UNIVERSITY OF CALIFORNIA

SANTA CRUZ

Structural Heterogeneity of Telomerase RNA

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

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL AND DEVELOPMENTAL BIOLOGY

by

Nicholas M. Forino

September 2023

The Dissertation of Nicholas M. Forino is approved:

_____
Professor Michael Stone, Chair

_____
Professor Melissa Jurica

_____
Professor Manuel Ares

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

TABLE OF CONTENTS

Page

LIST OF TABLES AND FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

ABSTRACT

Nicholas Forino

Structural Heterogeneity of Telomerase RNA

Eukaryotic chromosomes are linear DNA strands that terminate with telomeres, protective regions of repetitive sequences that are coated in DNA-binding proteins. Telomeres gradually shorten with each cell division until they reach a critical length that induces a 'telomere crisis', resulting in programmed cell death or senescence. The telomerase enzyme prevents telomere crisis by using a reverse transcriptase (TERT) and integral telomerase RNA (TR) to synthesize new telomeric DNA and counteract DNA loss. While telomerase is beneficial for preserving self-renewing tissues, it is also a hallmark of most human cancers.

Despite decades of work since telomerase's discovery in 1984, zero telomerase-targeting therapeutics are approved for clinical use today. A key obstacle in telomerase drug development was the lack of structural information available, but recent high-resolution cryogenic electron microscopy (cryo EM) models in combination with other structural studies have now eliminated this barrier. While a growing body of research is illuminating telomerase in higher molecular detail, much remains to be learned about the process of telomerase assembly; how the various protein and RNA components of this ribonucleoprotein (RNP) assemble to produce the active complex. The knowledge gained on this topic will accelerate therapeutic development and better enable us to understand diseases caused by telomerase deficiencies.

This thesis focuses on human telomerase biogenesis from an RNA structural perspective, highlighting the interplay between human TR (hTR) structure and protein-binding events that must be properly orchestrated to achieve the assembly of the active enzyme. I will focus first on in vitro studies that reveal hTR adopts heterogenous folds, and describe a discrete hTR conformational change that occurs upon TERT binding. Next, I will focus on how alternatively folded conformations of hTR persist in the cellular environment and their impact on the biogenesis of telomerase. In addition, I provide a detailed guide to designing, executing, and analyzing in-cell RNA chemical probing experiments from start to finish. Finally, I describe the future directions of my unfinished projects on the topics of hTR folding chaperones and disease-associated hTR mutations.

ACKNOWLEDGEMENTS

First and foremost, I must thank Michael Stone for his steadfast mentorship. His love and enthusiasm for science is infectious and provided the sustaining energy I needed during the height of the pandemic. Michael has always valued my growth as a scientist. Even when my interests took turns into uncharted territory for the lab, he encouraged my growth and highlighted my victories. Thank you for your unwavering patience, and showing me how to be a caring and effective mentor as well as a fearless experimentalist.

To everyone in the Stone Lab, thank you for fostering a creative and collaborative environment that I will always cherish. In particular, thank you Christina Palka for sharing your never-ending stoke for RNA with me, along with your technical wizardry. Thank you to my thesis committee members, Manny Ares and Melissa Jurica, for supporting my work and thoughtful advice as it has taken shape. Thank you Susan Carpenter for the important lesson that scientific research is an exercise of humility. I would also like to thank Jia Zheng Woo, Silvi Rouskin, Art Zaug, and Tom Cech for being collaborators of the highest caliber.

I must also thank my parents, Michael and Maria Forino, for their unfailing love and support throughout my life. Thank you for always cultivating my interests and always encouraging me strive for what will make me ultimately happiest in life.

Last but not least, I would like to thank Shanna, the love of my life. Life as your partner is such a privilege. Thank you for filling my world with light and love. I

hope the little telomerases in each of our cells keep us going for as long as possible to maximize our time together.

CHAPTER I

Introduction

Telomeres

Life as we know it requires the safekeeping of DNA, the blueprint molecule that originates all the biochemical processes separating organisms from inanimate matter. Eukaryotes partition their DNA across multiple linear strands, each molecule possessing two ends that are capped with protective structures called telomeres. Telomeres are regions of repetitive DNA (TTAGGG in humans) with a single-stranded 3' overhang, and are coated in a protein sheath called shelterin[1]. Telomeres guard eukaryotic genomes through two mechanisms: 1) hiding the 3' end from cellular DNA surveillance and repair machinery[2,3], and 2) providing a buffer region against naturally occurring DNA erosion that occurs from incomplete DNA replication every cell division[4]. This gradual loss of DNA causes the 'end replication problem', and if left unchecked culminates in a 'telomere crisis' wherein cells with short telomeres will either self-destruct or enter a senescent state[5,6]. While the removal of cells in crisis is normally beneficial, some cells escape this fate by expressing the telomerase enzyme, which lengthens their telomeres (**Fig. 1**). Telomerase's ability to preserve the proliferative capacity of cells is what enables self-renewing tissues such as stem cells and human cancers to divide without limit[6–8]. Deficiencies in telomerase production are one cause of a constellation of diseases collectively known as telomere syndromes[9–11]. The unique nature of telomerase to

cause disease whether it is inappropriately present (in cancer cells) or absent (in normal regenerating tissues) makes it an attractive therapeutic target.

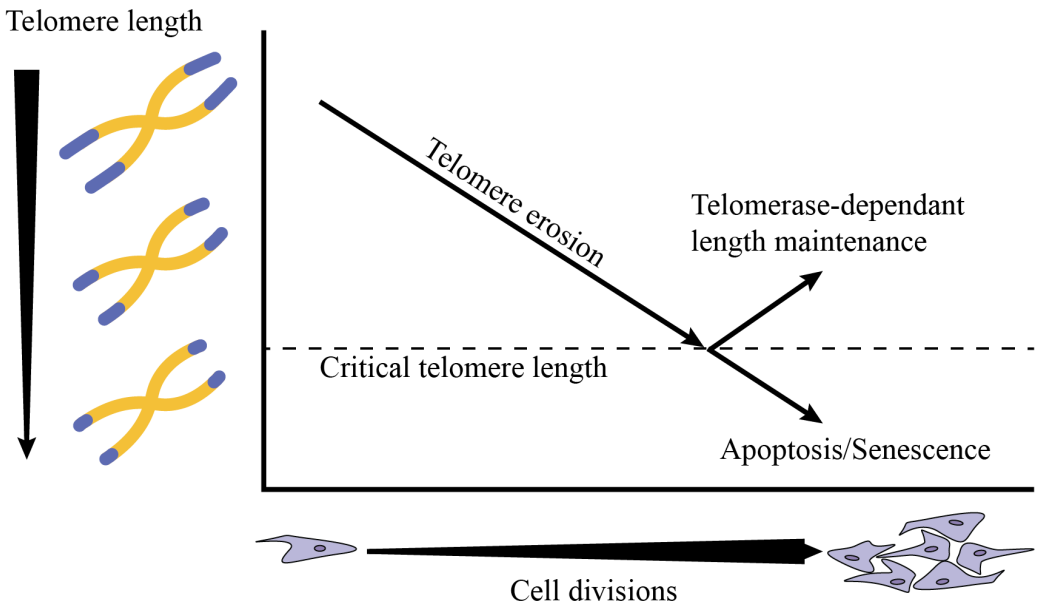


Figure 1: Biological effects of telomere erosion over cell divisions.

Telomere length steadily decreases over successive cell divisions until a critically short length threshold is reached. If telomeres continue to shorten, cells in telomere crisis either initiate apoptosis or enter a non-replicative senescent state. Telomerase lengthens telomeres, allowing cells to escape this fate.

Telomerase

Telomerase is a ribonucleoprotein (RNP) that mitigates telomere erosion by replenishing telomeric DNA. It is a reverse transcriptase (TERT) that carries its own RNA template supplied by the telomerase RNA (TR) subunit, which it uses to synthesize DNA at the 3’ end of telomeres. These protein and RNA components work in concert to perform telomerase’s unique catalytic cycle known as repeat addition processivity (RAP); telomerase binds the chromosome 3’ end, synthesizes telomeric

DNA repeats, and re-engages the substrate for subsequent repeat additions[12,13]. Additionally, telomerase contains a suite of species-specific proteins that control its maturation and cellular trafficking[14–18].

TERT contains four conserved domains: the Telomerase Essential N-terminal (TEN) domain, Reverse Transcriptase (RT) domain, RNA Binding Domain (RBD), and C-Terminal Extension (CTE) domain (**Fig. 2A**). TEN mediates contacts with both shelterin and the telomere DNA 3' end that enable telomerase recruitment to telomeres and the processive nature of telomerase RAP, respectively[19,20]. RBD establishes high affinity contacts with TR that are required for catalysis[21–23]. RT and CTE are the palm and thumb domains of telomerase, respectively[24].

TR possesses a conserved multi-domain architecture, but is variable in both sequence content and overall length across species[25,26]. Universally, TR is comprised of two core domains. The template-pseudoknot (t/PK) domain includes the substrate-binding template sequence and a pseudoknot motif. Downstream of the t/PK is the Stem Terminus Element (STE), which forms a short stem-loop in ciliates (designated as Stem Loop 4), or a three-way junction (TWJ) designated as the Conserved Regions 4/5 (CR4/5) in vertebrates[27,28]. Both core domains make essential contacts with TERT-RBD that are necessary for telomerase assembly and catalytic activity[29–32]. Vertebrate TRs also include a Small Cajal body RNA (ScaRNA) domain on their 3' end that binds biogenesis factors important for telomerase maturation and cellular localization[33,34]. Human TR (hTR), is 451 nucleotides long and is structurally itemized into 8 major stems (P1, P2, etc.) across its three domains (**Fig. 2B**).

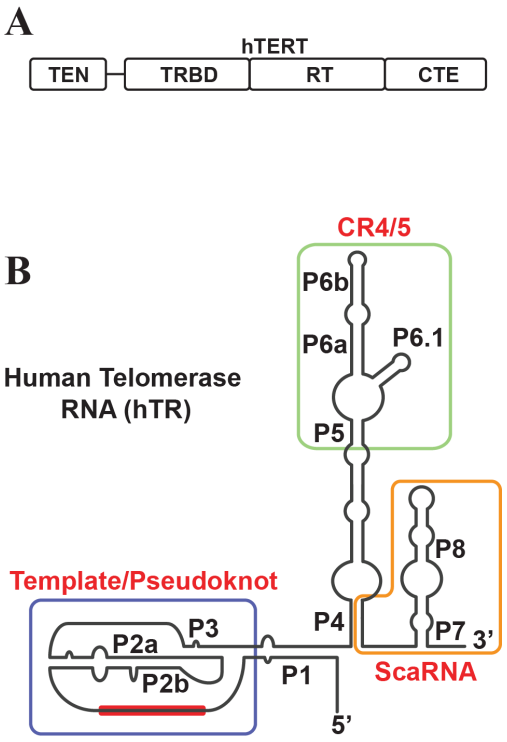


Figure 2: Domain architecture of hTERT and hTR.

(A) hTERT domain architecture: Telomerase Essential N-terminal domain (TEN), Reverse Transcriptase domain (RT), RNA Binding Domain (RBD), C-Terminal Extension (CTE).

(B) hTR domain architecture as determined by phylogenetic and mutagenesis studies[19,25-28]: Template/Pseudoknot (t/PK) with template sequence shown as a red bar, the TWJ Conserved Regions 4/5 (CR4/5), and small Cajal-body RNA domain (ScaRNA).

TR is more than simply a template sequence for TERT, as the evolution of its conserved structural architecture suggests. In fact, TR forms the scaffold upon which telomerase is built and spatially organizes its proteins. The cryo EM structure of the human and *Tetrahymena thermophila* holoenzymes reveal an intricately arranged architecture of TR-protein interactions. The human telomerase RNP has a bi-lobed shape (**Fig. 3**) in which the catalytic lobe (TERT, t/PK, and CR4/5) and the biogenesis lobe (ScaRNA domain, H/ACA RNP) are separated by flexible hTR linkers provided by the P1 and P4 stems[20,35–39]. Despite the advances in elucidating the final assembled telomerase RNP, there is a lack of structural information about how hTR transitions from a naked RNA molecule into active telomerase during the biogenesis pathway.

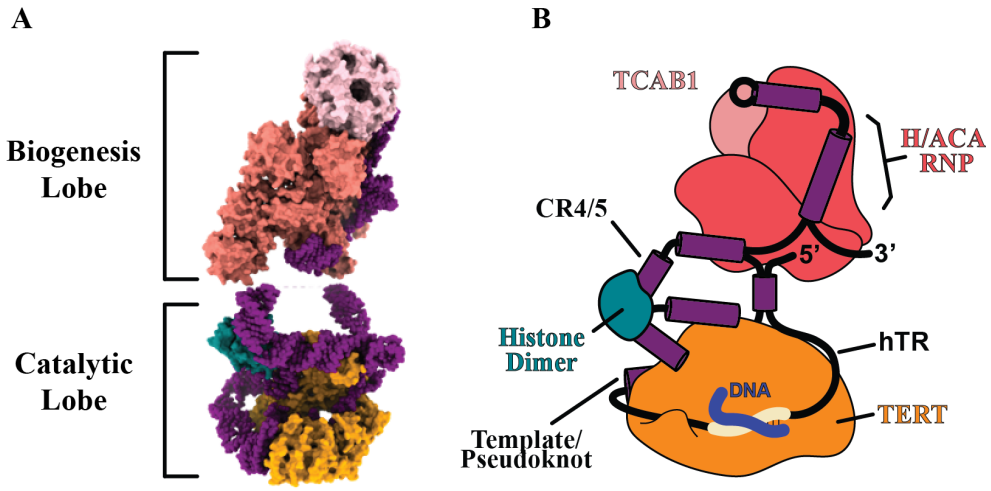


Figure 3: Bilobed architecture of the human telomerase RNP.

(A) Cryo EM structure[36] of human telomerase depicting the biogenesis and catalytic lobes. hTR (purple) organizes the RNP into two distinct protein hubs. (B) Schematic detailing arrangement of proteins and hTR domains in human telomerase. The H/ACA core tetramer is a shared protein complex that binds the small Cajal Body RNAs. Telomerase Cajal Body protein 1 (TCAB1) binds a specific hTR motif located in the ScaRNA domain known as the CAB box, and enables hTR to transit into nuclear Cajal Bodies[16-18]. A histone H2A/B dimer (teal) was discovered to be bound to the CR4/5 domain but its function is unknown[36].

Chapter II of this thesis focuses on the CR4/5 domain and its structure-function relationship required for TERT binding. Prior studies of CR4/5 from *Oryzias latipes* (medaka fish) revealed that major helical repositioning within the TWJ is necessary to accommodate and recognize TERT RBD[40,41], but it was unknown whether a similar rearrangement occurs within CR4/5 of human TR (hTR). We set out to query the in vitro structure of fish and human CR4/5 in the absence of TERT and to ascertain whether human CR4/5 is remodeled upon protein binding like its fish homolog. To accomplish this, we employed high throughput RNA chemical probing

and structure determination techniques in combination with thermodynamic RNA folding predictions tools to query the structure of TERT-free CR4/5 structure in vitro. We studied TERT-induced remodeling of human CR4/5 through single molecule Förster resonance energy transfer (FRET), which allowed us to infer changes in the overall helical arrangement of the CR4/5 TWJ motif. Our study revealed a unique structural heterogeneity that distinguishes the human CR4/5 secondary structure from fish CR4/5. Despite this, we discovered that the human RNA undergoes a similar structural remodeling of its TWJ architecture to that of *Oryzias latipes*, suggesting they share some degree of homology in their TERT-binding mechanisms.

Chapter III describes a study of hTR structure within living cells to further our understanding of telomerase biogenesis. Previous in-cell studies of hTR structure came to contradictory conclusions about whether this RNA is a pre-organized scaffold for telomerase proteins, or becomes remodeled with its core domains during telomerase assembly[42,43]. This study demonstrates the efficacy and experimental power of a recently developed chemical probing technique that leverages next-generation sequencing and data deconvolution analysis to illuminate RNA structural ensembles[44,45]. We found that most hTR molecules appear to be canonically folded in a manner consistent with being assembled into telomerase, but a small alternatively folded fraction of hTR persists in the steady state. Through mutagenesis we stabilized the alternate fold of CR4/5 and demonstrated that it negatively impacts telomerase biogenesis. Our findings suggest that alternate hTR folding represents a barrier in

telomerase biogenesis, and invites further exploration of potential hTR chaperones that contribute to telomerase biogenesis.

