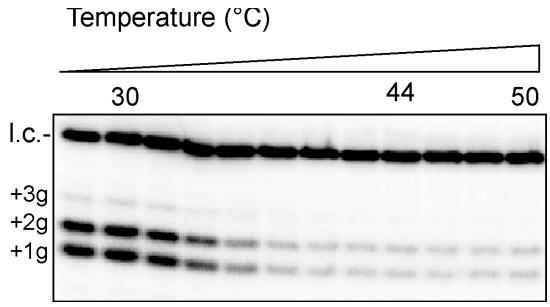
a. Template free telomerase requires both RNA and DNA strands that can form the duplex

Fig. 3.2 Template free telomerase utilizes duplex substrates and releases duplex product. a, Three pseudoknot constructs, hTR 64-184, 32-195 and 44-184, as shown in Fig. 3. are assembled with TERT and hTR CR4-CR5 domain (hTR 239-328). The reaction substrate was either 7bp duplex or the single stranded RNA or DNA, as indicated on the top of the gel. l.c., ^{32}P end labeled 15 nt DNA.

To insure the RNA and DNA oligos pre-annealed as a heteroduplex before telomerase reaction, we tested the UV melting curve of the duplex under the telomerase reaction condition (Fig. 3.2b). With the *Tm* around 40°C, most oligos should be annealed as heteroduplex under our reaction temperature at 25°C. Lastly, we carried out a series of template-free telomerase reactions under temperatures ranging from 25 to 50°C (Fig. 3.2c). At the higher temperatures (40-50°C), where RNA and DNA oligos are single stranded, the telomerase activity significantly decreased. The wild-type telomerase however is active at the temperatures up to 55°C (Fig. 3.2d), indicating the failure of reaction is due to the concentration of substrates in the duplex form, instead of the inactivity of the enzyme at high temperature. These data indicate that the RNA/DNA duplex, rather than single stranded oligos, are the substrate for template free telomerase.

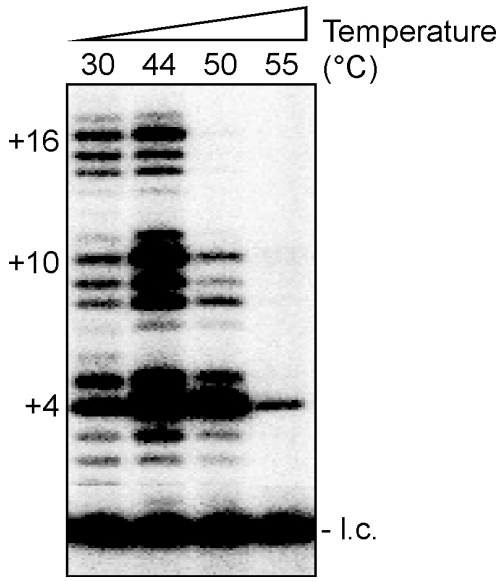


b. UV melting curve of RNA/DNA duplex 7 base pairs



c. Temperature gradient reactions of template free telomerase reaction

Fig.3.2 b, The OD260 of 7 bp duplex substrate from 5 to 65°C. Data points are connected with sigmoid dose response trend line. c, Template free telomerase using duplex 7bp was reacted under temperatures ranging from 25 to 50°C. Specific temperature, such as 30, 44 and 50°C, are labeled for comparison purpose.



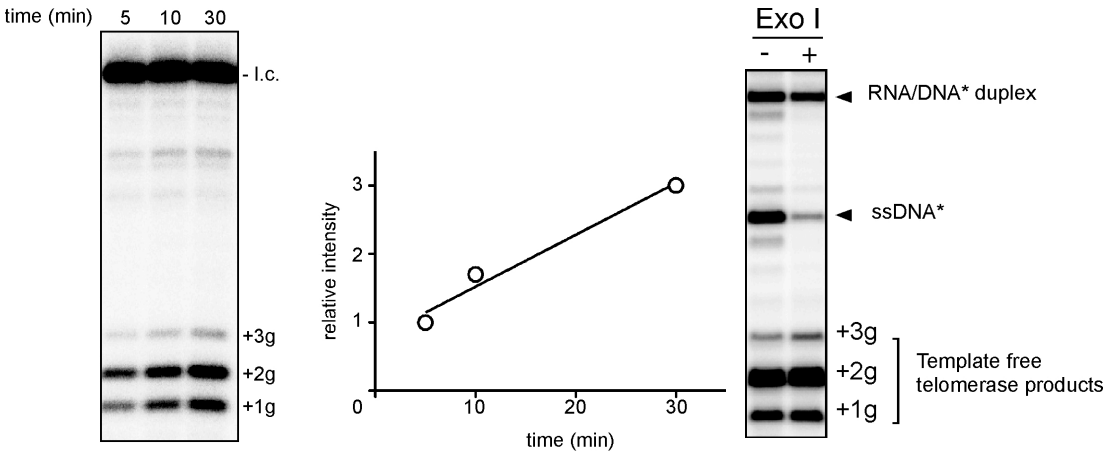
d. Temperature gradient reactions of wild-type telomerase reaction

Fig. 3.2 d, wild-type telomerase was assayed with single stranded telomeric primer $(TTAGGG)_3$ under different temperatures as indicated at the top of the gel. Major repeats of telomeric DNA extension product are labeled on the left of the gel, showing numbers of nucleotides added in each repeat. l.c.: loading control of end labeled 15 nt DNA oligo.

To confirm our findings using telomerase reconstituted in cells, we overexpressed hTR and hTERT to assemble "super-telomerase" as described in Cristofari et. al. (9). Instead of using full length hTR, we designed three mutated hTR genes corresponding to the three constructs used in the *in vitro* reconstituted telomerase (Fig. 3.1a). Northern blotting analysis indicated correct processing and equal levels of different hTR constructs (Fig. S3.2). We reasoned that the super-telomerase cell extract might contain other polymerases to react on RNA/DNA duplex, or single stranded oligos. An anti-FLAG immuno-precipitation was performed prior to telomerase reaction, since our hTERT construct contained a 3X FLAG tag at the N terminus. When supplied with 7 bp RNA/DNA substrate, cell reconstituted telomerase yielded similar results as *in vitro*

reconstituted telomerase. Template free telomerase (hTR 62-451) is more active than the telomerases with hTR tethered in the active site (hTR 1-451) (Fig. S3.2, compare lane 8 to lane 11). Addition of dCTP in the reaction result in one more nucleotide incorporation compared to ^{32}P dGTP only reaction (Fig. S3.2, compare lane 7 and 8). This result indicates template free telomerase reconstituted *in vivo* also reverse transcribed upon RNA/DNA duplex.

We next set out to investigate whether template free telomerase releases the product in a duplex form or as single stranded RNA and DNA oligos. It is tempting to think that telomerase would unwind the duplex within the active site after repeat synthesis because wild-type telomerase has to undergo a template translocation event to permit a processive reaction (28). We first conducted a time course reaction to confirm that template-free telomerase has enzyme turnover (Fig. 3.2e). The E. coli Exonuclease I (Exo I), which specifically cleaves single stranded DNA from 3' to 5' end, was mixed with the telomerase reaction. If telomerase releases its products as single stranded oligos, Exo I would digest the departed product DNA oligo. Our result showed that telomerase products as well as a DNA control pre-annealed with RNA oligo remain intact, while a single stranded DNA control was significantly digested after incubation (Fig. 3.2f). This data suggests that template-free telomerase can recognize substrates and release products in the duplex form.



e. The template free telomerase reaction has turnover

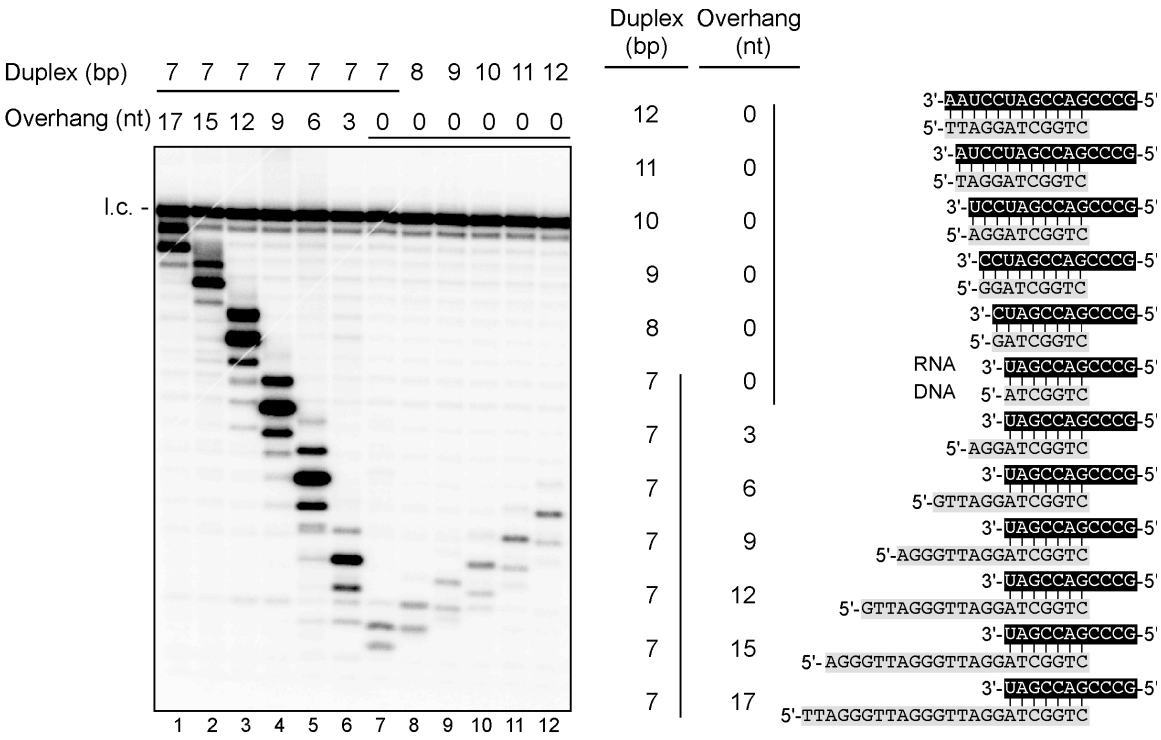
f. Exonuclease I digestion of template free telomerase reaction products

Fig. 3.2 e, (left panel) Template free telomerase reaction was carried out from 5 to 30 mins. Reaction products of 3 g extension are labeled on the right of the gel. (right panel) The relative intensity of the reaction product was plotted against time. f, template free telomerase reaction was carried out with or without Exonuclease I. RNA/DNA* duplex indicates the DNA was end labeled and serve as a control in double stranded form. RNA oligo: 5'-GCCCGACCCUAACUGA-3' DNA 5'-^{32}P-TTAGGGTTAGGGTTAGGTCAGTTAGGGTC-3' ssDNA*: an end labeled single stranded DNA control: 5'-^{32}P-TTAGGGTTAGGGTTAGGATCGGTC-3'. Template free telomerase substrate, RNA oligo: 5'-GCCCGACCGAU-3'; DNA oligo: 5'-AGGGTTAGGATCGGTC-3'.

3.4.3 Primer extension activity of template free telomerase using various duplex substrates

3.4.3.1 Duplex with 5' DNA overhang

Telomerase contains a TEN domain which binds to the 5' end of the single stranded telomeric DNA (11). We extended the 5' end of the DNA oligo by 3 to 17 nts of telomeric sequence overhanging the 7 bp duplex to test how this TEN-DNA interaction affects the template-free telomerase reaction. The 5' overhang greatly enhanced the overall activity even when it is only 3 nucleotides long (Fig. 3.3a, compare lane 1 to 6 with lane 7). The nucleotide addition processivity has also increased, as the ratio between +3g/+1g band gradually increased along with the increase of overhang length. This is probably due to the existence of the proposed proximal and distal DNA binding sites, including TEN domain, in the TERT protein (10,19). To test if the overhang increased the affinity between the enzyme and the substrate, we measured the apparent *Km* of 7 bp duplex with or without a 12 nt overhang. The overhang significantly decrease the *Km* by 30 fold compared to the duplex without the overhang, which has an apparent *Km* about 100 μM (Fig. S3.3). This phenomenon is consistent with the fact that telomerase has evolved specific motifs to recognize the single stranded DNA substrate, while the recognition of duplex substrate at the catalytic core has a much lower affinity to allow the processive telomerase reaction.

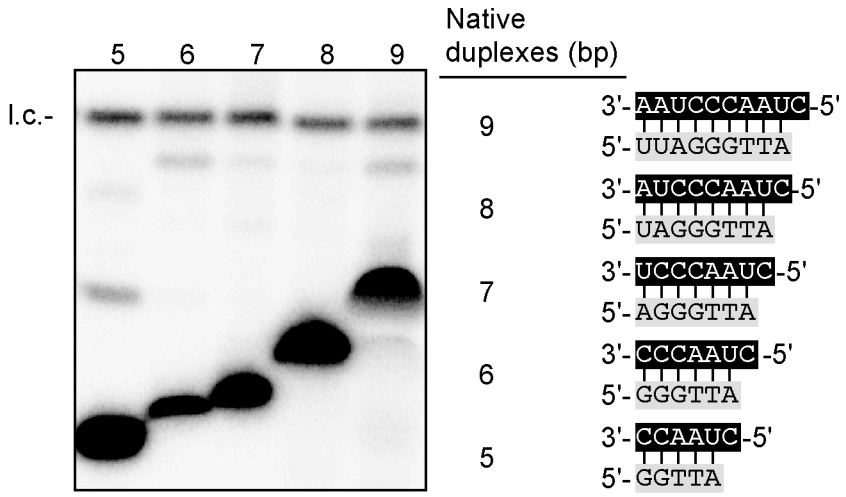


a. Template free telomerase reaction with various duplexes with different stem lengths and 5' DNA overhang lengths

Fig. 3.3 Template free telomerase reaction on different duplex substrates. a, (left panel) the activity assay of template free telomerase on substrates with different stem length and 5' DNA overhang length. (right panel) schematic of all the duplexes used in this study. The RNA sequences are highlighted with black backround while the DNA sequences are highlighted with grey background.

3.4.3.2 Various duplex lengths

When the wild-type human telomerase is performing reverse transcription, the potential base pairs between the RNA template and the DNA primer can be from 5 to 11. Therefore, we designed a set of duplexes to test if template free telomerase has length requirement for the duplex substrates. We first used the 7 to 12 base pairs non-native duplexes, which all have the same duplex sequence in the 3' DNA and 5' RNA portion. These duplexes all have higher Tm compared to duplex 7, most of which form duplex under our reaction condition. In this case, any reaction defect would represent inability of the telomerase to use available duplexes. Interestingly, all duplexes yield similar levels of activity (Fig 3.3a, lane 7 to 12). Template-free telomerase also show comparable level of activity with a 16 bp duplex, suggesting the active site of telomerase doesn't discriminate long duplexes. We then test if telomerase can use shorter duplexes with native sequence that are only 5 to 9 base pairs long. Due to the low *Tm* of these duplexes under standard conditions, we conducted the reaction at 4°C with higher concentration of K^+ and oligos to significantly increase the Tm (see Materials and Methods). Using these extreme conditions, we clearly show that template free human telomerase efficiently use native sequence duplex ranging from 5 to 9 base pairs (Fig. 3.3b).



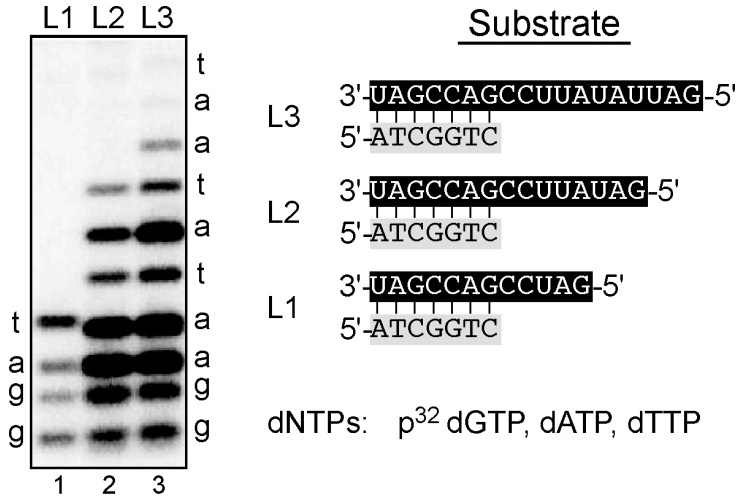
b. Template free telomerase reaction with telomeric duplexes ranging from 5 to 9 base pairs

Fig 3.3 b, (left panel) template free telomerase was assayed with 5 different telomeric duplexes from 5 to 9 base pairs. The reaction was carried out under a condition that has high salt and high duplex substrate concentration to assure short duplex formation (see Materials and Methods). (right panel) schematic of the duplex substrates used in this study.

3.4.3.3 Different RNA template lengths

Upon binding to the RNA/DNA duplex, conventional RTs can reverse transcribe a long strand of DNA along the RNA template. It is interesting to see how long telomerase can extend when a long RNA template is available. Template-free telomerase was able to reach the end of the template when it is 4 or 7 nt long (Fig. 3.3c, lane 1 and 2). However, the extension only goes up to 8 nt when the template is 11 nt long (Fig. 3.3c. lane 3). We also tested a different sequence and longer template and still observed an 8 nt extension (Fig. S3.4). All four nucleotides were included in our test, addition of dA seems to be less processive than the others, indicated by stronger stops in Fig. 3C lane 3. Nonetheless, the nucleotide addition processivity of all residues is low despite the template sequence.

Pausing/dissociation occurred at each position when a nucleotide was added. This processivity is rather low compared to retroviral RTs, such as HIV RT, which can synthesis 300 nucleotides along rA template without dissociation (29). Therefore, telomerase might have evolved with a special mechanism to use the short template in processive reaction cycles to add a large number of DNA repeats.

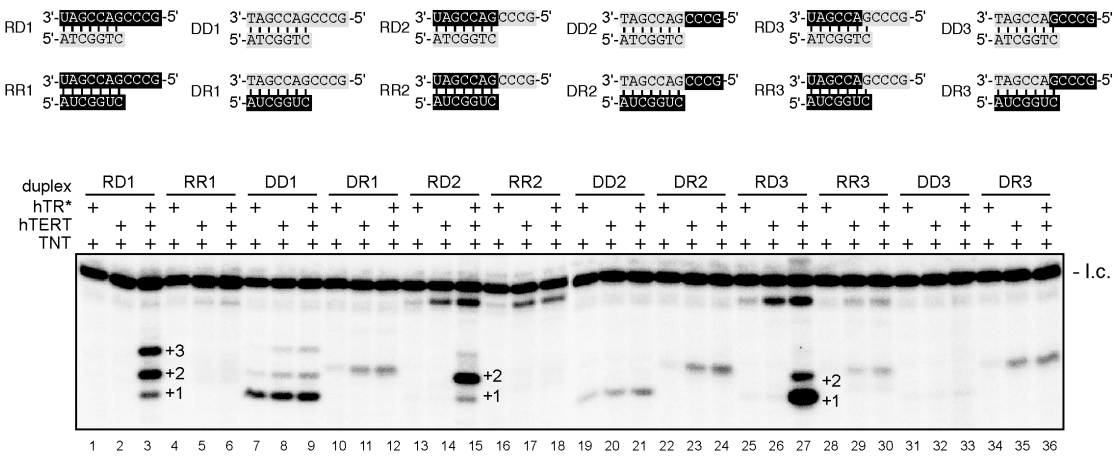


c. Template free telomerase reaction with substrates that have long RNA templates

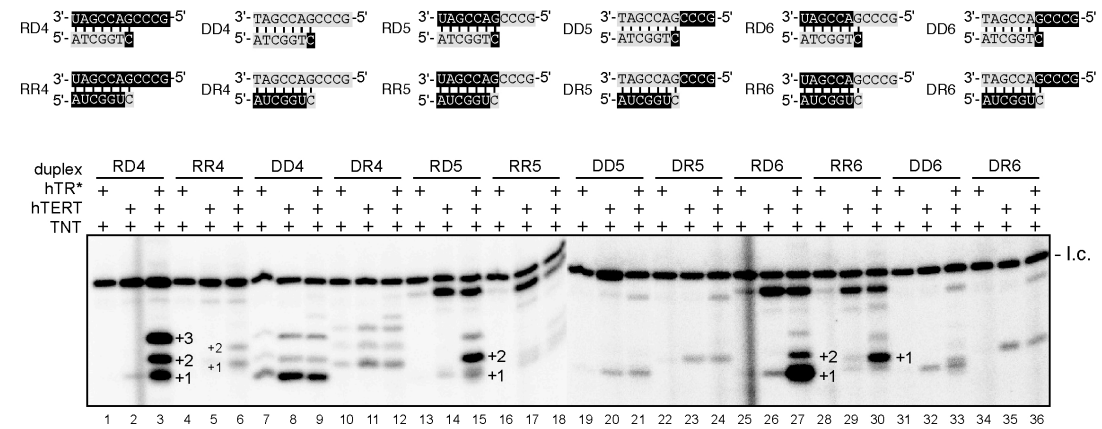
Fig. 3.3 c, (left panel) Template free telomerase extends duplex with long RNA templates (L1: 5 nt, L2:8 nt, L3:11 nt). The nucleotides added to the DNA are indicated on the sides of the gel. (right panel) Schematic of the substrates used in this reaction. Three dNTP substrates were used, with only ^{32}P dGTP as radioactive.

3.4.4 Comparison of different hybrids as substrates for template free telomerase

Telomerase has exhibited limited activity to act as an RNA-dependent RNA polymerase (RdRP) and a DNA-dependent DNA polymerase (23, 24). Conventional RTs also have both robust RNA-dependent and DNA dependent DNA polymerase activity. The RT domain of telomerase is highly conserved compared to other RTs, suggesting that the above mentioned polymerization activities could be performed by telomerase RT. Testing this idea was previously hampered by the nature of telomerase in that it carries an intrinsic RNA template. Taking advantage of the template free telomerase system, we directly compared the ability of telomerase to use RNA/DNA, RNA/RNA, DNA/DNA and DNA/RNA hybrids. All the hybrids were designed with the same sequences of ribo- or deoxyribo- nucleotides (Fig. 3.4a, RD1, RR1, DD1 and DR1). Template free telomerase was reconstituted *in vitro* followed by an anti-FLAG IP, in order to prevent background RNA or DNA polymerase activity from the TNT lysate (Promega manufacturer handbook). Surprisingly, only RNA/DNA duplex gave rise to telomerase specific reaction products (Fig 3.4a, lane 3).



a. Template free telomerase reaction with homo- or hetero- RNA/DNA duplexes and chimeric duplexes



b. Template free telomerase reaction with duplexes priming with chimeric oligos

Fig. 3.4 Template free telomerase reaction on different combinations of hybrid duplexes. The substrate duplexes are represented as two letter symbols plus one number. For example: RD1=RNA/DNA hybrid no.1, DR2=DNA/RNA hybrid no.2. The exact sequences are shown on the top of the gel. As previously stated, black background represents ribonucleotides, while the grey background represents deoxyribonucleotides. The oligo with both grey and black backgrounds stands for a chimeric oligo. The substrates were assayed with either template free telomerase RNA (hTR*), hTERT or the template free RNP complex. The telomerase specific addition of nucleotides is labeled as "+1, +2, +3" and so on.

We then swapped the template overhang in different hybrids, e.g., putting DNA template to RNA/DNA duplex stem and RNA template to DNA/DNA duplex stem (Fig. 3.4a, RD2, RR2, DD2 and DR2). In the third set of duplex hybrids, the chimerical sequence of templates was extended into the duplex stem. For example, the RD3 substrate will have a DNA-DNA base pair at the 3' end of the DNA, and the template overhang is all deoxyribonucleotides (Fig. 3.4a). Again, only RD2 and RD3 can be used by telomerase (Fig. 3.4a, lane 15 and 27). These data suggested that the duplex stem is the essential factor for telomerase recognition, only the RNA/DNA hybrid can be used. The nature of the template and the last base pair of the duplex stem are not important. Comparing the reaction pattern of three different RNA/DNA hybrids, the conclusion that DNA/DNA stem is not preferred by telomerase active site can be further supported. The RD1 produces three dG along three rC templates (Fig. 3.4a, lane 3). The RD2 only allowed two dG incorporated along two dC templates (Fig. 3.4a, lane 15), potentially because the telomerase active site does not allow more than two DNA/DNA base pairs in the hybrid. This is further supported in the RD3 substrate, where the duplex already contains a dG/dC base pair in the duplex stem. Addition of only one dG, making two DNA-DNA basepairs in the stem, strongly seized the extension (Fig. 3.4a, lane 27).

Both RNA and DNA can prime the RNA dependent DNA polymerase reaction for retroviral RTs (30). We thus try chimeric primer oligos in our duplexes to see if different residues prime the reaction differently. These chimeric oligos were base-paired with those chimeric template oligos used in the first set, forming 12 new chimeric hybrid duplex substrates (Fig. 3.4b). As expected, with the preferred RNA/DNA duplex

substrates, ribonucleotide residue primes the extension as efficiently as dexoyribonucleotide residues (compare lane 3, 15 and 27 from Fig. 3.4a and 3.4b). Surprisingly, two RNA/RNA hybrids become more favorable substrates when the last nucleotide in the primer oligo was changed from ribo- to deoxyribo- nucleotide (compare lane 6 and 30 from Fig. 3.4a and 3.4b). This suggests that although both nucleotides can prime DNA elongation, deoxyribonucleotides are still more favorable. This also indicates that RNA/RNA duplex is the second most favorable duplex substrate after RNA/DNA duplex.

Lastly, we tried NTPs and α ^{32}P GTP as building blocks with the duplexes we used in Fig. 3.4a. However, none of the duplexes show telomerase specific activity (Fig. S3.5), indicating the RNA polymerase activity of telomerase is much weaker than DNA polymerase activity.

3.5 Discussion

Telomerase is believed to be a single strand specific reverse transcriptase carrying an intrinsic template within its RNA moiety. Both the protein and the RNA components of the telomerase enzyme have been specifically evolved for utilizing single stranded telomeric DNA. TERT protein has evolved single stranded DNA binding motifs to interact with telomeric DNA (19). The telomerase RNA template is usually complementary to the telomeric DNA and is usually longer than one repeat (20). This helps with recognizing the telomeric ends by RNA/DNA base pairing. The other RTs do not have an intrinsic RNA template and they have to use duplex substrates. It is believed that telomerase is evolutionarily related to other reverse transcriptases, because it

contains the seven signature protein motifs in the TERT responsible for RT reactions. Due to the existence of the intrinsic RNA template in the TERT active site, the possible ability of the telomerase enzyme to react on duplex substrates has therefore been overlooked in the past.

In this study, we have specifically designed a template free telomerase to assay the enzyme's ability of using duplex substrates. Our biochemical analysis of template free telomerase has proven that the catalytic core of the telomerase enzyme retains the ability to use RNA/DNA duplex as other RTs (Fig. 3.1). The existence of the intrinsic RNA template does reduce the accessibility of the catalytic core to the duplex substrates. Both ends of the RNA template are tethered by linkers between template to P1 stem (5'end) and to P2a stem (3'end) (Fig 3.1b). When the RNA template is not tethered, the template free telomerase reaction on duplex substrate has an increased level of activity. The weak signal detected with hTERT assembled with hTR CR4-CR5 only indicates these two components are the minimal requirements for an active telomerase (Fig.3.1c). In the wild-type telomerase, the function of the pseudoknot fragment might be providing the template sequence and enhancing the enzyme activity, rather than supplying the catalytic property (4).

We have also provided evidence indicating that the telomerase catalytic core releases the product as a RNA/DNA duplex (Fig. 3.2e. 2f.) Although template free telomerase is a heavily modified artificial system, we believe that the duplex binding and releasing shed a light on understanding the natural telomerase reaction cycle. During a translocation event of a telomerase reaction, the RNA template and DNA product need to

dissociate and re-anneal for the next round of repeat synthesis. This duplex dissociation could happen within the catalytic core or elsewhere after being released from the catalytic core. Our data suggested that the strand separation between RNA template/DNA product hybrid happens outside of the catalytic core. However, we could not rule out the possibility that, in a real telomerase reaction cycle, long overhang of DNA primer and the tethering of the RNA template provide an anchor site for the interaction between duplex and the TERT protein. When a conformational change occurs during translocation, the RNA/DNA hybrid got pulled apart due to their interaction to protein respectively.

The ability of template free telomerase to extend different duplexes with various lengths can be explained by the recently solved *Tribolium* TERT crystal structure (Fig. 3.3a) (17). The beetle TERT resembles a ring structure with the catalytic site sitting in the center of the ring. The RNA/DNA duplex is possibly positioned in the center of the ring for catalytic reaction. Although human TERT contains an extra TEN domain compared to beetle TERT, it is highly possible that the ring structure also exists in human TERT. Thus, as long as the 3' of the DNA in the RNA/DNA duplex can be positioned in the active site, the length of the duplex substrate can be variable.

Our test of using GTP as substrates for template free telomerase is somewhat contradictary to the recent finding about RdRP activity from hTERT (25). The different reaction conditions and substrate nature, such as salt concentration, duplex substrate versus potential hairpin substrate and TNT reconstituted hTERT versus *E. coli* expressed hTERT, might explain the different results obtained. Nonetheless, it is reasonable that telomerase cannot use NTPs as efficiently as dNTPs, thus insuring the sequence integrity

of chromosome ends.

3.6 References

1. Meyerson, M., Counter, C. M., Eaton, E. N., Ellisen, L. W., Steiner, P., Caddle, S. D., Ziaugra, L., Beijersbergen, R. L., Davidoff, M. J., Liu, Q., Bacchetti, S., Haber, D. A., and Weinberg, R. A. (1997) *Cell* 90, 785-795

2. 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) *Science* 266, 2011-2015

3. Chen, J. L., and Greider, C. W. (2004) *Proc Natl Acad Sci U S A* 101, 14683-14684

4. Qiao, F., and Cech, T. R. (2008) *Nat Struct Mol Biol* 15, 634-640

5. Collins, K. (2008) *Mech Ageing Dev* 129, 91-98

6. Tesmer, V. M., Ford, L. P., Holt, S. E., Frank, B. C., Yi, X., Aisner, D. L., Ouellette, M., Shay, J. W., and Wright, W. E. (1999) *Mol Cell Biol* 19, 6207-6216

7. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) *Mol Cell* 17, 671-682

8. Leeper, T., Leulliot, N., and Varani, G. (2003) *Nucleic Acids Res* 31, 2614-2621

9. Cristofari, G., and Lingner, J. (2006) *EMBO J* 25, 565-574

10. Autexier, C., and Lue, N. F. (2006) *Annu Rev Biochem* 75, 493-517

11. Jacobs, S. A., Podell, E. R., and Cech, T. R. (2006) *Nat Struct Mol Biol* 13, 218-225

12. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2000) *Mol Cell* 6, 493-499

13. Moriarty, T. J., Huard, S., Dupuis, S., and Autexier, C. (2002) *Mol Cell Biol* 22, 1253-1265

14. Peng, Y., Mian, I. S., and Lue, N. F. (2001) *Mol Cell* 7, 1201-1211

15. Huard, S., Moriarty, T. J., and Autexier, C. (2003) *Nucleic Acids Res* 31, 4059-4070

16. 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) *Science* 277, 955-959

17. Gillis, A. J., Schuller, A. P., and Skordalakes, E. (2008) *Nature* 455, 633-637

18. Forstemann, K., and Lingner, J. (2005) *EMBO Rep* 6, 361-366

19. Finger, S. N., and Bryan, T. M. (2008) *Nucleic Acids Res* 36, 1260-1272

20. Chen, J. L., and Greider, C. W. (2003) *EMBO J* 22, 304-314

21. Abbotts, J., Bebenek, K., Kunkel, T. A., and Wilson, S. H. (1993) *J Biol Chem* 268, 10312-10323

22. Telestisky & Goff (1997) Reverse transcriptase and the generation of retroviral DNA. in book *Retroviruses*

23. Collins, K., and Greider, C. W. (1995) *EMBO J* 14, 5422-5432

24. Legassie, J. D., and Jarstfer, M. B. (2005) *Biochemistry* 44, 14191-14201

25. Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W. C., and Masutomi, K. (2009) *Nature* 461, 230-235

26. Miller, M. C., and Collins, K. (2002) *Proc Natl Acad Sci U S A* 99, 6585-6590

27. Lue, N. F., Bosoy, D., Moriarty, T. J., Autexier, C., Altman, B., and Leng, S. (2005) *Proc Natl Acad Sci U S A* 102, 9778-9783

28. Greider, C. W. (1991) *Mol Cell Biol* 11, 4572-4580

29. Huber, H. E., McCoy, J. M., Seehra, J. S., and Richardson, C. C. (1989) *J Biol Chem* 264, 4669-4678

30. Freed, E. O. (2001) *Somat Cell Mol Genet* 26, 13-33

CHAPTER 4

RNA/DNA DUPLEX BINDING IS AN ESSENTIAL STEP DURING TEMPLATE TRANSLOCATION

4.1 Abstract

Telomerase reverse transcriptase (RT) is unique in that it recognizes single stranded DNA substrates and synthesizes telomere repeats onto the chromosome termini using its intrinsic RNA template. RNA/DNA duplex binding has been implicated in template translocation efficiency thus contributing to telomerase processivity (Chapter 3). Here, we use the template free telomerase developed previously to react with telomeric RNA/DNA duplex substrates. Mutations within telomerase reverse transcriptase (TERT) that affect repeat addition processivity correlate with the enzyme's affinity toward duplex substrate, with the exception that N-terminal TERT mutant reduced processivity by hampering single stranded DNA binding. Telomerase catalytic core favors 5 and 6 base pairs duplex substrates. In consistency, a 5 bp duplex but not 7 bp duplex inhibits telomerase processivity in a wild-type telomerase reaction. Moreover, circular permutation of 7 base pair duplex indicates that telomerase active site has sequence preferences towards duplex substrates. Detailed nucleotide substitution revealed that two rA-dT basepairs within the telomeric duplex are responsible for the sequence specificity. Our data have not only demonstrated the catalytic core of telomerase specifically evolved for the recognition of RNA/DNA duplex with a certain length, but also indicated that duplex binding is an essential step during template translocation, which determines telomere repeat addition processivity.

4.2 Introduction

Human telomerase is capable of adding multiple telomere repeats upon one primer binding event. This repeat addition processivity relies on a template translocation mechanism whereby the RNA template dissociates and realigns relatively to the DNA substrate. After translocation, a 5 base-pair RNA/DNA duplex would form in the catalytic site of telomerase, with the 6 nt RNA template available for the next repeat synthesis. Theoretically, the RNA/DNA duplex inside the catalytic core can be extended up to 11 base pairs. The processive addition of telomeric DNA can be attributed to the intrinsic property of the telomerase core components. The realignment region and template boundary element in hTR facilitate template translocation by supplying a relocation site for the 3' end of the DNA and avoiding incorporation of non-telomeric sequence (1, 2). In the hTERT protein, mutations in Telomerase esstinal N-terminus (TEN), motif 3, Insertion in Finger Domain (IFD) and C-terminus extension (CTE) all affect telomerase processivity. While the TEN domain functions as an anchor site for upstream telomere substrate binding, motif 3-IFD-CTE might form a molecular clamp to facilitate duplex formation or binding during translocation (3).

To investigate if duplex binding is a key step during template translocation, we used template-free human telomerase designed in chapter 3. When supplied with pre-annealed telomeric RNA/DNA duplex as substrate, this system allows for direct investigation of telomerase's ability to bind the duplex. Here, we show that our template free telomerase reverse transcribed upon telomeric RNA/DNA substrate. More importantly, proposed molecular clamp mutations that decrease telomerase processivity

had lower affinity to duplex substrate compared to wild-type enzyme, while a mutant with augmented processivity shows higher affinity. Including a duplex substrate in wildtype telomerase reaction reduces its processivity severely, while single stranded RNA or DNA did not have the same effect. The TEN mutation decreased processivity through defective upstream binding, therefore leaving the duplex binding ability intact. Using this template free telomerase and circular permuted 7 base pair telomeric duplexes, we directly showed that telomerase preferred some of its native RNA/DNA duplex over other native sequences. These data support that duplex binding is a critical step during template translocation, thus contributing to telomerase repeat addition processivity.