Chapter IV is a technical guide on the design, execution, and analysis of in-cell chemical probing experiments that leverage next generation sequencing readouts. This powerful method produces highly desirable data for any molecular biologist interested in RNA structure, but oftentimes the computational knowledge required can represent a significant bottleneck toward progress. Within, I explain the bioinformatic methodology that enabled Chapter III of this thesis, and also share what I believe to be some best practices for data analysis and RNA structure visualization.

Chapter V is a summary of two ongoing projects further delving into additional aspects of hTR structure and function. One such area of study is the noncanonical role of histone proteins in RNA chaperoning. Telomerase cryoEM structures unexpectedly identified histone H2A/B dimers bound to hTR CR4/5[20,36]. I present evidence that H2A/B dimers specifically bind CR4/5 and influence its overall structural landscape, potentially readying the RNA for TERT binding. Another project involves structural studies of hTR in cells derived from patients harboring disease-associated hTR mutations[46]. The results gained from these experiments will add further detail to our understanding of homeostatic telomerase biogenesis and reveal new insights into the etiology of telomere syndromes.

References

1. Lange, T. de. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).

2. Arnoult, N. & Karlseder, J. Complex interactions between the DNA-damage response and mammalian telomeres. *Nat Struct Mol Biol* **22**, 859–866 (2015).

3. Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. & de Lange, T. p53- and ATM-Dependent Apoptosis Induced by Telomeres Lacking TRF2. *Science* **283**, 1321–1325 (1999).

4. Lingner, J., Cooper, J. P. & Cech, T. R. Telomerase and DNA End Replication: No Longer a Lagging Strand Problem? *Science* **269**, 1533–1534 (1995).

5. Hayflick, L. THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. *Exp Cell Res* **37**, 614–636 (1965).

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

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

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

9. Yamaguchi, H. *et al.* Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. *Blood* **102**, 916–918 (2003).

10. Vulliamy, T. J. & Dokal, I. Dyskeratosis congenita: the diverse clinical presentation of mutations in the telomerase complex. *Biochimie* **90**, 122–130 (2008).

11. Savage, S. A. Human telomeres and telomere biology disorders. *Prog Mol Biol Transl Sci* **125**, 41–66 (2014).

12. Greider, C. W. Telomerase Is Processive. *Molecular and Cellular Biology* **11**, 4572–4580 (1991).

13. Zhao, Y. *et al.* Processive and Distributive Extension of Human Telomeres by Telomerase under Homeostatic and Nonequilibrium Conditions. *Molecular Cell* **42**, 297–307 (2011).

14. Vogan, J. M. & Collins, K. Dynamics of Human Telomerase Holoenzyme Assembly and Subunit Exchange across the Cell Cycle. *J Biol Chem* **290**, 21320–21335 (2015).

15. Vogan, J. M. *et al.* Minimized human telomerase maintains telomeres and resolves endogenous roles of H/ACA proteins, TCAB1, and Cajal bodies. *eLife* **5**, e18221 (2016).

16. Egan, E. D. & Collins, K. Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747–1759 (2012).

17. Egan, E. D. & Collins, K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Mol. Cell. Biol.* **32**, 2428–2439 (2012).

18. Venteicher, A. S. *et al.* A Human Telomerase Holoenzyme Protein Required for Cajal Body Localization and Telomere Synthesis. *Science* **323**, 644–648 (2009).

19. Robart, A. R. & Collins, K. Human Telomerase Domain Interactions Capture DNA for TEN Domain-Dependent Processive Elongation. *Molecular Cell* **42**, 308–318 (2011).

20. Sekne, Z., Ghanim, G. E., van Roon, A.-M. M. & Nguyen, T. H. D. Structural basis of human telomerase recruitment by TPP1-POT1. *Science* **375**, 1173–1176 (2022).

21. Lai, C. K., Mitchell, J. R. & Collins, K. RNA Binding Domain of Telomerase Reverse Transcriptase. *Molecular and Cellular Biology* **21**, 990–1000 (2001).

22. Robart, A. R. & Collins, K. Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked subunit variants. *J. Biol. Chem.* **285**, 4375–4386 (2010).

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

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

25. Bhattacharyya, A. & Blackburn, E. H. Architecture of telomerase RNA. *EMBO J* **13**, 5721–5731 (1994).

26. Podlevsky, J. D. & Chen, J. J.-L. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol* **13**, 720–732 (2016).

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

28. Chen, J.-L. & Greider, C. W. An emerging consensus for telomerase RNA structure. *Proc Natl Acad Sci U S A* **101**, 14683–14684 (2004).

29. Gilley, D. & Blackburn, E. H. The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein. *Proc Natl Acad Sci U S A* **96**, 6621–6625 (1999).

30. Chen, J.-L. & Greider, C. W. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proceedings of the National Academy of Sciences* **102**, 8080–8085 (2005).

31. Sperger, J. M. & Cech, T. R. A Stem−Loop of Tetrahymena Telomerase RNA Distant from the Template Potentiates RNA Folding and Telomerase Activity. *Biochemistry* **40**, 7005–7016 (2001).

32. Autexier, C., Pruzan, R., Funk, W. D. & Greider, C. W. Reconstitution of human telomerase activity and identification of a minimal functional region of the human telomerase RNA. *The EMBO Journal* **15**, 5928–5935 (1996).

33. Jády, B. E., Bertrand, E. & Kiss, T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body–specific localization signal. *Journal of Cell Biology* **164**, 647–652 (2004).

34. Mitchell, J. R., Cheng, J. & Collins, K. A Box H/ACA Small Nucleolar RNA-Like Domain at the Human Telomerase RNA 3′ End. *Molecular and Cellular Biology* **19**, 567–576 (1999).

35. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190 (2018).

36. Ghanim, G. E. *et al.* Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* **593**, 449–453 (2021).

37. Wan, F. *et al.* Zipper head mechanism of telomere synthesis by human telomerase. *Cell Res* **31**, 1275–1290 (2021).

38. Liu, B. *et al.* Structure of active human telomerase with telomere shelterin protein TPP1. *Nature* **604**, 578–583 (2022).

39. Jiang, J. *et al.* Structure of Tetrahymena telomerase reveals previously unknown subunits, functions, and interactions. *Science* **350**, aab4070 (2015).

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

41. Huang, J. *et al.* Structural basis for protein-RNA recognition in telomerase. *Nat. Struct. Mol. Biol.* **21**, 507–512 (2014).

42. Antal, M., Boros, É., Solymosy, F. & Kiss, T. Analysis of the structure of human telomerase RNA in vivo. *Nucleic Acids Res* **30**, 912–920 (2002).

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

44. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**, 75–82 (2017).

45. Tomezsko, P. J. *et al.* Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**, 438–442 (2020).

46. Boyraz, B., Bellomo, C. M., Fleming, M. D., Cutler, C. S. & Agarwal, S. A novel TERC CR4/CR5 domain mutation causes telomere disease via decreased TERT binding. *Blood* **128**, 2089–2092 (2016).

CHAPTER II

Folding heterogeneity in the essential human telomerase RNA three-way junction

(originally published in the journal RNA)

Abstract

Telomeres safeguard the genome by suppressing illicit DNA damage responses at chromosome termini. To compensate for incomplete DNA replication at telomeres, most continually dividing cells, including many cancers, express the telomerase ribonucleoprotein (RNP) complex. Telomerase maintains telomere length by catalyzing de novo synthesis of short DNA repeats using an internal telomerase RNA (TR) template. TRs from diverse species harbor structurally conserved domains that contribute to RNP biogenesis and function. In vertebrate TRs, the conserved regions 4 and 5 (CR4/5) fold into a three-way junction (TWJ) that binds directly to the telomerase catalytic protein subunit and is required for telomerase function. We have analyzed the structural properties of the human TR (hTR) CR4/5 domain using a combination of in vitro chemical mapping, secondary structural modeling, and single-molecule structural analysis. Our data suggest the essential P6.1 stem–loop within CR4/5 is not stably folded in the absence of the telomerase reverse transcriptase in vitro. Rather, the hTR CR4/5 domain adopts a heterogeneous ensemble of conformations. Finally, single-molecule FRET measurements of CR4/5 and a mutant designed to stabilize the P6.1 stem demonstrate that TERT binding selects for a

structural conformation of CR4/5 that is not the dominant state of the TERT-free in vitro RNA ensemble.

Introduction

The ends of linear chromosomes in eukaryotic cells terminate with repetitive DNA sequences that bind to specialized proteins to form telomeres[1,2]. Telomeres protect coding DNA from degradation and distinguish chromosomal termini from double-stranded breaks to evade unwanted recognition by DNA damage response machineries[3-5] . With each round of cell division, the inability of the conventional replication machinery to completely copy the lagging strand template results in gradual telomere attrition. Ultimately the presence of a critically short telomere drives cells into permanent cell growth arrest or apoptosis[6,7]. However, cells that must retain high proliferative capacity maintain telomere length through the action of the telomerase reverse transcriptase[8-12]. Given the importance of maintaining telomere length in dividing cells, germ-line mutations in telomerase genes result in severe developmental defects[13-15]. In addition, telomerase contributes to the unchecked cell growth that is a hallmark of human cancers[16,17]. Therefore, efforts to better understand telomerase structure, function, and regulation have direct biomedical significance.

Telomerase is a multi-subunit ribonucleoprotein (RNP) complex that includes the catalytic telomerase reverse transcriptase (TERT) protein, telomerase RNA (TR), and several additional species-specific holoenzyme proteins that are necessary for

proper RNP biogenesis[18,19]. The TERT domain architecture is well-conserved across species and consists of the telomerase essential amino-terminal (TEN) domain, the telomerase RNA-binding domain (TRBD), the reverse transcriptase (RT) domain, and the carboxy-terminal extension (CTE) (**Fig. 1A**). In contrast, comparison of TRs across species ranging from yeasts to human reveals an exceedingly high degree of variation in both RNA length and sequence[20-,22]. Interestingly, in spite of this apparent evolutionary divergence, several conserved TR structural elements exist that are essential for enzyme assembly and function. These include the highly conserved template/pseudoknot (t/PK) domain and a stem-terminal element (STE) (**Fig. 1B**). In vertebrate TRs, the STE is thought to fold into an RNA three-way junction (TWJ) often referred to as the conserved regions 4/5 (CR4/5) domain (**Fig. 1C**). With regard to TR primary sequence, the CR4/5 domain is spatially separated from the RNA template that must necessarily reside in the TERT enzyme active site; yet, naturally occurring mutations in human telomerase RNA (hTR) CR4/5 can result in human diseases characterized by loss of telomerase function[13,14,23]. In hTR, the CR4/5 domain includes three RNA helices (P5, P6, and P6.1) joined together by an expanded RNA junction sequence (**Fig. 1C**).

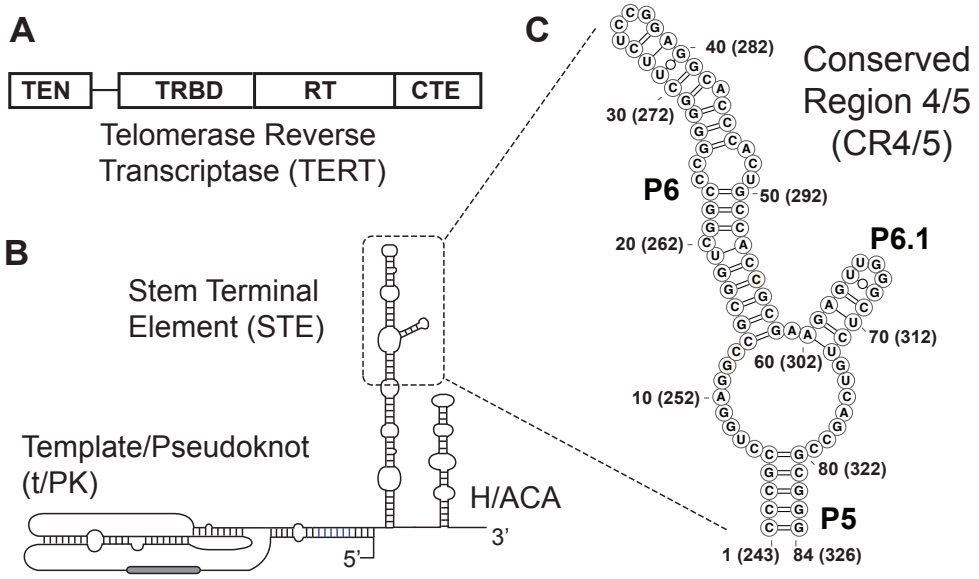


Figure 1: Conserved protein and RNA domains of the telomerase catalytic core.

(A) The conserved domain architecture of the telomerase reverse transcriptase (TERT) catalytic protein subunit, including the telomerase essential amino-terminal (TEN) domain, the RNA-binding domain (TRBD), the reverse transcriptase (RT) domain, and the carboxy-terminal extension (CTE). (B) The conserved domain organization of the human telomerase RNA (TR), including the template/pseudoknot (t/PK) domain, the stem terminal element (STE), and the H/ACA box motif. (C) Conserved regions 4 and 5 (CR4/5) domain of the human TR (hTR) comprised of stems P5, P6, and P6.1. Nucleotide numbering system used throughout the study is indicated together with the corresponding nucleotide numbering within full-length hTR in parentheses.

Detailed biochemical studies performed on vertebrate TR CR4/5 variants have shown that a stably formed P6.1 helix within the TWJ is essential for telomerase assembly and function[24-26]. Chemical and enzymatic RNA structure probing experiments of full-length hTR have reported a complex pattern of both reactivity and protection in the P6.1 stem and the adjacent junction region leading to mixed conclusions regarding the overall architecture of the TWJ region[27,28]. However, NMR studies of isolated P6.1 constructs have demonstrated that this RNA sequence is capable of adopting a stable stem–loop motif and is even further stabilized by pseudouridine modifications that may occur in some hTR molecules in vivo[28-30]. More recently, the human telomerase holoenzyme protein TCAB1 was implicated in mediating proper folding of the CR4/5 TWJ domain[31]. Protein–RNA cross-linking studies and an atomic-resolution structure of the medaka fish TR TWJ bound by its cognate TERT–TRBD revealed the molecular details of the TERT–RNA interaction[26,32]. Interestingly, the helical arrangement observed in the medaka protein–RNA complex was substantially altered when compared to the solution structure of the same RNA domain in the absence of protein[33]. Over the last several years, cryo-EM structures of the *Tetrahymena* and human telomerase RNPs were reported[34,35], providing additional details on the arrangement of protein and RNA domains within the fully assembled telomerase RNP complex. Both structures suggest that an apical stem–loop within the STE (P6.1 in hTR) lies at the interface of the TERT–CTE and TERT–TRBD domains, providing clues as to the essential requirement of the P6.1 stem–loop in coupling the two TERT domains during

telomerase assembly and/or function. Despite significant advances in structural studies on hTR, open questions remain regarding the predominant fold and stability of CR4/5 in its RNP unbound state and how the folding of this junction changes upon RNP assembly. Here, we set out to characterize the in vitro RNA folding properties of the hTR CR4/5 domain using a combination of chemical mapping and structural modeling, paired together with single-molecule Förster resonance energy transfer (smFRET) experiments. Chemical probing experiments using a variety of RNA modification reagents revealed a substantial degree of reactivity within the region of hTR CR4/5 expected to form the essential P6.1 stem–loop structure. Use of chemical reactivity data to guide computational modeling of CR4/5 structure reveals the hTR P6.1 stem is predicted to fold with much less confidence than the medaka P6.1 stem. To further characterize hTR CR4/5 structure, we systematically perturbed each nucleotide within the hTR CR4/5 domain, and queried the effects of each mutation on the chemical reactivity profile[36,37]. The results of these multidimensional chemical mapping (MCM) experiments reinforce the conclusion that the P6.1 stem–loop is not well ordered in vitro. Our use of smFRET to probe the tertiary conformational properties of the hTR CR4/5 domain revealed its heterogeneous RNA folding behavior, characterized by at least three distinct FRET states. The FRET profile of a CR4/5 mutant engineered to stabilize the canonical secondary structure of the P6.1 stem was comparatively enriched with a low FRET state, and the WT CR4/5 bound to TERT yielded a homogenous FRET profile consisting of a similar low FRET state. Collectively, our results suggest the majority of molecules in the in vitro CR4/5

structural ensemble do not possess a stably folded P6.1. Upon binding TERT, CR4/5 structural heterogeneity is suppressed and the domain adopts a more uniform conformation, likely the canonical TWJ including the essential P6.1 stem.