4.3 Materials and Methods

4.3.1 Oligos and duplex substrate preparation

Deoxyribo- or Ribo- nucleotide oligos are ordered from Integrated DNA technology Inc. Detail sequence of each oligo can be found in figures. To assemble RNA/DNA substrates, the complementary oligos are mixed in indicated concentrations in 1X annealing buffer (100 mM Tris-HCl (pH7.5), 500 mM NaCl, 50 mM EDTA). The mixture was heated at 70°C for 3 min and gradually cool down in room temperature.

4.3.2 Tm measurement

Duplex substrates were supplied in a 250 µL volume at 20 µM final concentration in 1X telomerase reaction buffer (50 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$, 1 mM Spermidine) and 1X annealing buffer. The OD-260 of the solution was measured along the temperature change from 70 to -1°C at 1°C intervals using Cary 300 UV-vis spectrophotometer (Varian tech.) with multicell holder thermal controllable

accessory. Data points were fitted to sigmoidal curve (variable slope) to calculate the annealing temperature value, Tm (Prism, Graphpad Software Inc.).

4.3.3 *In vitro* and *in vivo* reconstitution of template free telomerase

See section 2.3.3

4.3.4 Template free telomerase activity assay

Activity assay was modified from previously described protocol (Chapter 3). Basically, 1.5 μL of in vitro reconstituted telomerase was assayed in a 10 μL reaction with 50 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$, 1 mM Spermidine, 0.165 μM α^{32}P-dGTP (3000 Ci/mmole, 10 mCi/mL, PerkinElmer), and 20 μM duplex substrate. The reaction was carried on for 60 min at 4°C or 25°C and terminated by phenol/chloroform extraction, then ethanol precipitated. The product was applied to 18% polyacrylamide gel, and analyzed by Phosphor Imager. For the duplex competition of telomerase processivity assay, wildtype telomerase was reacted in 10 μL volumne with 50 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$, 1 mM Spermidine, 1 mM dTTP, 1 mM dATP, 2 μM dGTP, 0.165 μM α^{32}P-dGTP (3000 Ci/mmole, 10 mCi/mL, PerkinElmer), and 1.5 μL 1μM $(GGTTAG)_3$ telomeric primer. Competition duplexes or oligos were added to 200 μM as final concentration. The reaction condition and duplex concentration used in *Km* measurement and turnover measurement assays were specified in the figure legends. For processivity competition assay, the telomerase activity assay was altered to include 200 μM of processivity inhibitory duplexes. The reaction was carried out under 4°C for 2 hours. Termination of

the reaction and data analysis is the same as conventional telomerase activity assay described above.

4.3.5 Immuno-precipitation (IP) of N-FLAG tagged telomerase

For each 50μL 293 FT cell extract, 20μL of anti-FLAG M2 affinity agarose beads (Invitrogen) was centrifuged 5000 xg for 30 sec. The beads were washed twice with 500 μL TBS buffer (50 mM Tris HCl, 150 mM NaCl, pH 7.4). 50 μL 293 FT cell extract was added to the washed beads allow for binding at 4°C with gentle agitation. After 2 hours of IP, the beads were washed 3 times with 500 μL TBS and 1 time with 1X PE buffer (telomerase reaction buffer).

4.3.6 Western blotting

Immuno-precipitated telomerase holoenzyme with NFLAG-tagged hTERT from 50μL 293 FT cell lysate was heated at 95°C for 5 min in 1X Laemmli buffer (0.125M Tris-HCl, pH6.8, 2% SDS, 10% glycerol, 5% 2-mercaptoethanol and 0.0025% bromophenol blue), fractionated on a 6% or 8% SDS-PAGE gel, and electro-transferred onto the PVDF membrane. Blocking (overnight at 4ºC) and incubation with antibodies (1 hour at room temperature) were carried out in 5% nonfat milk/1X TTBS (20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.05% Tween 20). Anti-hTERT goat polyclonal antibody L-20 (Santa Cruz Biotechnology) was used as the primary antibodies. After incubation with the HRP-conjugated secondary antibody (Santa Cruz Biotechnology), the blots were developed using the Immobilon Western Chemiluminescent HRP substrate (Millipore), and the blot images were acquired and analyzed using a Gel Logic440 system (Kodak).

4.3.7 Northern blotting

Immuno-precipitated telomerase holoenzyme from 50μL post-transfection 293 FT cells was phenol/chloroform extracted and ethanol precipitated. The RNA was resolved on a 4% polyacrylamide/8M urea denaturing gel and electro-transferred to the Hybond-XL membrane (GE Healthcare). Preparation of the riboprobes and hybridization of the blot were carried out as described in chapter 5.

4.3.8 Terminal deoxynucleotidyl transferase (TdT) reaction

100 nM of DNA oligo is mixed in 25 μL reaction with 25 mM NaCacodylate (pH7.2), 1 mM CoCl2, 0.1 mM DTT and 10 U of TdT (USB scientific) and incubated at 37°C for 1 min. Reaction was terminated by phenol/chloroform extraction and ethanol precipitated.

4.4 Results

4.4.1 Template free telomerase extends telomeric duplex better than non-telomeric duplex substrate

Although telomerase is a single strand specific reverse transcriptase, the catalytic site of this enzyme still retains the ability to utilize duplex substrates similar to other RTs (Chapter 3). Our previous study focused on a non-telomeric duplex sequences. Due to the relative conserved GT rich telomere sequence in all the species identified to date and the absolute conserved 5'-GGTTAG-3' telomeric repeat in vertebrates, we reasoned that telomerase active site might have evolved to suit specific telomeric sequence. Therefore, we carried out template free telomerase reaction with both non-telomeric and telomeric duplex substrates side by side. As illustrated in chapter 3, the template free telomerase is

reconstituted *in vivo* as "super telomerase" using 293 FT cell transfection system (4). Instead of using the full length hTR gene, we introduced a hTR lacking the first 63 nucleotides including the template sequence. Important domains for catalytic activity, the triple helix/pseudoknot domain and three way junction/CR4-CR5 (5, 6), remain in the construct. The Sno/ScaRNA domain, that is responsible for hTR in vivo biogenesis is also present in the template free telomerase hTR (Fig.4.1).

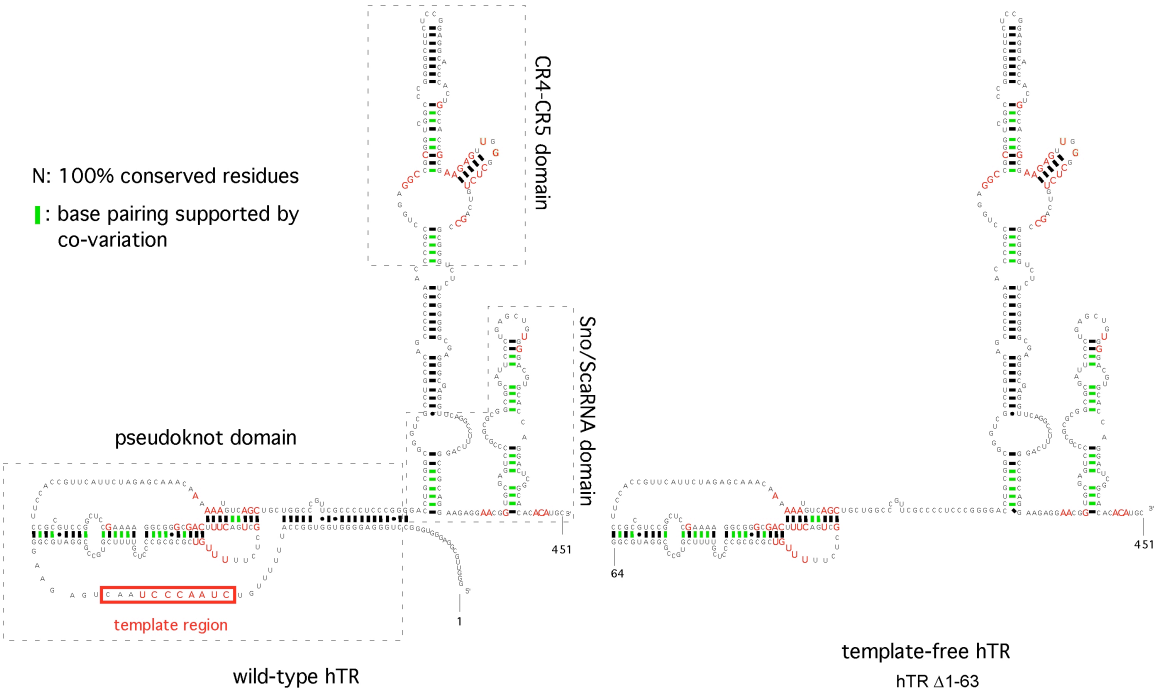


Fig.4.1 Secondary structure of human telomerase RNA and template free hTR construct. Full length hTR nt 1-451 (left) and template free hTR nt 64-451 (right) sequences are represented as secondary structures. The nucleotides highlighted in red stands for 100% conservation in all the vertebrate telomerase RNAs. The basepairs highlighted in green represent base pairing supported by phylogenetic comparative analysis. Three major domain structures are indicated in dash-line boxes in wildtype hTR structure. The starting and ending residues of the two hTR structures are labeled. For the *in vivo* reconstituted telomerase from 293FT cells, hTR 64-451 was introduced. For the *in vitro* reconstituted telomerase, template free pseudoknot domain together with CR4-CR5 domain were assembled with hTERT in Rabbit Reticulocyte Lysate to obtain active enzyme (see Materials and Methods).

The template free telomerase was extracted from 293 FT cells and immuno-precipitated with anti-FLAG agarose beads and then assayed for activity with both non-telomeric and telomeric sequence duplexes (see Materials and Methods). In consistence with the data presented in chapter 3, the activity observed is telomerase specific (Fig.4.2, lane 3, 4, 14 and 15). Firstly, 293 FT cell without over expression of neither hTERT nor hTR gene has no duplex extending activity (Fig. 4.2, lane 1 and 12). Secondly, transfection with only the hTERT gene didn't yield observable activity (Fig. 4.2, lane 2 and 13). Moreover, we have included a catalytic inactive hTERT D868N construct as a negative control. Over expression of hTERT D868N alone produced similar activity as hTERT only, indicating hTERT itself does not react on the duplex substrates (Fig. 4.2, compare lane 2 with 7 and 13 with 18). Thirdly, only when RNA and DNA oligos were presented together, namely, when duplex substrates are available, can telomerase extend the DNA oligo, thus excluding the possibility of terminal transferase activity (Fig. 4.2, lane 6, 7, 16 and 17). Lastly, the protruding RNA template sequence was designed to be 3'-CA-5', directing the consecutive synthesis of dG and dT residues. When dGTP and dTTP substrates are both included in the reaction, telomerase extend one more residue than dGTP only reaction (Fig. 4.2, compare lane 3 with lane 4 and lane 14 with lane 15). This result again indicated template free telomerase synthesized DNA product along the RNA template. The catalytic inactive construct did not show any comparable activity as the wild-type telomerase construct (Fig. 4.2, lane 8, 9, 19 and 20). Moreover, to assure the correct residues were incorporated into the DNA sequence, we included a terminal transferase reaction product on the same DNA oligos as control (lane 10 and 11). The

terminal deoxynucleotidyl transferase (TdT) incorporates dG residues onto the oligo, presumably having the first residue at the same position as the telomerase reaction product. Indeed, the two activities showed the same mobility in the polyacrylamide gel. As the second residue incorporated by TdT is still a dG, higher molecular weight than dT, we observed a lower mobility of product of TdT reaction compared to second band of the telomerase product. These data indicate that the correct residues were added into the DNA as predicted by RNA template.

To compare activity levels of non-telomeric and telomeric duplex sequences, the substrate used in chapter 3 and a newly design telomeric duplex with 5'-GGTTAGG-3' DNA sequence were compared. These two duplexes showed similar melting curves under UV melting Tm analysis. Thus, under our reaction condition in room temperature, the available amount of duplex substrates is equal. The telomeric duplex activity is 2 fold higher than the non-telomeric duplex.

To confirm the presence of both hTERT and hTR components in the complex that we pulled down with anti-FLAG agarose beads, western blotting and northern blotting analysis were performed (see Materials and Methods). The hTERT and hTR were present as expected (Fig.4.2, lower panel): when hTERT gene was transfected, protein was enriched in the IP'ed beads, and hTR Δ1-63 migrates faster than the full length construct (1-451). These data further confirmed that the template free telomerase reacts as a true RNP complex with the existence of both hTERT and hTR components.

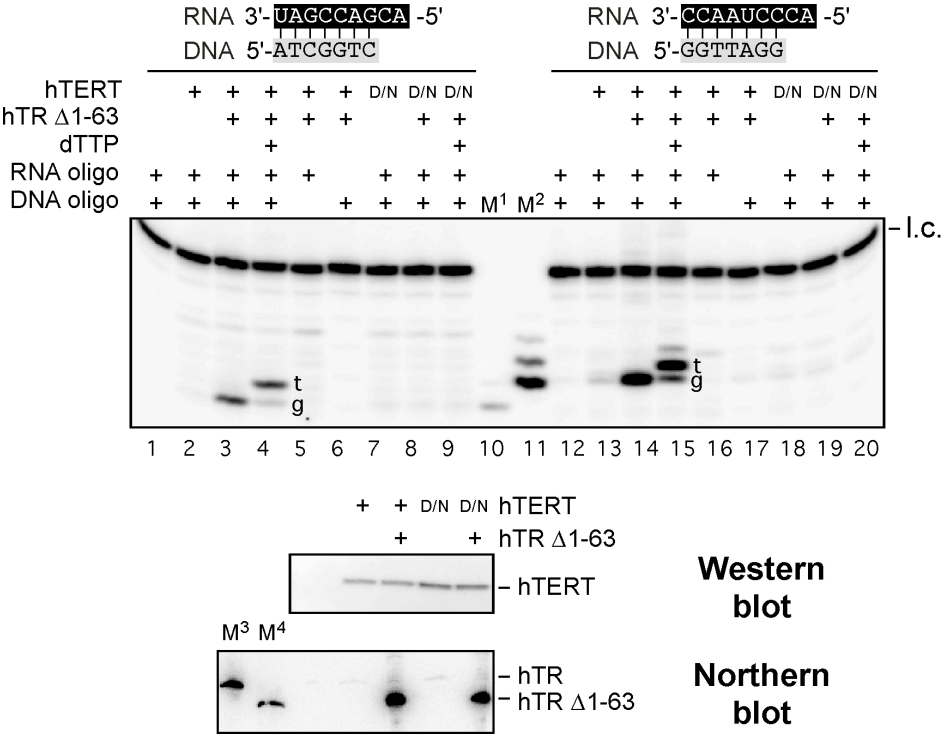


Fig. 4.2 Template free telomerase reaction with both non-telomeric and telomeric RNA/DNA duplexes. (Upper panel) Template free telomerase was reconstituted in 293 FT cells with over expression of hTERT and hTR Δ1-63 and assayed with non-telomeric and telomeric RNA/DNA duplexes (see materials and methods). Transfection with either the hTERT gene, hTR Δ1-63 gene or hTERT D868 catalytic inactive variant are indicated. Sequence of the duplex is indicated on the top of the gel. Characters with black background indicate RNA sequence, and characters with grey background denotes DNA sequence. Combinations of RNA oligo alone, DNA oligo alone or both, and presences of dTTP residue in the reaction are labeled on the top of the gel. M1: terminal transferase extension of DNA oligo: 5' ATCGGTC 3' as size marker. M2: terminal transferase extension of DNA oligo: 5' GGTTAGG 3'. (Middle panel) Western blot analysis of template free telomerase. Template free telomerase was immuno-precipitated by anti-FLAG agarose beads and immuno blotted with hTERT antibody (see Materials and Methods). l.c. A ^{32}P end labeled 15 nt DNA oligo as loading control. (Lower Panel) Northern blot analysis of IPed template free telomerase. Template free telomerase was immuno-precipitated by anti-FLAG agarose beads and hybridized with anti-hTR riboprobe (see Materials and Methods). M3, 1 ng *in vitro* transcribed full length hTR as size marker. M4, 1 ng *in vitro* transcribed hTR Δ1-63 as size marker.

4.4.2 Processivity defect mutants affected duplex binding

The RNA/DNA duplex binding was implicated in template translocation during wild-type telomerase reaction (Chapter 3). The mutations in motif 3, motif IFD and CTE domain of hTERT presumably decreased the enzyme's ability to bind RNA/DNA duplex, thus reducing telomerase processivity (3). To further address the connection between duplex binding and telomerase processivity, we chose 7 telomerase processivity mutants and assayed their duplex binding ability (Fig. 4.3). Mutant N95A (TEN) resides in the TEN domain, which is responsible for the upstream DNA binding away from the telomerase active site. Other mutants, N666A, R669A, L681A, D684A (motif 3), 790-794 VVIE-4A (IFD) and L980A (CTE) sit in the hypothetical molecular clamp surrounding the RNA/DNA duplex around the telomerase catalytic site (3). Among all the mutants, only D684A augmented telomerase processivity, thus serving as a positive control in this study. N95A mutant decreased processivity through a distinctive manner compared to molecular clamp mutants, i.e. defective upstream DNA binding versus duplex binding (molecular clamp mutants).

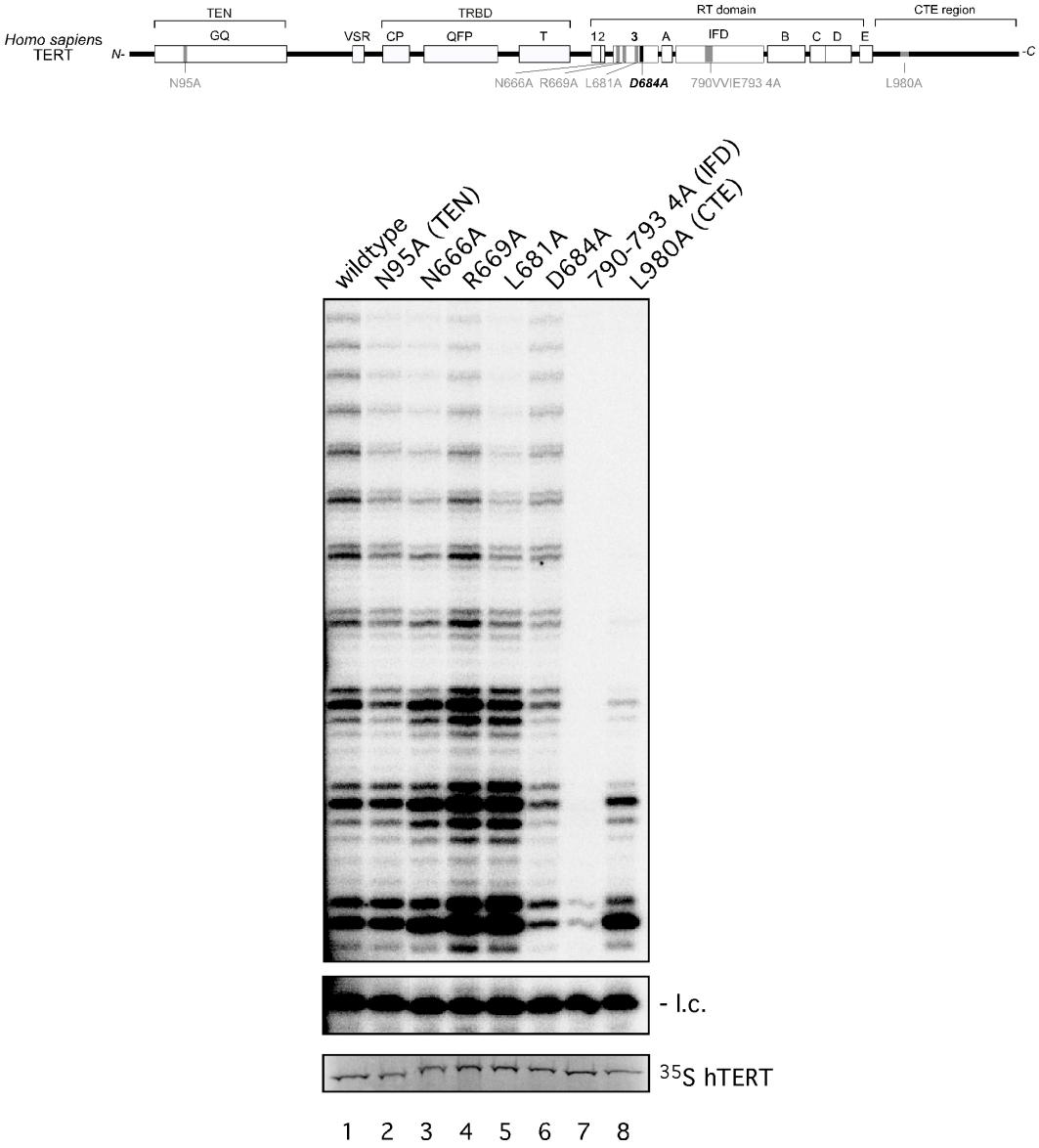


Fig. 4.3 Processivity defect mutants in motif 3, IFD and CTE domain. (Upper panel), schematic of mutation distribution in hTERT. Four major structural domains are depicted: Telomerase essential N-terminas domain (TEN), Telomerase RNA binding domain (TRBD), Reverse transcriptase domain (RT) and C-terminal extension domain (CTE). The signature motifs in each domain are also labeled. (lower panel) Wild-type and processivity mutant hTERTs were assembled with hTR nt 32-195 and nt 239-328 in RRL to reconstitute active telomerase. Telomerase activity was assayed using $(TTAGGG)_3$ primer (see Materials and Methods). l.c.: a ^{32}P end labeled 15 nt DNA oligo used as loading control. Expression level of hTERT was monitored using SDS PAGE analysis of ^{35}S Met incorporation into the protein.

We subsequently assembled template free telomerase using the processivity mutant hTERT proteins and performed activity assay with telomeric duplex substrates. Intriguingly, most low processive telomerase failed to extend duplexes, except for the N95A and R669A mutants (Fig. 4.4, lane 4-11 and lane 14-17). In contrast, hyper-processive mutant D684A is as active as wild-type enzyme in this assay. We reasoned that duplex affinity to the catalytic site might be responsible for these different levels of activity (Fig. S4.1). Hyper-processive D684A has only a 30 μM apparent *Km*, lower than the 58 μM of wild-type enzyme. The N95A TEN mutation reduced the DNA-protein interaction between the telomeric primer and hTERT. The higher probability of primer releasing contributes to the low processivity for this mutant (3). Therefore, the duplex binding ability remains intact. Indeed, the *Km* value for N95A is comparable to that of the wild-type. For the R669A mutant, higher *Km* (141 μM compared to 58μM of the wild-type) was observed, indicating lower affinity to the duplex substrates. Other lower processivity mutants did not have high enough activity to allow for *Km* determination, but they possibly have very low affinity to the duplex substrates. It is unexpected for the R669A to have a high activity level since the *Km* to duplex was very high for this mutant. We previously showed that it had almost two fold increase in turnover rate when using a short primer for activity assay (3). The faster turnover might be the reason behind the high activity observed here. When using a duplex substrate, R669A exhibited a 1.5 fold increase in enzyme turnover rate compared to wild-type (Fig. S4.2). Our assay using the template free telomerase and telomeric duplex thus provides the evidence for a direct correlation between processivity and the duplex binding event. This indicates that duplex

binding might be one essential step during template translocation, the key event determining telomerase processivity.

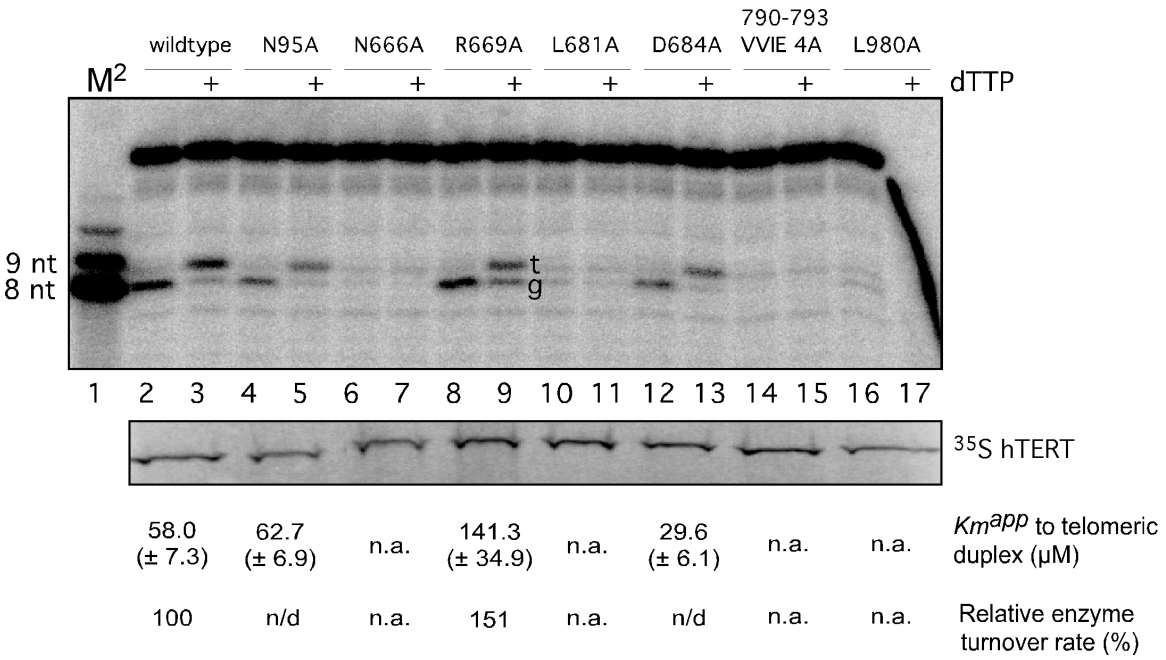


Fig. 4.4 Template free telomerase reaction with telomeric duplex. Wild-type and mutant template free telomerase was assembled in Rabbit Reticulocyte lysate and reacted with telomeric duplex with a 5'-GGTTAGG-3' DNA sequence. Processivity mutants are described in Fig. 4.3. Inclusion of dTTP substrate is indicated on the top of the gel. The relative amount of hTERT, with incorporation of ^{35}S methionine, of each mutant is indicated with SDS-PAGE analysis. The apparent *Km* of wild-type, N95A, R669A and D684A are shown on the bottom of the gel, together with the enzyme turnover rate of wildtype and R669A enzyme. The sample gel could be found in Fig. S4.1 and S4.2. n/d: not determined. n.a.: not applicable. M2: TdT reaction on primer GGTTAGG as marker.

4.4.3 Telomerase active site favors duplex ranging from 5 to 7 base pairs.

During the processive addition of telomeric DNA repeats in the wild-type telomerase reaction, the duplex length in the telomerase active site could range from 5 to 11 base pairs (Fig. 1.1). The duplex length could increase from 5 to 11 as the enzyme extends the telomeric DNA. However, it is also possible that telomerase is maintaining

constant base pairing in the catalytic site, i.e. unwinding upstream base pairs while extending downstream DNA. It was shown that a RNA/DNA duplex that is approximately 5 to 7 base pairs long was maintained in the yeast telomerase active site (7). The recently reported *Tribolium* TERT structure also suggested that the telomerase catalytic core could hold a 7 base pair duplex (8). To gain more detailed information regarding the human telomerase active site, we designed 6 duplexes with various length ranging from 5 to 10 basepairs and tested which length is the most suitable (Fig.4.5).

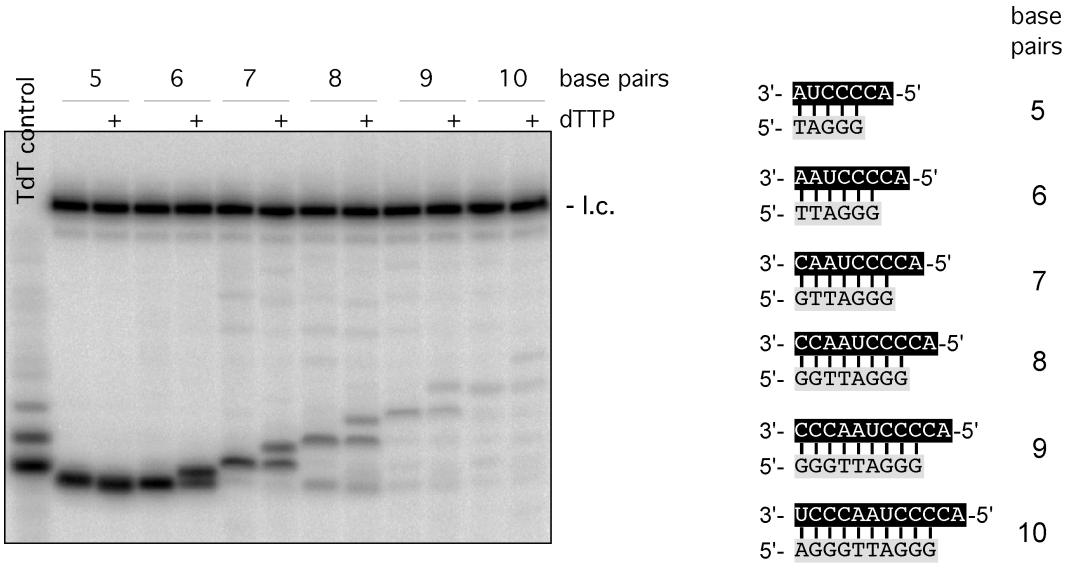


Fig. 4.5 Telomerase catalytic core favors duplex length from 5 to 7 base pairs. Template free telomerase was assembled in RRL and assayed for activity using telomeric duplexes ranging from 5 to 7 base pairs (sequences shown in the right panel, duplex lengths are indicated). RNA sequences are highlighted in black background, DNA sequences are highlighted in light grey background. TdT control: GTTAGGG oligo were extended using dGTP substrates by terminal transferase (see Materials and Methods), used as a size marker. l.c. loading control: a ^{32}P end labeled 15 nt DNA oligo nucleotides.

The duplex sequences were specifically designed to have three consecutive Gs at the 3' end of the DNA, so that even the shortest 5 bp duplex would have a Tm around 17°C under our reaction condition (Fig. S4.3). We then carried out the template free telomerase reaction using duplex 5-10 bp substrates at 4°C, where most of the substrate form duplex, thus allowed for fair comparison of all the substrates. To our surprise, the shortest 5 bp duplex produced the highest signal and the activity decreased gradually as the duplex length increased (Fig 4.5). While the duplex 5 and 6 generated higher activity, long duplex 9 and 10 gave significantly lower leveles of activity, indicating the catalytic core of human telomerase favors duplex ranging from 5 to 7 base pairs. However, our *Km* measurement suggested 5 bp duplex did not have higher affinity towards the template free telomerase than the 7 bp or 9 bp duplexes (Fig. S4.4). The high activity of template free telomerase reaction using 5 bp substrate could be attributed to the high enzyme turnover rate (Fig. S4.5). It is unlikely for telomerase to have distinct nucleotide addition rate upon different duplex substrate ending with the same residues. Thus, we speculate that higher turnover rate of the shorter duplex resulted from faster off-rate of their products releasing from telomerase active site. Faster off-rate of a short duplex is consistent with the idea that the duplex within telomerase active site is maintained in 5 to 7 base pairs, which allows easier duplex releasing when a complete repeat synthesis is finished.

Another interesting phenomenon is that only the shortest 5 bp duplex did not extend one more dT residue when dTTP was present in the reaction. The fact that telomerase extends 6 bp duplex and longer ones with higher nucleotide addition

processivity indicates that longer 5' DNA and/or 3' RNA protrudings in the duplex maybe responsible. We therefore created two more 5 bp duplexes with either a 2 nt DNA or RNA overhang. Apparently, the DNA overhang in the 5 bp duplex promotes nucleotide addition processivity, while the 2 nt RNA overhang does not (Fig. 4.6, compare lane 4 to 6). Similarly, a 1 nt DNA overhang also promotes nucleotide addition processivity (Fig. 4.6, lane 12). The data again suggested a 5 bp duplex is a better substrate than 7 bp, as neither the 2 nt DNA nor RNA overhang affected the activity of the 5 bp duplex (Fig. 4.6, compare lane 1-6 with lane 7-8).

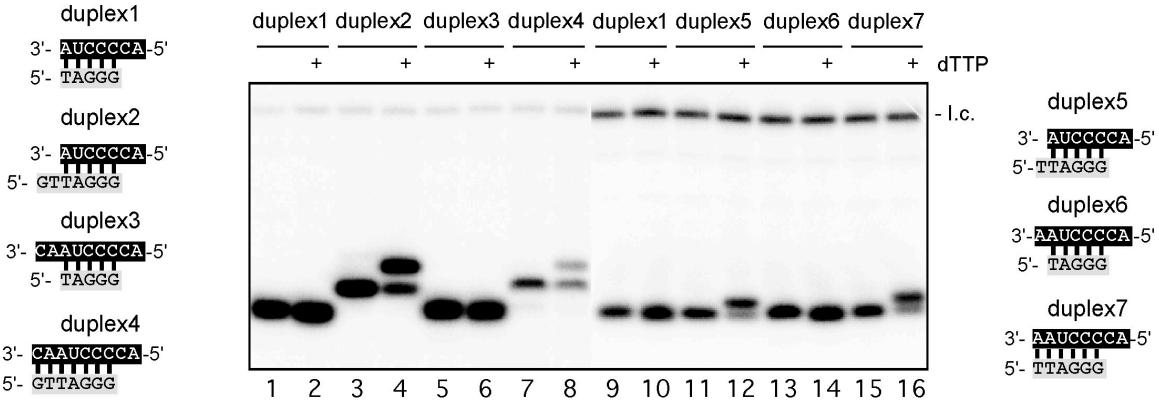


Fig. 4.6 DNA overhang promotes nucleotide addition processivity for duplex substrate. *In vitro* reconstituted template free telomerase was assayed with seven duplex substrates with or without overhangs. Including of dTTP and ^{32}P dGTP in the reactions is indicated on the top of the gel. Addition of dG and dT residues is labeled in the gel. l.c.: loading control, a ^{32}P end labeled 15 nt DNA oligonucleotides.

The correlation between low processivity and defective duplex binding indicates duplex binding is involved in the template translocation event of wild-type telomerase reaction. We designed a duplex competition experiment to explore this possibility. If duplex binding is indeed a step during translocation, an excess amount of duplex within the telomerase reaction would compete for the catalytic site and affect template translocation. In turn, the interference of template translocation will result in lower processivity. Intriguingly, a 5 base pair duplex severely decreased processivity, but the 7 base pair duplex did not (Fig. 4.7). This phenomenon is consistent with the fact that a 5 bp duplex fits the catalytic site better than the longer 7 bp duplex (Fig. 4.6). The lower processivity is indeed a result of duplex competition because high concentration of single stranded DNA or RNA did not reduce the processivity (Fig. 4.7, lane 5, 6 and 7). In the 5 base pair competition, the activity from single stranded duplex DNA could also be detected in the reaction as a distinctive banding pattern that is 3 nt different from the products of long telomeric substrate (Fig. 4.7, lane 5 and 7).