Results

Chemical probing of the telomerase RNA three-way junction

The TWJ motif is well conserved across many telomerase RNA systems, ranging from yeasts to vertebrates. Many of the RNA structural models that are used to generate hypotheses relating to telomerase function are derived from sequence covariation analysis[22] and/or the use of biochemical mutagenesis[24,25]. One challenge of methods such as sequence covariation analysis is that the resultant models may not accurately capture the structural properties of all RNA folding intermediates before it interacts with physiological binding partners. Indeed, studies of telomerase biogenesis indicate that hTR accumulates in subnuclear compartments prior to assembly with the TERT protein subunit[38,39], raising the distinct possibility that hTR may exist in various structural states prior to telomerase assembly. To better understand the structural properties of TRs prior to and during RNP biogenesis, we set out to analyze the secondary structural properties of telomerase TWJs from two vertebrate systems: medaka fish (*Oryzias latipes*) and human. The medaka TR TWJ serves as an important benchmark in our TR structural analyses because its atomic structure is well characterized in the absence and presence of the TERT–TRBD[26,33].

For each TR system, we used an isolated CR4/5 RNA fragment to facilitate in vitro structure probing. Notably, the isolated hTR CR4/5 domain used in our studies is sufficient to support telomerase function when reconstituted with the hTR t/PK domain and TERT protein[40] (**Supplemental Fig. S1**). Several sequence elements were added to the TR segment to assist in quantitative data analysis of chemical probing experiments (**Fig. 2A**). First, a primer binding site was appended to the RNA 3′-end for use in the reverse transcriptase reactions required to readout sites of RNA modification. Second, a short RNA hairpin structure flanked by unstructured "buffer" regions was added to serve as an internal normalization control when calculating chemical reactivities (see Materials and Methods for details)[41]. De novo structure predictions using only the RNA primary sequences as calculated on the RNAstructure web server[42] yielded the lowest free energy conformations with the expected stems that collectively form the TWJ fold (**Fig. 2B**). In the case of the hTR CR4/5 domain, RNAstructure predicted an additional cross-junction clamping helix not typically included in canonical representations of this region of hTR. Furthermore, multiple structures with nearly isoenergetic stability were also predicted, including conformations lacking the essential P6.1 stem–loop (**Supplemental Fig. S2**), highlighting the need for experimental data to validate specific RNA models.

To experimentally evaluate each of these CR4/5 structure predictions, we performed selective hydroxyl acylation analyzed by primer extension (SHAPE) experiments using 1-methyl-7-nitroisatoic anhydride (1M7), a fast-acting chemical modifier[43,44]. In addition, experiments were also performed using the base-specific

reagents dimethyl sulfate (DMS) or 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT), which primarily react with adenine/cytosine or guanine/uracil bases, respectively (**Supplemental Fig. S3**). Reactivity profiles obtained by all three chemical probing methods (DMS, CMCT, and 1M7) for the medaka CR4/5 yielded data that support the canonical base pairing arrangement expected for this TWJ fold, and are highly consistent with the reported solution structure of this same RNA fragment (**Fig. 2C; Supplemental Figs. S3 and S4**)[26]. In contrast, for the human CR4/5 domain, strong 1M7 reactivity was observed in the region expected to fold into the P6.1 stem (**Fig. 2D**). To test whether this discrepancy in SHAPE profiles of the human and medaka CR4/5 domains was due to unique structural interactions with magnesium, the SHAPE experiments were repeated across a titration of $MgCl_2$. Interestingly, the reactivity patterns did not show any detectable $MgCl_2$ dependence for either the medaka or human construct (**Supplemental Fig. S5**). The reactivity observed in the hTR P6.1 stem is unexpected given previous structural studies of isolated P6.1 constructs[29] and the established importance of the P6.1 stem–loop structure in promoting telomerase RNP assembly and function[24,25,26] but is consistent with previous studies that use chemical mapping to examine the CR4/5 in full-length hTR in vivo and in vitro[27,28]. Taken together, these data suggest that using primary sequence information alone, the RNAstructure folding algorithm effectively predicts a base pairing configuration suggested by the SHAPE data of the medaka TR TWJ. However, significant disparity between the sequence alone prediction and the SHAPE data are observed in the expanded junction/6.1 stem of the

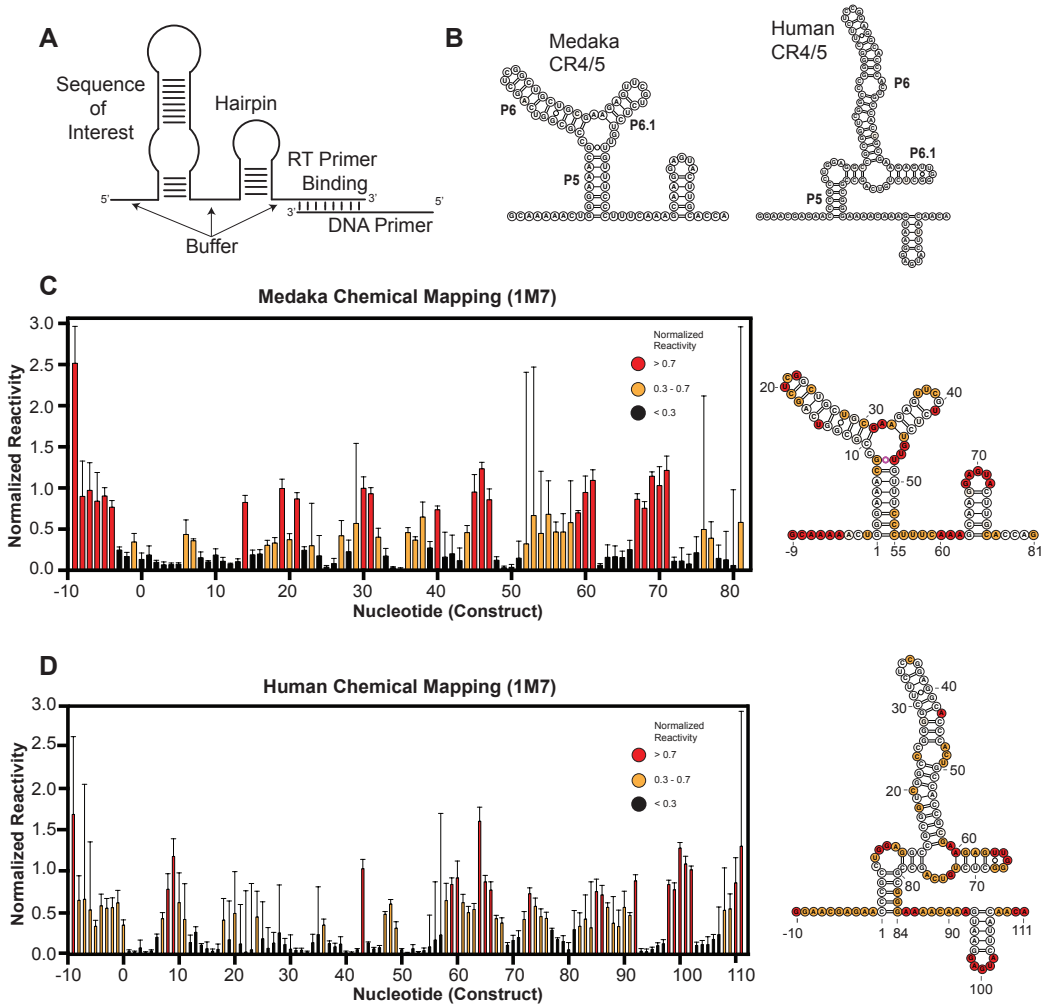


Figure 2: Chemical probing of the medaka and human CR4/5 domain.

(A) Cartoon schematic of general RNA construct design, including the RNA sequence of interest flanked by unstructured RNA buffer sequences, a normalization RNA hairpin, and a reverse transcriptase priming site. (B) Lowest energy predicted secondary structure of medaka (left) and human (right) CR4/5 domain using RNAstructure. (C) (left) Chemical mapping of the medaka CR4/5 domain by SHAPE (1M7 probing) at 1 mM MgCl2. (D) Chemical mapping of the human CR4/5 domain by SHAPE (1M7 probing) at 1 mM MgCl2. For both (C,D), color coding in the bar plot and structure schematic is as described in C. Plotted normalized reactivity values are color-coded (red >0.7, yellow 0.3–0.7, and black <0.3). Each bar plotted represents experiments conducted in triplicate or greater with the respective standard deviation as error bars (right). Color-coded schematic of the reactivity data is shown on the RNAstructure predicted secondary structure.

hTR CR4/5 domain. Thus, human CR4/5 displays a complex folding behavior that confounds RNAstructure predictions in the absence of chemical probing data.

SHAPE-guided modeling of human CR4/5 does not support formation of the P6.1 stem

RNAstructure calculates the lowest free energy structures using thermodynamic parameters that are dynamically sampled against databases of structures with well-characterized stabilities[42]. Experimentally derived chemical probing data significantly improves the predictive power of the RNAstructure folding algorithm[45]. SHAPE reactivities are used to calculate a pseudoenergy change term (ΔG_{SHAPE}) at each nucleotide i using the formula $\Delta G_{SHAPE}(i) = m$ ln(SHAPE reactivity(i) + 1) + b, which is then utilized as a nearest neighbor free energy term for structure prediction[46]. The slope and intercept parameters m and b, respectively, were empirically parameterized against the 23S rRNA and produce accurate (>89% correct base pairs) predictions even when varied within a large "sweet spot" of absolute values[46]. Importantly, the slope parameter m can be increased to disfavor the prediction of helices containing reactive nucleotides.

Using this approach, we performed SHAPE experiments of the previously mentioned medaka and human CR4/5 constructs, then generated SHAPE-guided structure models while increasing the slope parameter within its accurate range (1.8–5 kcal/mol). In our analysis, we used the Biers component of the HiTRACE software package to implement RNAstructure with a nonparametric bootstrapping function to

estimate confidence values for each RNA helix in the predicted structures[36,37]. The bootstrapping function iteratively subsamples the reactivity data with replacement, then runs the RNAstructure algorithm. The collection of bootstrapping-derived structures is then used to calculate the frequency of each RNA helix present across all computationally derived replicates. In this way, the resulting bootstrap value for any given helix provides a metric to evaluate its predictive confidence. It is important to note that bootstrap values are a statistical tool to analyze computational prediction methods, and should not be interpreted as an indicator of the equilibrium conformation(s) present for a particular RNA of interest.

As expected, the addition of the ΔG_{SHAPE} constraints to predictions of the medaka TR CR4/5 yields the canonical TWJ fold with each of the expected helices being called with high confidence as the SHAPE slope parameter was increased (**Fig. 3A**). Bootstrap-calculated confidence in the P6 stem slightly decreases at higher SHAPE slope (>4 kcal/mol) because of the presence of moderate SHAPE reactivity at nucleotides known to be base paired in the crystal structure (G14, C16, A17). Overall, this result indicated that addition of experimentally derived data does not cause the RNAstructure algorithm to significantly deviate in its prediction of the lowest energy conformation for the medaka TR CR4/5. In the case of the hTR CR4/5, the inclusion of ΔG_{SHAPE} constraints in structure calculation recaptures a lowest energy conformation in which the P5, P6, and normalization hairpin are called with high confidence. In contrast, the confidence value of the P6.1 stem significantly decreases as SHAPE slope increases, consistent with the high levels of SHAPE reactivity in this

region disfavoring the prediction of a stem–loop motif (**Fig. 3B**). These data-driven structure predictions indicate the hTR CR4/5 domain likely does not adopt its expected TWJ motif in the absence of telomerase-associated proteins in vitro.

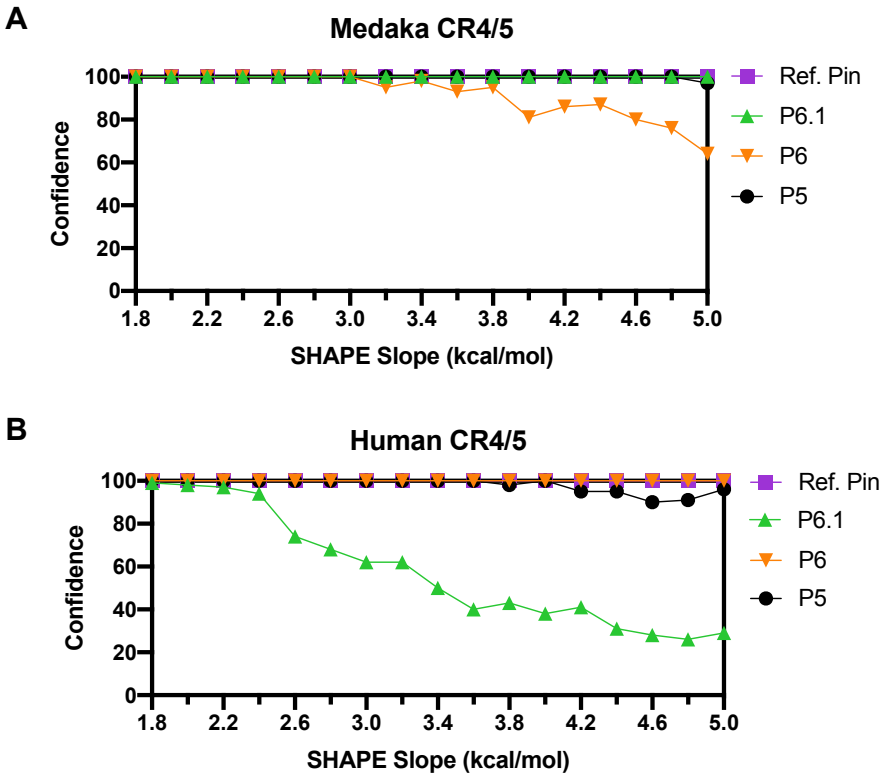


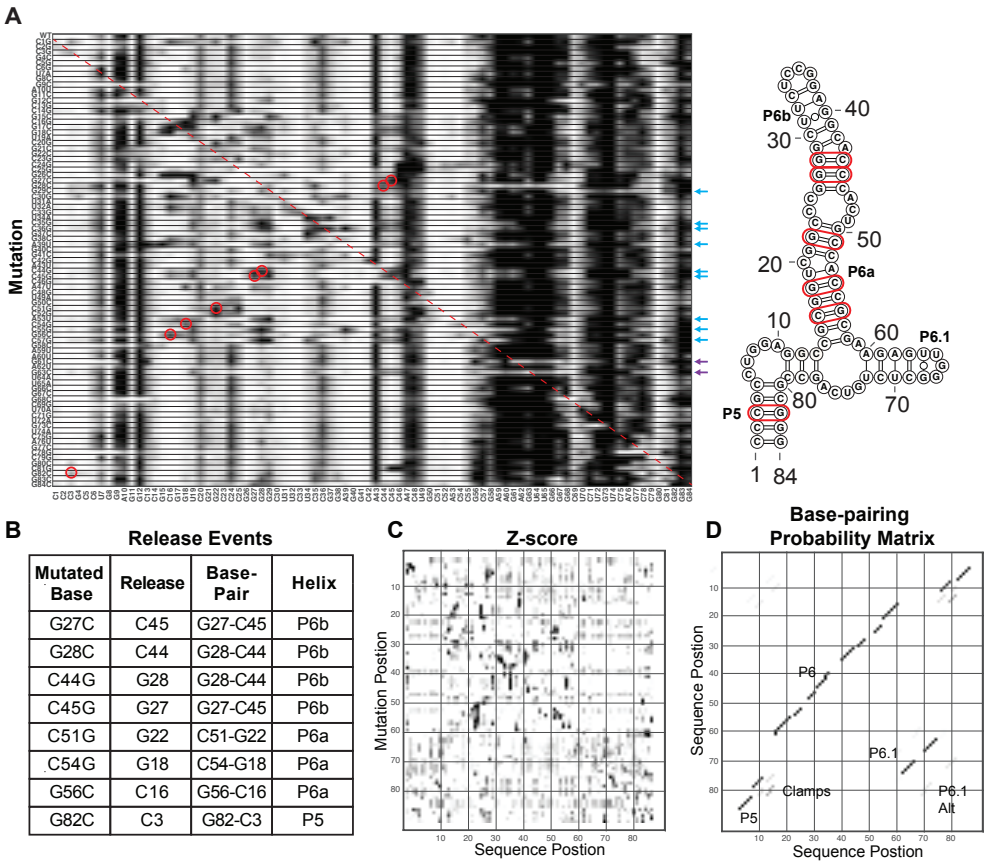
Figure 3: Data-guided RNA secondary structure prediction of medaka and human CR4/5 domains.

SHAPE (1M7) reactivity data were used as weights to guide RNA structure prediction for medaka (A) and human (B) CR4/5 domains. Using the Biers package of HiTRACE, RNAstructure models of each RNA domain were calculated with 100 bootstrap replicates, while varying the SHAPE slope parameter in intervals of 0.2 kcal/mol. The abundance of each helical RNA element (Confidence) derives from the bootstrap replicates and is plotted for each respective value of SHAPE slope. Ref. Pin refers to Reference hairpin.