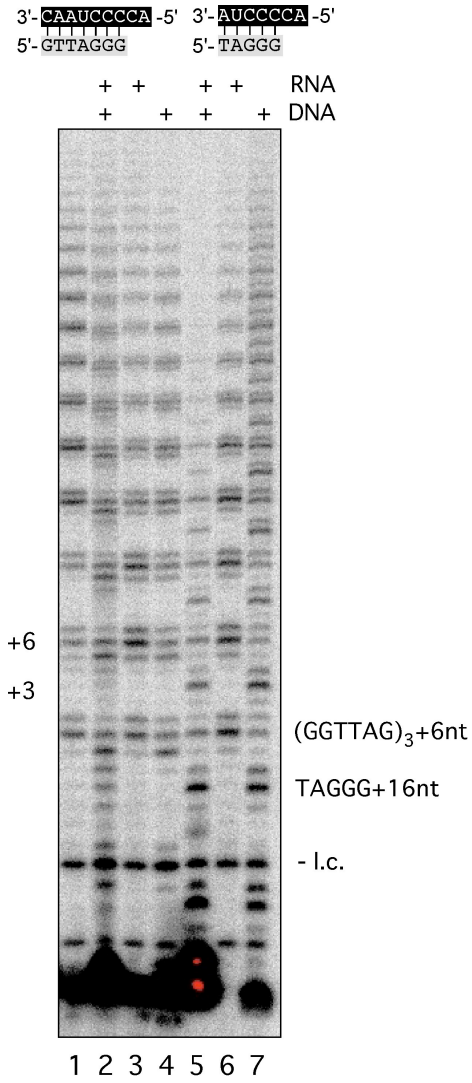


Fig. 4.7 Processivity competition by 5 and 7 base pair duplexes. Wild-type telomerase was reconstituted in RRL and assayed for processivity using 100nM 5'-$(GGTTAG)_3$-3' telomeric primer. 200μM of duplex 5bp or 7bp competitor was added when telomerase reaction was initiated (see Materials and Methods). Single stranded RNA or DNA oligos used to assemble duplexes were added as control competitors in other separated reactions as indicated on the top of the gel. Different banding pattern of 5'-TAGGG-3' and 5'-$(GGTTAG)_3$-3' substrates are labeled on the right. l.c.: ^{32}P end labeled 18 nucleotides DNA oligo.

4.4.5 Telomerase catalytic core exhibits sequence specificity towards telomeric duplexes.

Other than the length of the duplexes, the sequence of duplex would also vary in the active site during wild-type telomerase reaction as the enzyme is incorporating

different nucleotides into the duplex. To test whether duplex with different sequences would yield different results in the template free reaction, we designed six circular per-mutated telomeric duplex substrates of the 5'-TTAGGG-3' sequence. The RNA template overhang is 5'-UUC-3'. The first rC directs incorporation of a dG radioactive signal in the DNA strand. Consecutive rU templates allows monitoring nucleotide addition processivity of different duplexes. The template free telomerase reaction on six different duplexes yielded very distinctive reaction patterns. CP1 and CP2 are the best substrate for the telomerase active site as they showed the strongest signal and had lower *Km* values compared to others (Fig. 4.8 and Fig S4.6). On the other hand, nucleotide addition processivity also varied significantly among 6 different substrates. CP5 duplex, with native telomeric repeat ending, has no nucleotide added at all. The CP4 product provided a hint on the failure of extending CP5 duplex already, as the CP4 product ends at a native telomeric repeat position with the sequence 5'-GGTTAG-3', which is the exact sequence of CP5 duplex.

Although it allows direct test of the interaction between duplex substrate and telomerase active site, template free telomerase is indeed a heavily mutated system. The sequence specificity observed in the template free telomerase reaction might play a very trivial role in wildtype telomerase, where the RNA template is tethered on both ends and the DNA oligo has a long 5' overhang. Therefore, we assembled telomerase with 6 circular permutated template sequences within the full length hTR. Surprisingly, the duplex sequence is still a major determinant for complete telomeric repeat synthesis. In spite of the different template sequences, all 6 circular permutated telomerases have the

banding pattern stopped after adding 5'-GGTTAG-3' (Fig. 4.9). We then used 6 circular permutated telomeric primers to react with the circular permutated template mutants. The same primer has the same banding pattern when using different template mutants, suggesting sequence determinant of repeat synthesis has stronger effect than the P1 stem defined template boundary (Fig. S4.7).

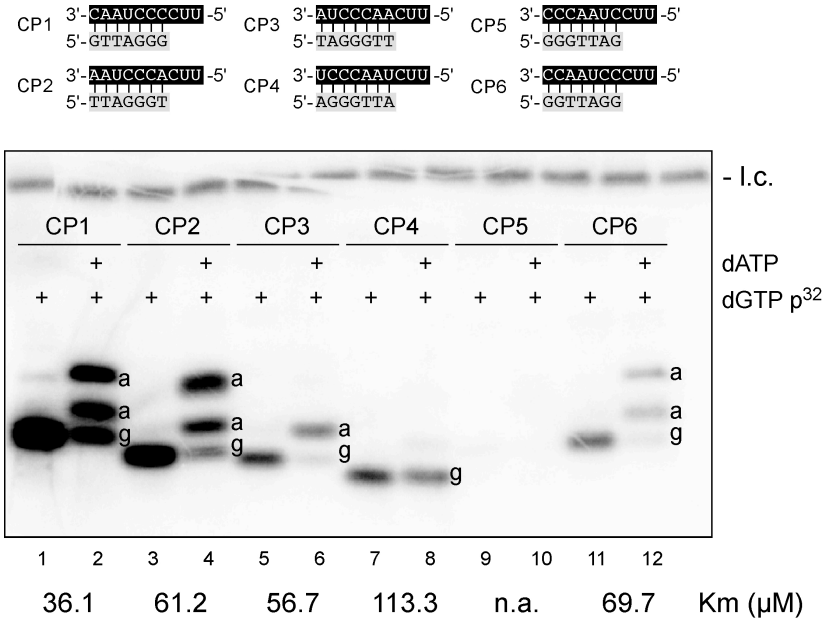


Fig. 4.8 Telomerase active site shows sequence specificity. Template free telomerase was reconstituted in RRL and assayed for activity using six circular permutated telomeric sequence duplexes. Substrate sequences (CP1-6) are depicted on the top of the gel. Black background highlights RNA sequences, while grey background highlights DNA sequences. The apparent *Km* value of each substrate is indicated at the bottom of the gel. Original gel image could be found in Fig. S4.6. Addition of dATP and ^{32}P dGTP in the reaction is indicated as "+". Incorporation of dG or dT residues are labeled as "g" and "a" on the right of the signal in the gel. l.c. 15 nt DNA oligo with ^{32}P end labeling as loading control.

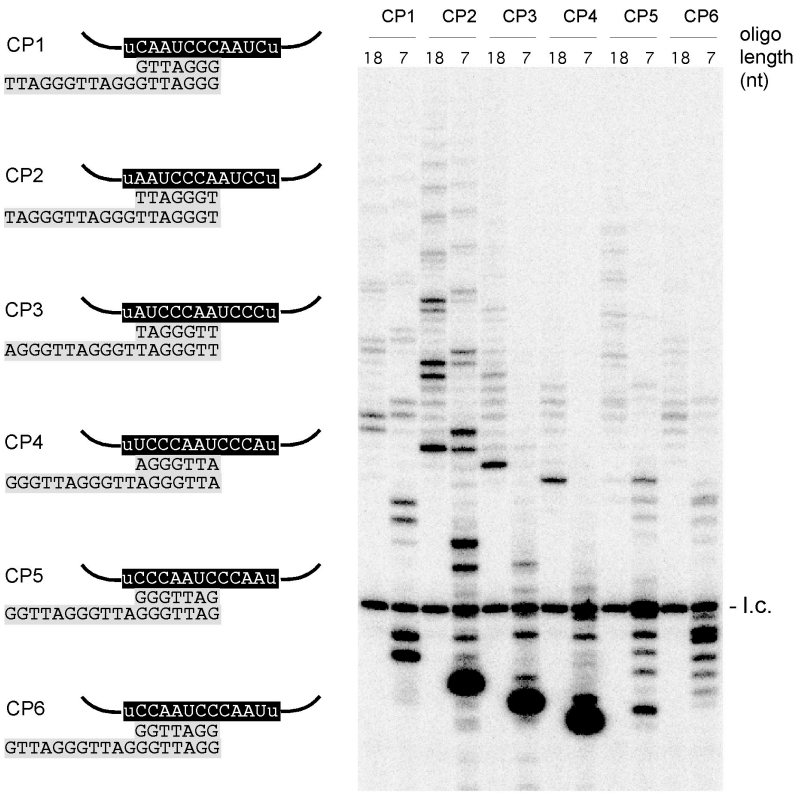


Fig. 4.9 Circular permutated template sequences in telomerase RNA. Six circular permutated hTR pseudoknot fragments (nt 26-195) were assembled with hTERT and CR4-CR5 domain (nt 239-328) and assayed for activity with their corresponding telomeric primers. (left panel) Annealing of either 7nt or 18nt primer to template results in 4 nucleotides of space for primer extension. (right panel) activity assay of template circular permutated telomerase. l.c. 15 nt ^{32}P end labeled DNA oligo nucleotides.

The distinctive duplex sequences result in different nucleotide addition processivity for CP1 (3 nt) and CP4 (1nt). To determine which base pair(s) in CP4 duplex is responsible for the low nucleotide addition processivity, we swapped three base pairings between CP4 and CP1 duplexes. The result clearly indicates that the middle 3 base pairs encode a signal for the stop after a 5'-GGTTAG-3' repeat (Fig. 4.10, duplex CP 4.a). We then mutated the 3 base pairings one by one to further explore the exact stop signal within this region. Interestingly, mutating the first rA-dT base pair of the two consecutive A-T base pairs in CP4 duplex resulted in one more dA extension when dATP

is included in the reaction (Fig. 4.10, CP4.d). The fact that CP4.d could not be extended by two dA residues might be because the remaining rA-dT now functions as the stop signal. Therefore, telomerase active site can accommodate 4 base pairs after encountering the first rA-dT base pair in the RNA/DNA duplex.

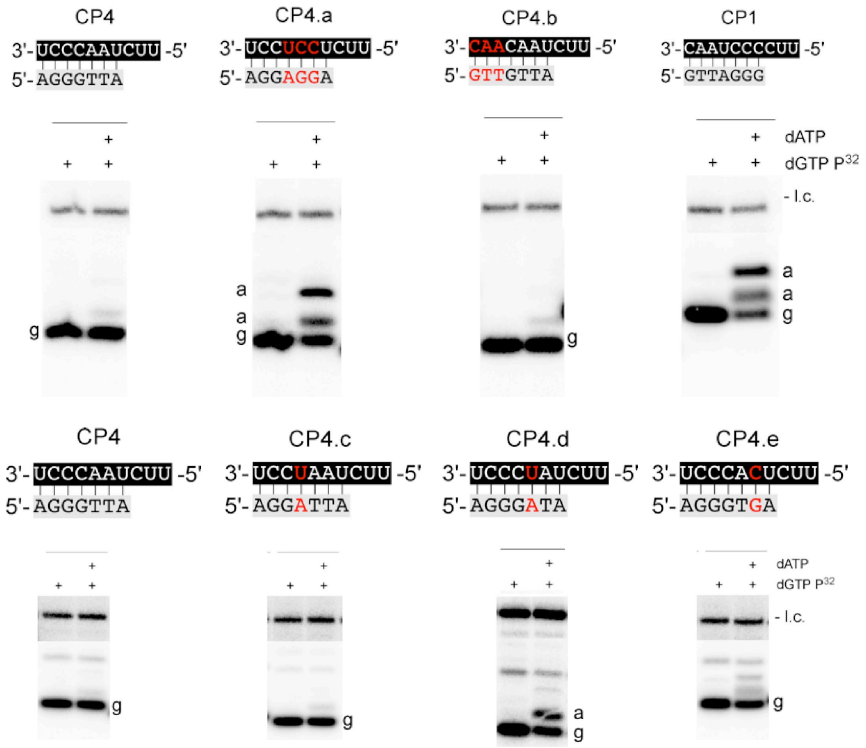


Fig. 4.10 Sequence determinant of nucleotide addition processivity in the duplex substrate. Duplex substrates are depicted on top of the gel. Black background represents RNA sequences and grey background represents DNA sequences. The red font highlights the mutated sequence within CP4 duplex substrate. Addition of dATP and ^{32}P dGTP nucleotides in the reaction is indicated. The incorporated nucleotides (a or g) are labeled alongside the gel.

4.5 Discussion

Template free telomerase provided a unique system to specifically look at the template translocation step of processive addition of telomere repeats by the telomerase enzyme. The telomerase specific activity on duplex substrate indicates the catalytic core of the enzyme obtained the ability to utilize duplex other than single stranded DNA. Furthermore, the catalytic core does favor telomeric duplex over non-telomeric duplex (Fig. 4.2). The mutations in molecular clamp surrounding RNA/DNA duplex that reduced telomerase processivity have lower affinity to the duplex substrates, while the hyper-processive molecular clamp mutant increases the affinity between template free telomerase and the duplex (Fig. 4.3 and 4.4). On the other hand, the hypo-processivity mutant in TEN domain did not decrease processivity through low duplex binding affinity but through upstream DNA-protein interaction outside the active site. Also, short duplexes ranging from 5 to 7 bp are better substrates for telomerase than 8 to 10 bp duplexes (Fig. 4.5). Single stranded DNA overhang facilitates duplex movement in catalytic site, thus promoting nucleotide addition processivity (Fig. 4.6). Using a 5 bp duplex to compete with the telomerase processive reaction severely reduced the processivity while the 7 bp duplex competitor did not (Fig. 4.7). This suggests that a constant base pairing is maintained in the active site of human telomerase, similar to the number estimated in yeast telomerase and tribolium telomerase (7, 8). The data also suggested duplex binding and releasing is indeed happening during template translocation, thus allowing duplex competitor to reduce the processivity.

The sequence specificity of telomerase active site is very unexpected. However, the stopping pattern of nucleotide addition of six circular per-mutated duplexes strongly implies a signal is encoded in the 5'-GGTTAG-3' complete telomeric sequence to finish a single repeat synthesis (Fig. 4.8, 4.10). This data is somewhat contradictory to the template definition element study reported previously (2). Detailed *Km* studies shed a light on the reason behind telomerase active site sequence specificity. The affinity between duplex and active site decreases gradually towards the complete telomeric repeat 5'-GGTTAG-3'. The CP5 duplex affinity was not tested due to extremely low activity. But we speculate this substrate has the lowest affinity as indicated by the trend. Fitting this phenomenon into wildtype telomerase reaction makes more sense. When a repeat synthesis is initiated (sequence with CP6:GTTAGG, CP1:TTAGGG), the duplex needs to sit stably in the active site and allows for following nucleotide addition. When the repeat synthesis is finished (CP5:GGTTAG), loosen duplex easily falls off the telomerase active site and initiate duplex separation for next round of repeat synthesis.

The DNA overhang increaseing nucleotide addition processivity is also not completely unexpected. A 5bp duplex with 1 or 2nt DNA overhang mimics the substrate formed within the active site in a wild-type telomerase reaction. Without DNA overhang, the duplex might get stuck in the active site and result in an unprocessive or inactive enzyme.

4.6 References

1. Chen, J. L., and Greider, C. W. (2003) *EMBO J* 22, 304-314

2. Chen, J. L., and Greider, C. W. (2003) *Genes Dev* 17, 2747-2752

3. Xie, M., Podlevsky, J. D., Qi, X., Bley, C. J., and Chen, J. J. (2010) *Nucleic Acids Res* 38, 1982-1996

4. Cristofari, G., and Lingner, J. (2006) *EMBO J* 25, 565-574

5. Tesmer, V. M., Ford, L. P., Holt, S. E., Frank, B. C., Yi, X., Aisner, D. L., Ouellette, M., Shay, J. W., and Wright, W. E. (1999) *Mol Cell Biol* 19, 6207-6216

6. Qiao, F., and Cech, T. R. (2008) *Nat Struct Mol Biol* 15, 634-640

7. Forstemann, K., and Lingner, J. (2005) *EMBO Rep* 6, 361-366

8. Gillis, A. J., Schuller, A. P., and Skordalakes, E. (2008) *Nature* 455, 633-637

CHAPTER 5

STRUCTURE AND FUNCTION OF THE SMALLEST VERTEBRATE TELOMERASE RNA FROM TELEOST FISH

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.

Author contributions: M. Xie cloned teleost fish TRs and zebrafish TERT, performed fish telomerase activity assays. A. Mosig identified fish telomerase RNA genes from fish genome database. X. Qi cloned Fugu TERT, performed Northern blotting for fish TR and constructed the phylogenetic tree. Y. Li cloned medaka TERT. P.F. Stadler and J.J. Chen conceived the project.

5.1 Abstract

Telomerase extends chromosome ends by copying a short template sequence within its intrinsic RNA component. Telomerase RNA (TR) from different groups of species varies dramatically in sequence and size. We report here the bioinformatic identification, secondary structure comparison, and functional analysis of the smallest known vertebrate TRs from five teleost fishes. The teleost TRs (312–348 nucleotides) are significantly smaller than the cartilaginous fish TRs (478–559 nucleotides) and tetrapod TRs. This remarkable length reduction of teleost fish TRs correlates positively with the genome size, reflecting an unusual structural plasticity of TR during evolution. The teleost TR consists of a compact three-domain structure, lacking most of the sequences in regions that are variable in other vertebrate TR structures. The medaka and fugu TRs, when assembled with their telomerase reverse transcriptase (TERT) protein counterparts, reconstituted active and processive telomerase enzymes. Titration analysis of individual RNA domains suggests that the efficient assembly of the telomerase complex is influenced more by the telomerase reverse transcriptase (TERT) binding of the CR4–CR5 domain than the pseudoknot domain of TR. The remarkably small teleost fish TR further expands our understanding about the evolutionary divergence of vertebrate TR.

5.2 Introduction

Telomeres are specialized DNA-protein complexes that cap chromosome ends and are important for genome stability and cellular proliferation (1). Telomeres consist of repetitive DNA sequences and a variety of telomere-associated proteins. The length of telomeric DNA in most eukaryotes is maintained by telomerase, a specialized reverse

transcriptase that synthesizes telomeric DNA repeats at chromosome ends to counterbalance the natural shortening that occurs during DNA replication. Telomerase, a ribonucleoprotein (RNP) enzyme, consists of at least two essential corecomponents, the catalytic protein component telomerase reverse transcriptase (TERT), and the telomeraseRNA (TR) that provides a template for telomeric DNA synthesis. TR is remarkably variable in size, sequence, and even secondary structure between different groups of eukaryotes. To date, TR sequences have been identified in 28 ciliates, 14 yeasts, and 38 vertebrates. Due to the lack of sequence similarity between groups of species, the TR secondary structures were determined independently for each of these three groups (2). The vertebrate TR secondary structure is composed of three highly conserved structural domains: the pseudoknot/template domain, the CR4–CR5 domain, and the scaRNA domain (3–5). The pseudoknot/template domain contains a template region for telomeric DNA synthesis, and a conserved pseudoknot structure essential for telomerase activity. The CR4–CR5 domain together with the pseudoknot/template domain are both required for reconstituting active telomerase *in vitro* (6). However, their mechanistic roles are unclear. The scaRNA domain is crucial for the 3-end processing of TR and telomerase RNP biogenesis *in vivo* (3, 7). Whereas TRs from 34 tetrapods and 4 cartilaginous fishes share this three-domain structure (4), they have not yet been identified from teleost fish that comprises near half of the extant vertebrate species. Teleost fish is the most diverse group among vertebrates (8), and is distinct from the cartilaginous fish. The teleost and tetrapods (including amphibian, reptile, birds, and mammals)diverged from each other around 450 million years ago. Since then, teleost fish

have undergone genome duplication and rediploidization, resulting in an amazing level of genomic diversity. The relatively faster evolution rate and the consequent diversity in teleost fish offer an attractive model for evolutionary studies. Identification of TR from teleost fish using degenerate PCR or BLAST search has, however, not been successful due to a high degree of sequence variation in TR. Here we report the identification of TRs from five teleost fish, *D*anio rerio, *Oryzias latipes, Gasterosteus aculeatus, Takifugu rubripes* and *Tetraodon nigroviridis*, using a novel bioinformatics method. To structurally and functionally characterize the teleost TR, we have cloned *TR* as well as *TERT* protein genes from medaka, fugufish, and zebrafish, and reconstituted telomerase activity for medaka and fugufish. The structural and functional analyses of the teleost fish telomerase enzyme provide important new insights into the evolution of the vertebrate telomerase RNP.

5.3. Materials and Methods

5.3.1 Bioinformatics Search of Teleost Fish TR Sequences

A sequence search was performed using fragrep2. The input pattern, shown in supplemental Fig. S1, consists of eight positionspecific weight matrices (PWMs). The quality of match between aPWMand a DNA sequence is measured as a fraction of similarity above an unavoidable background (9). The computational approach and implementation details of fragrep2 are described in detail in Mosig *et al.* (10). Our search pattern was generated by annotating the eight conserved regions in the TR alignment published in Chen *et al.* (4), and converted to a fragrep2 search pattern using the aln2pattern tool and both fragrep2 and aln2pattern (available www.bioinf.uni-leipzig.de/

Software). The initial search pattern (Fig. S1) resulted in a single plausible hit in the medaka genome (assembly MEDAKA1). A BLAST search using medaka sequence as query against other teleost fish genomes revealed homologs in the stickleback (assembly BROAD S1), fugu (assembly FUGU 4.0), and tetraodon (assembly TETRAODON 7). Based on the four teleost TR sequences, a modified and less stringent search pattern was generated, with which we found 79 candidate sequences in the zebrafish genome (assembly Zv6). These were screened using INFERNAL (11) and the secondary structure annotated TR alignment from the Rfam data base (12), resulting in a single sequence that fit well with other teleost candidates and the previously known vertebrate TR sequences. The alignment of all 43 known vertebrate TR sequences can be obtained from the Telomerase Data base (telomerase.asu.edu).

5.3.2 Genomic DNA and Total RNA Isolation

For isolation of genomic DNA and total RNA, medaka fish (*O. latipes*) were purchased from Aquatic Eco-Systems (Apopka, FL), and Zebrafish (*D. rerio*) were obtained from Dr. Yung Chang (Arizona State University, AZ) or purchased from Aquatical Tropicals, Inc. (Plant City, FL). Green spotted pufferfish (*T. nigroviridis*) were purchased from AquariumFish.net. Liver tissue of fugu (*T. rubripes*) fish was obtained from Dr. Shugo Watabe (University of Tokyo, Japan). Genomic DNA was isolated from 50 to 100 mg of fish tissue using the DNAzol reagent (Invitrogen) following the manufacturer's instruction. Stickleback (*G. aculeatus*) genomic DNA was a generous gift from Dr. David Kingsley (Stanford University). Total RNA was isolated from 100 to 200 mg of gill or liver tissues using 1 ml of TRIzol reagent (Invitrogen) following the

manufacturer's instructions. Concentrations of DNA and RNA samples were determined by *A*260 measurement using the Nanodrop ND-1000 spectrophotometer (Nano-Drop Technologies).

5.3.3 Sequencing and Cloning of TR Genes

To verify the sequences, teleost fish TR genes were PCR amplified from genomic DNA and the PCR products were sequenced directly. *The verified sequences of five teleost fish TR genes were deposited into GenBankTM with* the following accession numbers: EF569636 (*D. rerio*), EF569637 (O. latipes), EF569638 (T.rubripes), EF680233 (T. nigroviridis), and EF680234 (G. aculeatus).

For medaka, zebrafish, and fugu, the PCR products of TRgenes were cloned into the EcoRV site of the pZero vector(Invitrogen) to generate pMedaka-TR, pZebrafish-TR, and pFugu-TR. Plasmids were sequenced to confirm sequence accuracy of the cloned TR genes.

5.3.4 Identification and Cloning of Teleost Fish TERT Genes

To reconstitute telomerase activity, we cloned TERT genes from medaka, zebrafish, and fugu. The fugu TERT (AY861384) and medaka TERT (DQ248968) gene sequences have been previously identified and were available from GenBank (13). The zebrafish TERT gene was identified in this study via a BLAST search of the zebrafish genome data base using the fugu TERT protein sequence as query. The exact 5'- and 3'-ends of the full-length zebrafish TERT cDNA sequence were determined by the 5'- and 3'-rapid amplification of cDNA ends (RACE) using a SMART-RACE cDNA Amplification Kit (Clontech). The cDNA sequence was determined by direct sequencing

of the reverse transcriptase-PCR products. The sequence of zebrafish TERT gene has been deposited into GenBank with accession number EF202140.

To clone the TERT genes, the coding sequences of medaka and zebrafish TERT genes were PCR amplified from the cDNA samples prepared from total RNA samples using Thermoscript reverse transcriptase (Invitrogen) and an oligo(dT)18 reverse primer. The fugu TERT cDNA was PCR amplified from a cDNA library obtained from Dr. Byrappa Venkatesh (Institute of Molecular and Cell Biology, Singapore). ThePCRproducts of the medaka, zebrafish, and fugu TERT cDNAs were cloned into the pCITE vector for in vitro synthesis of the recombinant TERT proteins.

5.3.5 *In Vitro* Transcription of TR

RNA was prepared by T7 *in vitro* transcription using PCR DNA products as template as described previously (14, 15).

5.3.6 Northern Blotting Analysis

Twenty micrograms of total RNA was resolved on a 4% polyacrylamide, 8 M urea denaturing gel and electrotransferred to Hybond-XL membrane (Amersham Biosciences) at 0.5 A for 1 h. The membrane was UV cross-linked and prehybridized at 65 °C for 30 min in 20 ml of UltraHyb hybridization buffer (Ambion). Riboprobes with sequences complementary to the target RNA were generated by *in vitro* transcription from a PCR DNA template that contained the T7 promoter and labeled internally with [α-^{32}P]UTP using a MaxiScript kit (Ambion). After incubation at 37 °C for 1 h, 1μl of RNase-free DNase I (2 units/μl) was added to the reaction, followed by a 20-min incubation at 37 °C to remove the DNA template. Riboprobes were then purified using

microspin G-25 columns (GE Healthcare). The membrane was hybridized at 65 °C overnight in 20 ml of UltraHyb buffer with the riboprobe added to 1x1000,000 cpm/ml. The hybridized membrane was washed twice in 20 ml of 1X SSC (3.0 M NaCl and 0.3 M sodium citrate, pH 7.0), 0.2% SDS for 10 min at 65 °C, and twice in 20 ml of 0.2 SSC, 0.1% SDS for 30 min at 65 °C. The blot was analyzed using a phosphorimager, Bio-Rad FX Pro.

5.3.7 In Vitro Reconstitution of Telomerase

Human, medaka, fugu, and zebrafish telomerases were reconstituted using the TNT (transcription and translation) Quick Coupled rabbit reticulocyte lysate system (Promega). Briefly, recombinant TERT protein was synthesized in 10 µl of rabbit reticulocyte lysate at 30 °C for 60 min following the manufacturer's instructions. To assemble the telomerase complex, *in vitro* synthesized TR was added to the TNT reaction of TERT synthesis, and incubated at 30 °C for 30 min. For the titration experiments of individual RNA domains, the pseudoknot/template or CR4–CR5 RNA fragment was added to a saturated 3 µM, whereas the other RNA fragment was added to various concentrations as indicated.

5.3.8 Conventional Telomerase Activity Assay

Enzymatic activity of *in vitro* reconstituted telomerase was analyzed using a direct primer extension assay. A 10µl reaction was carried out with 3 µl of *in vitro* reconstituted telomerase sample in the presence of 1PE buffer (50mM Tris-HCl, pH 8.3, 50mM KCl, 2mM dithiothreitol, 3 mM $MgCl_2$, and 1 mM spermidine), 1 mM dATP, 1 mM dGTP, 1 mMdTTP, and 2 pmol of 5-^{32}P-end labeled $(TTAGGG)_3$ telomere primer at 30 °C for 2 h.

The products were subjected to phenol/chloroform extraction and ethanol precipitation, followed by 10% denaturing PAGE. Gels were dried, and products were detected and analyzed using a Bio-Rad FX Pro Imager. For each reaction, activity was determined by measuring the total intensity of extended telomere substrate, correcting for background, and normalizing against unextended primer (loading control). Relative activities were obtained by dividing the activity of each reaction by that of the reaction with saturated concentration of RNA fragments. For the titration assay, the relative activities were plotted against concentrations of RNA fragment and the nonlinear regression curve fitting was carried out using the one-site binding (hyperbola) equation, $Y = B_{max} \times X/(K_d + X)$ (Prism 5, Graphpad Software, Inc.).

5.4 Results

5.4.1 A novel Bioinformatics approach to identify TR sequences

Despite significant efforts to clone TRs from a diverse array of vertebrate species, TR sequences have not been identified from teleost fish (4). Computational searches for TR candidates using the Basic Local Alignment Search Tool (BLAST) of the sequenced teleost fish genomes have been unsuccessful. The inability to identify TR sequences in teleost fish using either degenerate PCR or BLAST presumably stems from the fact that vertebrate TRs are conserved only in eight relatively short regions (called Conserved Region 1–8, or CR1– CR8) that are interrupted by highly variable sequences with a large number of indels (4).

To identify TR sequences, we employed an improved homology search tool, fragrep2, to search teleost fish genomes. The original version of the fragrep program implements a

specialized algorithm for homology search that considers gap-free sequence patterns separated by variable-length regions of nonaligned sequence (16). This approach has been demonstrated to work well for genomewide searches of non-coding RNAs (16, 17). However, it had not been successful in finding teleost fish TRs. This is because even the relatively well conserved blocks, *i.e.* CR1–CR8, contained too many variations to be well represented by a single consensus sequence. To circumvent this, in fragrep2, we have replaced consensus sequences by PWMs to search for matched DNA sequences (10). As shown in supplemental Fig. S5.1, the initial search pattern contains a collection of PWMsas well as minimal and maximal distances between these PWM blocks.

Using this new approach, we successfully found a TR candidate in the medaka genome. Homologs of this medaka sequence could then be readily found by means of BLAST in stickleback, fugu, and Tetraodon genomes. All four sequences are flanked upstream by an ADP-ribosylation factor and downstream by homologs of human LASP1 and/or PLXDC2 (Table S5.1). Based on the alignment of the four teleost fish sequences, we modified the search pattern and were able to retrieve a single convincing candidate from the zebrafish genome using fragrep2. Surprisingly, the genomic location of the zebrafish TR candidate is neither syntenic with that of the other teleost sequences nor with the human locus (Table S5.1). All five teleost TR genes were PCR amplified from genomic DNA samples and the PCR DNA products were sequenced directly to verify the sequences identified from the genome databases (see “Materials and Methods”).

5.4.2 Unique transcription elements of Fish TR Genes

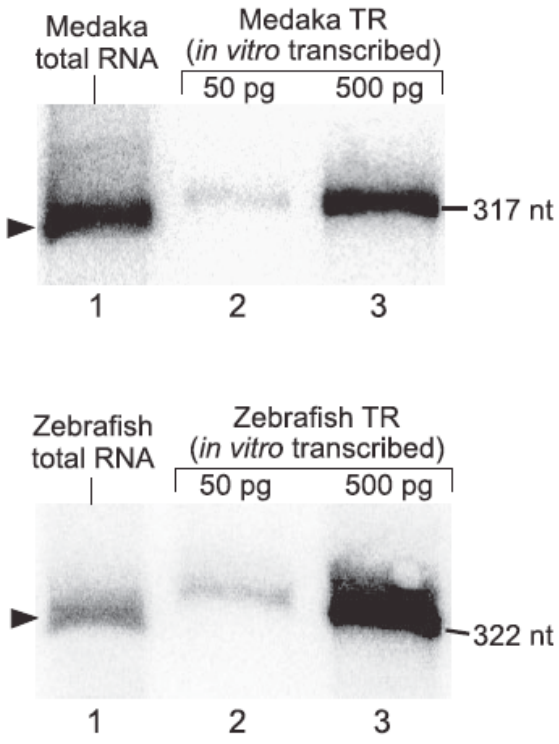
Analysis of genomic sequences upstream of the fish TR-coding sequences revealed transcriptional elements typical of an RNA polymerase II promoter: a conserved TATA box-like and a CCAAT box element (Fig. S5.2). This suggests that, like other vertebrate TRs, teleost TRs are products of RNA polymerase II. Interestingly, a putative CRE-BP1/c-Jun binding element, located between the TATA and CCAAT boxes, is conserved in both teleost and cartilaginous fishes, and some amphibians (bullfrog and horned frog) (Fig. S5.2). This data suggest an evolutionary change in transcriptional regulation of the *TR* gene along the tetrapod lineage.

5.4.3 The compact size of teleost fish TR

To confirm the presence of the identified teleost TR transcripts in cells, we performed Northern blotting analysis to detect the endogenous TRs. The medaka and zebrafish TRs were each detected as a single band on the Northern blot (Fig. 5.1a, *lane 1*). Based on the Northern result, the size of the endogenous medaka and zebrafish TRs are estimated to be slightly smaller than the *in vitro* transcribed RNA markers that are 317 and 322 nt, respectively (Fig. 5.1a, compare lanes 1 and *2*).

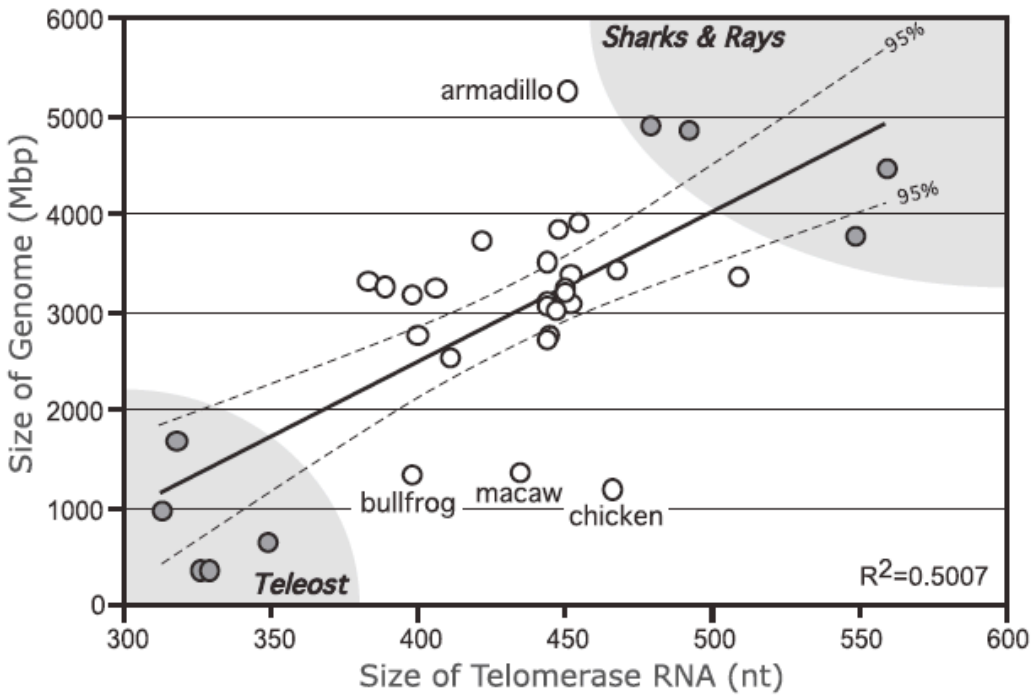
To determine the actual size of the endogenous TR, we mapped the 5'-ends of medaka and zebrafish TRs by 5'-RACE. The results showed that the 5'-ends of both medaka and zebrafish TRs lie 14 nucleotides upstream of the template sequence. Assuming that the 3'-end of the fish TR is located, like other vertebrate TRs, 3 residues downstream of the box ACA motif, the medaka and zebrafish TRs are predicted to be 312 and 317 nt long, respectively, consistent with the sizes observed from the Northern analysis. Based on

sequence alignment, the other three teleost TR homologs are predicted to be 348 (stickleback), 325 (fugu), and 328 nt (Tetraodon). This makes teleost TRs the smallest among all known vertebrates, as the size of previously known vertebrate TRs ranges from 382 to 559 nt (4).



a. Northern blotting analysis of medaka and zebrafish TRs

Fig. 5.1 a, Twenty micrograms of total RNA (lane 1), and 50 (lane 2) or 500 pg (lane 3) of *in vitro* transcribed medaka or zebrafish TRs were electrophoresed on 4% denaturing polyacrylamide gels. Blots were each hybridized with riboprobes specific to each TR. Endogenous TR bands are indicated by solid triangles. The *in vitro* transcribed medaka TR (317 nt) and zebrafish TR (322 nt) serve as markers for size estimation and mass quantitation. The levels of endogenous TR in liver cells were quantitated to be 508 and 110 pg per 20 g of total RNA for medaka and zebrafish, respectively.



b. positive correlation between the TR size and the genome size

Fig.5.1 continued. b, The genome sizes (Mbp) were derived from C-values (pg) obtained from The Animal Genome Size Data base (www.genomesize.com). The sizes of TRs are based on data from Chen *et al.* (4) and this study. Five teleost and four cartilaginous (sharks and rays) fishes are clustered into two separated groups at the *lower-left* and *higher-right ends* of the graph, respectively. The 95% confidence band (*dashed*) of the linear regression line (*solid*) is shown. The *p* value is <0.0001.