Multidimensional chemical mapping supports hTR CR4/5 structural heterogeneity

To further probe the structure of the hTR CR4/5 domain, we performed multidimensional chemical mapping (MCM)[36]. This systematic mutagenesis approach permits rapid chemical probing analysis of a panel of RNA mutant constructs designed to explicitly test for the presence of Watson–Crick base pairing in a proposed RNA secondary structural model[36,37]. If a mutation is made to a base that is engaged in a base pair, then one expects the release of the interacting partner that consequently becomes accessible to the SHAPE probe.

To probe for such specific release events, we generated a set of 84 mutants across the entire hTR CR4/5 construct. The chemical reactivity profiles of all RNA variants were stacked vertically to generate a reactivity tapestry (**Fig. 4A**). Signals on the diagonal of the reactivity tapestry represent release events at the engineered site of mutation (**Fig. 4A, red dotted line**). Signals that deviate from the wild-type reactivity profile indicate changes in reactivity that result from each individual mutation. Many of the single-mutant reactivity profiles revealed complex structural rearrangements beyond the simple base pair release event principle. However, visual inspection of the data reveals multiple features in the reactivity tapestry that support specific base pairs present within the hTR CR4/5 (**Fig. 4A, red circles**, and **Fig. 4B**). For example, the G27C and G28C mutations each resulted in increased reactivity at positions C45 and C44, respectively, providing support for these base pairs being present within the P6b stem (**Fig. 4B**). Similarly, the C44G and C45G mutations resulted in release events in



Mutated Base	Release	Base-Pair	Helix
G27C	C45	G27-C45	P6b
G28C	C44	G28-C44	P6b
C44G	G28	G28-C44	P6b
C45G	G27	G27-C45	P6b
C51G	G22	C51-G22	P6a
C54G	G18	C54-G18	P6a
G56C	C16	G56-C16	P6a
G82C	C3	G82-C3	P5

Figure 4: Mutate-and-map profile of the human CR4/5 domain.

(A) Systematic mutations were introduced at each base within the hTR CR4/5 domain as indicated (A → U, U → A, G → C, and C → U). The structure of each mutant was interrogated by SHAPE (1M7), and the resultant reactivity profiles were stacked to create a reactivity tapestry that permits visual comparison of the chemical reactivity at each nucleotide across all mutants. The red dashed line corresponds to the position of expected signal of enhanced reactivity at the site of the base substitution. Specific sites of enhanced reactivity ("release events") are circled in red. Positions of validated base pairing interactions are highlighted in red in the secondary structure model shown to the right. Mutation positions with the P6 stem (blue arrows) and P6.1 stem (purple arrows) that induce large-scale changes in the reactivity patterns are indicated. (B) Summary of specific mutations and sites of correlated enhancements of chemical reactivity of CR4/5 base pairs that these data support. (C) Bootstrap support values are plotted in a base pair probability matrix represented in gray scale. High confidence stems give rise to dark and symmetric signals. (D) Representative alternative hTR CR4/5 junction structure predictions from mutate-and-map experiments.

G28 and G27, respectively, providing independent support for these same base pairs in the P6b stem. Increased reactivity was also observed for certain mutations within the P6a stem; for example, the C51G, C54G, and G56C mutations each caused increased signal at positions G22, G18, and C16, respectively. Lastly, the G82C mutation located within the P5 stem resulted in increased reactivity at position C3, providing support for this specific base pairing interaction. Notably, the high baseline reactivity observed in the hTR CR4/5 junction and P6.1 stem–loop region precludes unambiguous visual analysis of the MCM data. However, we found that mutations introduced at the base of the P6a stem (A53U, C55G, and C57G) and several mutations in P6b had the unexpected effect of causing substantial structural rearrangements in the CR4/5 domain, evidenced by reduced reactivity in the junction region and increased reactivity within the P6 stem (**Fig. 4A, blue arrows**). Other notable global folding changes were observed for single G → C substitutions located within the P6.1 stem, such as G61C and G63C, which both induce the CR4/5 domain to fold into an extended two-helix junction (**Fig. 4A, purple arrows**, and **Supplemental Fig. S6**).

To achieve a quantitative analysis across the entire reactivity tapestry we generated a *Z*-score plot, where individual *Z*-scores report on the statistical significance of deviations in the reactivity level for a given nucleotide compared across all RNA constructs (**Supplemental Fig. S7**). *Z*-score values are then used as a pseudoenergy term to guide structure prediction by RNAstructure within the Biers component of the HiTRACE software package[37]. As with the SHAPE reactivity-

guided RNAstructure calculations, the *Z*-score data can be used to perform bootstrapping analysis as a measure of confidence in each predicted helical segment and to generate a base pair probability matrix (**Fig. 4C**). The results of the *Z*-score analysis are consistent with the presence of structures other than the canonical P5, P6, and P6.1 stems in the CR4/5 structure ensemble. For example, in multiple Z-score-driven structures, an alternative P6.1 stem (P6.1 alt) was predicted in addition to several mutually exclusive cross-junction clamping helices (**Fig. 4D**). Taken together, the results of the MCM experiments provide additional experimental evidence for base pairing interactions in the P6a, P6b, and P5 stems, and support the notion that the junction region and P6.1 stem–loop may adopt noncanonical base pairing configurations.

Single-molecule analysis reveals CR4/5 folding heterogeneity and remodeling upon telomerase RNP assembly

Results from our ensemble chemical probing experiments suggest that the human CR4/5 domain exhibits folding heterogeneity, particularly in the junction region that is proximal to the functionally essential P6.1 stem–loop. To directly detect hTR CR4/5 folding heterogeneity in the presence and absence of TERT protein, and to understand how heterogeneity of hTR CR4/5 secondary structure affects its tertiary conformation, we used a single-molecule Förster resonance energy transfer (smFRET) technique. Single-molecule FRET measures RNA conformation(s) as the distance-dependent energy transfer between a FRET donor (Cy3) and an acceptor

(Cy5) dye incorporated into the RNA. FRET probes were strategically incorporated at positions U32 (Cy5) and U70 (Cy3) to establish a dye pair that reports on the physical proximity of the P6 and P6.1 stem–loops (Supplemental Fig. S1). Using this design principle, we created two different FRET constructs: a WT CR4/5 domain and a mutant CR4/5 designed to encourage P6.1 folding (Mut CR4/5) (**Fig. 5A**). This mutant CR4/5 contains junction linker regions consisting only of adenines intended to constrain its folding landscape to favor the formation of the P6.1 stem. One-dimensional chemical probing with 1M7 and SHAPE-guided modeling of Mut CR4/5 supported the notion that the P6.1 stem within the Mut CR4/5 construct forms more readily compared to WT CR4/5 (**Fig. 5B**) (**Supplemental Fig. S8**). Importantly, both the WT and Mut CR4/5 constructs reconstitute active telomerase complexes in vitro (**Supplemental Fig. S9**). We note that while the Mut CR4/5 construct appeared to show a slight decrease in telomerase reconstitution efficiency, the assembled RNP complexes displayed quantitatively indistinguishable repeat addition processivity values as measured by direct primer extension assays (**Supplemental Fig. S9**). Single-molecule measurements were made using a solution confocal fluorescence microscope, in which FRET values are extracted from individual freely diffusing molecules as they traverse through the excitation beam. We then collected several thousand FRET values from free and TERT-bound CR4/5 molecules, compiled them into histograms, and fit the data with Gaussian functions to approximate distinct FRET populations.

The WT CR4/5 domain exhibits a substantially heterogeneous FRET profile consisting of at least three unique FRET populations, with the majority of molecules falling into populations centered at higher FRET (~0.75 and ~0.9) along with a minor population at lower FRET (~0.3) (**Fig. 5C, top panel**). This observation is consistent with our chemical probing data, which suggests the P6.1 stem is not a stably folded motif and that this region of CR4/5 displays structural heterogeneity. Molecules reporting high FRET values likely exist in a conformation in which the P6.1 nucleotides are in close proximity to P6b, while lower FRET states indicated conformations of CR4/5, in which the P6.1 nucleotides are distal from P6b in tertiary space. We then investigated how stabilizing the secondary structure of the P6.1 stem would affect the structural heterogeneity of CR4/5. Whereas the FRET distribution of WT CR4/5 is predominantly represented by two populations reporting higher FRET (0.75 and 0.9) range and marginally low FRET (~0.3) population, the FRET distribution of Mut CR4/5 appears significantly less heterogeneous, comprised mostly of molecules falling into a single low FRET population (**Fig. 5C, middle panel**). These observations suggest that stabilizing the P6.1 stem constrains overall structural heterogeneity of CR4/5, shifting the folding landscape toward a low FRET conformation.

Next, we measured the FRET properties of the WT CR4/5 domain after reconstitution with TERT and the hTR template/pseudoknot (t/PK) domain into catalytically active telomerase RNP complexes. Assembly of WT CR4/5 into telomerase RNPs essentially abolishes the apparent heterogeneity of the CR4/5

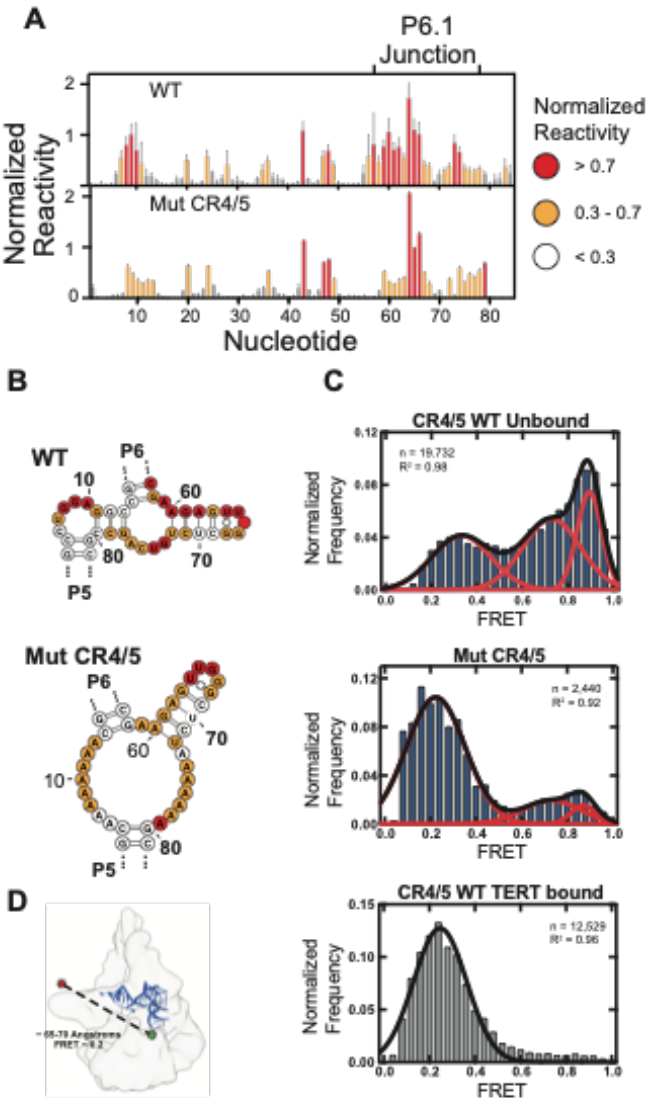


Figure 5: Single-molecule FRET analysis of hTR CR4/5 domain.

(A) SHAPE reactivity profiles of WT (top plot) and Mut (bottom plot) CR4/5. (B) RNAstructure predictions of WT (top) and Mut (bottom) CR4/5 with overlaid SHAPE reactivity (C) smFRET histograms collected from freely diffusing WT (top), Mut (middle), and TERT-bound (bottom) CR4/5. Red and black lines depict Gaussian functions manually fit to the data with associated R-squared. (D) Cryo-EM density of human telomerase[35] with a docked medaka CR4/5 crystal structure[33]. Approximate dye locations and inter-dye distance are indicated.

domain and yields a single low FRET population (~0.3) (**Fig. CB, bottom panel**). This finding suggests that upon telomerase assembly, and consequently the folding of the P6.1 stem motif, the P6.1 and P6b stems are stabilized at an increased distance from each other. The estimated distance (~65–70 angstroms) between the FRET dyes in an assembled state is consistent with the respective dye label positions modeled in the human telomerase cryo-EM structure (**Fig. 5D**)[35]. This result lends additional support to a human CR4/5 structural transition upon binding to the TERT protein as was proposed for the medaka CR4/5 domain[26,33].

Discussion

Telomerase RNPs derived from diverse organisms must assemble upon highly structured telomerase RNA (TR) scaffolds[18,47]. TRs possess a multidomain architecture conserved from unicellular ciliates to humans and serve to nucleate the assembly of telomerase complexes through interactions with the telomerase reverse transcriptase (TERT) and other lineage-specific proteins[20-22,48]. Despite their essential role in telomerase assembly, it remains unclear how TRs transition from their initial protein-free conformations to the intricate tertiary structures seen in active telomerase complexes[34,35] and how the nucleotides in the junction of the human CR4/5 affect the structural architecture. In the present study, we use a novel combination of SHAPE-guided RNA modeling and smFRET to demonstrate that the essential P6.1 stem of hTR CR4/5 is not stably folded in vitro and exists as a structural ensemble that is remodeled by the binding of TERT.

The stem terminal element ([STE] stem–loop IV in *Tetrahymena* TR and CR4/5 in hTR) makes a high affinity interaction with TERT[32] and, when mutated, abrogates telomerase biogenesis[24,25], precipitating human disease. In the *Tetrahymena* telomerase RNP, the TR stem–loop IV binds the assembly factor p65, which stabilizes a bent-helix conformation that places the apical loop at the interface of the TRBD and CTE domains of *Tetrahymena* TERT, potentially stabilizing the architecture of TERT[34,49-52]. Similarly, folding of the TR pseudoknot requires interactions with *Tetrahymena* TERT to stably form and support catalytic activity of telomerase[53]. In hTR, the H/ACA box proteins (Dyskerin, NOP10, NHP2, and GAR1) regulate telomerase biogenesis and may play a similar role in facilitating the CR4/5 to adopt a conformation that engages the TRBD–CTE interface[31,54]. Structural studies of the smaller *Oryzias latipes* (medaka) CR4/5 revealed protein-induced rearrangements of the TWJ motif, rotating the P6.1 stem nearly 180 degrees around the axis of P5 and P6 to clamp upon the TERT RNA-binding domain (TRBD)[26,33]. Presumably, the hTR CR4/5 adopts a similar RNP assembled conformation given it shares invariant nucleotides comprising the P6.1 region and most of the TWJ motif, a notion consistent with the medium-resolution cryo-EM structure of human telomerase[35]. NMR studies demonstrate that pseudouridinylation of the P6.1 stem may alter the structural stability of the P6.1[28,30,31]; however, the precise role of posttranscriptional modification of hTR in RNA folding and telomerase biogenesis is not firmly established. Moreover, the human TWJ is expanded by ten nucleotides compared to its medaka counterpart and therefore

traverses a more complex folding landscape to arrive at its functional RNP state. The role of this expanded junction in human hTR folding has remained enigmatic.

Chemical mapping has been previously used to qualitatively infer hTR structure in its protein-bound and -unbound states[27,28]. Here, we use chemical data to guide in silico predictions that suggest hTR CR4/5 adopts noncanonical TWJ folds in the absence of TERT protein. Our analysis produces a secondary structure model of medaka CR4/5 consistent with its atomic resolution model (**Fig. 2**). In contrast, our analysis of human CR4/5 suggests that the P6.1 stem is not stably formed, as it is predicted with notably less abundance as the free energy penalty for its reactive nucleotides are increased within the previously established accurate range of values (Fig. 3)[46]. An exhaustive mutate-and-map strategy[36,37] of hTR CR4/5 identified base pairing signatures between specific nucleotides in P5 and P6, but was unable to detect Watson–Crick base pairing between nucleotides proposed to form P6.1 (**Fig. 4**). Notably, two mutations we analyzed by mutate-and-map, C45G and G63C (C287G and G305C, respectively, in full-length hTR) either drastically decrease or abolish 1M7 reactivity of the P6.1 region. C45G (C287G) is a patient-derived hTR mutation in the CR4/5 P6b stem that disrupts RNP assembly and induces aplastic anemia[13]. The G63C (G305C) mutation resides in P6.1 and disrupts the same G–C pair as a mutation associated with dyskeratosis congenita[55]. RNAstructure predictions of these mutants based solely on primary sequence reveal a non-TWJ conformation, in which nucleotides from the P6.1 region pair with nucleotides from the P6a stem (Supplemental Fig. S6). These non-TWJ conformations are consistent with our

mutate-and-map data, shedding light on the etiology of diseases arising from mutations in hTR that affect CR4/5 architecture.

The sequence of the P6.1 stem and junction region are strictly conserved across vertebrate TRs. Covariance patterns in CR4/5 suggest evolutionary pressure to maintain the P5 and P6 stems[22], whereas the P6.1 stem lacks any instances of covarying base pairs. Yet, it is known that a stable P6.1 stem is required for TERT binding[24-26,32]. The extreme sequence conservation within P6.1 stem–loop and junction region of hTR suggests the presence of a selective pressure other than preservation of RNA structure alone. Our in vitro smFRET data demonstrate that the majority of the free CR4/5 RNA is in a structure that is not the state found in the TERT-bound state. Using the engineered poly(A) CR4/5 mutant, we observe that the highly conserved nucleotides in the junction are critical in setting up RNA architecture at the junction and that in the absence of competing structures the RNA fold resembles that of the TERT-bound state. This data provides an explanation for the seemingly discordant findings that the junction region is highly conserved but lacks sequence covariation. Namely, that the P6.1 stem is characteristic of the functionally "assembled" state of hTR CR4/5, but in the absence of TERT the junction adopts conformations other than the canonical P6.1 stem (**Fig. 6**). While the specific identities and functional role(s) of alternate CR4/5 folds remain to be determined, it is conceivable that junction nucleotides may be conserved to preserve RNA structural plasticity required for RNP assembly, as well as to mediate sequence-specific protein interactions that may or may not be present in the fully assembled

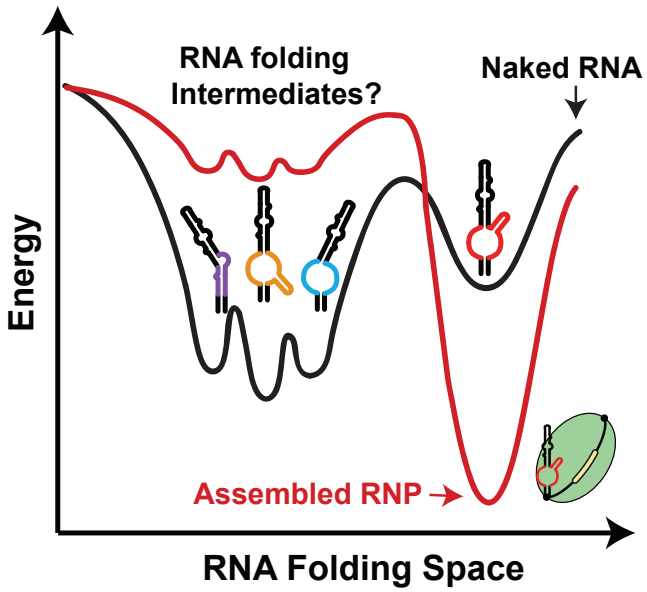


Figure 6: Model describing functional role of CR4/5 folding heterogeneity in human telomerase biogenesis.

A schematic depicting a hypothetical folding landscape of the hTR CR4/5 domain. Energy valleys represent unique conformations available to CR4/5. The 'depth' of a valley is a conceptual proxy for the stability and relative abundance of a particular conformation. In the folding landscape of the 'naked' CR4/5 (black line), there exists a diverse ensemble of TWJ conformations with a small contingent of molecules adopting a fold representative of the canonical P6.1 stem (red TWJ). Upon RNP assembly, the CR4/5 folding landscape becomes dominated by one predominant CR4/5 conformation (red line) because of TERT-induced remodeling of the CR4/5 structure.

RNP complex. Another interpretation is that CR4/5 structural heterogeneity limits telomerase assembly, and requires interactions with telomerase proteins to properly assemble the active enzyme. Recently, a study showed that the binding of telomerase-associated protein TCAB1 to hTR positively influences the folding of the P6.1 and P6b stems in an hTR construct lacking the template/pseudoknot domain[31]. Thus, TCAB1 may enforce proper CR4/5 folding either through direct protein–RNA

interactions or potentially mediating access to other binding partners via trafficking hTR to Cajal bodies[56]. Future studies using methods such as DMS-MapSeq will permit investigation of how hTR folds in vivo during various stages of telomerase biogenesis, as well as how RNA modifications affect hTR folding[57].

Materials and Methods

Preparation of RNAs for chemical probing and in vitro telomerase reconstitution

Design and synthesis of RNA chemical probing constructs

Constructs for RNA chemical probing contained the RNA of interest (medaka CR4/5 [nt 170–220] and hTR CR4/5 [nt 243–326]) with additional flanking sequences for normalization purposes in data analysis (described below) and for reverse transcriptase binding[41]. RNA constructs for chemical probing were iteratively queried on the RNAstructure web server[42] and redesigned to discourage base pairing of the flanking sequences with the RNA of interest. Each RNA construct was synthesized by in vitro transcription. The DNA templates were assembled from DNA oligonucleotides designed using the Primerize tool[58] and synthesized by IDT (**Supplemental Fig. S10; Supplemental Table S2**). In the event that a complete DNA template could not be synthesized by one primer assembly reaction using Phusion polymerase (NEB), a "two-piece" scheme was used, in which the products of two separate primer assemblies were used to generate the complete DNA product.

In vitro transcription of RNAs

RNA constructs for chemical probing and fragments used for in vitro telomerase reconstitution (hTR CR4/5 [nt 239–328] and hTR t/PK [nt 32–195]) were in vitro transcribed using homemade T7 RNA polymerase[59] in RNA polymerase reaction buffer (40 mM Tris-HCl, pH 7.9, 28 mM $MgCl_2$, 90 mM DTT, 2 mM spermidine, 1.5

mM each NTP, and 40 U RNasin Plus [Promega]). The reaction was incubated overnight at 37°C followed by the addition of 10 units of TURBO DNase (Thermo Fisher) for 15 min at 37°C. RNA was phenol–chloroform extracted and ethanol precipitated prior to denaturing urea polyacrylamide gel electrophoresis (PAGE) purification. RNAs used in mutate-and-map experiments were transcribed in parallel on 96-well plates and purified using AMPure XP beads (Agencourt). RNA quality was then checked diagnostically by denaturing urea PAGE.

Structural modeling of RNAs guided by chemical probing data

Chemical probing of RNAs

Chemical probing and mutate-and-map experiments were carried out as described previously[41,60,61]. Briefly, 1.2 pmol of RNA was denaturated at 95°C in 50 mM Na-HEPES, pH 8.0, for 3 min, and folded by cooling to room temperature over 20 min, and adding $MgCl_2$ to the desired concentration (1–10 mM). RNA was aliquoted in 15 µL volumes into a 96-well plate and mixed with nuclease-free H_2O (control), or chemically modified in the presence of 5 mM 1-methyl-7-nitroisatoic anhydride (1M7)[44], 25 mM 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT, Sigma-Aldrich), or 0.25% dimethyl sulfate (DMS, Sigma-Aldrich) for 10 min at room temperature. Mutate-and-map experiments utilized only 1M7 as the chemical modifier and at a 10 mM $MgCl_2$. Chemical modification was stopped by adding 9.75 µL quench and purification mix (1.53 M NaCl, 1.5 µL washed oligo-dT beads, Ambion), 6.4 nM FAM-labeled, reverse transcriptase primer (sequence

in Supplemental Table S1), and 2.55 M Na-MES for 1M7 and CMCT reactions, or 50% 2-mercaptoethanol for DMS reactions. RNA in each well was purified by bead immobilization on a magnetic rack and two washes with 100 µL 70% ethanol. RNA was then resuspended in 2.5 µL nuclease-free water prior to reverse transcription.

Reverse transcription of modified RNAs and cDNA purification

RNA was reverse transcribed from annealed fluorescent primer in a reaction containing 1× First Strand Buffer (Thermo Fisher), 5 mM DTT, 0.8 mM dNTP mix, and 20 U of SuperScript III Reverse Transcriptase (Thermo Fisher) at 48°C for 30 min. RNA was hydrolyzed in the presence of 200 mM NaOH at 95°C for 3 min, then placed on ice for 3 min and quenched with 1 volume 5 M NaCl, 1 volume 2 M HCl, and 1 volume 3 M sodium acetate. cDNA was purified on magnetic beads as described previously, then eluted by incubation for 20 min in 11 µL Formamide-ROX350 mix (1000 µL Hi-Di Formamide [Thermo Fisher] and 8 µL ROX350 ladder [Thermo Fisher]). Samples were then transferred to a 96-well plate in "concentrated" (4 µL sample + 11 µL ROX mix) and "dilute" (1 µL sample + 14 µL ROX mix) for saturation correction in downstream analysis. Sample plates were sent to Elim Biopharmaceuticals for analysis by capillary electrophoresis.

Analysis of capillary electrophoresis data with HiTRACE

Capillary electrophoresis runs from chemical probing and mutate-and-map experiments were analyzed with the HiTRACE MATLAB package[62]. All of the raw

data presented in the current study are freely available on the RNA Mapping Database (RMDB IDs: M2CR45_1M7_0000, MCR45_1M7_000, HCR45_1M7_000)[63]. Lanes of similar treatment groups (e.g., 1M7 modified) were aligned together, bands fit to Gaussian peaks, background subtracted using the no-modification lane, corrected for signal attenuation, and normalized to the internal hairpin control. The end result of these steps is a numerical array of "reactivity" values for each RNA nucleotide that can be used as weights in structure prediction. For mutate-and-map data sets, each nucleotide is assigned a *Z*-score, calculated as its average reactivity across all mutants divided by the standard deviation[36]. Nucleotides with overall high reactivity across the mutants (average of 0.8 or higher) are ignored in *Z*-score calculation.

Data-guided RNA structure prediction

Data-guided secondary structure modeling was performed using the Biers MATLAB package (https://ribokit.github.io/Biers/). Briefly, the Fold function of the RNAstructure suite applied reactivity values as pseudoenergy modifiers to calculate the minimum free energy structure of CR4/5 RNA. Bootstrapping analysis of data-guided structure prediction was performed as described previously[36,37]. For mutate-and-map data sets, *Z*-scores were used as pseudoenergy modifiers to calculate a base pairing probability matrix with RNAstructure and to run bootstrapping analysis with Biers. Secondary structures were visualized using the VARNA applet[64].

Methods for SHAPE modeling

SHAPE-guided predictions were performed with the HiTRACE MATLAB package. The "rna_structure" script was run, while varying the SHAPE slope parameter argument (use the command "open rna_structure" for help) from 1.8 to 5.0 kcal/mol. One hundred bootstrap replicates were performed for each prediction run. Then the results were visualized using the "output_varna" command to produce RNA secondary structure models. For each prediction run, we queried the percent abundance of each canonical human CR4/5 helical element (P5, P6a, P6b, P6.1) and the embedded reference hairpin from among the bootstrapped models. The percent abundance (or bootstrap confidence) of each RNA helix was then recorded under the associated SHAPE slope parameter used to calculate the predictions.

Telomerase expression and purification

In vitro reconstitution of human telomerase

Human telomerase was reconstituted in rabbit reticulocyte lysate (RRL) using the TNT Quick Coupled Transcription/Translation System (Promega) as described previously[65,66]. In LoBind tubes (Eppendorf), 200 µL of TnT quick mix was combined with 5 µg of pNFLAG-hTERT plasmid as well as 1 µM of in vitro transcribed and unlabeled hTR t/PK and CR4/5 fragments. Less abundant dye-labeled CR4/5 was added at 0.1 µM. The reaction was incubated for 3 h at 30°C. 5 µL of 0.5 M EDTA, pH 8.0, were then added to chelate Mg^{2+} ions present in the lysate. Human telomerase was immunopurified via the amino-terminal FLAG tag on hTERT using

αFLAG M2-agarose beads (Sigma-Aldrich). Beads contained in 50 µL bead slurry were first washed three times with wash buffer (50 mM Tris-HCl, pH 8.3, 3 mM $MgCl_2$, 2 mM DTT, 100 mM NaCl) with 30 sec centrifugation steps at 2350 rcf at 4°C after each wash. The beads were then blocked twice in blocking buffer (50 mM Tris-HCl, pH 8.3, 3 mM $MgCl_2$, 2 mM DTT, 500 µg/mL BSA, 50 µg/mL glycogen, 100 µg/mL yeast tRNA) for 15 min under gentle agitation at 4°C followed by 30 sec centrifugation at 2350 rcf and removal of the supernatant. After blocking, the beads were resuspended in 200 µL blocking buffer and added to the telomerase reconstitution reaction in RRL. The beads and lysate were incubated for 2 h at 4°C under gentle agitation. The beads were then pelleted for 30 sec at 2350 rcf and at 4°C and the supernatant was discarded. The beads were then washed three times in wash buffer containing 300 mM NaCl followed by three wash steps in wash buffer containing 100 mM NaCl. A 30 sec centrifugation at 2350 rcf at 4°C was performed between each wash cycle. To elute the enzyme, the beads were incubated in 60 µL elution buffer (50 mM Tris-HCl, pH 8.3, 3 mM $MgCl_2$, 2 mM DTT, 750 µg/mL 3× FLAG peptide, 20% glycerol) under gentle agitation at 4°C for 1 h. After elution, the beads were removed by centrifugation at 10,000 rcf through Nanosep MF 0.45 µm filters. 5 µL aliquots were prepared in LoBind tubes (Eppendorf), flash frozen in liquid nitrogen, and stored at −80°C until use.

Telomerase activity assays

^{32}P-end-labelling of DNA primers

A total of 50 pmol of DNA primer was labeled with γ-^{32}P ATP using T4 polynucleotide kinase (NEB) in 1× PNK buffer (70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT) in 50 µL reaction volume. The reaction was incubated for 1 h at 37°C followed by heat inactivation of T4 PNK at 65°C for 20 min. Centrispin columns (Princeton Separations) were used to purify labeled primer.

Primer extension assays

Telomerase activity assays of in vitro reconstituted human telomerase were performed using 5 µL purified telomerase in a 15 µl reaction volume brought to 1× activity buffer concentrations (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 1 mM $MgCl_2$, 2 mM DTT, 50 nM ^{32}P-end-labled primer, and 10 µM of each dATP, dTTP, and dGTP). Reactions were incubated for 90 min at 30°C and quenched with 200 µL 1× TES buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.1% SDS). DNA products were then phenol–chloroform extracted and ethanol precipitated. DNA pellets were resuspended in 1× formamide gel loading buffer (50 mM Tris Base, 50 mM boric acid, 2 mM EDTA, 80% [v/v] formamide, 0.05% [w/v] each bromophenol blue and xylene cyanol), and resolved on a 12% denaturing urea PAGE gel. The gel was then dried and exposed to a storage phosphor screen (GE Healthcare) and scanned using a Typhoon scanner (GE Healthcare). Band intensities were quantified using SAFA and ImageJ[67,68]. The “fraction left behind” (FLB) for a given lane was calculated by

summing each repeat addition processivity (RAP) band and all RAP bands below it divided by the total RAP band intensity counts for that lane. The natural logarithm of (1-FLB) was then plotted against repeat number and fitted by linear regression. The slope value of the linear fit was used to determine processivity $R_{1/2}$ values from $-\ln(2)/\text{slope}$[69]. Total activity was calculated in ImageJ by taking the total intensity of each lane and normalizing to the wild-type lane.

Preparation of dye-labeled hTR CR4/5 for single-molecule experiments

Synthesis of dye-labeled hTR CR4/5 RNA

Synthetic CR4/5 (hTR 239–330) was ordered from Dharmacon as two separate oligonucleotides: Fragment 1 (hTR 239–278) and Fragment 2 (hTR 279–330), each harboring a site-specific aminoallyl modification at the five position of uracil base as indicated in Supplemental Table S1. Oligonucleotides were deprotected in deprotection buffer (100 mM acetic acid, pH 3.6) following the manufacturer's instructions, then ethanol precipitated in the presence of 300 mM sodium acetate, pH 5.2. To enable RNA ligation, Fragment 2 was phosphorylated using T4 PNK (NEB), phenol–chloroform extracted, and ethanol precipitated in the presence of sodium acetate. A total of 10 nmol of each RNA fragment was brought to 100 μL in 0.1 M sodium bicarbonate, pH 9.0, and mixed with an equal volume of a Cy3 or Cy5 Amersham mono-reactive dye pack in DMSO (GE Healthcare). The labeling mix was incubated at 37°C in the dark for 2 h, then ethanol precipitated. Pellets were

resuspended in 60 µL buffer A (0.1 M triethylammonium acetate [TEAA], pH 7.5), and HPLC purified on a reversed phase C8 column (Agilent Technologies).