Teleost fishes have notably small genomes, whereas the cartilaginous fishes have relatively large genomes (18). Intriguingly, teleost fishes with smaller genomes have the smallest TRs, whereas cartilaginous fishes with larger genomes have the largest TRs (from 478 to 559 nt) among vertebrates. By plotting the TR size over the genome size, we found a positive correlation between the size variation of TR and the genome size with an R2 value of 0.5007 and a p value <0.0001 (Fig. 5.1B). This strong correlation suggests that the size variation of fish TR resulted from evolution of the fish genome.

5.4.4 Secondary structure of teleost fish

To determine whether these small teleost TRs share a similar secondary structure with other vertebrate TRs, we constructed secondary structure models for teleost fish TRs using phylogenetic comparative analysis. The primary sequences of the five teleost TRs identified were aligned manually as described previously (4). The eight conserved regions CR1–CR8 found previously in 35 vertebrate TRs are largely conserved in the teleost TRs (Fig. 5.2). Because of their small size and the presence of the CR sequences, teleost fish TR sequences can be readily aligned without much ambiguity. The aligned sequences were analyzed for covariations to derive a conserved secondary structural model for the teleost TR (Fig. 5.3a and b). Homologous to the structures of other vertebrate TRs, the proposed teleost structure contains 11 helices (P1, P2a, P2b, P3, P4, P5, P6, P6.1, P7a, P7b, and P8) grouped into three separate structural domains: the pseudoknot/template domain, the CR4–CR5 domain, and the snoRNA domain (Fig. 5.3a and b). All helices, except for the P6.1 and P7a, were supported with at least one covariation per helix. All five teleost TRs share a similar secondary structure with

variation mostly in the hypervariable region between the P4 and P5 helices (Fig. 5.3a. b. and supplemental Fig. S5.3).

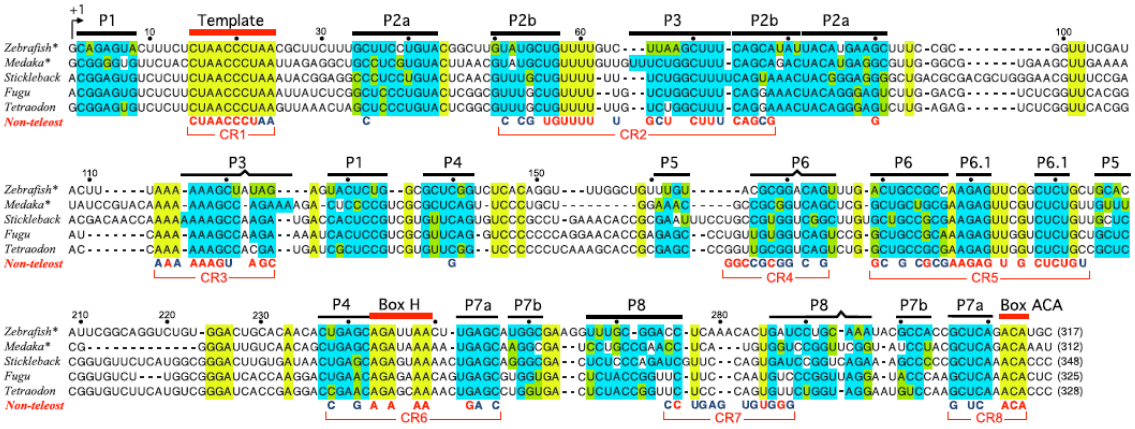
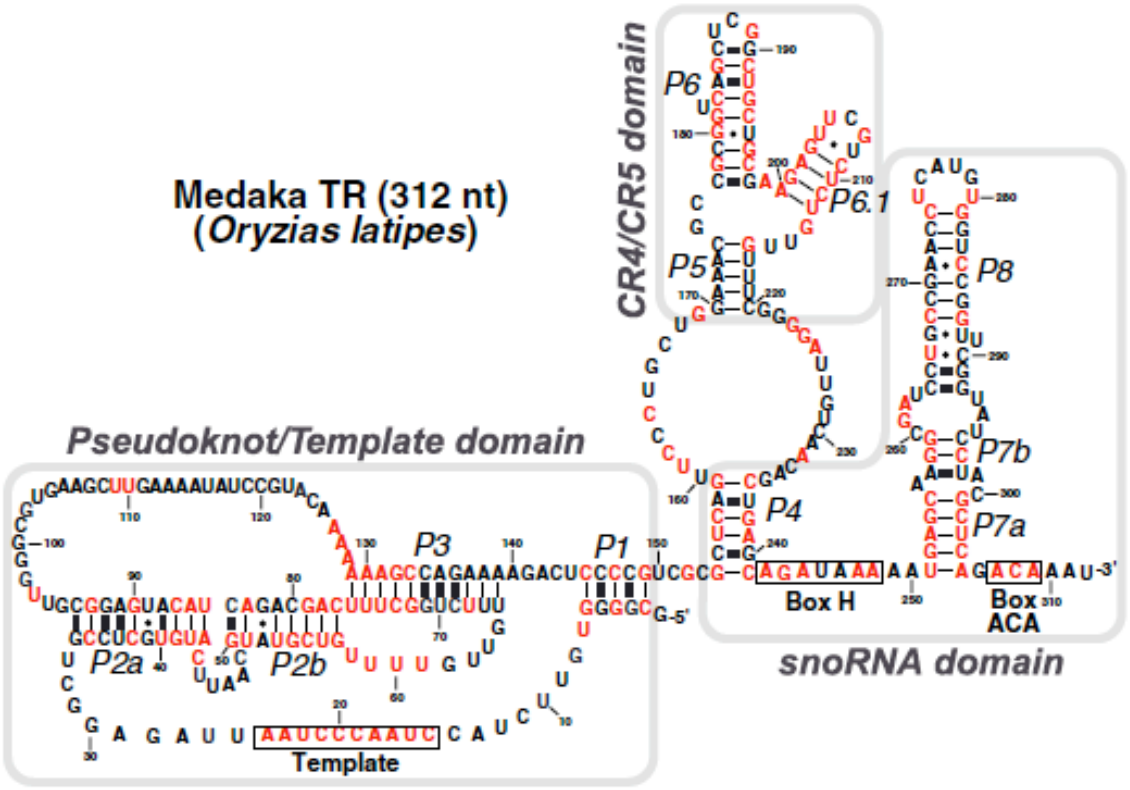
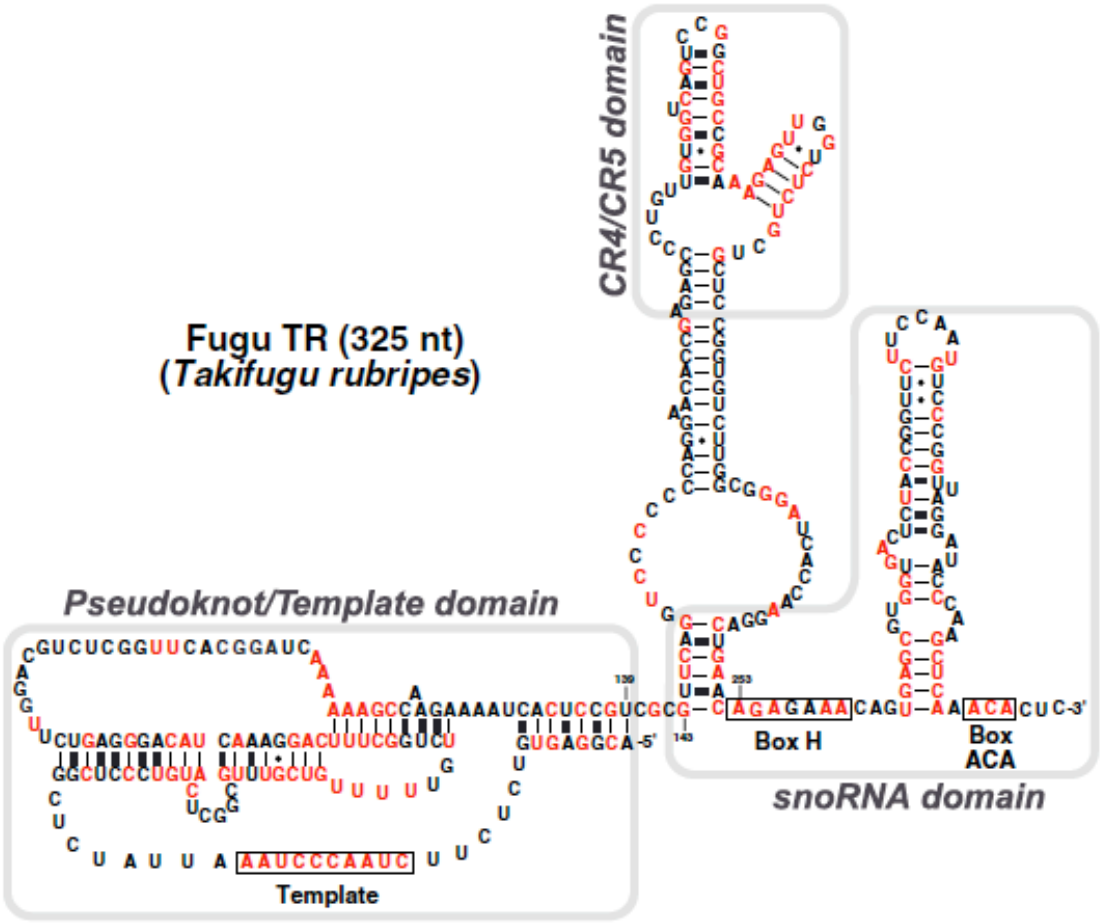


Fig. 5.2 Sequence alignment of teleost fish TR. The alignment includes TR sequences from zebrafish (*D. rerio*), medaka (*O. latipes*), stickleback (*G. aculeatus*), fugu (*T. rubripes*), and tetraodon (*T. nigroviridis*). Residues that are 100 (red) or 80% (blue) conserved in non-teleost vertebrates TRs (Chen et al. (4)) are shown below the alignment. The eight conserved regions (CRs) are indicated with red brackets. Black lines above the alignment indicate helices (P1–P8) in the secondary structures. Conserved motifs, i.e. the template, box H and box ACA, are indicated with red lines above the aligned sequences. Residues shaded in blue indicate conserved nucleotides that form Watson-Crick base pairings, whereas the ones shaded in green indicate nucleotides that co-vary and maintain base pairing. The residues shaded in yellow are located in the single-stranded regions and are universally conserved among the five teleost fishes. Dashes (-) denote alignment gaps. Every tenth nucleotide of the zebrafish sequence is marked with dots above the alignment. The size of each RNA is indicated at the end of the respective sequence. Asterisks (*) indicate organisms for which the 5'-end of the RNA was determined by 5'-RACE.



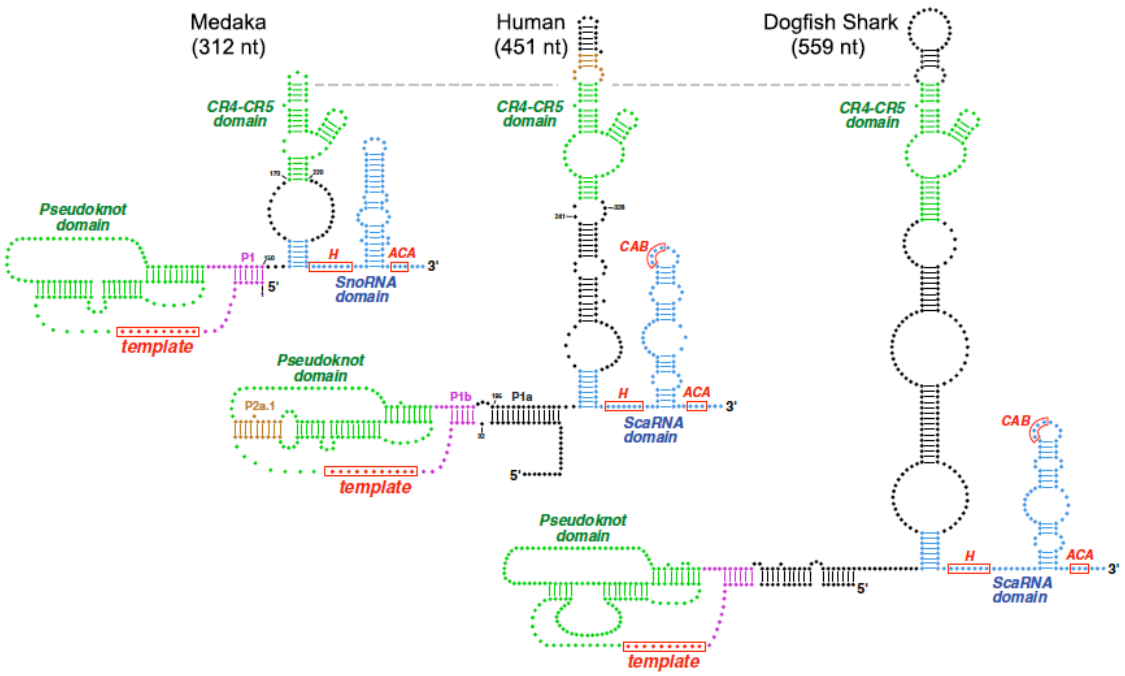
a. secondary structures of medaka TR

Fig.5.3. a, Residues conserved in all five teleost TRs are shown in red. Three structural domains (pseudoknot/template, CR4–CR5, and snoRNA) are *outlined* and labeled. On the medaka TR structure, 11 helices (P1, P2a, P2b, P3, P4, P5, P6, P6.1, P7a, P7b, and P8) and every tenth nucleotide of the sequence are labeled. The template region, box H, and ACA motifs are indicated by black boxes.



b. secondary structures of fugu TR

Fig.5.3. continued b, Residues conserved in all five teleost TRs are shown in red. Three structural domains (pseudoknot/template, CR4–CR5, and snoRNA) are outlined and labeled. The template region, box H, and ACA motifs are indicated by black boxes.



c. Vertebrate TRs share a conserved secondary structures

Fig.5.3. continued c, comparison of secondary structures of medaka, human, and shark TRs. The pseudoknot and CR4–CR5 domains are shown in green, whereas the scaRNA domain (or snoRNA domain in the teleost TR) is shown in cyan. The structural determinants (the P1 helix) for template boundary definition are shown in magenta. In the human TR structure, the mammal-specific structural elements required for activity are shown in brown.

Being the smallest, the teleost TR resembles the essential core of vertebrate TR (Fig. 5.3c). It contains shorter linker sequences between the three conserved domains. The commonalities and differences of the vertebrate TR structures are discussed in detail below.

5.4.4.1 Pseudoknot/Template Domain

Pseudoknot/template domain consists of a highly conserved pseudoknot structure, the template sequence, and the P1 helix that defines the boundary of the RNA template. The pseudoknot structure consists of the P2a–P2b and P3 helices that are universally present in vertebrate TRs (Fig. 5.3c). The mammalian pseudoknot, however, contains an additional helix P2a.1 that extends the P2a helix (Fig. 3*B*, *human TR*). This mammal-specific P2a.1 helix is essential for human telomerase activity and is possibly involved in binding to the TERT protein (19). In teleost TR, the P2a and P2b helices are separated by a conserved asymmetric (0/6) internal loop (Fig. 5.3a and b), whereas, in other groups of vertebrates, this internal loop contains a varying number of residues.

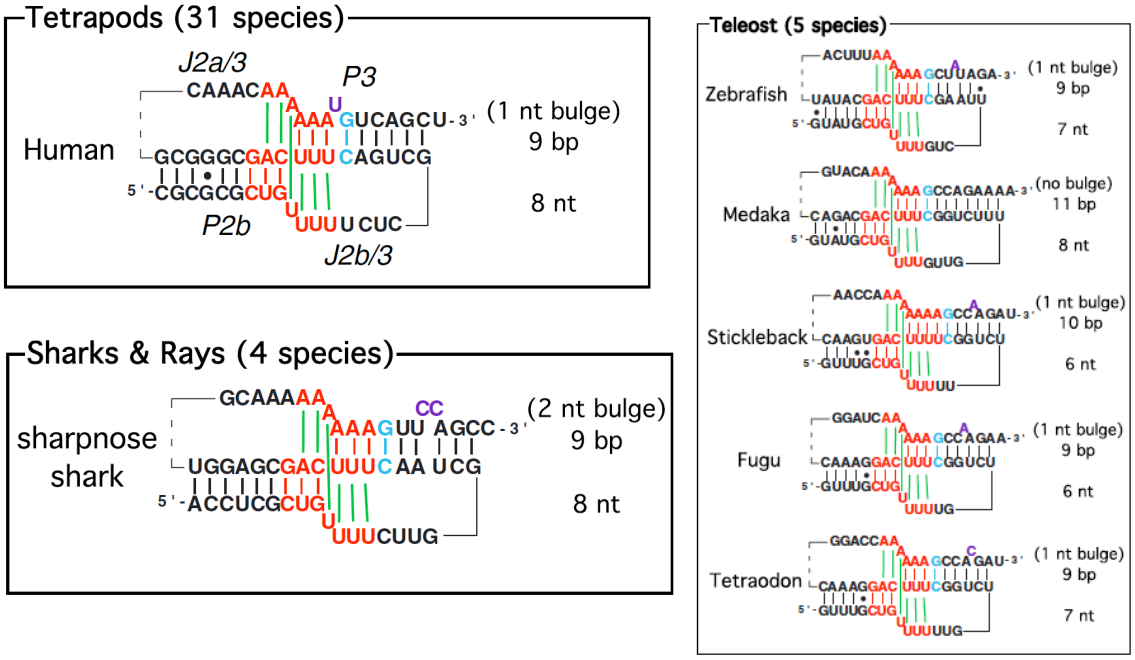
The P3 helix, in tetrapods, is conserved as a 9-base pair helix with a single nucleotide bulge (Fig. 5.4a, tetrapods). The shark and ray P3 helix has the same length but with a 2-nucleotide bulge at a different position (Fig. 5.4a, sharks and rays). Medaka TR interestingly lacks any bulge in its P3 helix, whereas other teleost TRs have a 1-nucleotide bulge at the position identical to the sharks. Notably, the lack of a bulge in the medaka P3 helix seems to be compensated by extensions of the P3 helix and J2b/3 loop (Fig. 5.4a, medaka). The variation of the size and position of the bulge in the P3 helix suggests that it might not be a critical element for the function or structure of the

pseudoknot structure. Deletion of the bulge in the human P3 helix results in a minor reduction of telomerase activity (20, 21). The real role of the P3 bulge has yet to be revealed. Based on an NMR solution structure, the pseudoknot of human TR forms a triple helix that involves 5 base triples and a base pair at the junction of P2b and P3 helices (21). The sequences that form the triple helix are absolutely conserved even in teleost TR, confirming its critical role in telomerase function (Fig. 5.4a). In contrast, the distal portion of the P3 helix and the J2b/3 are less conserved, and are slightly variable in length and sequence (Fig. 5.4a, teleost panel).

In all vertebrates, except for some rodents, TRs possess a long-range interacting P1 helix upstream of the template region (Fig. 5.3). In human TR, the P1 helix consists of two individual helices, P1a and P1b, separated by an internal loop. The teleost P1 helix is substantially shorter, containing only the P1b equivalent portion while lacking the P1a portion. The integrity of P1b helix and its distance from the template defines the boundary of the RNA template (22). In human telomerase, disruption of the P1b helix alters the template boundary, resulting in template usage outside of the normal template. Likewise, disruption of the P1 helix in medaka TR also altered the template boundary and result in decreased processivity (Fig.S5.5). This supports the notion that the P1 helix is also the element for template boundary definition in teleost telomerase.

Comparing pseudoknot fragments from all vertebrate species show that the linker J2a/3 between P2a and P3 stem is extremely flexible (4). The correct folding of pseudoknot structure is presumably important for telomerase catalysis. The minimal length allowed in J2a/3 linker might therefore indicate the distance between P2a stem and

P3 stem when pseudoknot domain folds into a three dimensional structure. The data suggested that 8 nucleotides between P2a and P3 stem is shorter than the minimal length as the activity decreased greatly, while 12, 16 and 20 nucleotides J2a/3 still permit wild-type level of activity (Fig. S5.11).



a. comparison of the triple helix region within the pseudoknot domain

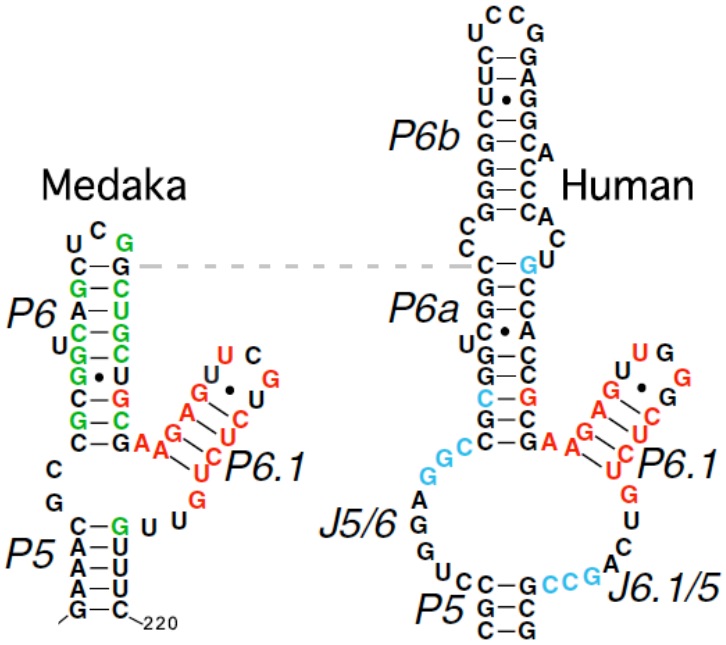
Fig.5.4 Structural comparison of the pseudoknot and CR4–CR5 domains, and sequence alignment of the CR7 domains. a, comparison of the triple helix region within the pseudoknot domain. The schematic of the triple helix region from human (tetrapods), sharpnose shark (cartilaginous), and five teleost are shown, based on an NMR structure reported previously (21). For human, structural elements, etc. P2b, P3, J2a/3, J2b/3, are labeled. The triple helix forming sequence (red) conserved in all species, the bulge (purple) on the P3 helix, and the conserved G-C base pair (cyan) close to the triple helix are highlighted. The green bars indicate the hoogsteen base pair. The size of the bulge, p3 stem, and J2b/3 loop are indicated to the right of the schematics. The dashed line in J2a/3 represents omitted sequences.

5.4.4.2 CR4-CR5 Domain

The CR4–CR5 domain, in addition to the pseudoknot/template domain, is a structural element essential for *in vitro* telomerase activity. The P6 and P6.1 helices in this domain are universally present in all known vertebrate TRs (Fig. 5.3b). Remarkably, the sequence (5'-AAGAGNUNGNCUCUG-3') of the P6.1 stem-loop is highly conserved even in the teleost fish. It was previously thought that the invariant sequence of the P6.1 helix loop was due to a biased sequence collection that resulted from the PCR amplification strategy used for cloning most of the vertebrate TRs (4). This PCR strategy presumably amplified only the TR sequences with conserved sequence in the P6.1 stem-loop, part of the annealing site of the PCR reverse primers. However, all five teleost fish TRs were identified through bioinformatic searches, instead of PCR. The structure, not the sequence, of the P6.1 helix is known to be important for telomerase activity *in vitro* as compensatory mutations that maintain the helical structures of P6.1 do not reduce activity of reconstituted telomerase (15). Surprisingly, similar compensatory mutations of P6.1 helix resulted in reduced telomerase activity reconstituted *in vivo* (23). The absolute sequence conservation in the P6.1 helix suggests that, in addition to its based-paired structure, the sequence of this helix might be also important for the *in vivo* function of telomerase.

The teleost TR, lacking the distal stem-loop P6b, consists of a shorter P6 (i.e. homologous to the P6a in human TR), P.6.1, and P5 helices in the CR4–CR5 domain (Fig. 5.4b). Whereas the P6b helix is dispensable in teleost fish and some tetrapods such as turtle and frog, the proximal part of the P6b stem-loop is required for human

telomerase activity (24). The single stranded regions, J5/6 and J6.1/5, at the three-way junction between P5, P6, and P6.1 helices are relatively more variable in teleost than in other vertebrates. Although its essential role in telomerase function is evident, the mechanistic role of the CR4–CR5 domain remains to be uncovered.

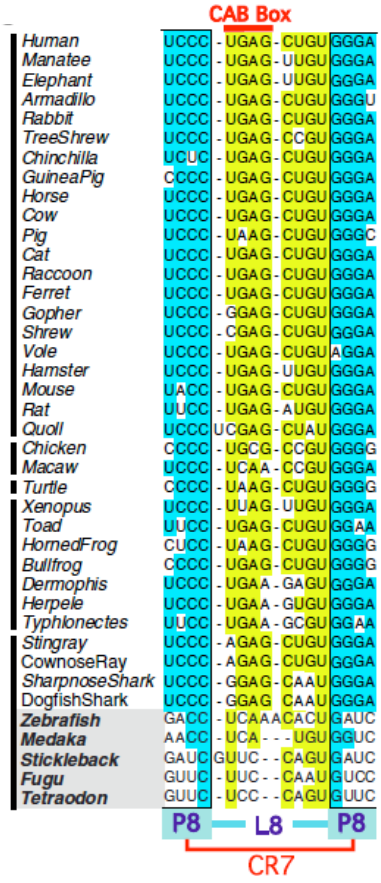


b. comparison of medaka and human CR4–CR5 secondary structure.

Fig.5.4 continued. b, Helixes P5, P6a, P6b, and P6.1 are labeled. Residues in *red* indicate conserved nucleotides in all vertebrates. Nucleotides in *green* indicate conservation in 5 teleost. Whereas nucleotides in *blue* indicate conservation in other vertebrates excluding teleost.

5.4.4.3 SnoRNA/ScaRNA domain

The 3'-portion of vertebrate TR contains a unique secondary structure (hairpin-hinge-hairpin tail) and sequence motifs (box H and ACA) that are critical for TR biogenesis and shared by the box H/ACA snoRNAs (7). Most vertebrate TRs contain an additional motif called the CAB box that is shared by the small Cajal body RNAs (scaRNAs) (3). Whereas the box H and ACA are important for RNA localization to nucleoli, the CAB box is important for localization of the RNA to the Cajal body where RNP complex assembly is thought to take place (25). Interestingly, teleost TR lacks an obvious CAB box (UGAG) in the CR7 region (Fig. 5.4c). The lack of CAB box implies that teleost TR might not localize to the Cajal body. Because the Cajal body has been suggested to play a role in telomerase regulation and telomere recruitment (26), it would be interesting to understand TR localization in teleost and its correlation with the regulation of telomerase function.



c. Teleost TR lacks an obvious CAB box motif (UGAG).

Fig.5.4 continued. c, The five teleost TR sequences shaded in gray are aligned manually with the alignment of 35 non-teleost TR sequences derived from Chen et al. (4). The conserved CAB box is indicated with red lines above the aligned sequences. Residues identical to human sequence are shaded in blue (helix P8) or yellow (loop L8). Dashes (-) denote alignment gaps.

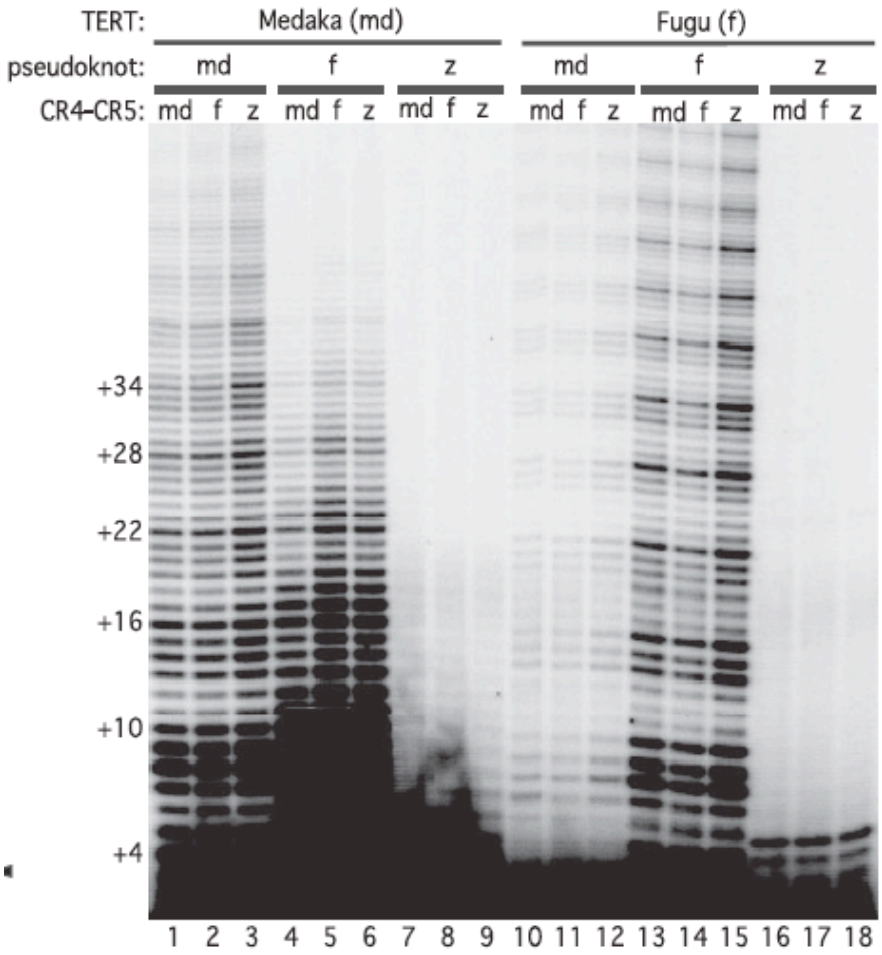
5.4.5 Medaka and Fugu Telomerases Reconstituted in Vitro Are Active and Processive

Telomerase activity reconstituted *in vitro* requires both the TR component and the catalytic TERT protein. To functionally characterize the structural elements of teleost TR, we reconstituted telomerase from *in vitro* synthesized TERT protein and TR (see "Experimental Procedures"). Active telomerases were successfully reconstituted for medaka and fugu, confirming the authenticity of the teleost telomerase components cloned (Fig. 5.5). As predicted from the presence of the 4-nucleotide alignment sequence in their RNA templates, the reconstituted medaka telomerases are processive, generating a typical 6-nucleotide ladder pattern of the elongated products (Fig. 5.5a).

a. Reconstitution of medaka telomerase activity

Fig.5.5 Activity assay of *in vitro* reconstituted teleost telomerase. a, reconstitution of medaka telomerase activity. Telomerase reconstitution was carried out in 10 l of rabbit reticulocyte lysate in the presence of L-[^{35}S]methionine with TR alone (*lane 1*), TERT alone (*lane 2*), TR TERT (*lane 3*), or TR TERT treated with RNaseA (*lane 4*). Reconstituted telomerase were then assayed using the conventional direct assay (see "Materials and Methods"). The *bottom panel* shows the SDS-PAGE analysis of [^{35}S]methionine-labeled medaka TERT protein from each reaction.

Vertebrate TERT protein possesses two RNA-binding sites that bind independently to the CR4–CR5 and pseudoknot domains of the TR. As shown previously, human TERT is functionally compatible with the mouse CR4–CR5 domain but not the mouse pseudoknot domain (14). In this study, we also showed that the medaka and fugu TERT proteins reconstituted telomerase activity with CR4–CR5 RNA fragments, but not the pseudoknot domain, from other teleost fish species (Fig. 5.5b, lanes 1–6 and 10–15) or even distantly related vertebrates such as human, quoll, chicken, turtle, frog, and shark (Fig. S5.6). This difference in cross-species compatibility indicates that the CR4–CR5 domain is functionally more conserved across a wide variety of species than the pseudoknot domain. Unlike the fugu TERT, the medaka TERT assembled with the fugu pseudoknot RNA to reconstitute telomerase activity with a low processivity (Fig. 5.5b, lanes 4–6), suggesting a more relaxed RNA binding specificity of the medaka TERT protein. However, the pseudoknot fragment of zebrafish TR failed to generate telomerase activity when assembled with medaka or fugu TERT proteins (Fig. 5.5b, lanes 7–9 and 16–18), suggesting a cross-species incompatibility of the zebrafish pseudoknot with the TERT protein.



b. medaka and fugu TERT proteins synthesized *in vitro* were assembled with *in vitro* transcribed pseudoknot/template and CR4–CR5 RNA fragments of medaka (md), fugu (f), or zebrafish (z).

Fig.5.5 continued. b, The RNA fragments, medaka pseudoknot (1–150), medaka CR4–CR5 (154 –241), fugu pseudoknot (1–139), fugu CR4–CR5 (143–253), zebrafish pseudoknot (1–134), and zebrafish CR4–CR5 (137–242) were assembled in different combinations with either medaka or fugu recombinant TERT protein as indicated *above* the gel. The assembled telomerases were analyzed for activity using a conventional telomerase assay. The *numbers* on the left (4, 10,16,22,28,34 etc.) indicate the number of nucleotides added to the primer for each major band seen.

To analyze activity of zebrafish telomerase, we thus identified and cloned zebrafish TERT cDNA (see "Materials and Methods"). Unexpectedly, the *in vitro* synthesized zebrafish TERT protein failed to reconstitute a detectable activity when assembled with zebrafish, medaka, or fugu TRs (Fig. S5.7). Based on the alignment of TERT amino acid sequences, the cloned zebrafish TERT protein was unlikely to be an alternative splicing variant, as it contained all essential motifs. The possibility of mutations in the cloned zebrafish TERT gene was ruled out as identical sequences were found from two individual zebrafish obtained from different sources. Whereas gene duplication is relatively common in teleost, more rigorous BLAST searches of the zebrafish genome did not reveal any other candidate sequences for the TERT gene.Wespeculate that the *in vitro* synthesized zebrafish TERT protein, unlike the medaka and fugu TERT proteins, might not fold correctly as the recombinant zebrafish TERT protein migrated faster than expected on SDS-PAGE (Fig. S5.8).

5.4.6 The CR4–CR5 Domain Is the Main Determinant in TR for Functional Binding to Medaka TERT

During reconstituting teleost fish telomerase, we observed a significantly lower activity of the reconstituted enzyme using two RNA fragments than that of the enzyme reconstituted using the full-length RNA (Fig. S5.9). To determine which RNA fragment was responsible for the lower reconstituted activity, we carried out the *in vitro* reconstitution with titrations of each of the two RNA fragments as well as the full-length RNA. We define the median effective concentration (or EC_{50}) as the RNA concentration required to generate 50% of the saturated activity of reconstituted telomerase. It is noteworthy that this EC_{50} value measured in this assay is related only to the functional binding (or assembly) of the RNA fragment to the TERT protein, excluding nonspecific or non-functional bindings. A lower EC_{50} value of the RNA indicates that the RNA assembles more efficiently with the TERT protein to generate active telomerase. Remarkably, the CR4–CR5 fragments and the full-length TR gave rise to comparable EC_{50} values. The medaka CR4–CR5 and fulllength RNAs had similar EC_{50} values of 87.4 and 85.9 nM, respectively, whereas the human CR4–CR5 and full-length RNAs had EC_{50} values of 203.9 and 241.6 nM, respectively (Fig. 5.6). In comparison, the medaka and human pseudoknot RNA fragments had high EC_{50} values of 506.2 and 523.5 nM, respectively (Fig. 5.6). The reduction of reconstituted activity at high concentrations of the full-length TR might be due to the multimerization or aggregation of TR as previously reported (27). Our result indicates that the CR4–CR5 domain is the main determinant for efficient binding and assembly of TR to the TERT protein.

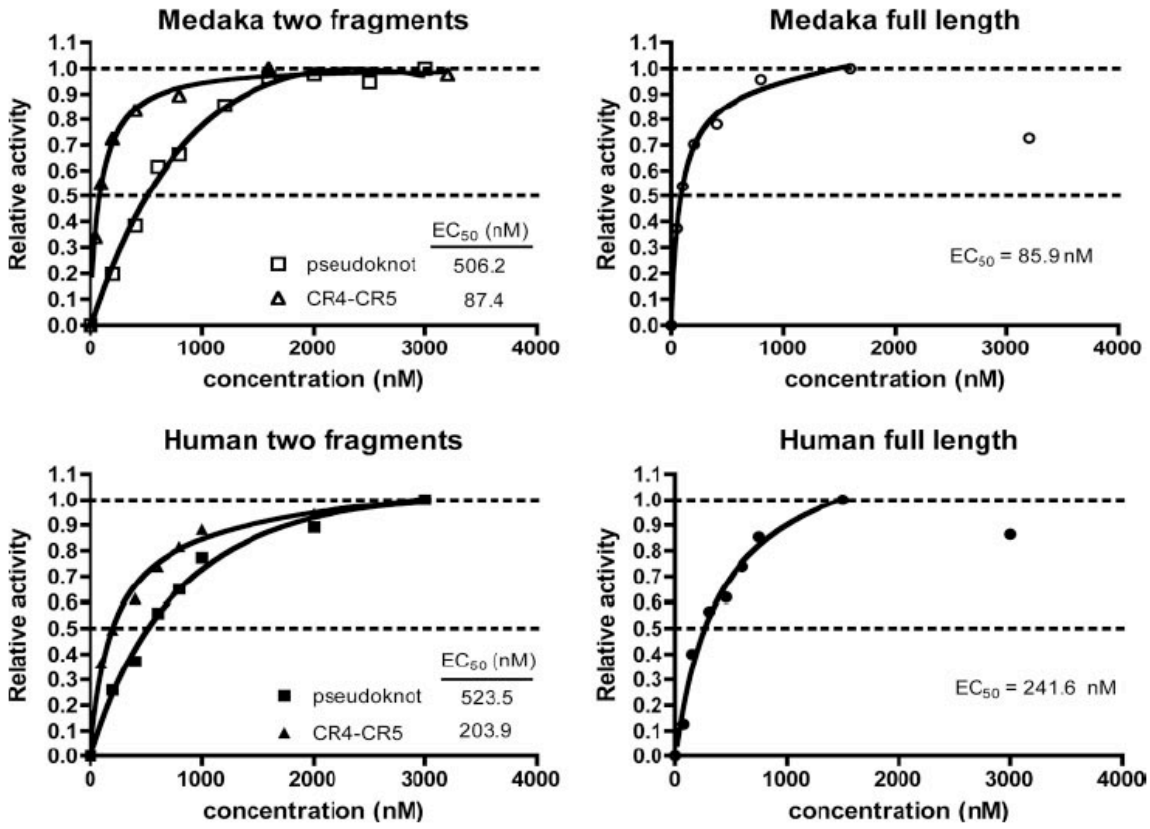


Fig. 5.6 Effective concentrations of the pseudoknot and CR4–CR5 domains to assemble active telomerase in vitro. Titration experiments were performed with pseudoknot and CR4–CR5 RNA fragments or fulllength TR alone for reconstituting medaka (upper panel) and human (lower panel) telomerase enzymes. Various concentrations of pseudoknot or CR4–CR5 RNA fragments were assembled with the other RNA fragment at a saturated 3μM and the in vitro synthesized TERT protein, followed by the conventional telomerase assay. The pseudoknot (medaka, nt 1–150 and human, nt 32–195) and CR4–CR5 (medaka, nt 170–220 and human, nt 241–328) RNA fragments were titrated as indicated. The relative activity represents the ratio of total activity of each reaction over the total activity of the reaction with saturated concentrations of both RNA fragments. The median effective concentration (EC_{50}) values of each RNA fragment are indicated.

5.5 Discussion

Unlike the TERT, TR is prominently divergent in size, sequence, and even structure. In this study, by using a novel bioinformatics approach, we have successfully identified TR sequences from five teleost genomes. The structural and functional analyses of teleost fish telomerase provide important insights into the structural evolution of vertebrate TR as well as the co-evolution of the TR and TERT protein.

5.5.1 Fast Evolution of TR Structure and Size

Because of the various numbers of species-specific structural elements, the size of TR is remarkably variable, up to 1 order of magnitude, from 150 nt in ciliates to 1500 nt in yeasts. From the evolutionary point of view, the emergence or disappearance of structural elements in TR over a short evolutionary time scale is rather intriguing. The unusual plasticity of TR structure was likely facilitated by the non-lethal and progressive nature of the consequences of TR mutations. In organisms with long telomeres, the impact of telomerase mutations is delayed for a number of generations (28). Such delay could allow an accumulation of secondary mutations, some of which might compensate for the initial deleterious mutation, eventually leading to emergence of novel structural elements in TRs.

A possible scenario for the emergence of new structural elements is the insertion of a transposable element into the *TR* gene during evolution. For example, the scaRNA or snoRNA domains in the vertebrate TR is absent in both the ciliate and yeast TRs, and has been acquired during evolution along the vertebrate lineage. As some snoRNA and scaRNA contain characteristics of retrotransposons (29), it is possible that a transposition

event may have occurred and fused a mobile scaRNA gene with an ancestral TR gene. Because most vertebrates, including the early branched cartilaginous fish, contain the scaRNA specific motif (CAB box), we propose that it was a scaRNA, rather than a snoRNA, that was inserted into the vertebrate *TR* gene. Teleost fish and some bird TRs that lack an obvious CAB box, might have subsequently evolved to function without a CAB box motif. Notably, other scaRNAs, *e.g.* U100, from teleost fish contain a conserved CAB box sequence (30). Identification of TRs from early branching chordates such as the sea squirt will provide crucial clues about the origin of the vertebrate-specific structural domains. Based on the phylogenetic tree derived from the aligned TR sequences, tetrapods, teleost fishes, and cartilaginous fishes are grouped into three monophyletic clades (Fig. 5.7), representing three separated evolutionary lineages that lead to three distinct size groups of TR molecules. Cartilaginous and teleost fish TRs evolved in opposite directions toward size expansion and reduction, respectively, corresponding to their genome size evolution. The small sizes of teleost genomes are mainly due to the low abundance of transposable elements and the significant reduction in intron size (31). Our data suggest that genome compression affected not only the intergenic or intronic DNA sequences but also the RNA genes. Similarly, teleost RNaseP RNA is about 50 nt shorter than the 350-nt long human RNaseP RNA.

Interestingly, teleost fish TR appears to be more divergent than cartilaginous fish TR from tetrapod TR (Fig. 5.7). This is consistent with a recent comparative genomic study that showed a higher degree of sequence conservation between the human and elephant shark genomes than that of human and teleost fish genomes (32, 33). It is

generally believed that the teleost fish has experienced a genome duplication after diverging from tetrapod lineage and before the fish radiation (34). However, no extra *TR* gene or pseudoknot gene was found in the 5 teleost fish species, suggesting either the teleost *TR* gene was not duplicated or the duplicated TR copy has been lost from the common ancestor of teleost fish.

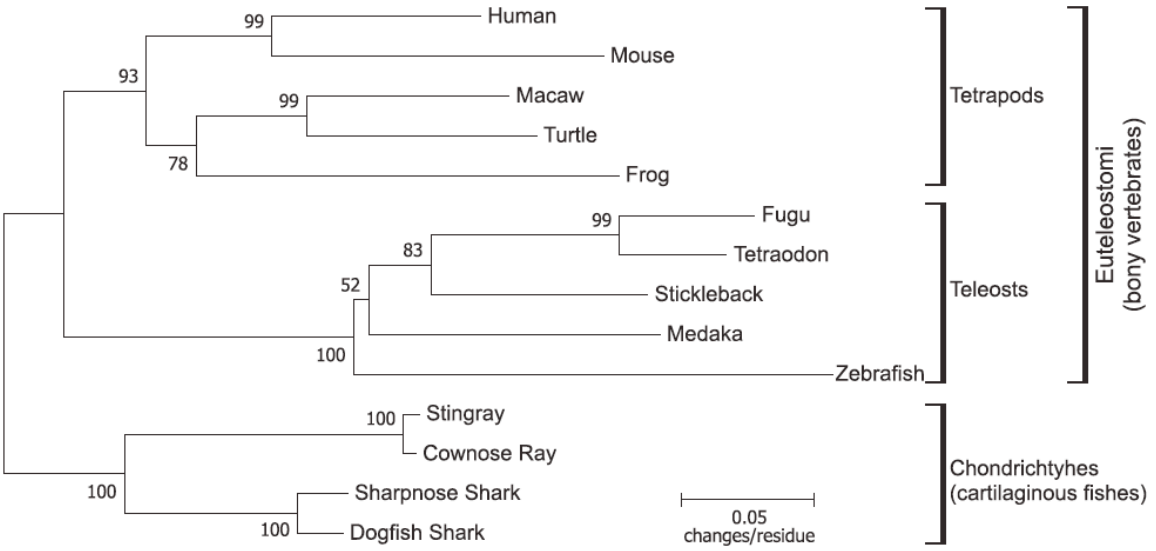


Fig. 5.7 The neighbor-joining tree inferred from the vertebrate TR sequences. The tree was derived using the neighbor-joining method from the aligned TR sequences of 14 vertebrates including 5 tetrapods (human, mouse, macaw, turtle, and frog), 5 teleost fishes (fugu, tetraodon, stickleback, medaka, and zebrafish), and 4 cartilaginous fishes (stingray, cownose ray, sharpnose shark, and dogfish shark). The phylogenetic tree was constructed using the program MEGA3.1 (37). The number next to each node indicates a value as a percentage of 1000 bootstrap replicates. Branch lengths are proportional to the number of residue changes. Scale bar indicates an evolutionary distance of 0.05 nucleotide substitution per position in the sequence.

5.5.2 Co-evolution of the TR and TERT Protein

During structural diversification, the function of the telomerase RNP has to be conserved through co-evolution between the RNA and protein components, which can be reflected by the interspecies compatibility of the components. For example, the CR4–CR5 RNA fragments from distantly related species such as human were able to reconstitute telomerase activity with medaka TERT (Fig. 5.5 and Fig. S5.6). In contrast, the pseudoknot/template RNA domain appears to be incompatible even between closely related species (e.g. between medaka and fugu, or between human and mouse), suggesting a faster rate of co-evolution between the pseudoknot RNA domain and the TERT protein.

The triple helix within the pseudoknot domain contains invariant sequences and is one of the most conserved structural elements in vertebrate TRs (Fig. 5.4a). As the triple helix seems to be an ancient feature conserved in many species (2, 21, 35), it is, thus, unlikely to be responsible for the interspecies incompatibility of the pseudoknot domains. The distal helix of P3 stem and J2b/3 loop, on the other hand, demonstrate some extent of variation among vertebrate species (Fig. 5.4a). Swapping the whole pseudoknot structure (P3, P2b, and J2b/3) between medaka and fugu TRs did not improve their inter-species compatibility (Fig. S5.10).

The teleost CR4–CR5 domain is considerably smaller than other vertebrates as it lacks the distal P6b helix. Nonetheless, the smaller medaka CR4–CR5 RNA fragment (50 nt) exceeds its human counterpart (89 nt) in effectiveness of reconstituting telomerase activity *in vitro* (Fig. 5.6). The higher assembly efficiency is likely due to a higher

binding affinity between the medaka TERT protein and the CR4–CR5 RNA fragment, which would require substantial co-evolution between the medaka TERT protein and the TR. Because the P6b helix in the CR4–CR5 domain of human TR is essential for binding to the human TERT protein (24), the human TERT might have evolved with an additional binding pocket for the P6b helix.

Whereas we were able to reconstitute activity from medaka and fugu telomerases, it is unclear why the zebrafish TERT failed to reconstitute detectable telomerase activity. Among the five teleost species studied, zebrafish branches out early and is more divergent than the other four teleost fishes (36).

In summary, the identification of teleost TR and characterization of its structure and function reveal an unusual divergence of vertebrate TR. The novel bioinformatic tool fragrep2 is an effective approach to find notoriously divergent TR sequences in eukaryotic genomes. The small teleost fish TR and the large cartilaginous fish TR reflect the unusual plasticity of TR structure during evolution. Teleost fish telomerase is very processive and contains a functional P1 helix that defines the template boundary. The conservation of the structure and function of teleost fish telomerase supports the use of teleost fish as a model organism for the study of telomerase biology.

5.6 References

1. Ferreira, M. G., Miller, K. M., and Cooper, J. P. (2004) *Mol. Cell* 13, 7–18

2. Chen, J.-L., and Greider, C. W. (2004) *Proc. Natl. Acad. Sci. U.S. A.* 101,14683–14684

3. Jady, B. E., Bertrand, E., and Kiss, T. (2004) *J. Cell Biol.* 164, 647–652

4. Chen, J.-L., Blasco, M. A., and Greider, C. W. (2000) *Cell* 100, 503–514

5. Chen, J.-L., and Greider, C. W. (2004) *Trends Biochem. Sci.* 29, 183–192

6. Tesmer, V. M., Ford, L. P., Holt, S. E., Frank, B. C., Yi, X., Aisner, D. L.,Ouellette, M., Shay, J. W., and Wright, W. E. (1999) *Mol. Cell. Biol.* 19, 6207–6216

7. Mitchell, J. R., Cheng, J., and Collins, K. (1999) *Mol. Cell. Biol.* 19, 567–576

8. Nelson, J. S. (2006) *Fishes of the World*, 4th Ed., Wiley, New York

9. Kel, A. E., Gossling, E., Reuter, I., Cheremushkin, E., Kel- Margoulis, O. V., and Wingender, E. (2003) *Nucleic Acids Res.* 31, 3576–3579

10. Mosig, A., Chen, J.-L., and Stadler, P. F. (2007) *Lect. Notes Comput. Sci.* 4645, 335–345

11. Eddy, S. R. (2006) Cold Spring Harbor Symp. Quant. Biol. 71, 117–128

12. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003) Nucleic Acids Res. 31, 439–441

13. Yap, W. H., Yeoh, E., Brenner, S., and Venkatesh, B. (2005) *Gene* (*Amst*.) 353, 207–217

14. Chen, J.-L., and Greider, C. W. (2003) *EMBO J.* 22, 304–314

15. Chen, J.-L., Opperman, K. K., and Greider, C. W. (2002) *Nucleic Acids Res.*30, 592–597

16. Mosig, A., Sameith, K., and Stadler, P. F. (2006) *Genomics Proteomics Bioinformatics* 4, 56–60

17. Backofen, R., Bernhart, S. H., Flamm, C., Fried, C., Fritzsch, G., Hackermuller, J., Hertel, J., Hofacker, I. L., Missal, K., Mosig, A., Prohaska, S. J., Rose, D., Stadler, P. F., Tanzer, A., Washietl, S., and Will, S. (2007) *J. Exp. Zool. Part B* 308, 1–25

18. Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch,I.J., Murray, B. G., Kapraun, D. F., Greilhuber, J., and Bennett, M. D. (2007) *Nucleic Acids Res.* 35, D332–D338

19. Ly, H., Blackburn, E. H., and Parslow, T. G. (2003) *Mol. Cell. Biol.* 23,6849–6856

20. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H., and James, T. L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 16998–17003

21. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) *Mol. Cell* 17, 671–682

22. Chen, J.-L., and Greider, C. W. (2003) *Genes Dev.* 17, 2747–2752

23. Ly, H., Calado, R. T., Allard, P., Baerlocher, G. M., Lansdorp, P. M., Young,N. S., and Parslow, T. G. (2005) *Blood* 105, 2332–2339

24. Mitchell, J. R., and Collins, K. (2000) *Mol. Cell* 6, 361–371

25. Richard, P., Darzacq, X., Bertrand, E., Jady, B. E., Verheggen, C., and Kiss,T. (2003) *EMBO J.* 22, 4283–4293

26. Tomlinson, R. L., Ziegler, T. D., Supakorndej, T., Terns, R. M., and Terns, M. P. (2006) Mol. Biol. Cell 17, 955–965

27. Ren, X., Gavory, G., Li, H., Ying, L., Klenerman, D., and Balasubramanian, S. (2003) Nucleic Acids Res. 31, 6509–6515

28. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) *Cell* 91, 25–34

29. Weber, M. J. (2006) *PLoS Genet.* 2, e205

30. Vitali, P., Royo, H., Seitz, H., Bachellerie, J. P., Huttenhofer, A.,and Cavaille, J. (2003) *Nucleic Acids Res.* 31, 6543–6551

31. Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J. P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K. J., McEwan, P., Bosak, S., Kellis, M., Volff, J. N., Guigo, R., Zody, M. C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E. S., Weissenbach, J., and Roest Crollius, H. (2004) *Nature* 431, 946–957

32. Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., Lee, A. P., Johnson, J., Dandona, N., Viswanathan, L. D., Tay, A., Venter, J. C., Strausberg, R. L., and Brenner, S. (2006) *Science* 314, 1892

33. Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., Lee, A. P., Johnson, J., Dandona, N., Viswanathan, L. D., Tay, A., Venter, J. C., Strausberg, R. L., and Brenner, S. (2007) *PLoS Biol.* 5, e101

34. Meyer, A., and Van de Peer, Y. (2005) *Bioessays* 27, 937–945

35. Shefer, K., Brown, Y., Gorkovoy, V., Nussbaum, T., Ulyanov, N. B., and Tzfati, Y. (2007) *Mol. Cell. Biol.* 27, 2130–2143

36. Benton, M. J., and Donoghue, P. C. (2007) *Mol. Biol. Evol.* 24, 26–53

37. Kumar, S., Tamura, K., and Nei, M. (2004) *Brief Bioinform.* 5, 150–163

CHAPTER 6

CONCLUSION

Processive addition of telomeric repeats by the enzyme telomerase is a complicated but delicate process that involves a series of well-coordinated movements within the telomerase core components. The detailed steps behind telomerase processivity remain to be revealed. However, the fact that telomerase can perform multiple repeat addition before complete dissociation from the substrate suggests that a template translocation event must occur between each repeat addition to provide an empty template for the following repeat synthesis.

Many conserved motifs in telomerase core components were indicated to involve in processivity. Our work based on a comprehensive sequence alignment has identified a processive-telomerase specific TERT motif, name motif 3, in the RT domain. This motif has been overlooked in the past because of the inefficient alignment algorithm and the lack of TERT sequences. We have associated telomerase processivity defect with its duplex binding ability as motif 3 low processivity mutants failed to extend short telomeric primers. Because the short primer lacks an upstream region to interact with other parts of the protein, defect of using short primer thus imply a defect of binding RNA template/DNA primer in the telomerase active site. One other intriguing phenomenon observed in motif 3 study is that telomerase repeat addition processivity and addition rate are functionally independent from each other. Motif 3 also contributes to repeat addition rate by regulating the translocation rate.

To take one step further. A novel template free telomerase was designed to directly test duplex binding by telomerase active site. Comparison between template free telomerase and other RTs using duplex substrate indicates that telomerase catalytic core has adapted for short repeat synthesis. More importantly, processivity defect mutants, whose translocation efficiency is significantly lower than wild-type, showed lower affinity towards telomeric RNA/DNA duplex. Using a 5 bp RNA/DNA duplex in wildtype telomerase reaction largely reduced telomerase processivity. This data strongly argue for duplex releasing/binding as a critical step during template translocation. We thus described the translocation event to a greater detail.

Vertebrate telomerase is a great system to study telomerase processivity because of its conserved 6 nucleotides repeat sequence and its relatively processive reaction. Teleost fish telomerase components were identified to broaden the knowledge of vertebrate telomerases. Teleost fish is a fast evolving group since branching from the tetrapod lineage. The telomerase RNA represents some unique properties compared to other vertebrates. Notebably, fish TR is the smallest vertebrate TR identified to date, yet still retained all the essential conserved regions important for function, providing a valuable model for structural study of telomerase RNP.

REFERENCES

Chapter 1 References

1. Watson, G., and Paigen, K. (1972) *Nat New Biol* 239, 120-122

2. 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) *Science* 266, 2011-2015

3. Muller, HJ. (1938) *Collecting Net* 13, 181-198

4. McClintock, B. (1939) *Proc Natl Acad Sci U S A* 25, 405-416

5. Blackburn, E. H., and Gall, J. G. (1978) *J Mol Biol* 120, 33-53

6. Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981) *Proc Natl Acad Sci U S A* 78, 3015-3019

7. Walmsley, R. W., Chan, C. S., Tye, B. K., and Petes, T. D. (1984) *Nature* 310, 157-160

8. Greider, C. W., and Blackburn, E. H. (1985) *Cell* 43, 405-413

9. Greider, C. W., and Blackburn, E. H. (1987) *Cell* 51, 887-898

10. Greider, C. W., and Blackburn, E. H. (1989) *Nature* 337, 331-337

11. Romero, D. P., and Blackburn, E. H. (1991) *Cell* 67, 343-353

12. Singer, M. S., and Gottschling, D. E. (1994) *Science* 266, 404-409

13. 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., and et al. (1995) *Science* 269, 1236-1241

14. Chen, J. L., Blasco, M. A., and Greider, C. W. (2000) *Cell* 100, 503-514

15. Dandjinou, A. T., Levesque, N., Larose, S., Lucier, J. F., Abou Elela, S., and Wellinger, R. J. (2004) *Curr Biol* 14, 1148-1158

16. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T. R. (1997) *Science* 276, 561-567

17. 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) *Science* 277, 955-959

18. Autexier, C., and Lue, N. F. (2006) *Annu Rev Biochem* 75, 493-517

19. Lundblad, V., and Szostak, J. W. (1989) *Cell* 57, 633-643

20. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) *Cell* 91, 25-34

21. Shay, J. W., and Wright, W. E. (2005) *Carcinogenesis* 26, 867-874

22. Hayflick, L. (1998) *Keio J Med* 47, 174-182

23. Collins, K., and Mitchell, J. R. (2002) *Oncogene* 21, 564-579

24. Li, S., Crothers, J., Haqq, C. M., and Blackburn, E. H. (2005) *J Biol Chem* 280, 23709-23717

25. 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) *Science* 279, 349-352

26. Blasco, M. A. (2005) *Nat Rev Genet* 6, 611-622

27. Armanios, M. (2009) *Annu Rev Genomics Hum Genet* 10, 45-61

28. Leeper, T., Leulliot, N., and Varani, G. (2003) *Nucleic Acids Res* 31, 2614-2621

29. Leeper, T. C., and Varani, G. (2005) *RNA* 11, 394-403

30. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) *Mol Cell* 17, 671-682

31. Chen, Y., Fender, J., Legassie, J. D., Jarstfer, M. B., Bryan, T. M., and Varani, G. (2006) *EMBO J* 25, 3156-3166

32. Richards, R. J., Theimer, C. A., Finger, L. D., and Feigon, J. (2006) *Nucleic Acids Res* 34, 816-825

33. Jacobs, S. A., Podell, E. R., and Cech, T. R. (2006) *Nat Struct Mol Biol* 13, 218-225

34. Rouda, S., and Skordalakes, E. (2007) *Structure* 15, 1403-1412

35. Gillis, A. J., Schuller, A. P., and Skordalakes, E. (2008) *Nature* 455, 633-637

36. Greider, C. W. (1991) *Mol Cell Biol* 11, 4572-4580

37. Prowse, K. R., Avilion, A. A., and Greider, C. W. (1993) *Proc Natl Acad Sci U S A* 90, 1493-1497

38. Chen, J. L., and Greider, C. W. (2003) *EMBO J* 22, 304-314

39. Lue, N. F., and Peng, Y. (1997) *Nucleic Acids Res* 25, 4331-4337

40. Fulton, T. B., and Blackburn, E. H. (1998) *Mol Cell Biol* 18, 4961-4970

41. Cohn, M., and Blackburn, E. H. (1995) *Science* 269, 396-400

42. Hammond, P. W., and Cech, T. R. (1997) *Nucleic Acids Res* 25, 3698-3704

43. Hammond, P. W., and Cech, T. R. (1998) *Biochemistry* 37, 5162-5172

44. Sun, D., Lopez-Guajardo, C. C., Quada, J., Hurley, L. H., and Von Hoff, D. D. (1999) *Biochemistry* 38, 4037-4044

45. Maine, I. P., Chen, S. F., and Windle, B. (1999) *Biochemistry* 38, 15325-15332

46. Hardy, C. D., Schultz, C. S., and Collins, K. (2001) *J Biol Chem* 276, 4863-4871

47. Jarstfer, M. B., and Cech, T. R. (2002) *Biochemistry* 41, 151-161

48. Aigner, S., and Cech, T. R. (2004) *RNA* 10, 1108-1118

49. Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., and Lei, M. (2007) *Nature* 445, 506-510

50. Autexier, C., and Greider, C. W. (1995) *Genes Dev* 9, 2227-2239

51. Hammond, P. W., Lively, T. N., and Cech, T. R. (1997) *Mol Cell Biol* 17, 296-308

52. Lue, N. F. (2005) *J Biol Chem* 280, 26586-26591

53. Romi, E., Baran, N., Gantman, M., Shmoish, M., Min, B., Collins, K., and Manor, H. (2007) *Proc Natl Acad Sci U S A* 104, 8791-8796

54. Lue, N. F., and Li, Z. (2007) *Nucleic Acids Res* 35, 5213-5222

55. Moriarty, T. J., Marie-Egyptienne, D. T., and Autexier, C. (2004) *Mol Cell Biol* 24, 3720-3733

56. Moriarty, T. J., Ward, R. J., Taboski, M. A., and Autexier, C. (2005) *Mol Biol Cell* 16, 3152-3161

57. Wyatt, H. D., Lobb, D. A., and Beattie, T. L. (2007) *Mol Cell Biol* 27, 3226-3240

58. Zaug, A. J., Podell, E. R., and Cech, T. R. (2008) *Nat Struct Mol Biol* 15, 870-872

59. Finger, S. N., and Bryan, T. M. (2008) *Nucleic Acids Res* 36, 1260-1272

60. Huard, S., Moriarty, T. J., and Autexier, C. (2003) *Nucleic Acids Res* 31, 4059-4070

61. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2000) *J Biol Chem* 275, 24199-24207

62. Lue, N. F., Lin, Y. C., and Mian, I. S. (2003) *Mol Cell Biol* 23, 8440-8449

Chapter 2 References

1. Shawi, M. and Autexier, C. (2008) 129, 3-10.

2. Armanios, M.Y., Chen, J.J.L., Cogan, J.D., Alder, J.K., Ingersoll, R.G., Markin, C., Lawson, W.E., Xie, M., Vulto, I., Phillips, J.A., 3rd *et al.* (2007) *N. Engl. J. Med.*, 356, 1317-1326.

3. Walne, A.J. and Dokal, I. (2008) *Mechanisms of ageing and development*, 129, 48-59.

4. Autexier, C. and Lue, N.F. (2006) *Annu. Rev. Biochem.*, 75, 493-517.

5. Jacobs, S.A., Podell, E.R. and Cech, T.R. (2006) *Nat. Struct. Mol. Biol.*, 13, 218-225.

6. Osanai, M., Kojima, K.K., Futahashi, R., Yaguchi, S. and Fujiwara, H. (2006) *Gene*, 376, 281-289.

7. Xiong, Y. and Eickbush, T.H. (1990) *EMBO J.*, 9, 3353-3362.

8. Greider, C.W. (1991) *Mol. Cell. Biol.*, 11, 4572-4580.

9. Morin, G.B. (1989) *Cell*, 59, 521-529.

10. Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F. and Chen, J.J.L. (2008) *J. Biol. Chem.*, 283, 2049-2059.

11. Cohn, M. and Blackburn, E.H. (1995) *Science (New York, N.Y*, 269, 396-400.

12. Prowse, K.R., Avilion, A.A. and Greider, C.W. (1993) *Proc. Natl. Acad. Sci. USA*, 90, 1493-1497.

13. Lue, N.F. and Peng, Y. (1997) *Nucleic Acids Res.*, 25, 4331-4337.

14. Bosoy, D. and Lue, N.F. (2004) *Nucleic Acids Res.*, 32, 93-101.

15. Wang, F., Podell, E.R., Zaug, A.J., Yang, Y., Baciu, P., Cech, T.R. and Lei, M. (2007) *Nature*, 445, 506-510.

16. Aigner, S. and Cech, T.R. (2004) *RNA*, 10, 1108-1118.

17. Sun, D., Lopez-Guajardo, C.C., Quada, J., Hurley, L.H. and Von Hoff, D.D. (1999) *Biochemistry*, 38, 4037-4044.

18. Harrington, L.A. and Greider, C.W. (1991) *Nature*, 353, 451-454.

19. Zaug, A.J., Podell, E.R. and Cech, T.R. (2008) *Nat. Struct. Mol. Biol.*, 15, 870-872.

20. Lue, N.F., Lin, Y.C. and Mian, I.S. (2003) *Mol. Cell. Biol.*, 23, 8440-8449.

21. Bryan, T.M., Goodrich, K.J. and Cech, T.R. (2000) *J. Biol. Chem.*, 275, 24199-24207.

22. Huard, S., Moriarty, T.J. and Autexier, C. (2003) *Nucleic Acids Res.*, 31, 4059-4070.

23. Chen, J.-L. and Greider, C.W. (2003) *EMBO J.*, 22, 304-314.

24. Gavory, G., Farrow, M. and Balasubramanian, S. (2002) *Nucleic Acids Res*, 30, 4470-4480.

25. Lai, C.K., Miller, M.C. and Collins, K. (2003) *Mol. Cell*, 11, 1673-1683.

26. Moriarty, T.J., Marie-Egyptienne, D.T. and Autexier, C. (2004) *Mol. Cell. Biol.*, 24, 3720-3733.

27. Drosopoulos, W.C., Direnzo, R. and Prasad, V.R. (2005) *J. Biol. Chem.*, 280, 32801-32810.

28. Podlevsky, J.D., Bley, C.J., Omana, R.V., Qi, X. and Chen, J.J.L. (2008) *Nucleic Acids Res*, 36, D339-343.

29. Ge, L. and Rudolph, P. (1997) *Biotechniques*, 22, 28-30.

30. Alder, J.K., Chen, J.J.L., Lancaster, L., Danoff, S., Su, S.C., Cogan, J.D., Vulto, I., Xie, M., Qi, X., Tuder, R.M. *et al.* (2008) *Proc. Natl. Acad. Sci. USA*, 105, 13051-13056.

31. Cristofari, G., Adolf, E., Reichenbach, P., Sikora, K., Terns, R.M., Terns, M.P. and Lingner, J. (2007) *Mol. Cell*, 27, 882-889.

32. Cristofari, G. and Lingner, J. (2006) *EMBO J.*, 25, 565-574.

33. Li, Y., Yates, J.A. and Chen, J.J.L. (2007) *Gene*, 400, 16-24.

34. Gillis, A.J., Schuller, A.P. and Skordalakes, E. (2008) *Nature*, 455, 633-637.

35. Malik, H.S. and Eickbush, T.H. (1998) *Mol Biol Evol*, 15, 1123-1134.

36. Zimmerly, S., Hausner, G. and Xc, W. (2001) *Nucleic Acids Res*, 29, 1238-1250.

37. Arkhipova, I.R. (2006) *Syst Biol*, 55, 875-885.

38. Liang, J., Yagasaki, H., Kamachi, Y., Hama, A., Matsumoto, K., Kato, K., Kudo, K. and Kojima, S. (2006) *Haematologica*, 91, 656-658.

39. Yamaguchi, H., Calado, R.T., Ly, H., Kajigaya, S., Baerlocher, G.M., Chanock, S.J., Lansdorp, P.M. and Young, N.S. (2005) *N. Engl. J. Med.*, 352, 1413-1424.

40. Moriarty, T.J., Ward, R.J., Taboski, M.A. and Autexier, C. (2005) *Mol. Biol. Cell*, 16, 3152-3161.

41. Wyatt, H.D., Lobb, D.A. and Beattie, T.L. (2007) *Mol. Cell. Biol.*, 27, 3226-3240.

42. Greider, C.W. (1995) In Blackburn, E. H. and Greider, C. W. (eds.), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 35-68.