Ligation of synthetic RNA fragments

To generate a CR4/5 RNA (hTR 239–328) with fluorescent dyes at positions U274 and U312, a splinted ligation reaction[70] containing 800 pmol of Cy3-labeled Fragment 2 (hTR 279-330), 1600 pmol of Cy5-labeled Fragment 1 (hTR 239–278), 1600 pmol of DNA splint (sequence: 5′-AGTGGGTGCCTCCGGAGAAGCCCCGGGCCGAC-3′) in 0.5× T4 DNA ligase buffer (NEB) was brought to 100 µL volume and incubated at 95°C for 5 min and at 30°C for 10 min. A total of 100 µL ligation mix (1.5× T4 DNA ligase buffer, 4000 U T4 DNA ligase [NEB], 2 mM ATP and 1 U/ µL RNAsin Plus [Promega]) was added to the reaction and incubated at 30°C for 18 h. A total of 10 U of TURBO DNase (Thermo Fisher Scientific) was added and the reaction incubated at 37°C for 15 min. The RNA was phenol–chloroform extracted and ethanol precipitated prior to PAGE purification.

Single-molecule experiments

Slide preparation and imaging

Glass micro slides (Gold Seal) were washed by hand with Alconox detergent and warm water, then dried with nitrogen. Sample channels were constructed with Parafilm strips and a plasma-cleaned glass coverslip (Fisher Scientific). Channels

were blocked with 10 mg/mL BSA (NEB) for 1 h and washed with imaging buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 1 mM $MgCl_2$, 1 mg/ml BSA, 8% glucose, and [±]-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid [Trolox] at saturation). Trolox-containing imaging buffer was generally filtered (0.2 μm) before and after adjusting the pH to 8.3 with NaOH. For imaging, 0.01 volumes of "Gloxy" solution (10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 200 μg/mL catalase, 100 mg/mL glucose oxidase) were added to the imaging buffer.

Confocal microscopy of doubly labeled CR4/5 RNA and human telomerase

Data was acquired with a confocal fluorescence microscope with 200-pM-labeled hTR CR4/5 and 50-fold diluted aliquots of in vitro reconstituted labeled human telomerase. A green laser (532 nm) set to 100 μW was used to excite the Cy3 donor dye within the slide channel, and fluorescence from a ~100 nm^3 volume was collected through a pinhole and passed on to a dichroic mirror to separate green and red wavelengths. Red and green light were individually detected by avalanche photodiode detectors (APDs) and written to a data file using custom LabView software. Data was collected for 30 min, usually capturing fluorescence from thousands of individual molecules.

Analysis of single-molecule data

Using custom MATLAB scripts, the data was thresholded to include only molecules with Cy5 fluorescence one standard deviation above the mean intensity detected by

the red (637 nm) APD, as well as corrected for direct Cy5 excitation by green light and dichroic mirror breakthrough. FRET efficiency was calculated in MATLAB with the equation

$$FRET=IA/(IA+ID),$$

where I_A and I_D are acceptor and donor intensity, respectively. Histograms were generated using GraphPad Prism. Gaussian approximation of FRET populations was performed by fitting each histogram with a nonlinear regression model, in which the mean of each Gaussian function was constrained to values determined by visual approximation.

References

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

2. Erdel, F. *et al.* Telomere Recognition and Assembly Mechanism of Mammalian Shelterin. *Cell Rep* **18**, 41–53 (2017).

3. J, M. H. The remaking of chromosomes. *Collecting Net* **8**, 198 (1938).

4. McClintock, B. The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis. *Proc Natl Acad Sci U S A* **25**, 405–416 (1939).

5. de Lange, T. Shelterin-Mediated Telomere Protection. *Annu Rev Genet* **52**, 223–247 (2018).

6. Hayflick, L. THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. *Exp Cell Res* **37**, 614–636 (1965).

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

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

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

10. Kolquist, K. A. *et al.* Expression of TERT in early premalignant lesions and a subset of cells in normal tissues. *Nat Genet* **19**, 182–186 (1998).

11. Wright, D. L. *et al.* Characterization of telomerase activity in the human oocyte and preimplantation embryo. *Mol Hum Reprod* **7**, 947–955 (2001).

12. Roth, A. *et al.* Telomerase levels control the lifespan of human T lymphocytes. *Blood* **102**, 849–857 (2003).

13. Yamaguchi, H. *et al.* Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. *Blood* **102**, 916–918 (2003).

14. Vulliamy, T. J. & Dokal, I. Dyskeratosis congenita: the diverse clinical presentation of mutations in the telomerase complex. *Biochimie* **90**, 122–130 (2008).

15. Savage, S. A. Human telomeres and telomere biology disorders. *Prog Mol Biol Transl Sci* **125**, 41–66 (2014).

16. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).

17. Blasco, M. A. Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet* **6**, 611–622 (2005).

18. Egan, E. D. & Collins, K. Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747–1759 (2012).

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

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

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

22. Chen, J.-L. & Greider, C. W. An emerging consensus for telomerase RNA structure. *Proc Natl Acad Sci U S A* **101**, 14683–14684 (2004).

23. Alder, J. K. *et al.* Diagnostic utility of telomere length testing in a hospital-based setting. *Proc Natl Acad Sci U S A* **115**, E2358–E2365 (2018).

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

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

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

27. Antal, M., Boros, É., Solymosy, F. & Kiss, T. Analysis of the structure of human telomerase RNA in vivo. *Nucleic Acids Res* **30**, 912–920 (2002).

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

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

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

31. Chen, L. *et al.* An activity switch in human telomerase based on RNA conformation and shaped by TCAB1. *Cell* **174**, 218-230.e13 (2018).

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

33. Huang, J. *et al.* Structural basis for protein-RNA recognition in telomerase. *Nat Struct Mol Biol* **21**, 507–512 (2014).

34. Jiang, J. *et al.* Structure of telomerase with telomeric DNA. *Cell* **173**, 1179-1190.e13 (2018).

35. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).

36. Kladwang, W., VanLang, C. C., Cordero, P. & Das, R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat Chem* **3**, 954–962 (2011).

37. Tian, S., Cordero, P., Kladwang, W. & Das, R. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **20**, 1815–1826 (2014).

38. Etheridge, K. T. *et al.* The nucleolar localization domain of the catalytic subunit of human telomerase. *J Biol Chem* **277**, 24764–24770 (2002).

39. Zhu, Y., Tomlinson, R. L., Lukowiak, A. A., Terns, R. M. & Terns, M. P. Telomerase RNA Accumulates in Cajal Bodies in Human Cancer Cells. *Mol Biol Cell* **15**, 81–90 (2004).

40. Tesmer, V. M. *et al.* Two Inactive Fragments of the Integral RNA Cooperate To Assemble Active Telomerase with the Human Protein Catalytic Subunit (hTERT) In Vitro. *Mol Cell Biol* **19**, 6207–6216 (1999).

41. Kladwang, W. *et al.* Standardization of RNA Chemical Mapping Experiments. *Biochemistry* **53**, 3063–3065 (2014).

42. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).

43. Mortimer, S. A., Trapnell, C., Aviran, S., Pachter, L. & Lucks, J. B. SHAPE-Seq: High-Throughput RNA Structure Analysis. *Curr Protoc Chem Biol* **4**, 275–297 (2012).

44. Turner, R., Shefer, K. & Ares, M. Safer one-pot synthesis of the 'SHAPE' reagent 1-methyl-7-nitroisatoic anhydride (1m7). *RNA* **19**, 1857–1863 (2013).

45. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101**, 7287–7292 (2004).

46. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A* **106**, 97–102 (2009).

47. Zappulla, D. C. & Cech, T. R. RNA as a flexible scaffold for proteins: yeast telomerase and beyond. *Cold Spring Harb Symp Quant Biol* **71**, 217–224 (2006).

48. Egan, E. D. & Collins, K. Specificity and Stoichiometry of Subunit Interactions in the Human Telomerase Holoenzyme Assembled In Vivo. *Mol Cell Biol* **30**, 2775–2786 (2010).

49. O'Connor, C. M. & Collins, K. A Novel RNA Binding Domain in Tetrahymena Telomerase p65 Initiates Hierarchical Assembly of Telomerase Holoenzyme. *Mol Cell Biol* **26**, 2029–2036 (2006).

50. Stone, M. D. *et al.* Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).

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

52. Singh, M. *et al.* Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol Cell* **47**, 16–26 (2012).

53. Mihalusova, M., Wu, J. Y. & Zhuang, X. Functional importance of telomerase pseudoknot revealed by single-molecule analysis. *Proc Natl Acad Sci U S A* **108**, 20339–20344 (2011).

54. Egan, E. D. & Collins, K. An Enhanced H/ACA RNP Assembly Mechanism for Human Telomerase RNA. *Mol Cell Biol* **32**, 2428–2439 (2012).

55. Vulliamy, T. J. *et al.* Differences in Disease Severity but Similar Telomere Lengths in Genetic Subgroups of Patients with Telomerase and Shelterin Mutations. *PLoS One* **6**, e24383 (2011).

56. Laprade, H. *et al.* Single-Molecule Imaging of Telomerase RNA Reveals a Recruitment-Retention Model for Telomere Elongation. *Mol Cell* **79**, 115-126.e6 (2020).

57. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**, 75–82 (2017).

58. Tian, S., Yesselman, J. D., Cordero, P. & Das, R. Primerize: automated primer assembly for transcribing non-coding RNA domains. *Nucleic Acids Res* **43**, W522–W526 (2015).

59. Rio, D. C. Expression and purification of active recombinant T7 RNA polymerase from E. coli. *Cold Spring Harb Protoc* **2013**, pdb.prot078527 (2013).

60. Kladwang, W. & Das, R. A mutate-and-map strategy for inferring base pairs in structured nucleic acids: proof of concept on a DNA/RNA helix. *Biochemistry* **49**, 7414–7416 (2010).

61. Cordero, P., Kladwang, W., VanLang, C. C. & Das, R. The Mutate-and-Map Protocol for Inferring Base Pairs in Structured RNA. *Methods Mol Biol* **1086**, 53–77 (2014).

62. Yoon, S. *et al.* HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics* **27**, 1798–1805 (2011).

63. Cordero, P., Lucks, J. B. & Das, R. An RNA Mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics* **28**, 3006–3008 (2012).

64. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).

65. Weinrich, S. L. *et al.* Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. *Nat Genet* **17**, 498–502 (1997).

66. Jansson, L. I. *et al.* Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proc Natl Acad Sci U S A* **116**, 9350–9359 (2019).

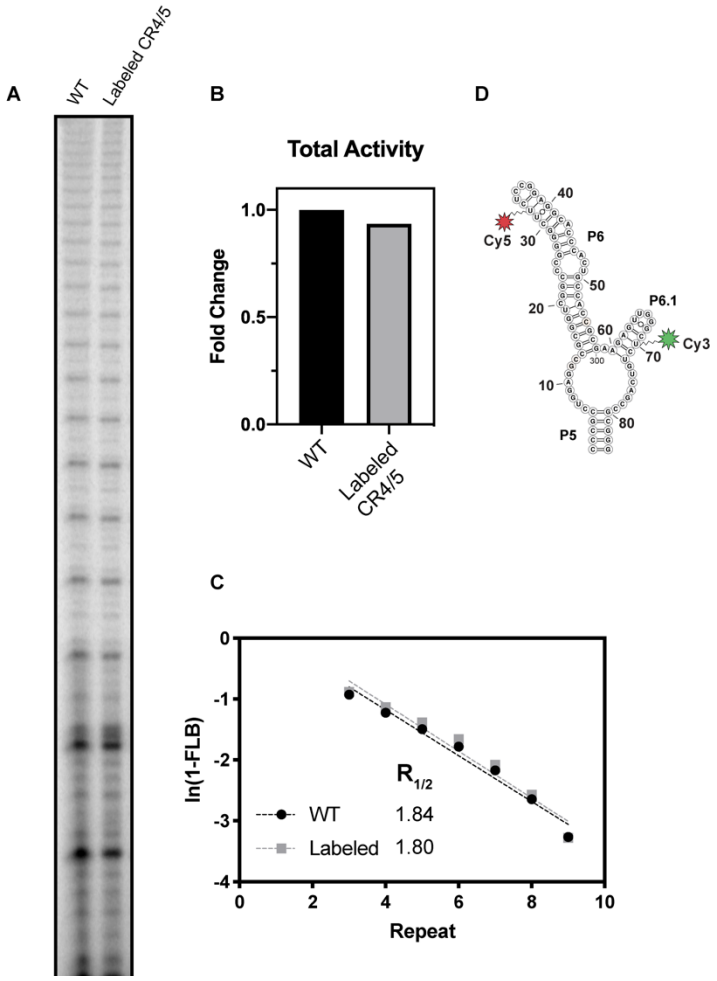
67. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D. & Altman, R. B. SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11**, 344–354 (2005).

68. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of Image Analysis. *Nat Methods* **9**, 671–675 (2012).

69. Latrick, C. M. & Cech, T. R. POT1–TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *EMBO J* **29**, 924–933 (2010).

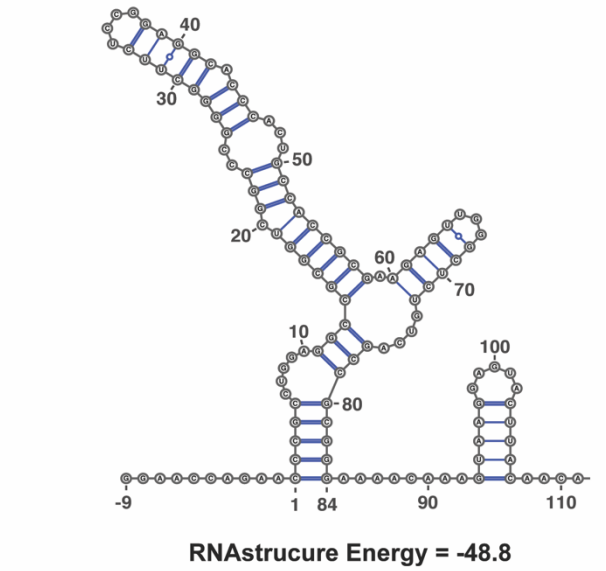
70. Akiyama, B. M. & Stone, M. D. Assembly of complex RNAs by splinted ligation. *Methods Enzymol* **469**, 27–46 (2009).

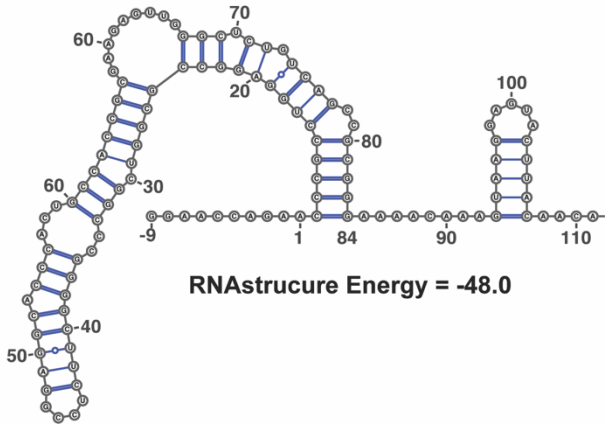
Supplementary material



Supplementary Figure S1. Primer extension assay of telomerase reconstituted with WT CR4/5 and dye-labeled CR4/5.

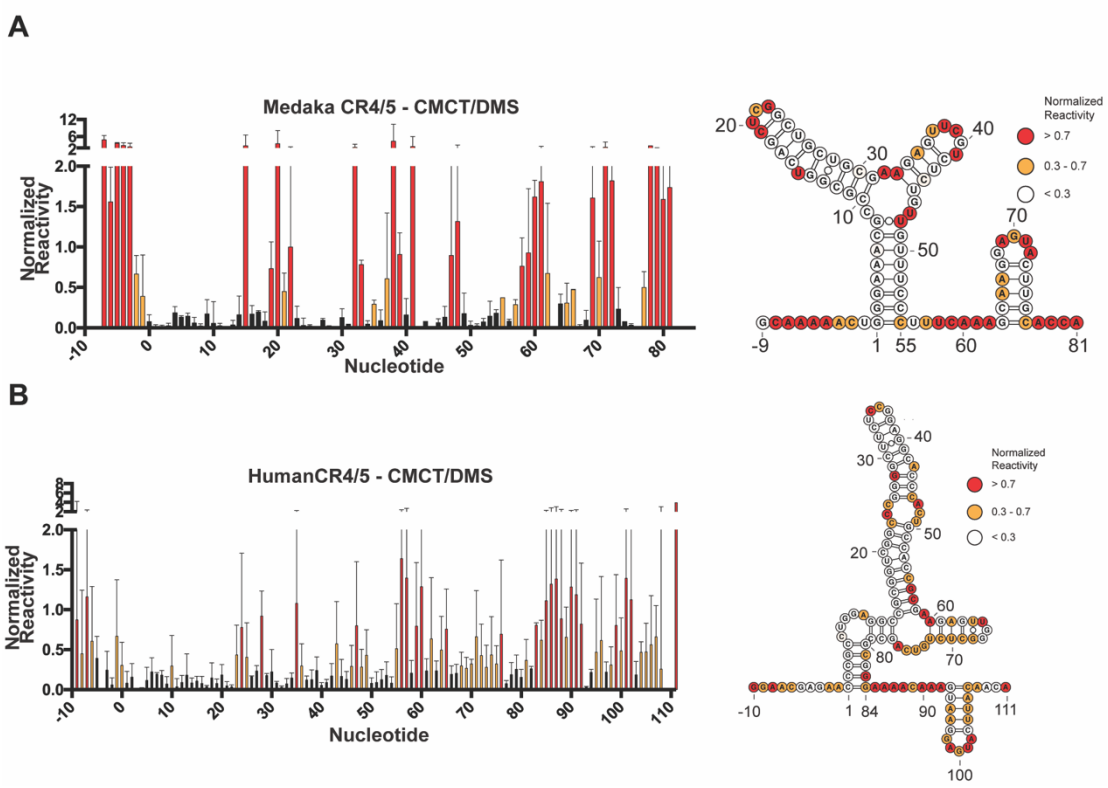
(A) Gel illustrating activity and processivity of reconstituted telomerase with WT CR4/5 and dye-labeled CR4/5 (B) Quantification of total activity. (C) Quantification of repeat addition processivity for telomerase enzymes reconstituted with unlabeled and dye-labeled CR4/5 domain, with indicated $R_{1/2}$ values extracted from the fits to plot of main repeat versus ln(1-FLB), where FLB is Fraction Left Behind (see Materials and Methods for reconstitution and data analysis details). (D) CR4/5 dye labeled positions.





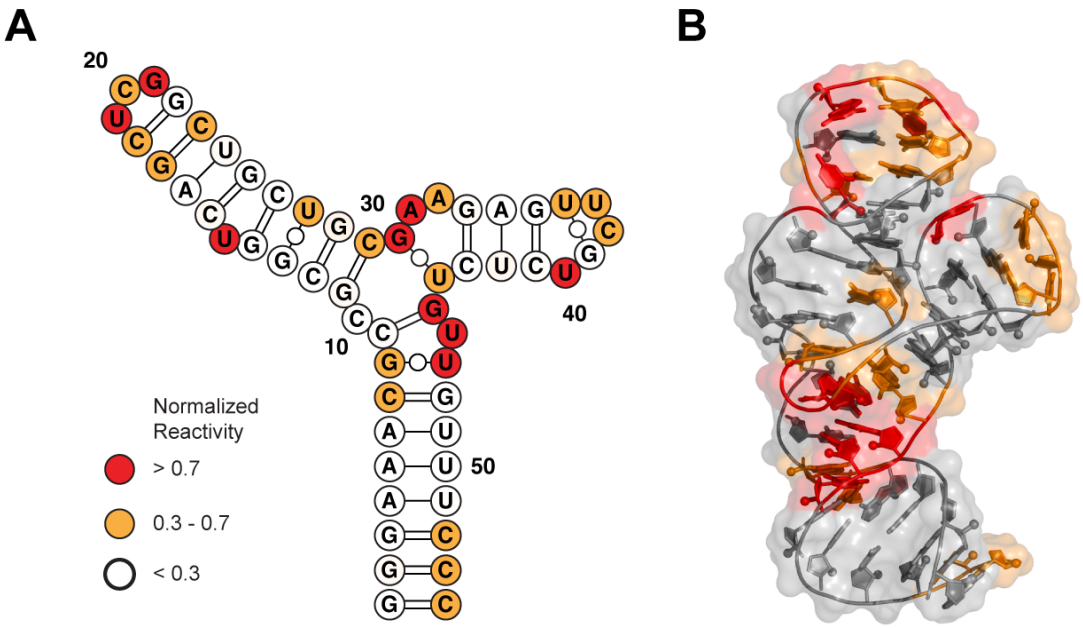
Supplementary Figure S2. Comparison of structures of the human CR4/5 domain predicted using the RNAstructure software package.

Similar energy values are observed for different structures predicted for the human CR4/5 domain.



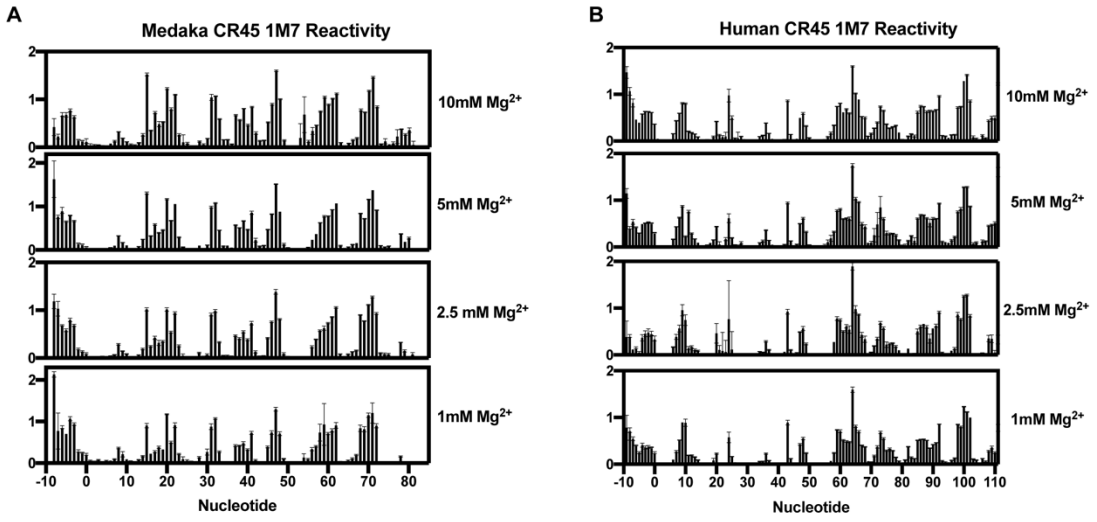
Supplementary Figure S3. Chemical mapping of medaka and human CR4/5 domains using merged CMCT/DMS data.

Chemical mapping of the medaka (**A**) and human (**B**) CR4/5 domain using merged CMCT/DMS data. Data were collected on TR molecules folded in the presence of 10mM $MgCl_2$. Plotted normalized reactivity values are color-coded (red > 0.7, yellow 0.3 - 0.7, and white < 0.3). Bar plot represents experiments conducted in triplicate and error bars are the standard deviation of the three replicates. Color-coded schematic of the reactivity data is shown on the RNAstructure predicted secondary structure.



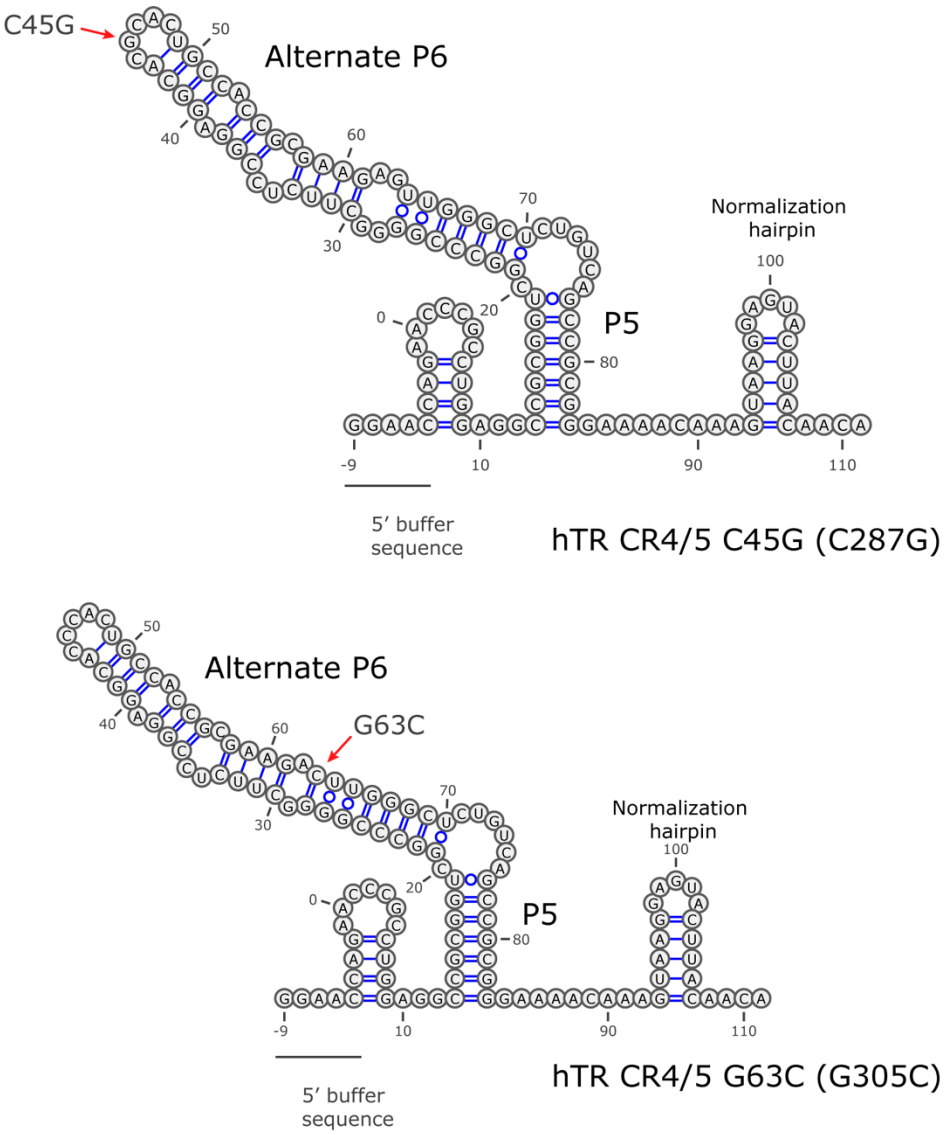
Supplementary Figure S4. Chemical mapping of medaka CR4/5 using 1M7.

Chemical mapping data for the medaka CR4/5 domain collected in the presence of 1mM $MgCl_2$ (**A**) mapped onto NMR solution secondary structure and (**B)** tertiary structure (PyMOL) (PDB: 2MHI).



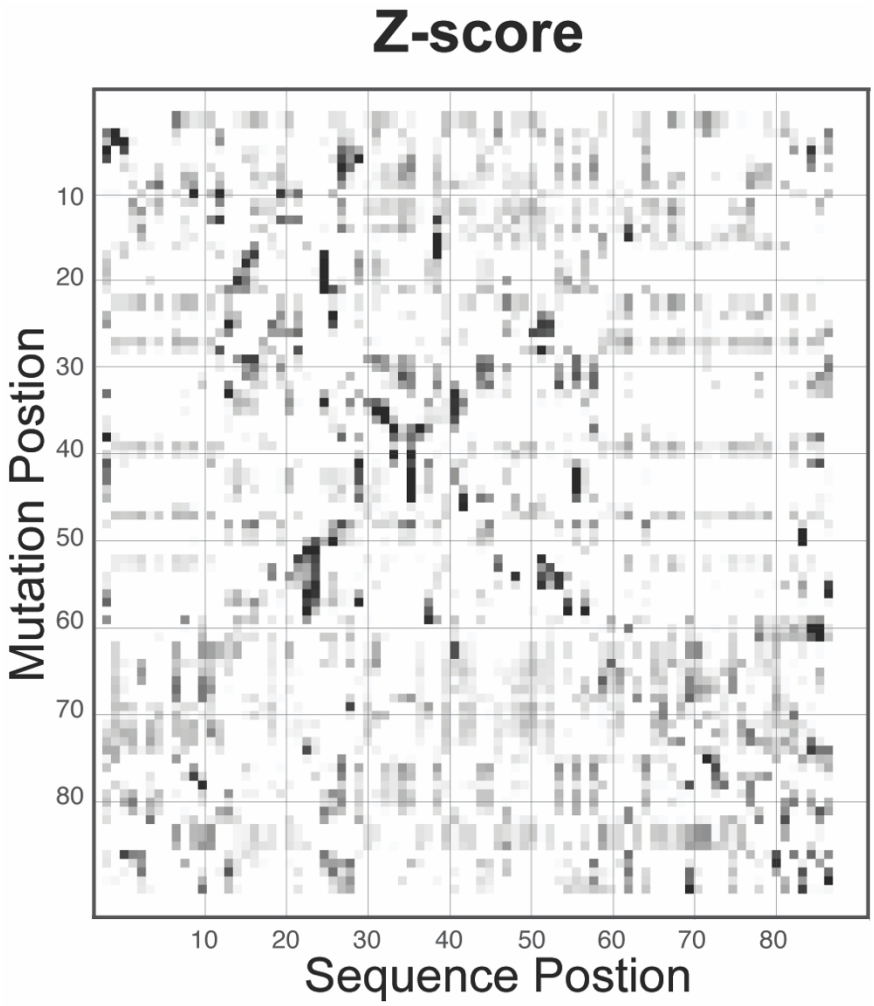
Supplementary Figure S5. $MgCl_2$ titration of human and medaka CR4/5 using 1M7.

SHAPE chemical mapping data for Medaka (**A**) and human (**B**) is plotted as a function of the indicated $MgCl_2$ concentrations. Bar graphs from three replicates are shown plotted with standard deviation.



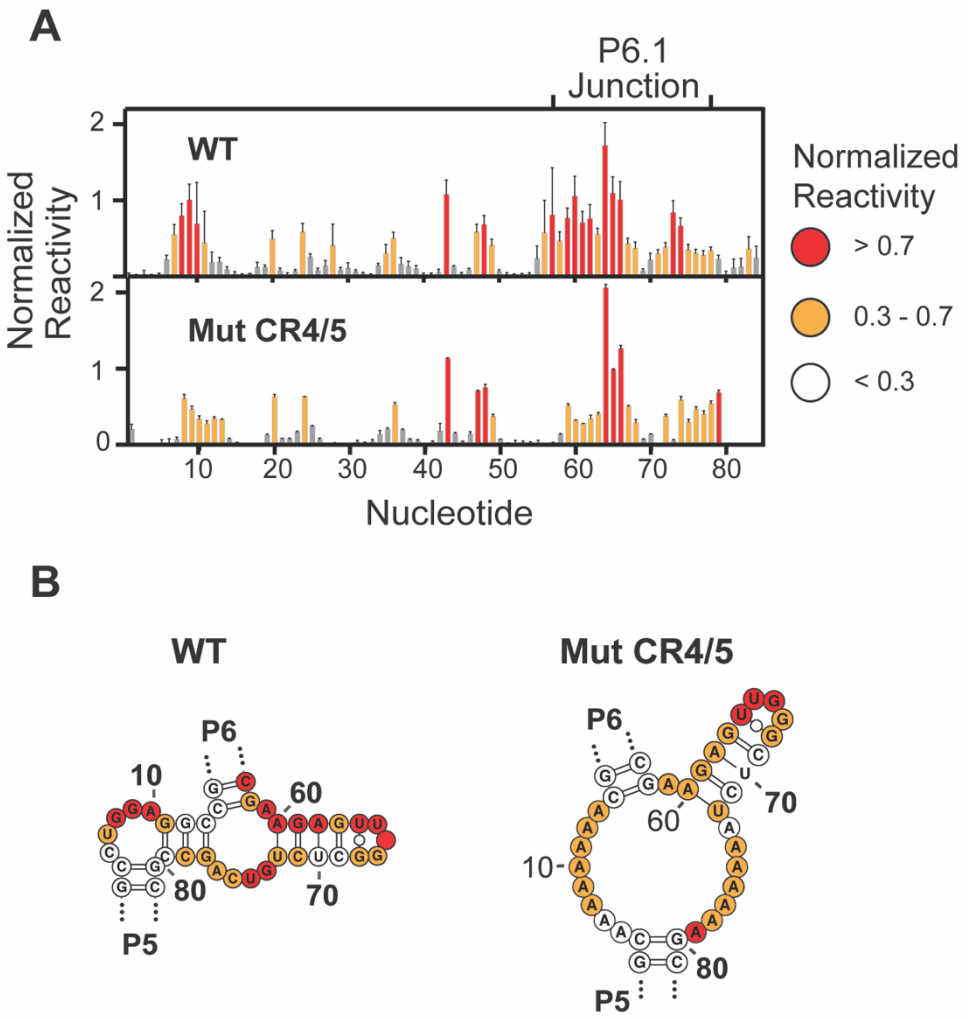
Supplementary Figure S6. Data-driven structure prediction of select CR4/5 mutants G61C and G63.