43. Forstemann, K. and Lingner, J. (2005). *EMBO reports*, 6, 361-366.

44. Finger, S.N. and Bryan, T.M. (2008) *Nucleic Acids Res.*, 36, 1260-1272.

Chapter 3 References

1. Meyerson, M., Counter, C. M., Eaton, E. N., Ellisen, L. W., Steiner, P., Caddle, S. D., Ziaugra, L., Beijersbergen, R. L., Davidoff, M. J., Liu, Q., Bacchetti, S., Haber, D. A., and Weinberg, R. A. (1997) *Cell* 90, 785-795

2. 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) *Science* 266, 2011-2015

3. Chen, J. L., and Greider, C. W. (2004) *Proc Natl Acad Sci U S A* 101, 14683-14684

4. Qiao, F., and Cech, T. R. (2008) *Nat Struct Mol Biol* 15, 634-640

5. Collins, K. (2008) *Mech Ageing Dev* 129, 91-98

6. Tesmer, V. M., Ford, L. P., Holt, S. E., Frank, B. C., Yi, X., Aisner, D. L., Ouellette, M., Shay, J. W., and Wright, W. E. (1999) *Mol Cell Biol* 19, 6207-6216

7. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) *Mol Cell* 17, 671-682

8. Leeper, T., Leulliot, N., and Varani, G. (2003) *Nucleic Acids Res* 31, 2614-2621

9. Cristofari, G., and Lingner, J. (2006) *EMBO J* 25, 565-574

10. Autexier, C., and Lue, N. F. (2006) *Annu Rev Biochem* 75, 493-517

11. Jacobs, S. A., Podell, E. R., and Cech, T. R. (2006) *Nat Struct Mol Biol* 13, 218-225

12. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2000) *Mol Cell* 6, 493-499

13. Moriarty, T. J., Huard, S., Dupuis, S., and Autexier, C. (2002) *Mol Cell Biol* 22, 1253-1265

14. Peng, Y., Mian, I. S., and Lue, N. F. (2001) *Mol Cell* 7, 1201-1211

15. Huard, S., Moriarty, T. J., and Autexier, C. (2003) *Nucleic Acids Res* 31, 4059-4070

16. 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) *Science* 277, 955-959

17. Gillis, A. J., Schuller, A. P., and Skordalakes, E. (2008) *Nature* 455, 633-637

18. Forstemann, K., and Lingner, J. (2005) *EMBO Rep* 6, 361-366
19. Finger, S. N., and Bryan, T. M. (2008) *Nucleic Acids Res* 36, 1260-1272

20. Chen, J. L., and Greider, C. W. (2003) *EMBO J* 22, 304-314

21. Abbotts, J., Bebenek, K., Kunkel, T. A., and Wilson, S. H. (1993) *J Biol Chem* 268, 10312-10323

22. Telestisky & Goff (1997) Reverse transcriptase and the generation of retroviral DNA. in book *Retroviruses*

23. Collins, K., and Greider, C. W. (1995) *EMBO J* 14, 5422-5432

24. Legassie, J. D., and Jarstfer, M. B. (2005) *Biochemistry* 44, 14191-14201

25. Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W. C., and Masutomi, K. (2009) *Nature* 461, 230-235

26. Miller, M. C., and Collins, K. (2002) *Proc Natl Acad Sci U S A* 99, 6585-6590

27. Lue, N. F., Bosoy, D., Moriarty, T. J., Autexier, C., Altman, B., and Leng, S. (2005) *Proc Natl Acad Sci U S A* 102, 9778-9783

28. Greider, C. W. (1991) *Mol Cell Biol* 11, 4572-4580

29. Huber, H. E., McCoy, J. M., Seehra, J. S., and Richardson, C. C. (1989) *J Biol Chem* 264, 4669-4678

30. Freed, E. O. (2001) *Somat Cell Mol Genet* 26, 13-33

Chapter 4 References

1. Chen, J. L., and Greider, C. W. (2003) *EMBO J* 22, 304-314

2. Chen, J. L., and Greider, C. W. (2003) *Genes Dev* 17, 2747-2752

3. Xie, M., Podlevsky, J. D., Qi, X., Bley, C. J., and Chen, J. J. (2010) *Nucleic Acids Res* 38, 1982-1996

4. Cristofari, G., and Lingner, J. (2006) *EMBO J* 25, 565-574

5. Tesmer, V. M., Ford, L. P., Holt, S. E., Frank, B. C., Yi, X., Aisner, D. L., Ouellette, M., Shay, J. W., and Wright, W. E. (1999) *Mol Cell Biol* 19, 6207-6216

6. Qiao, F., and Cech, T. R. (2008) *Nat Struct Mol Biol* 15, 634-640
7. Forstemann, K., and Lingner, J. (2005) *EMBO Rep* 6, 361-366

8. Gillis, A. J., Schuller, A. P., and Skordalakes, E. (2008) *Nature* 455, 633-637

Chapter 5 References

1. Ferreira, M. G., Miller, K. M., and Cooper, J. P. (2004) *Mol. Cell* 13, 7–18

2. Chen, J.-L., and Greider, C. W. (2004) *Proc. Natl. Acad. Sci. U.S. A.* 101,14683–14684

3. Jady, B. E., Bertrand, E., and Kiss, T. (2004) *J. Cell Biol.* 164, 647–652

4. Chen, J.-L., Blasco, M. A., and Greider, C. W. (2000) *Cell* 100, 503–514

5. Chen, J.-L., and Greider, C. W. (2004) *Trends Biochem. Sci.* 29, 183–192

6. Tesmer, V. M., Ford, L. P., Holt, S. E., Frank, B. C., Yi, X., Aisner, D. L.,Ouellette, M., Shay, J. W., and Wright, W. E. (1999) *Mol. Cell. Biol.* 19, 6207–6216

7. Mitchell, J. R., Cheng, J., and Collins, K. (1999) *Mol. Cell. Biol.* 19, 567–576

8. Nelson, J. S. (2006) *Fishes of the World*, 4th Ed., Wiley, New York

9. Kel, A. E., Gossling, E., Reuter, I., Cheremushkin, E., Kel- Margoulis, O. V., and Wingender, E. (2003) *Nucleic Acids Res.* 31, 3576–3579

10. Mosig, A., Chen, J.-L., and Stadler, P. F. (2007) *Lect. Notes Comput. Sci.* 4645, 335–345

11. Eddy, S. R. (2006) Cold Spring Harbor Symp. Quant. Biol. 71, 117–128

12. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003) Nucleic Acids Res. 31, 439–441

13. Yap, W. H., Yeoh, E., Brenner, S., and Venkatesh, B. (2005) *Gene* (*Amst*.) 353, 207–217

14. Chen, J.-L., and Greider, C. W. (2003) *EMBO J.* 22, 304–314

15. Chen, J.-L., Opperman, K. K., and Greider, C. W. (2002) *Nucleic Acids Res.*30, 592–597

16. Mosig, A., Sameith, K., and Stadler, P. F. (2006) *Genomics Proteomics Bioinformatics* 4, 56–60

17. Backofen, R., Bernhart, S. H., Flamm, C., Fried, C., Fritzsch, G., Hackermuller, J., Hertel, J., Hofacker, I. L., Missal, K., Mosig, A., Prohaska, S. J., Rose, D., Stadler, P. F., Tanzer, A., Washietl, S., and Will, S. (2007) *J. Exp. Zool. Part B* 308, 1–25

18. Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch,I.J., Murray, B. G., Kapraun, D. F., Greilhuber, J., and Bennett, M. D. (2007) *Nucleic Acids Res.* 35, D332–D338

19. Ly, H., Blackburn, E. H., and Parslow, T. G. (2003) *Mol. Cell. Biol.* 23,6849–6856

20. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H., and James, T. L. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 16998–17003

21. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) *Mol. Cell* 17, 671–682

22. Chen, J.-L., and Greider, C. W. (2003) *Genes Dev.* 17, 2747–2752

23. Ly, H., Calado, R. T., Allard, P., Baerlocher, G. M., Lansdorp, P. M., Young,N. S., and Parslow, T. G. (2005) *Blood* 105, 2332–2339

24. Mitchell, J. R., and Collins, K. (2000) *Mol. Cell* 6, 361–371

25. Richard, P., Darzacq, X., Bertrand, E., Jady, B. E., Verheggen, C., and Kiss,T. (2003) *EMBO J.* 22, 4283–4293

26. Tomlinson, R. L., Ziegler, T. D., Supakorndej, T., Terns, R. M., and Terns, M. P. (2006) Mol. Biol. Cell 17, 955–965

27. Ren, X., Gavory, G., Li, H., Ying, L., Klenerman, D., and Balasubramanian, S. (2003) Nucleic Acids Res. 31, 6509–6515

28. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) *Cell* 91, 25–34

29. Weber, M. J. (2006) *PLoS Genet.* 2, e205

30. Vitali, P., Royo, H., Seitz, H., Bachellerie, J. P., Huttenhofer, A.,and Cavaille, J. (2003) *Nucleic Acids Res.* 31, 6543–6551

31. Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J. P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K. J., McEwan, P., Bosak, S., Kellis, M., Volff, J. N., Guigo, R., Zody, M. C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E. S., Weissenbach, J., and Roest Crollius, H. (2004) *Nature* 431, 946–957

32. Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., Lee, A. P., Johnson, J., Dandona, N., Viswanathan, L. D., Tay, A., Venter, J. C., Strausberg, R. L., and Brenner, S. (2006) *Science* 314, 1892

33. Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., Lee, A. P., Johnson, J., Dandona, N., Viswanathan, L. D., Tay, A., Venter, J. C., Strausberg, R. L., and Brenner, S. (2007) *PLoS Biol.* 5, e101

34. Meyer, A., and Van de Peer, Y. (2005) *Bioessays* 27, 937–945

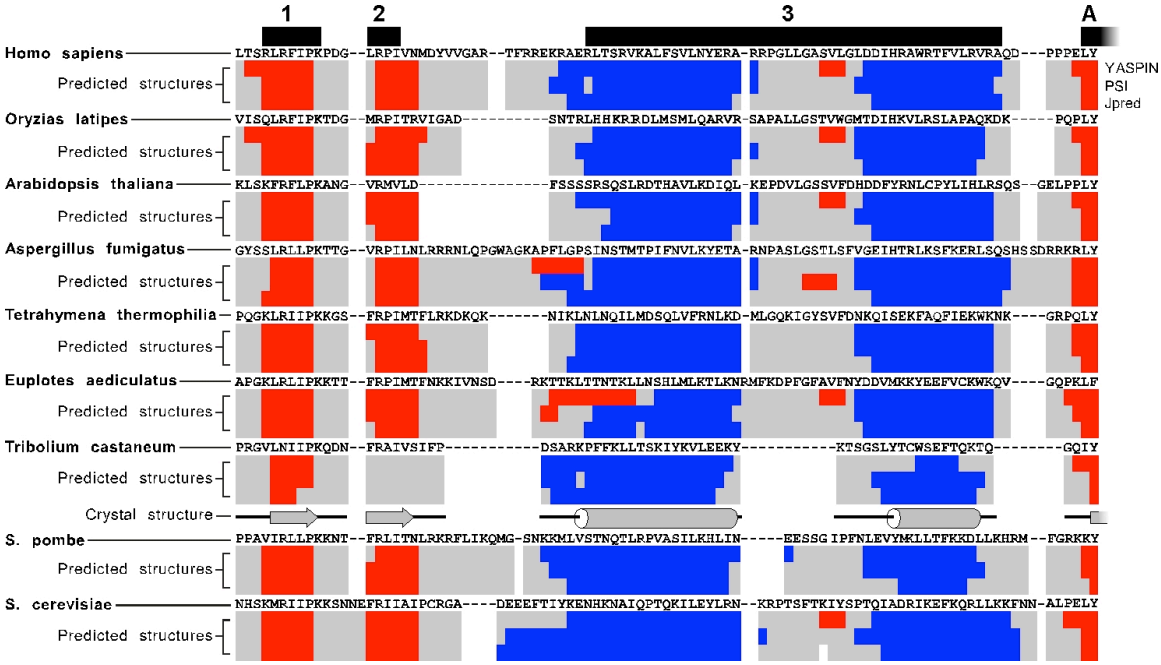
35. Shefer, K., Brown, Y., Gorkovoy, V., Nussbaum, T., Ulyanov, N. B., and Tzfati, Y. (2007) *Mol. Cell. Biol.* 27, 2130–2143

36. Benton, M. J., and Donoghue, P. C. (2007) *Mol. Biol. Evol.* 24, 26–53

37. Kumar, S., Tamura, K., and Nei, M. (2004) *Brief Bioinform.* 5, 150–163

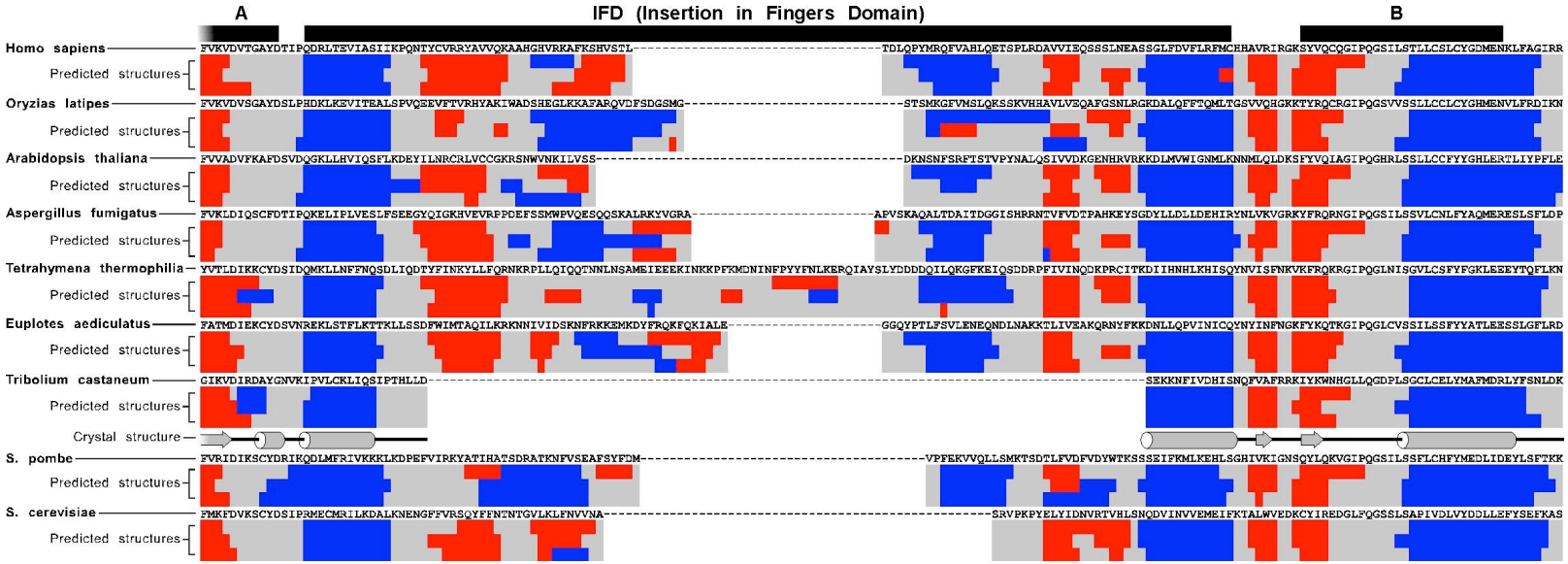
APPENDIX A

SUPPLEMENTAL INFORMATION FOR CHAPTER 2



a. Secondary structure prediction of TERT motif 1, 2, 3, A and B from representative species (*Homo sapiens, Oryzias latipes, Arabidopsis thaliana, Aspergilus fumigatus, Tetrahymena thermaphila, Euplotes aediculatus, Tribolium castaneum, Schizosaccharomyces pombe and Saccharomyces cerevisiae*).

Fig S2.1 Secondary structure prediction of TERT motifs 3 and IFD. a, Predicted secondary structures (α- helices colored blue, β-sheets colored red, and random coil colored grey) and the crystal structure of *Tribolium* TERT (α-helices denoted by cylinders, β-sheets by arrows, and random coil by a black line) are shown below the amino acid sequence. The primary sequence of the entire RT domain (motif 1-E) was input in three online secondary structure prediction algorithms (YASPIN<http://zeus.cs.vu.nl/programs/yaspinwww>, PSI <http://bioinf.cs.ucl.ac.uk/psipred/psiform.html> and JPred <http://www.compbio.dundee.ac.uk/~www-jpred>), and the output were combined. The predicted structures of motif 1-B are shown.



b. Secondary structure prediction of TERT motif A, IFD and B from representative species.

Fig. S2.1 Secondary structure prediction of TERT motifs 3 and IFD. b, Predicted secondary structures (α- helices colored blue, β-sheets colored red, and random coil colored grey) and the crystal structure of *Tribolium* TERT (α-helices denoted by cylinders, β-sheets by arrows, and random coil by a black line) are shown below the amino acid sequence. The primary sequence of the entire RT domain (motif 1-E) was input in three online secondary structure prediction algorithms (Y ASPIN<http://zeus.cs.vu.nl/programs/yaspinwww>, PSI <http://bioinf.cs.ucl.ac.uk/psipred/psiform.html> and JPred <http://www.compbio.dundee.ac.uk/~www-jpred>), and the output were combined.

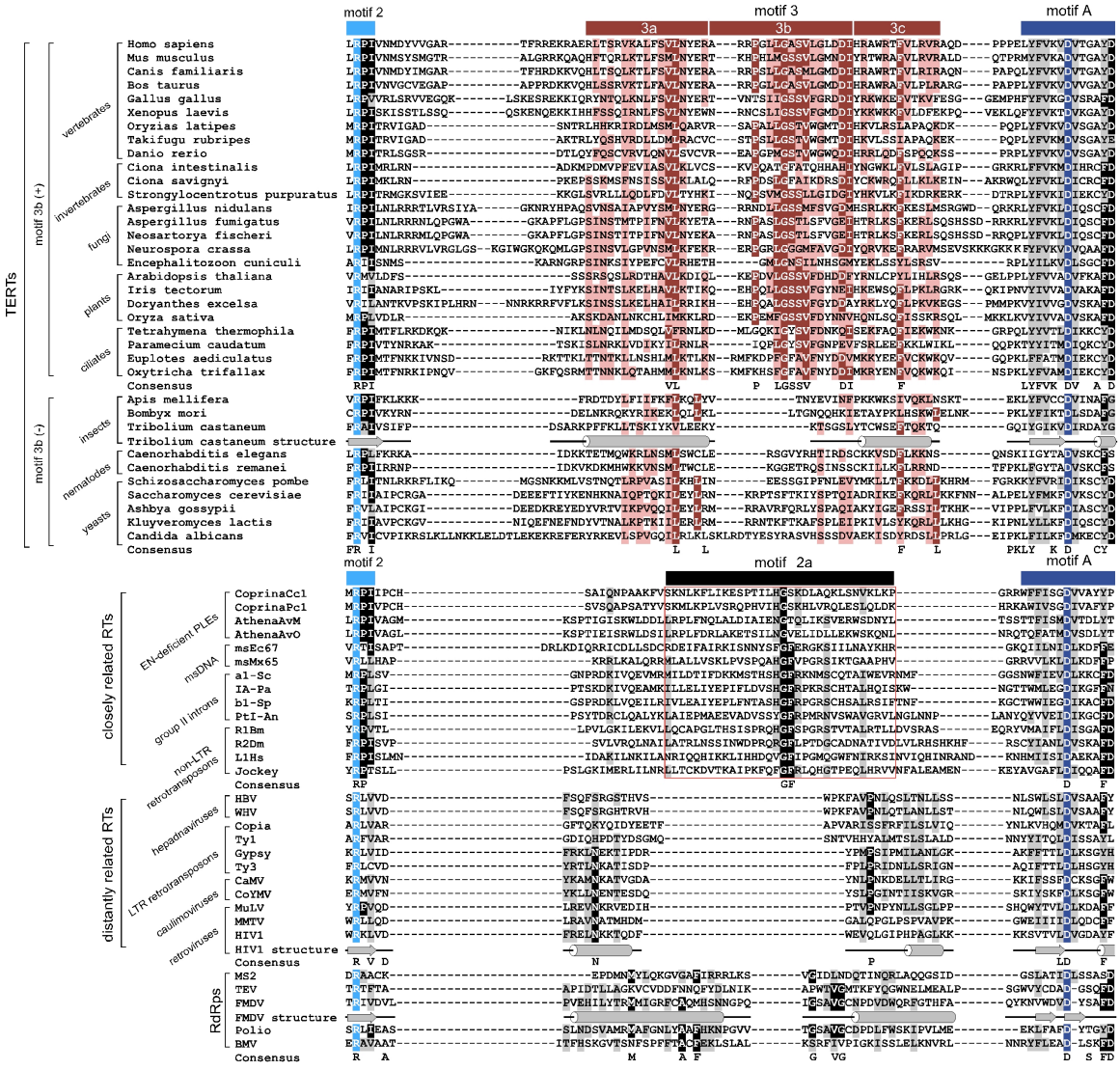


Fig. S2.2 Multiple sequence alignment from motif 2 to motif A of TERT, other RTs and RdRp. The TERT sequences includes additional species along with the representative species from Fig. 5.1b. The RT sequences are grouped as two monophyletic clades based on their evolutionary distance to TERT. The secondary structures (arrows: β-sheets, cylinders: α-helices) shown below the sequences are derived from the respective crystal structures of *Tribolium castaneum* TERT (3DU6), HIV1 RT (1HYS) and FMDV RdRp (1WNE). Motif 2 and A have identity/similarity set at 55% conservation, shaded black/grey with invariant residues are shaded in dark cyan and dark blue respectively. Red/light red shading indicates the motif 3 TERT identity/similarity. An analogous helix-turn-helix sequence within closely realted RTs has been previously identified as motif 2a. The identity/similarity for the residues in motifs 2 and A are derived from all sequences, while the linker/motif 3 identity/similarity is derived from each grouping. Consensus residues with 55% identity are shown below each group.

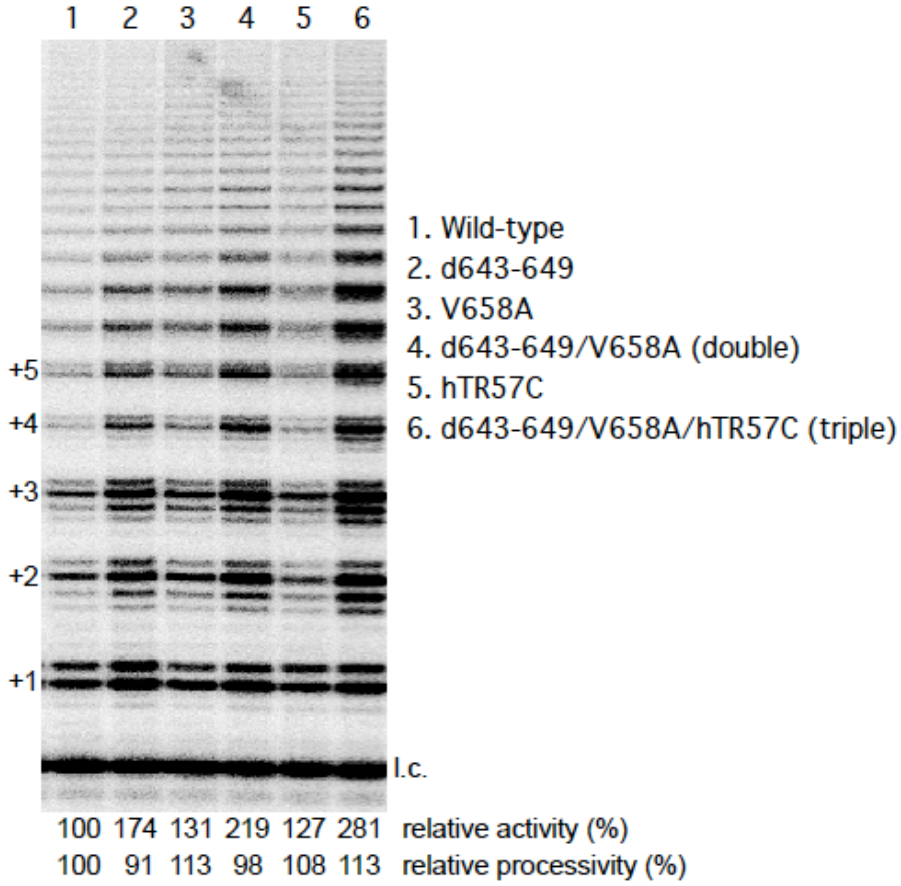


Fig. S2.3 Additive effects of motif 3 mutations on telomerase activity and processivity. Mutant telomerase were reconstituted in RRL and assayed by using the conventional telomerase activity assay. The telomerase mutants that contain different combinations of the motif 3 mutations (del-643-649, V658A and the double mutation del-643-649/V658A) or the hTR template mutation (57C) were assayed as indicated. The relative activity and processivity of different mutants are shown below the gel. l.c.: loading control. The signal of each repeat added was normalized with dGTP incorporated and the signal of first repeat. Log_{10}[normalized intensity] was then plotted against repeat number. Processivity was derived using equation processivity=- ln2/(2.303k), where k is the slope of each line.

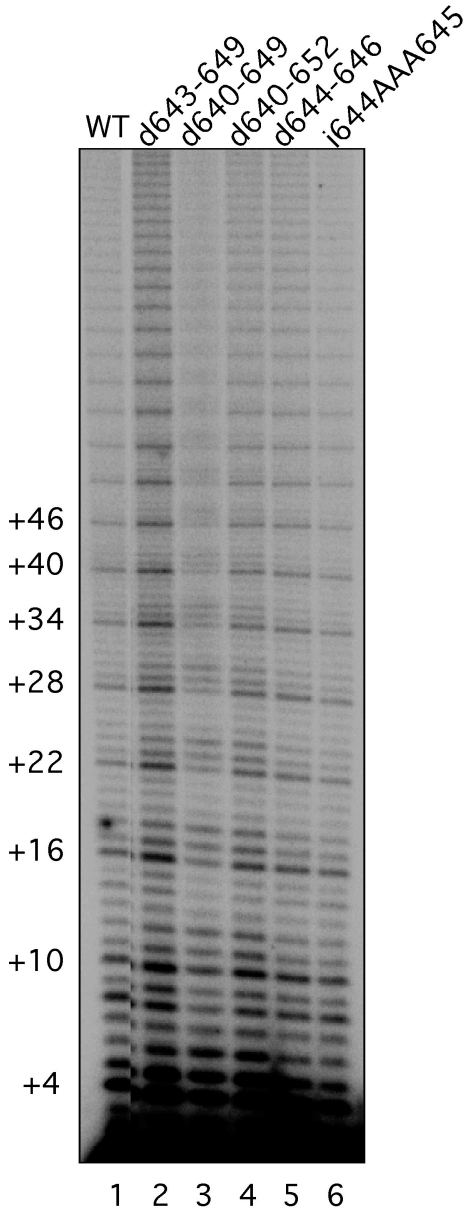


Fig. S2.4 The effect of motif3 N-terminal linker mutation on telomerase activity. Telomerase were assembled and assayed for activity as described in Materials and methods. hTERT d643-649: deletion of amino acid 643-649. i644AAA645: insertion of three alanines between amino acids 644 and 645.

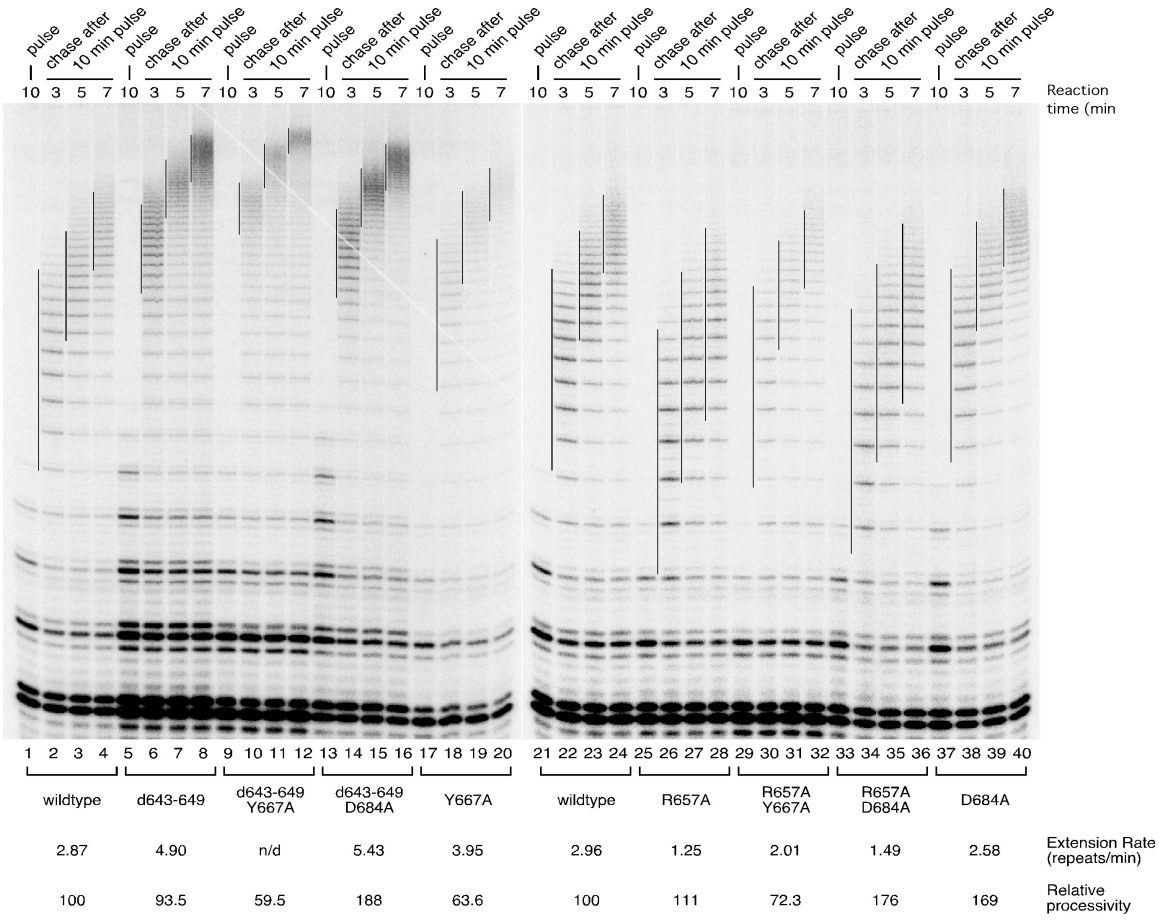
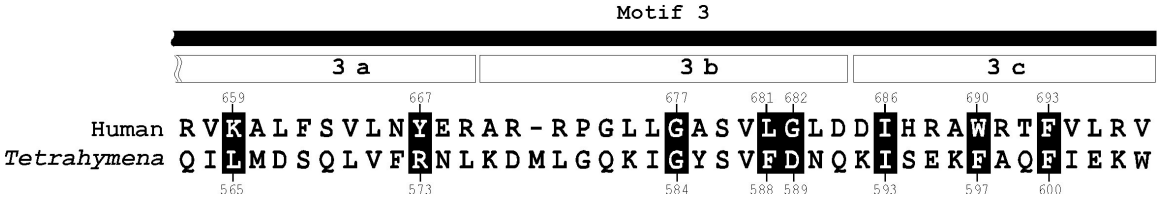
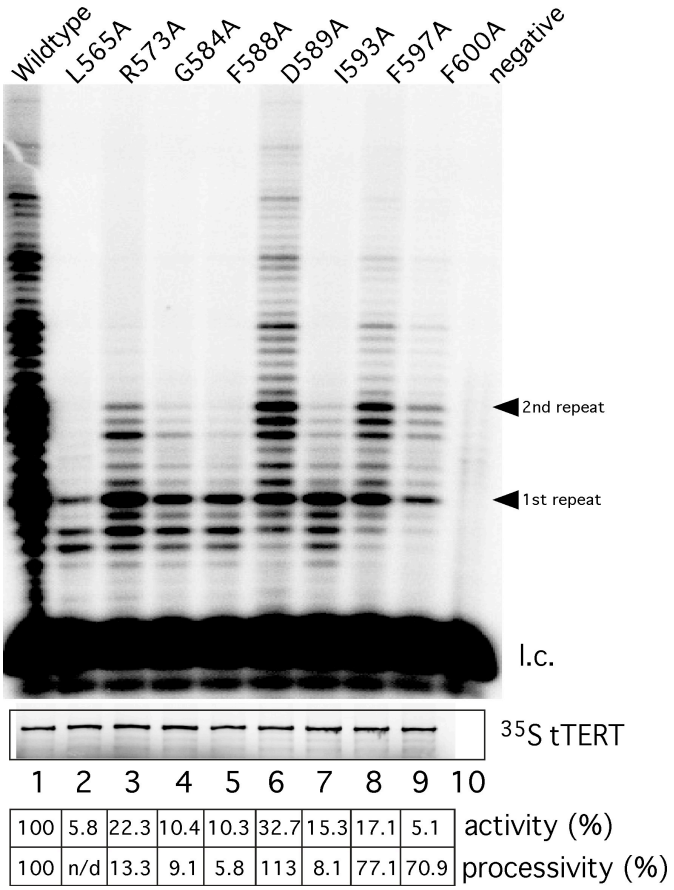


Fig. S2.5 Telomerase processivity and repeat addition rate are functionally separated. Four telomerase mutants, d643-649 with high repeat addition rate, R657A with low repeat addition rate, Y667A with low processivity and D684A with high processivity were combined into double mutation constructs. The single mutation or double mutation constructs were incubated with $(TTAGGG)_3$ primer in the pulse reaction for 5 min in which the [α-^{32}P]dGTP is incorporated to the newly synthesized telomere repeats. After 5 min of pulse reaction, non-radioactive dGTP was added to 100 μM to initiate the chase reactions and the reactions were terminated at different time points (2-10 min). The vertical lines on the gel denote the major bands of telomere products synthesized and labeled in the initial 5-min pulse reaction, and extended in the following chase reactions. Numbers on the right (+1, +2, +3 etc.) indicate the number of repeats added to the telomeric primer. Repeat-extension rate, expressed as repeats per minute, of each enzyme were calculated (see Materials and Methods) and indicated below the gel.

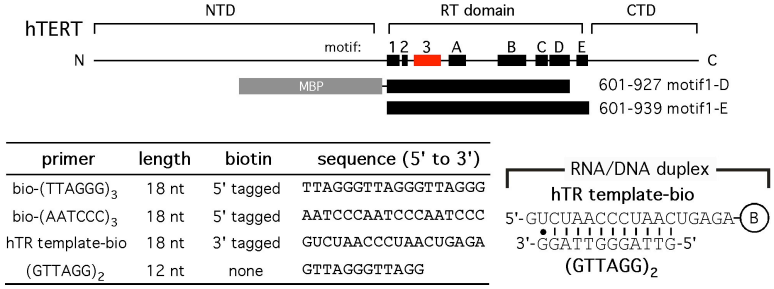


a. Sequence alignment of motif 3 from human and *Tetrahymena* TERTs.



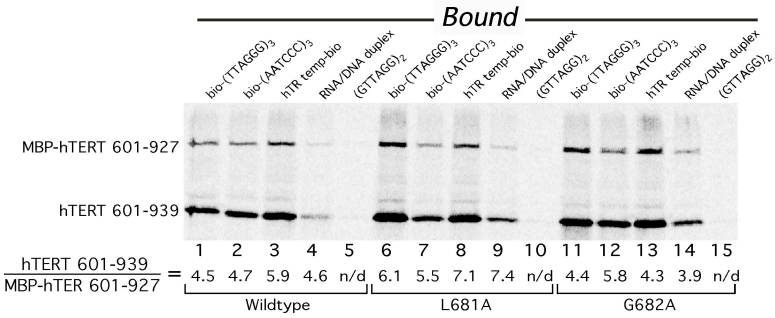
b. Activity assay of *Tetrahymena* TERT motif 3 mutants. Telomerase reconstituted *in vitro* were subjected to activity assay.

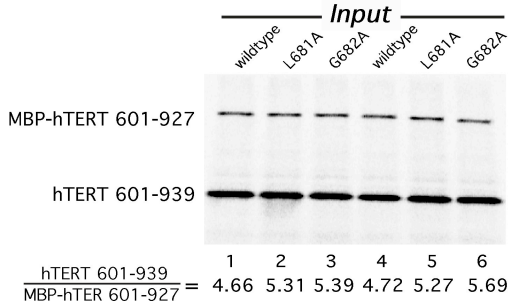
Fig. S2.6 Activity assay of *in vitro* reconstituted *Tetrahymena* telomerase. a, Homologous mutant positions emphasized in black, and indicated by the numbers above and below the sequence. N-terminus of motif 3a was not listed in detail. b, Telomerase reconstituted *in vitro* were subjected to activity assay. l.c.: loading control, [^{32}P] labeled 15mer DNA oligonucleotide. Below the gel, the [^{35}S] methionine labeled TERTs analyzed by SDS-PAGE for quantitation are shown. Quantitation of telomerase activity and processivity performed relative to wildtype. A ^{32}P-labeled 15 mer DNA oligonucleotide used as a loading control (l.c.). n/d: not determined.



primer	length	biotin	sequence (5' to 3')
bio-$(TTAGGG)_3$	18 nt	5' tagged	TTAGGGTTAGGGTTAGGG
bio-$(AATCCC)_3$	18 nt	5' tagged	AATCCCAATCCCAATCCC
hTR template-bio	18 nt	3' tagged	GUCUAACCCUAACUGAGA
$(GTTAGG)_2$	12 nt	none	GTTAGGGTTAGG

a. The RT fragments (601-639) of human TERT protein that contain the motif 3 mutations were synthesized *in vitro* and analyzed for binding to various biotin-conjugated oligonucleotides.





b. TERT protein binding assay with biotin-labeled oligonucleotides

Figure S2.7. Interactions between the RT domain of TERT, and ssDNA, ssRNA or RNA/DNA duplex. a, The sequences, length and biotin tag of oligonucleotides used in the TERT binding assay are listed in the table. The RNA/DNA duplex formed from the hybridization of the ssRNA (hTR-template-bio) and the ssDNA $(GTTAGG)_2$ is shown. An internal control, human TERT fragment (601-927) fused with a N-terminal MBP tag, was used in the binding assay. b, The RT fragments (hTERT-601-939) containing motif 3 mutants (L681A and G682A), together with the internal control, MBP-hTERT-601-927, were pulldown using various biotin-label oligonucleotides, biotin-$(TTAGGG)_3$, biotin-$(AATCCC)_3$, hTR-temp-bio, RNA/DNA duplex [hTR-temp-bio/$(GTTAGG)_2$] or $(GTTAGG)_2$ (a negative control). The input protein mixtures of hTERT 601-939 and MBP-hTERT 601-927 used in each reaction were analyzed by SDS-PAGE. The ratios between hTERT-601-939 and MBP-hTERT-601-927 bound to each biotin-labeled oligo are indicated below the gel.

APPENDIX B

SUPPLEMENTAL INFORMATION FOR CHAPTER 3

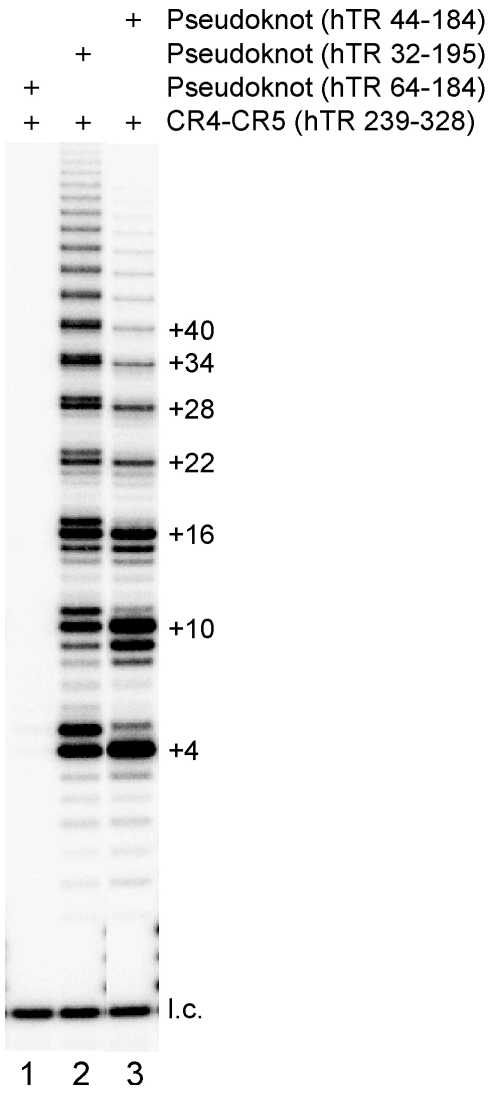
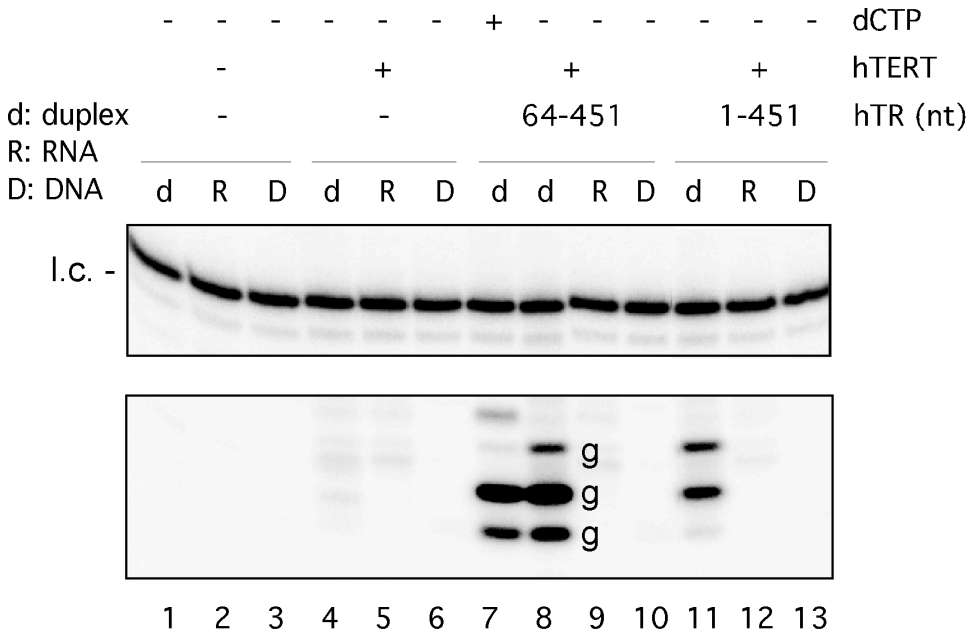
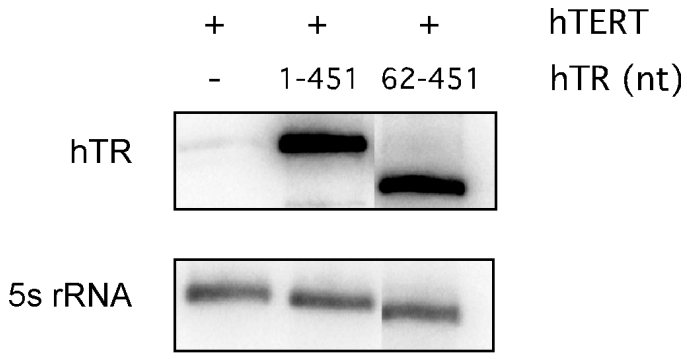


Fig. S3.1 Template free telomerase and wild-type telomerase reaction with single stranded telomeric DNA substrate. Three different telomerase RNA pseudoknot constructs with (hTR 32-195, hTR 44-184) or without template sequence (hTR 64-184) (see Fig. 3.1a for structure and sequence), are reconstituted with hTERT and hTR CR4-CR5 for active enzymes. These enzymes are the same batch as the ones used in Fig. 3.1 $(TTAGGG)_3$ primer was used. The major telomerase repeats added are labeled on the right, the numbers indicate the number of nucleotides incorporated. l.c.: loading control, a ^{32}P end labeled 15 nt DNA oligo.

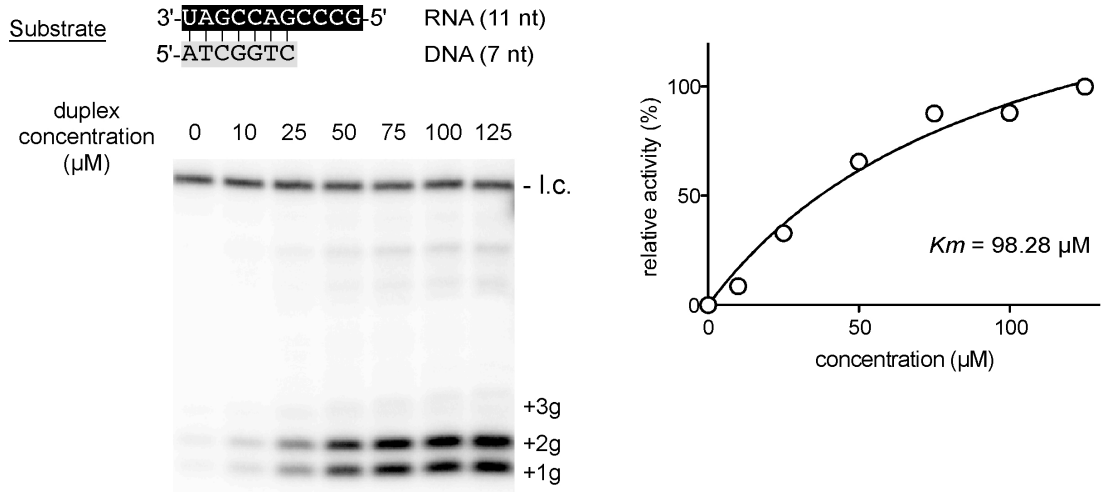


a. Template free telomerase reconstituted in 293 FT cells react with RNA/DNA duplex substrate.

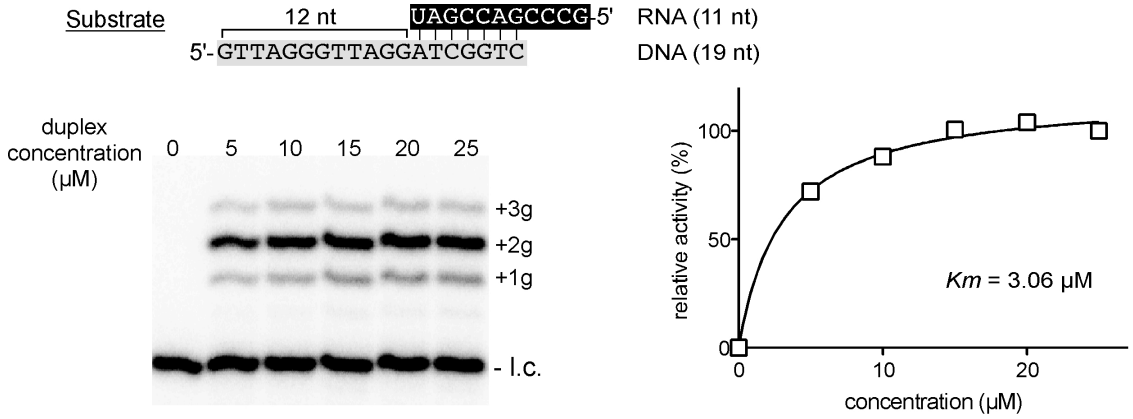


b. Northern blot of telomerase RNA expressed in 293 FT cells

Fig. S3.2 *In vivo* reconstituted telomerase react on duplex as well. a, Template free hTR (64-451) and full length hTR (1-451) are overexpressed in 293 cells with hTERT. The reconstituted telomerase was IPed and assayed for activity with duplex (d), RNA oligo (R) or DNA oligo (D) respectively. dCTP is included in the substrate in one of the reactions as indicated. The dG residues extended by telomerase are labeled as "g". b, Total RNA from 293 FT cells transfected with different RNA constructs were extracted and blotted with anti hTR riboprobe. 5s rRNA serves as a loading control.



a. Km measurement of template free telomerase with RNA/DNA duplex substrate without overhang.



b. Km measurement of template free telomerase with RNA/DNA duplex substrate with 12nt DNA overhang.

Fig. S3.3 *Km* measurement of template free telomerase with duplex substrate with or without DNA overhang. The template free telomerase was assayed with a series of different concentrations of duplex substrates. Addition of 3 G residues by template free telomerase is all considered as telomerase activity. The signal for each substrate concentration was expressed as relative activity to the highest duplex concentration (100%). Relative activity was plotted against primer concentration and fit with the Michaelis Menten curve to find the *Km* values.

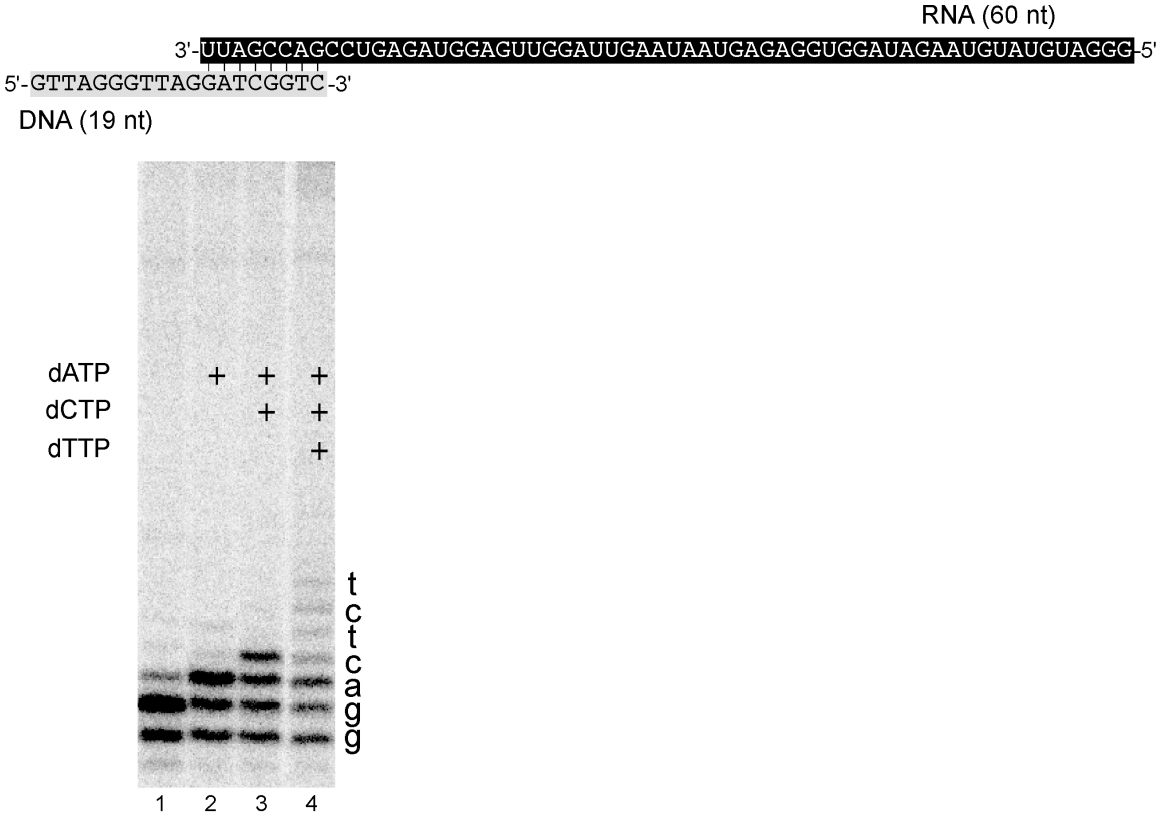


Fig. S3.4 Template free telomerase extends DNA primer along long RNA template. A substrate with 60 nt free RNA template and 12 nt DNA oligo was used. ^{32}P dGTP is presented in all the reactions. The addition of dATP, dCTP and dTTP in the reaction is indicated on the top of each lane. The predicted sequence added to the DNA primer is shown on the right of the gel.

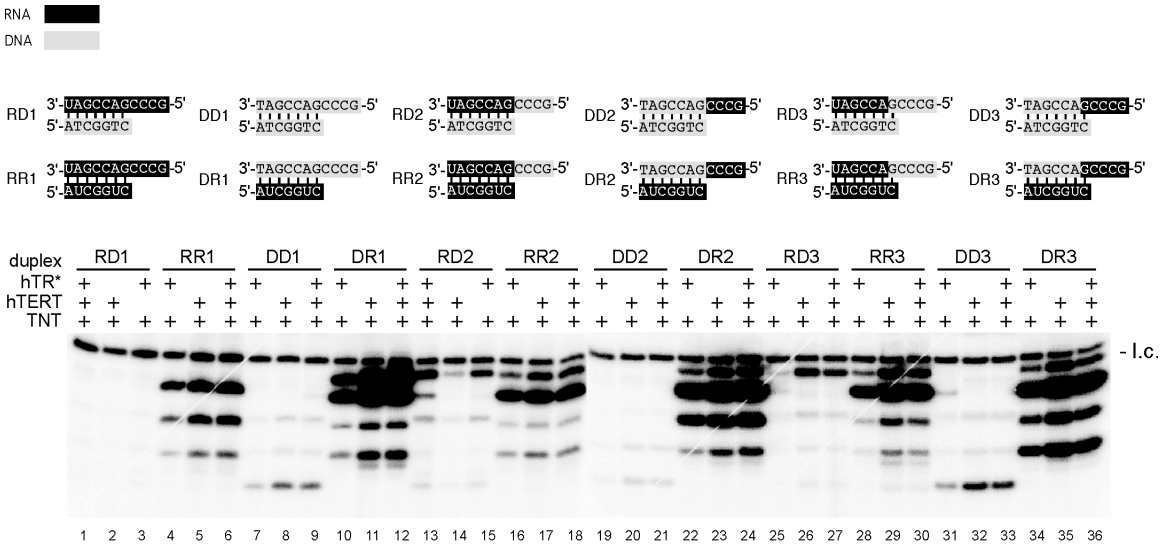


Fig. S3.5 Template free telomerase react with various duplex substrates and the ^{32}P GTP building blocks. The substrate duplexes are represented as two letter symbols plus one number. For example: RD1=RNA/DNA hybrid no.1, DR2=DNA/RNA hybrid no.2. The exact sequences are shown on the top of the gel. Black background represents ribonucleotides, while the grey background represents deoxyribonucleotides. The oligo with both grey and black backgrounds stands for a chimeric oligo. The substrates were assayed with either template free telomerase RNA (hTR*), hTERT or the template free RNP complex.

APPENDIX C

SUPPLEMENTAL INFORMATION FOR CHAPTER 4

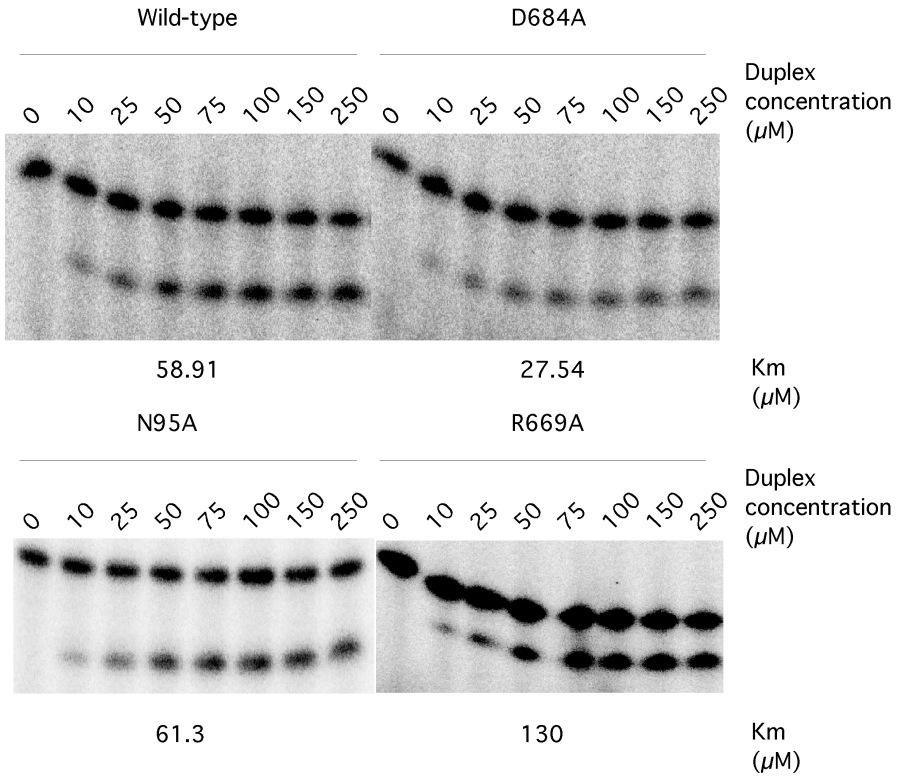


Fig. S4.1 Apparent *Km* measurement of processivity mutant template free telomerase toward CP6 duplex substrate. Processivity mutants were expressed in RRL and assembled with template free hTR pseudoknot fragment (nt 64-184) and CR4-CR5 fragment (nt 239-328). Sequence of duplex CP6 can be found in Fig. 4.2. Duplex were added in gradient concentrations as indicated on the top of the gel. The apparent *Km* was analyzed by fitting relative intensity of each concentration into Michaelis Menten curve. The *Km* value shown was the actual data for this specific run of experiment.

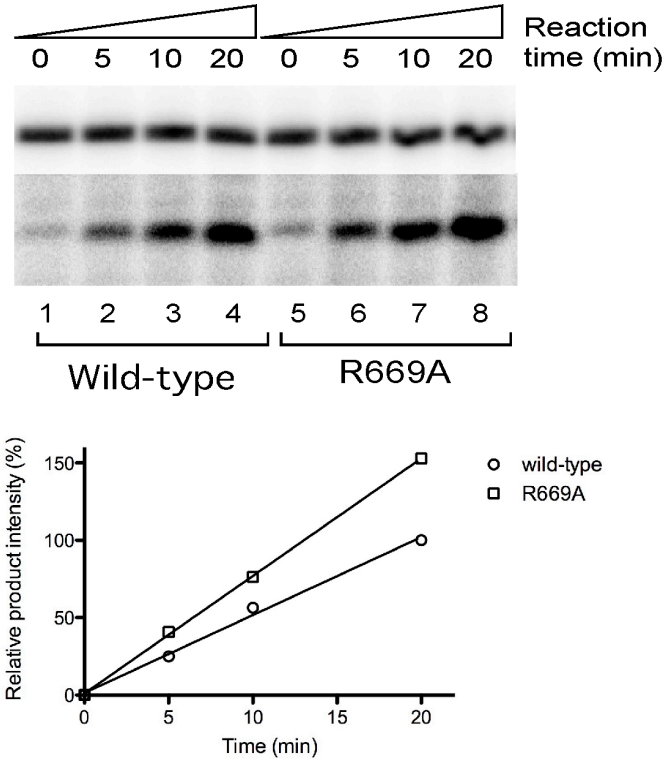


Fig. S4.2 Turnover rate of wildtype and R669A template free telomerase using telomeric CP6 duplex substrate. Wild-type and processivity mutant R669A TERT proteins were expressed in RRL and assemble with template free hTR pseudoknot fragment (nt 64-184) and CR4-CR5 fragment (nt 239-328). Sequence of duplex CP6 can be found in Fig. 4.2. Template free telomerase was pre-incubated with 20 μM duplex substrate at room temperature for 5 min. 0.165 μM ^{32}P dGTP was added to initiate the reaction. The reactions were terminated at various time points as indicated. Reaction signals were expressed in relative to the wildtype 20 min reaction (100%). Data points were plotted against time and fit with linear curve. The slopes of the lines represent enzyme turnover rate.

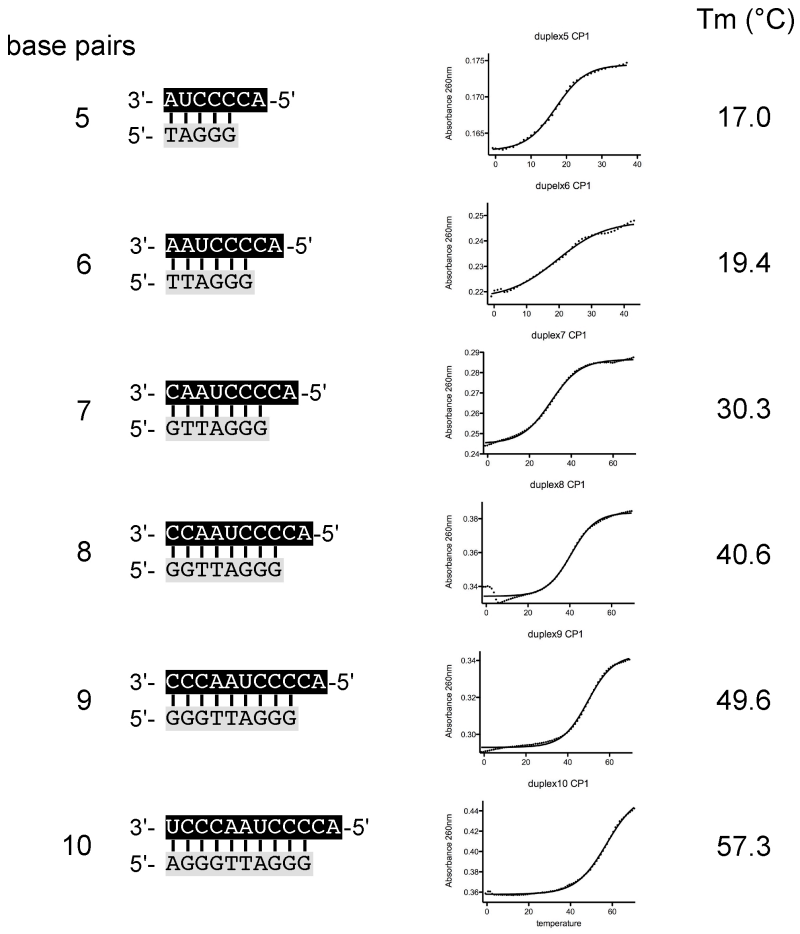


Fig. S4.3 Tm values of telomeric duplex 5 to 10 base pairs. 20μM of each duplex were mixed in telomerase reaction buffer and heated to 75°C for 5 mins. OD260nm were collected in 1°C intervals while the temperature gradually decreases to -1°C. OD260 was plotted against the temperature and a sigmoid curve was fitted. The procedure was carried out in Cary 300 UV-vis spectrophotometer (described in Materials and Methods). Sequences of each duplex are depicted on the left. The estimated Tm values were labeled on the right of the melting curve graph.

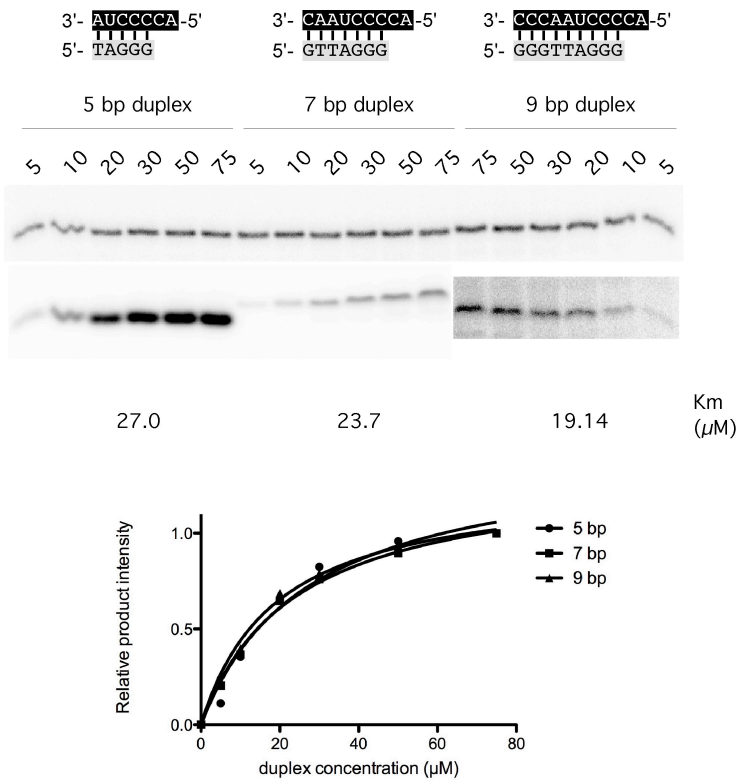


Fig. S4.4 Apparent *Km* measurement of template free telomerase toward telomeric duplex substrate ranging from 5 to 9 base pairs. Template free telomerase was expressed in RRL and assemble with template free hTR pseudoknot fragment (nt 64-184) and CR4-CR5 fragment (nt 239-328). Sequence of duplexes 5 to 9 base pairs can be found on top of the gel. Reactions were carried out at 4°C. Duplex were added in gradient concentrations as indicated. The apparent *Km* was analyzed by fitting relative intensity of each concentration into Michaelis Menten curve.

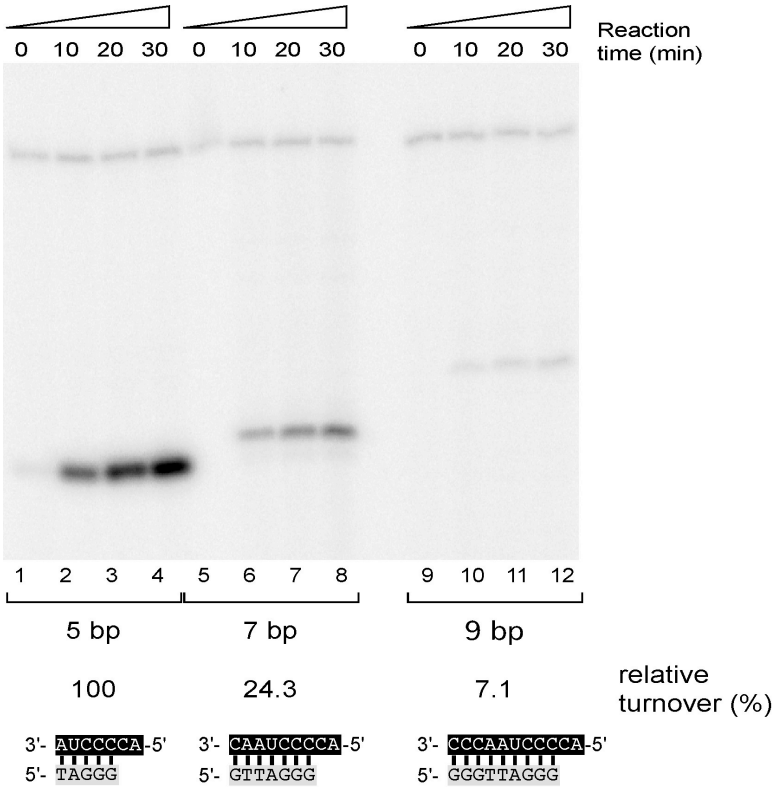


Fig. S4.5 Turnover rate of template free telomerase reacting on duplex substrate ranging from 5 to 9 base pairs. Wild-type TERT was expressed in RRL and assemble with template free hTR pseudoknot fragment (nt 64-184) and CR4-CR5 fragment (nt 239-328). Sequence of duplex 5 to 9 bps can be found at the bottom of the gel. Template free telomerase was pre-incubated with 20 μM duplex substrate at room temperature for 10 min and in 4°C for 20 min. At 4°C, 0.165 μM ^{32}P dGTP was added to initiate the reaction. The reactions were terminated at various time points as indicated. Reaction signals were expressed in relative to the wildtype 30 min reaction (100%). Data points were plotted against time and fit with linear curve. The slopes of the lines represent enzyme turnover rate.

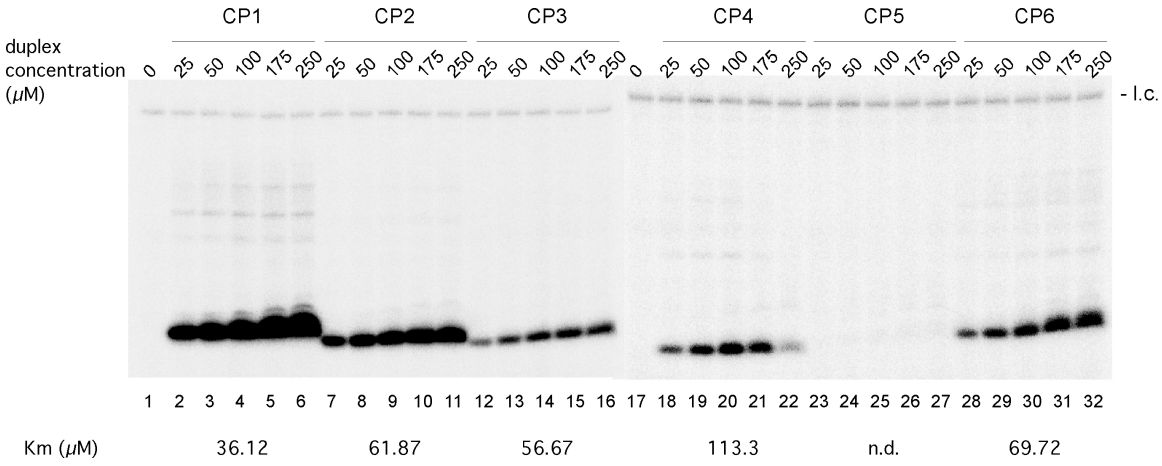


Fig. S4.6 Apparent *Km* measurement of template free telomerase toward six 7 base pair circular per-mutated telomeric duplex substrate. Template free telomerase was expressed in RRL and assemble with template free hTR pseudoknot fragment (nt 64-184) and CR4-CR5 fragment (nt 239-328). Sequence of six CP duplexes can be found in Fig. 4.8. Duplex were added in gradient concentrations as indicated on the top of the gel. The apparent *Km* was analyzed by fitting relative intensity of each concentration into Michaelis Menten curve. The *Km* value shown was the actual data for this set of experiment.

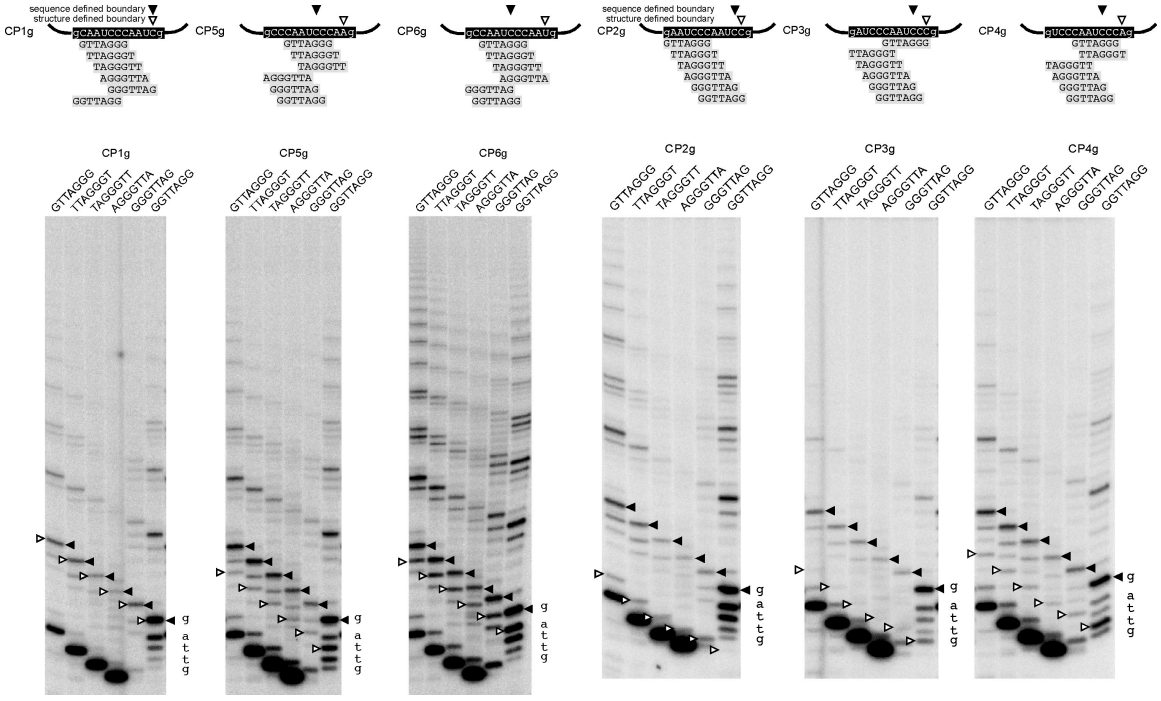


Fig. S4.7 Circular permutated hTR template mutant telomerase reaction with 7 nt circular permutated telomeric primers. Circular permutation of hTR template was generated in hTR nt 26-195 with 45g and 57g double mutations to exclude base pairing between telomeric sequence and the non-template region of the RNA. hTR nt 26-195 and nt 239-328 were assembled with hTERT to form active telomerase (Materials and Methods). (upper panel) The putative annealing scheme of circular permutated telomeric primer to each mutant hTR template was depicted on top of the gel. The corresponding stop patterns in the gel are labeled on the schematic as solid triangle (sequence determined boundary) and open triangle (structure determined boundary). (lower panel) activity assay of mutant telomerases using 6 circular permutated telomeric primers. Primers used in each reaction are indicated. Incorporation of nucleotides of the 5'-GGTTAGG-3' are labeled on the right of the gel. Banding patterns determined by duplex sequence (solid triangle) and structure (open triangle) are labeled on the side.

APPENDIX D

SUPPLEMENTAL INFORMATION FOR CHAPTER 5

Species	Assembly	Coordinates	Upstream gene	Downstream gene
Zebrafish	Zv6	Chr:25 10426622-10426938 (+)	TMEM178	HPCAL4
Medaka	MEDAKA1	Chr:8 5329954-5331990 (-)	Arf-like	-, LASP1
Stickleback	BROAD1	Group:XI 6569094-6569441 (+)	Arf-like	PLXDC2, LASP1
Fugu	FUGU 4.0	Scaffold_15 1952644-1952968 (+)	Arf-like	PLXDC2
Tetraodon	Tetraodon-7	Chr:3 13151278-13151605 (+)	Arf-like	PLXDC2
Human		Chr:3 170965092-170965529 (-)	ARPM1	MDS1

Supplemental table. S5.1 Genomic locations of teleost telomerase RNAs

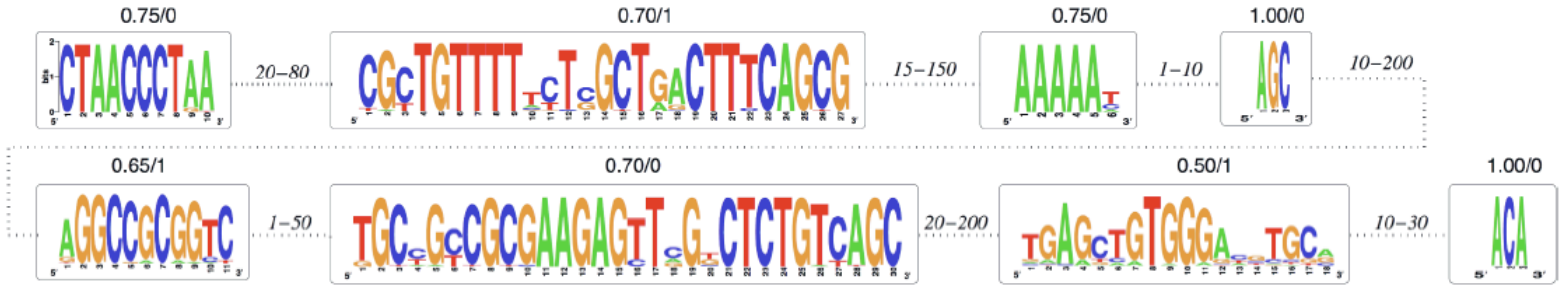


Fig. S5.1 Searching patterns for vertebrate telomerase RNA. Initial search pattern derived from the Chen et al., 2000 alignment. Each conserved sequence block is represented by a PWM displayed here as its sequence logo drawn using WebLogo (Crooks et al., Genome Res 14, 1188-90, 2004). A PWM is a matrix of score values that gives a weighted match to any given residue in the sequence. It calculates scores at each position independently from the residues at other positions. Three of the eight PWMs were allowed to match with 1 indel, and seven intervening intervals with prescribed length range. The numbers above the block are the required pwmatch-score (Kel et al., Nucleic Acids Research 31, 3576-9, 2003) and the maximal number of deletions in the genomic DNA. Dotted lines between the blocks represent unmatched genomic DNA, the numbers above the lines defining the minimal and maximal length of the intervening sequence, respectively. Starting with the most informative PWM, fragrep2 matches the PWM to the genomic DNA sequence and then collects combinations matches that obey the length restrictions on the intervening sequence using a dynamic programming algorithm.

Fig. S5.2 Cis-acting regulatory elements in the promoter region of fish telomerase RNA genes. The sequences were aligned based on the CCAAT element, TATA element and template sequence of the telomerase RNA gene. The putaive CRE-BP1/c-Jun element is predicted by using the Transcription Element Search System (TESS) at http://www.cbil.upenn.edu/tess/. The CCAAT, CRE-BP1/c-Jun, TATA element and the template region are shaded in blue, orange, green and red, respectively. The transcription initiation site of the telomerase RNA genes is indicated with an arrow above the sequence alignment.

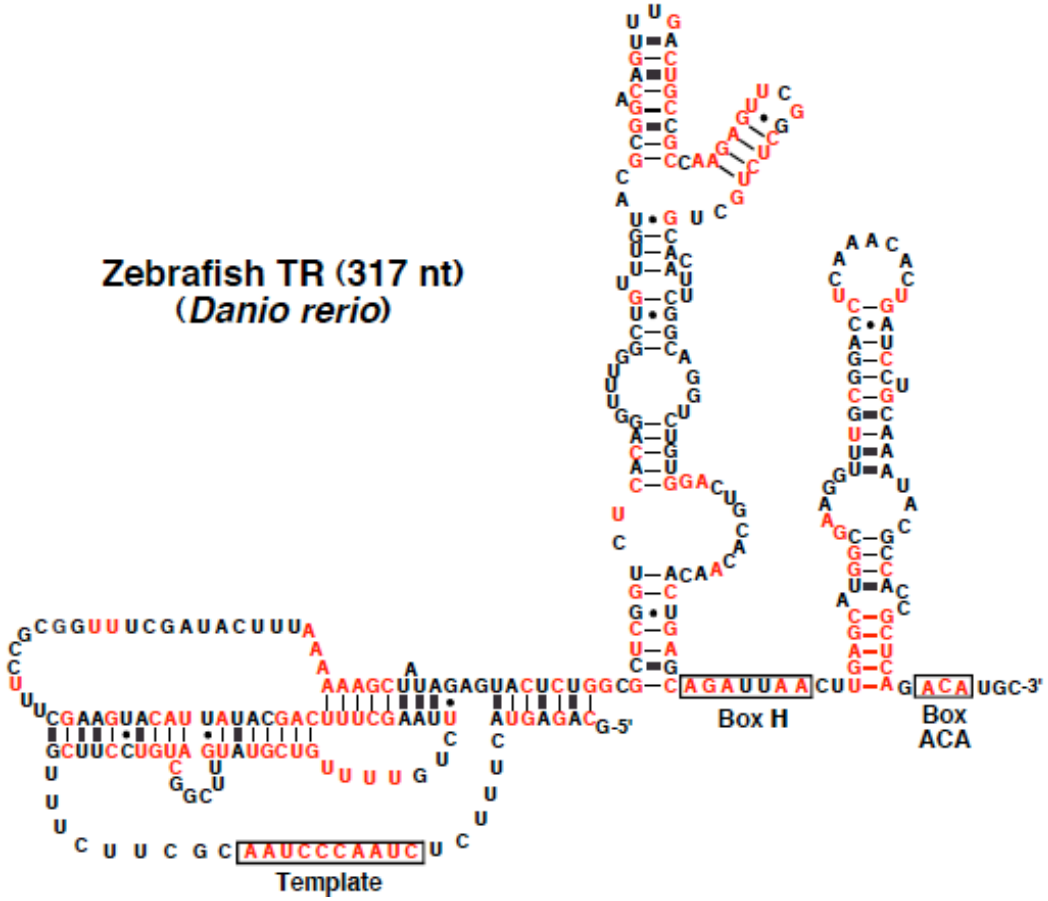


Fig. S5.3 a, Secondary structure models of zebrafish telomerase RNA. Residues conserved in all five teleost TRs are shown in red. Conserved nucleotides in the template region, box H and ACA motifs are indicated in black boxes.

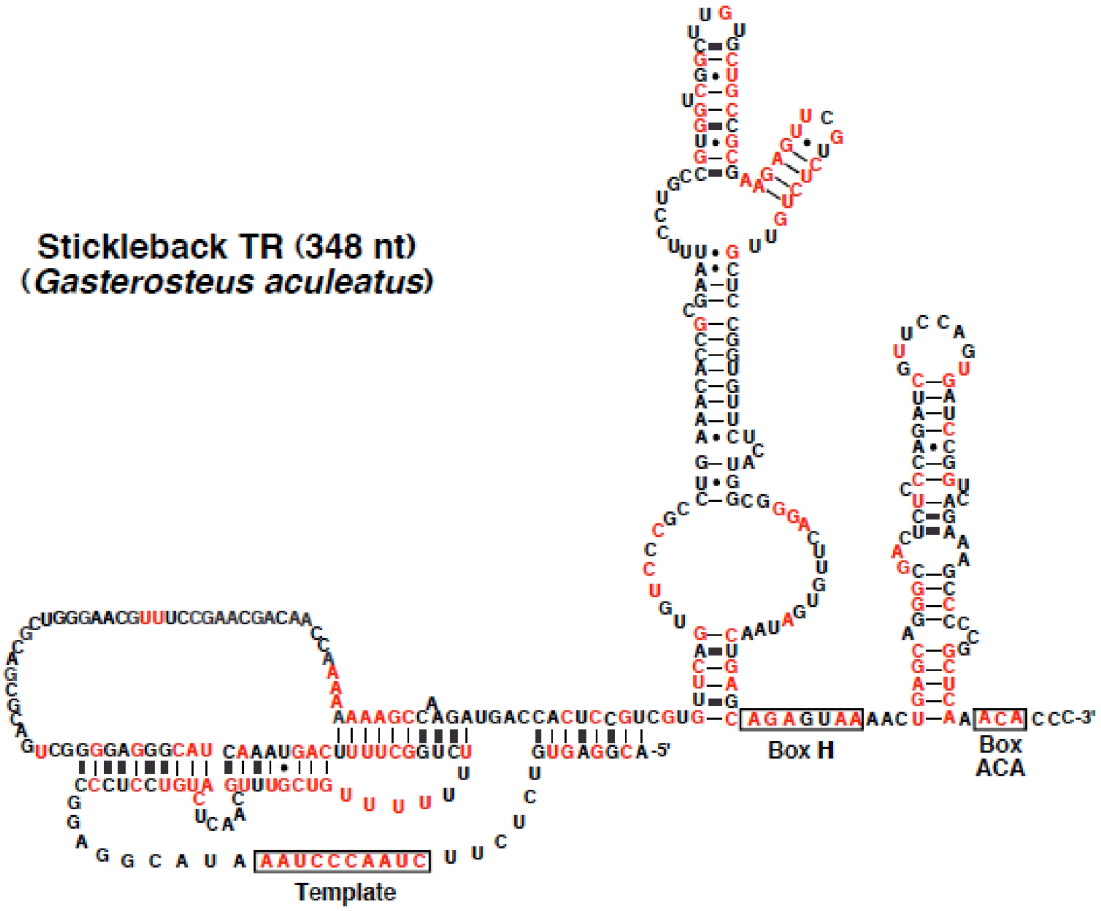


Fig. S5.3 b, Secondary structure models of stickleback telomerase RNA. Residues conserved in all fivet eleost TRs are shown in red. Conserved nucleotides in the template region, box H and ACA motifs are indicated in black boxes.

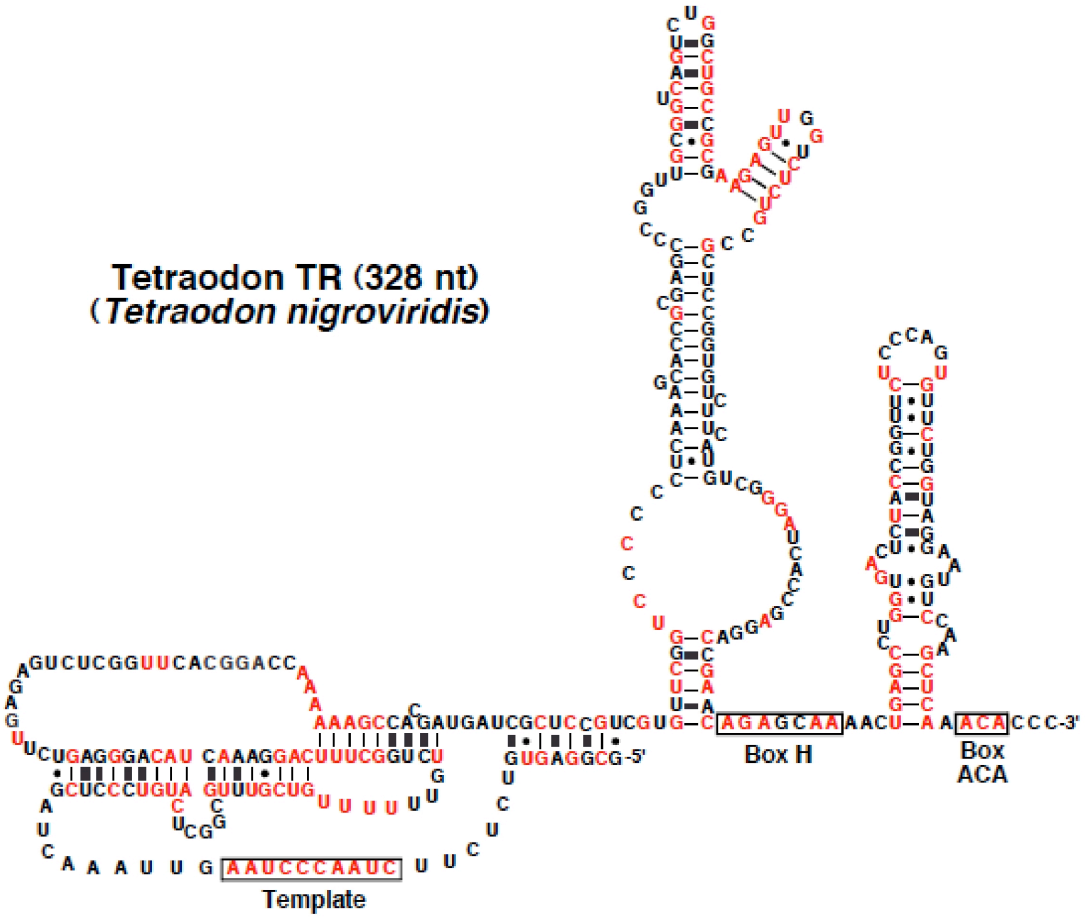


Fig. S5.3 c, Secondary structure models of Tetraodon telomerase RNAs. Residues conserved in all five teleost TRs are shown in red. Conserved nucleotides in the template region, box H and ACA motifs are indicated in black boxes.

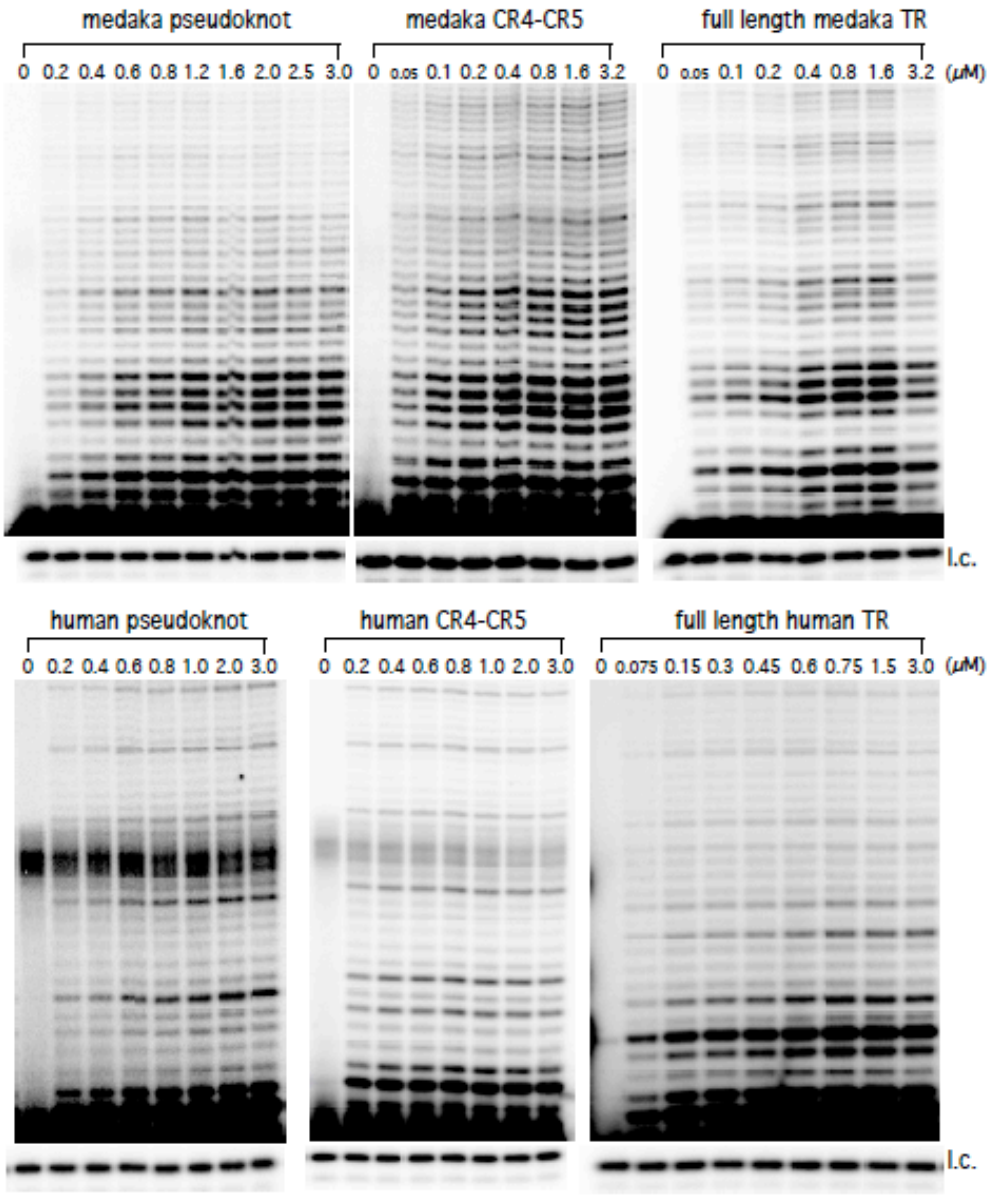


Fig. S5.4 Reconstitution of medaka and human telomerase activity with titration of full length telomerase RNA (TR) or TR fragments. For the titration of each TR fragment, the other essential fragement (either the pseudoknot or CR4-CR5 fragments) were added to a saturated concentration at 3 μM. Loading controls (l.c.) were shown at the bottom of the gel.

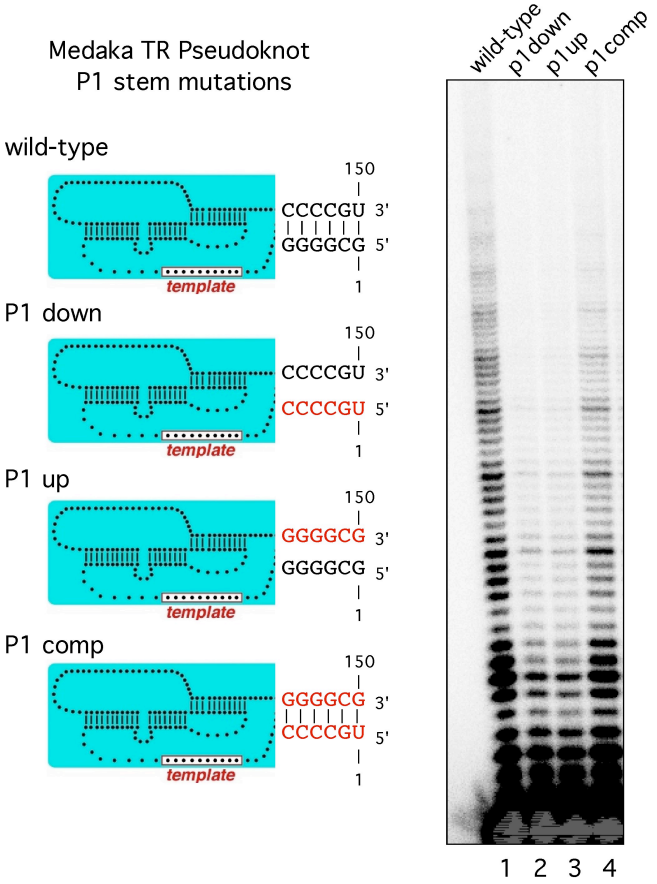


Fig. S5.5 Medaka TR P1 stem mutation decreased telomerase processivity. (left panel) the schematic of Medaka TR P1 stem mutations used in the activity assay. The major portion of medaka pseudoknot is represented using dots and cyan background. The P1 stem is expressed using the actual nucleotide sequence. The wildtype sequence is shown in black font. Mutations are shown as red letters. P1 down, the 5' side of the P1 stem was mutated to disrupt the base pairing. P1 up, the 3' side of the P1 stem was mutated to disrupt the base pairing. P1 comp, both 5' and 3' sides of the P1 stem were mutated and the base pairings are restored. (right panel) medaka TR pseudoknots were assembled with mdTERT and CR4-CR5 domain to reconstitute activity. Various mdTR pseudoknots used are indicated on the top of the gel.

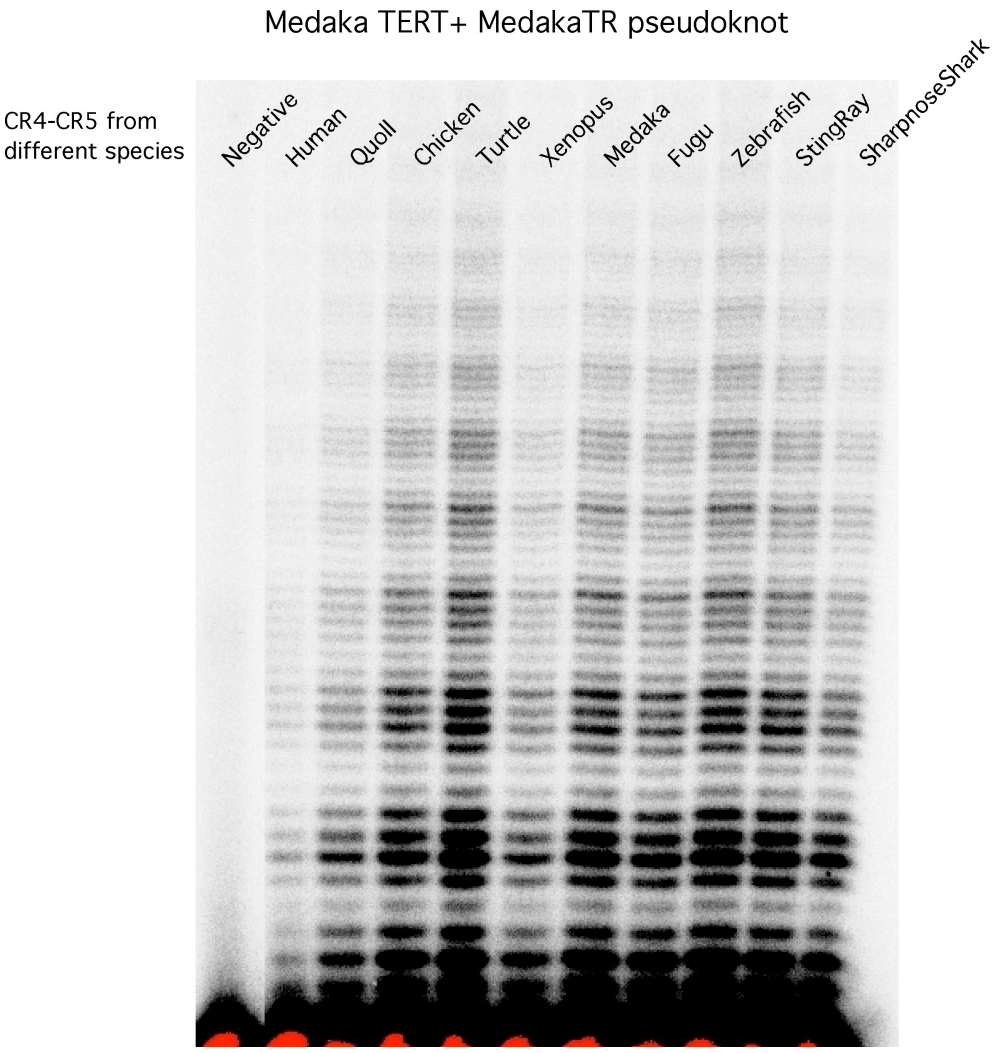


Fig. S5.6 Cross species compatibility of vertebrate CR4-CR5 fragments. Telomerase were assembled with mdTERT and mdTR pseudoknot domain with CR4-CR5 domains (the three way junction structure starting from P5 stem) from various species.

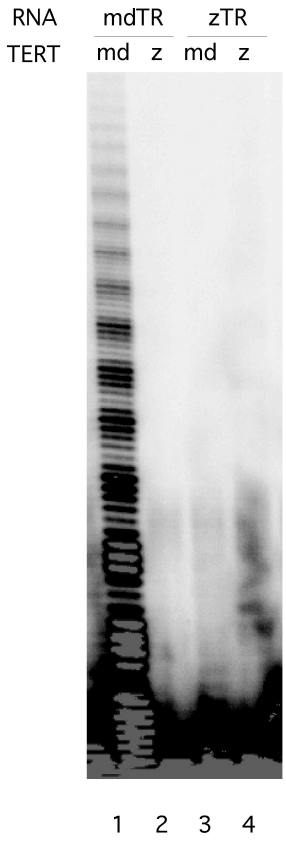


Fig. S5.7 Undetectable zebrafish telomerase activity. TERT from medaka (md) fish and zebrafish (z) were assembled with medaka TR (mdTR) and zebrafish TR (zTR) with different combinations and assayed for activity as indicated on top of the gel.

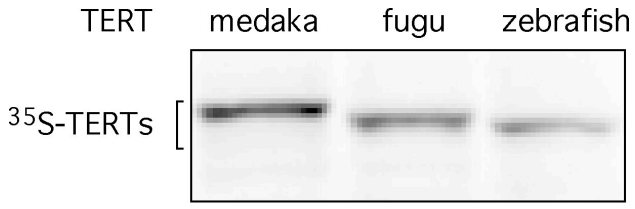


Fig. S5.8 SDS-PAGE analysis of ^{35}S Methionine labeled fish TERTs. Zebrafish TERT (1098 a.a.) migrates faster than medaka TERT (1091 a.a.) and Fugu TERT (1075 a.a.).

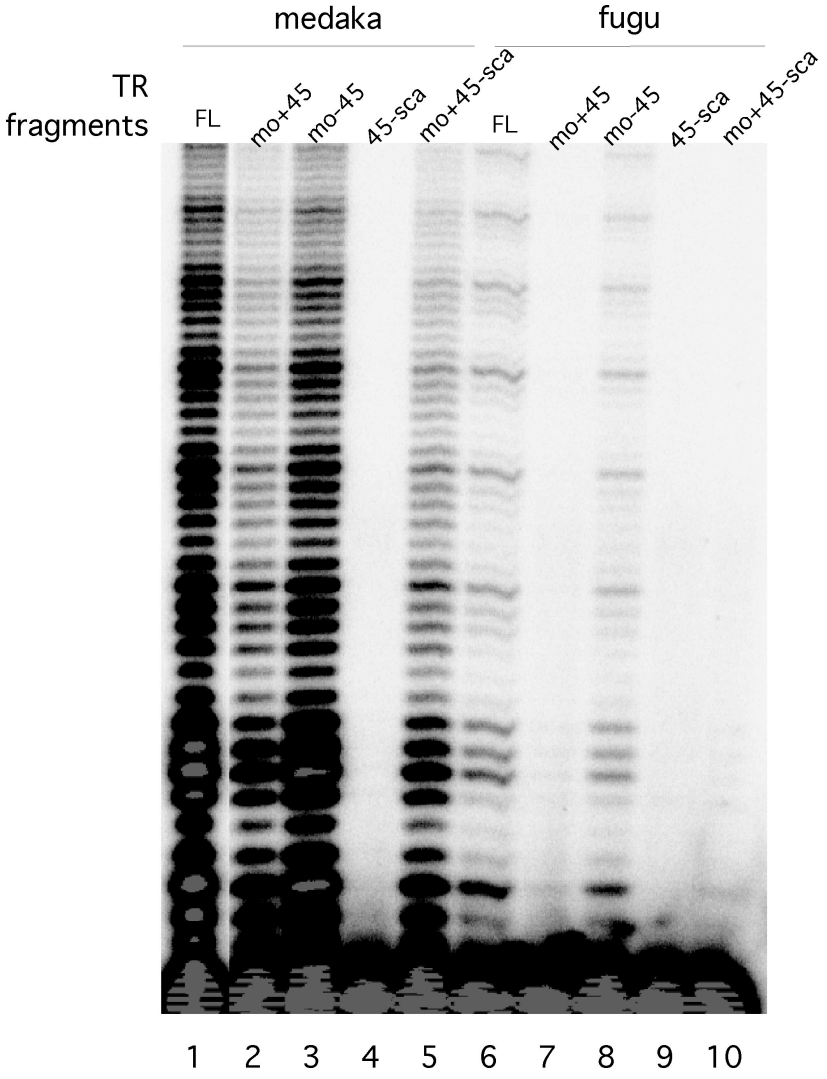
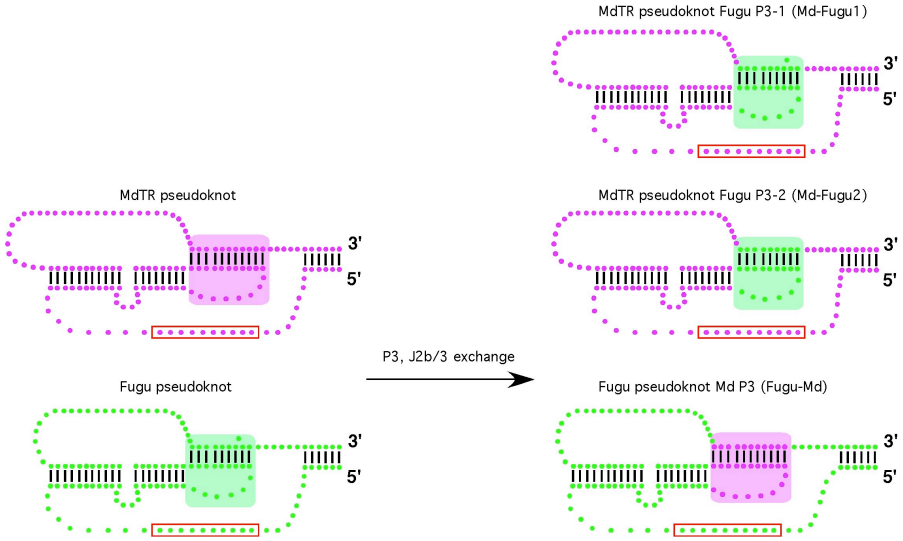


Fig. S5.9 Fish telomerase assembled using full length TR is more active than telomerase assembled with pseudoknot and CR4-CR5 domains separated. The TR concentration used to assemble fish telomerase was 200 nM. The combinations of TR fragments are labeled as the following. FL: full length. mo+45: two TR pieces with pseudoknot domain and CR4-CR5 domain separated. mo-45: one TR piece from pseudoknot to CR4-CR5 domain. 45-sca: one TR piece including CR4-CR5 domain and scaRNA domain. mo+ 45-sca: two TR pieces with one pseudoknot domain and one including CR4-CR5 and scaRNA domain.



a. P3 and J2b/3 swapping between medaka TR and fugu TR pseudoknot domains.

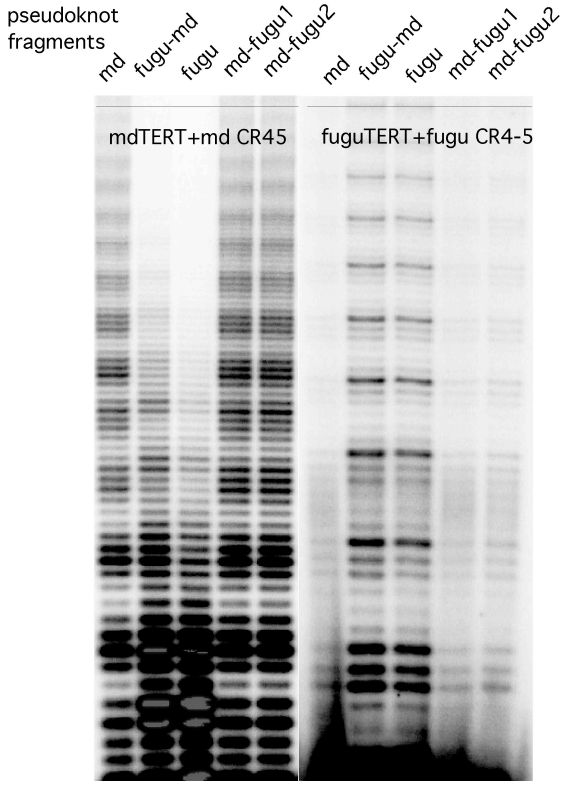


Fig. S5.10 b. Activity assay of fish telomerase with pseudoknots containing P3 and J2b/3 swapped between medaka TR (purple) and fugu (green). Pseudoknot fragments used (md, fugu-md, fugu, md-fugu1 and md-fugu2) are depicted on the upper panel.

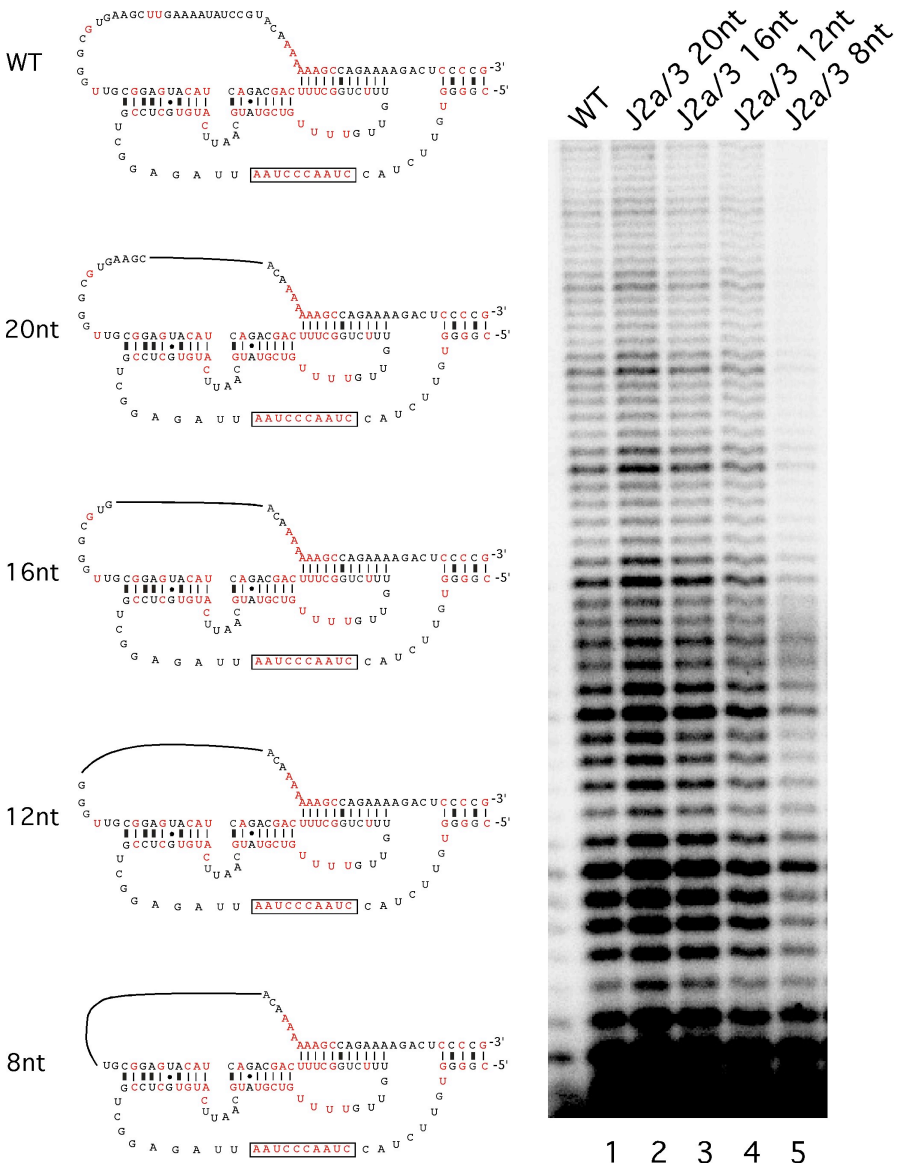


Fig. S5.11 Effective J2a/3 linker length on medaka TR pseudoknot domain. Medaka TR pseudoknot domain J2a/3 linker is shorten to various lengths as indicated on the left panel. J2a/3 mutant pseudoknots are assembled with mdTERT and CR4-CR5 domain to reconstitute telomerase and assayed for activity (right panel).

APPENDIX E

CO-AUTHOR APPROVAL

I verify that the following co-authors have approved of my use of our publication in my dissertation.

Julian Chen (Arizona State University)

Xiaodong Qi (Arizona State University)

Yang Li (Arizona State University)

Chris Bley (Arizona State University)

Joshua Podlevsky (Arizona State University)

Axel Mosig (Shanghai Institutes for Biological Sciences)

Peter Stadler (University of Leipzig)