The G61C mutation induced an alternate predicted conformation of the CR4/5 domain that resulted in the 5' buffer annealing to the 3WJ junction region and part of P5 forming a base pairing interaction with the bases normally involved in the P6.1 stem-loop. The G63C mutant caused a dramatic rearrangement in the predicted structure resulting in a totally different structure, with the P6, P5 and 3WJ region all taking on new base-pairing partners. In both constructs the normalization hairpin formed correctly and the mutation made is indicated with a red arrow.



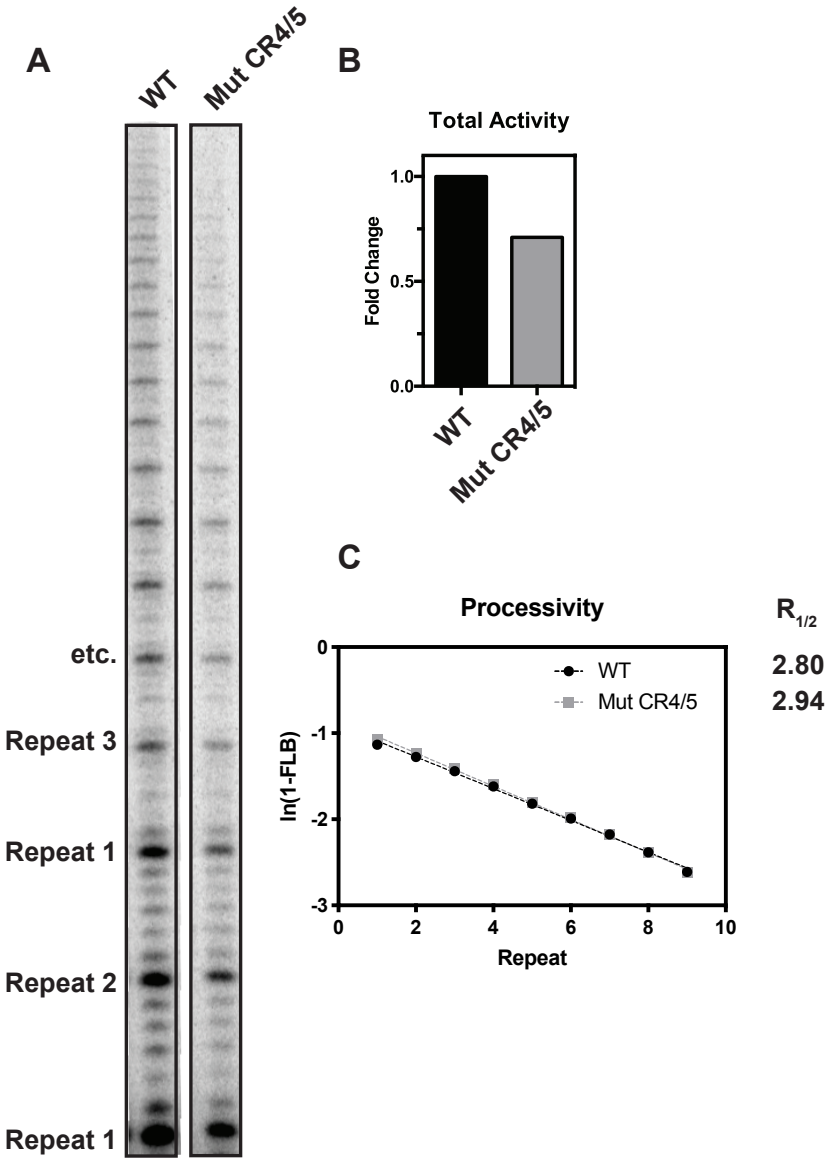
Supplementary Figure S7. Z-score plot from mutate-and-map experiments on the hTR CR4/5 domain.

A Z-score plot indicating patterns of mutation-induced changes in chemical reactivity at each nucleotide position across all RNA mutants. The Z-scores are used as weights to guide structure prediction in the RNAstructure software (see Materials and Methods for details).



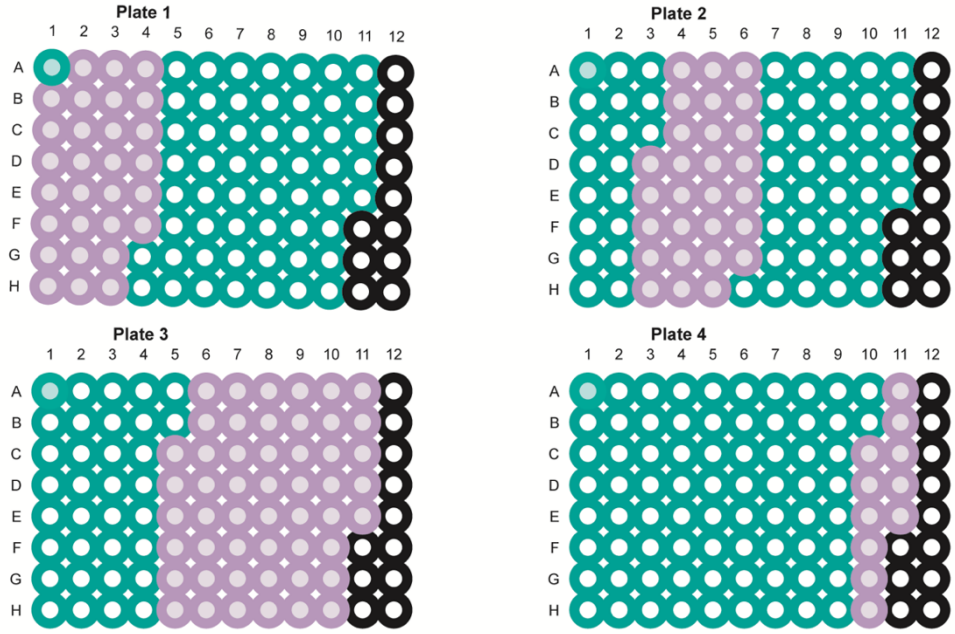
Supplementary Figure S8. SHAPE reactivity of wild-type and mutant hTR CR4/5 domains.

(A) 1M7 reactivity data for WT (top panel) and Mut CR4/5 (bottom panel) is shown for data collected in the presence of 10mM $MgCl_2$. Bar graphs are data averages from triplicate experiments with standard deviation plotted as error bars. (B) SHAPE-guided secondary structure predictions of the hTR WT and Mut CR4/5 domains with reactivity data mapped onto the structural model (Note: only the junction region is shown although the entire hTR CR4/5 construct was included in the modeling).



Supplementary Figure S9. Primer extension assay of WT and Mut CR4/5.

(A) Gel illustrating activity and processivity of reconstituted telomerase with WT CR4/5 and Mut CR4/5 (B) Quantification of total activity. (C) Quantification of repeat addition processivity for telomerase enzymes reconstituted with WT and Mut CR4/5 domain, with indicated R1/2 values extracted from the fits to plot of main repeat versus ln(1-FLB), where FLB is Fraction Left Behind (see Materials and Methods for reconstitution and data analysis details).



Supplementary Figure S10. Plate layout for assembling the 84 Mutate-and-Map constructs.

Plate 1, 2, 3, and 4 are all illustrated. For each RNA four primers were used (one from each plate) for the in the primer assembly reaction. By matching the identities of the wells (A1, B1, etc) on each plate, a full-length mutated construct could be assembled (ie - using A1 primers from all plates will make WT CR4/5, using B1 primers from all plates will make mutant C1G, using C1 primers from all will make mutant C2G). Green corresponds to a WT primer for that plate and purple corresponds to a mutant. Only mutant primers (purple) are listed in the M2 oligo table and the primer listed as A1 on each plate lists the WT primer.

Supplementary Table 1 - Oligonucleotides used in study (except mutate-and-map)

Construct	Sequence
Human CR4/5	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTG
Human CR4/5	GGGTGCCTCCGGAGAAGCCCCGGGCCGACCGCGGCCTCCAGGCGGGTTCTCGTT
Human CR4/5	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGCTCTGTCAGCCGCGGGA
Human CR4/5	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTACTTTGTTTTCCCGCGGCTGA
Medaka CR4/5	TTCTAATACGACTCACTATAGCAAAAACTGGGAAACGCCGCGGTCA
Medaka CR4/5	AGGGAAACAACAGAGACGAACTCTTCGCAGCAGCCGAGCTGACCGCGGCGTT
Medaka CR4/5	TCGTCTCTGTTGTTTCCCTTTCAAAGCAAGGAGTACTTGCACCAGACACAACACAAT
Medaka CR4/5	GCTGTAAATTGTGTTGTGTCTGGTGCAAGT
Mut-1 (lock 7)	TTCTAATACGACTCACTATAGGAACGAGAACCCGCAAAAAAAACGCGGTCGGC
Mut-1 (lock 7)	GCGGTGGCAGTGGGTGCCTCCGGAGAAGCCCCGGGCCGACCGCGTT
Mut-1 (lock 7)	CCACTGCCACCGCGAAGAGTTGGGCTCTAAAAAAAGCGGGAAAACAAAGTAAGG
Mut-1 (lock 7)	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTACTTTGTTTTCCCGCTTTTTTTA
RT FAM primer (medaka)	/56-FAM/AAAAAAAAAAAAAAAAAAAAGCTGTAAATTGTGTTGTGTC
RT FAM primer (human)	/56-FAM/AAAAAAAAAAAAAAAAAAAAGTTGTTGTTGTTGTTTCTTT
Telo primer for primer extension	TTAGGGTTAGGGTTAGGG
Dot Blot Hybridization Oligo	CGG TGG AAG GCG GCA GGC CGA GGC
hTR CR4/5 Fragment 1	GAACCCCGCCUGGAGGCCGCGGUCGGCCCGGGGCU(5-LC-N-U)CUCC3
hTR CR4/5 Fragment 2	GGAGGCACCCACUGCCACCGCGAAGAGUUGGGC(5-LC-N-U)CUGUCAGCCGCGGGUCUC3
DNA splint for CR4/5 construction	AGTGGGTGCCTCCGGAGAAGCCCCGGGCCGAC3

Supplementary Table 2 - DNA primers used to make Mutate-and-Map templates

PLATE 1		
WellPosition	**Name**	**Sequence**
A01	Lib1-WT	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGAGGCCGCGGTCGGCCCGGGG
B01	Lib1-C31G	TTCTAATACGACTCACTATAGGAACGAGAAGCCGCC TGGAGGCCGCGGTCGGCCCGGGG
C01	Lib1-C32G	TTCTAATACGACTCACTATAGGAACGAGAACGCGCC TGGAGGCCGCGGTCGGCCCGGGG
D01	Lib1-C33G	TTCTAATACGACTCACTATAGGAACGAGAACCGGCC TGGAGGCCGCGGTCGGCCCGGGG
E01	Lib1-G34C	TTCTAATACGACTCACTATAGGAACGAGAACCCCCCT GGAGGCCGCGGTCGGCCCGGGG
F01	Lib1-C35G	TTCTAATACGACTCACTATAGGAACGAGAACCCGGC TGGAGGCCGCGGTCGGCCCGGGG
G01	Lib1-C36G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCG TGGAGGCCGCGGTCGGCCCGGGG
H01	Lib1-T37A	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC AGGAGGCCGCGGTCGGCCCGGGG
A02	Lib1-G38C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TCGAGGCCGCGGTCGGCCCGGGG
B02	Lib1-G39C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGCAGGCCGCGGTCGGCCCGGGG
C02	Lib1-A40T	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGTGGCCGCGGTCGGCCCGGGG
D02	Lib1-G41C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGACGCCGCGGTCGGCCCGGGG
E02	Lib1-G42C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGAGCCCGCGGTCGGCCCGGGG
F02	Lib1-C43G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGAGGGCGCGGTCGGCCCGGGG
G02	Lib1-C44G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGAGGCGGCGGTCGGCCCGGGG
H02	Lib1-G45C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGAGGCCCCGGTCGGCCCGGGG
A03	Lib1-C46G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGAGGCCGGGGTCGGCCCGGGG
B03	Lib1-G47C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCC TGGAGGCCGCCGTCGGCCCGGGG

C03	Lib1-G48C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGCTCGGCCCGGGG
D03	Lib1-T49A	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGACGGCCCGGGG
E03	Lib1-C50G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTGGGCCCGGGG
F03	Lib1-G51C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCCGCCCGGGG
G03	Lib1-G52C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGCCCCGGGG
H03	Lib1-C53G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGGGCCGGGG
A04	Lib1-C54G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGGCGCGGGG
B04	Lib1-C55G	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGGCCGGGGG
C04	Lib1-G56C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGGCCCCGGG
D04	Lib1-G57C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGGCCCGCGG
E04	Lib1-G58C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGGCCCGGCG
F04	Lib1-G59C	TTCTAATACGACTCACTATAGGAACGAGAACCCGCCTGGAGGCCGCGGTCGGCCCGGGC
PLATE 2		
WellPosition	**Name**	**Sequence**
A01	Lib1-WT	GGGTGCCTCCGGAGAAGCCCCGGGCCGA
D03	Lib1-T49A	GGGTGCCTCCGGAGAAGCCCCGGGCCGT
E03	Lib1-C50G	GGGTGCCTCCGGAGAAGCCCCGGGCCCA
F03	Lib1-G51C	GGGTGCCTCCGGAGAAGCCCCGGGCGGA
G03	Lib1-G52C	GGGTGCCTCCGGAGAAGCCCCGGGGCGA
H03	Lib1-C53G	GGGTGCCTCCGGAGAAGCCCCGGCCCGA
A04	Lib1-C54G	GGGTGCCTCCGGAGAAGCCCCGCGCCGA
B04	Lib1-C55G	GGGTGCCTCCGGAGAAGCCCCCGGCCGA

C04	Lib1-G56C	GGGTGCCTCCGGAGAAGCCCGGGGCCGA
D04	Lib1-G57C	GGGTGCCTCCGGAGAAGCCGCGGGCCGA
E04	Lib1-G58C	GGGTGCCTCCGGAGAAGCGCCGGGCCGA
F04	Lib1-G59C	GGGTGCCTCCGGAGAAGGCCCGGGCCGA
G04	Lib1-C60G	GGGTGCCTCCGGAGAACCCCCGGGCCGA
H04	Lib1-T61A	GGGTGCCTCCGGAGATGCCCCGGGCCGA
A05	Lib1-T62A	GGGTGCCTCCGGAGTAGCCCCGGGCCGA
B05	Lib1-C63G	GGGTGCCTCCGGACAAGCCCCGGGCCGA
C05	Lib1-T64A	GGGTGCCTCCGGTGAAGCCCCGGGCCGA
D05	Lib1-C65G	GGGTGCCTCCGCAGAAGCCCCGGGCCGA
E05	Lib1-C66G	GGGTGCCTCCCGAGAAGCCCCGGGCCGA
F05	Lib1-G67C	GGGTGCCTCGGGAGAAGCCCCGGGCCGA
G05	Lib1-G68C	GGGTGCCTGCGGAGAAGCCCCGGGCCGA
H05	Lib1-A69T	GGGTGCCACCGGAGAAGCCCCGGGCCGA
A06	Lib1-G70C	GGGTGCGTCCGGAGAAGCCCCGGGCCGA
B06	Lib1-G71C	GGGTGGCTCCGGAGAAGCCCCGGGCCGA
C06	Lib1-C72G	GGGTCCCTCCGGAGAAGCCCCGGGCCGA
D06	Lib1-A73T	GGGAGCCTCCGGAGAAGCCCCGGGCCGA
E06	Lib1-C74G	GGCTGCCTCCGGAGAAGCCCCGGGCCGA
F06	Lib1-C75G	GCGTGCCTCCGGAGAAGCCCCGGGCCGA
G06	Lib1-C76G	CGGTGCCTCCGGAGAAGCCCCGGGCCGA
PLATE 3		

WellPosition	**Name**	**Sequence**
A01	Lib1-WT	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
C05	Lib1-T64A	ACCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
D05	Lib1-C65G	TGCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
E05	Lib1-C66G	TCGGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
F05	Lib1-G67C	TCCCGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
G05	Lib1-G68C	TCCGCAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
H05	Lib1-A69T	TCCGGTGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
A06	Lib1-G70C	TCCGGACGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
B06	Lib1-G71C	TCCGGAGCCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
C06	Lib1-C72G	TCCGGAGGGACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
D06	Lib1-A73T	TCCGGAGGCTCCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
E06	Lib1-C74G	TCCGGAGGCAGCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
F06	Lib1-C75G	TCCGGAGGCACGCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
G06	Lib1-C76G	TCCGGAGGCACCGACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
H06	Lib1-A77T	TCCGGAGGCACCCTCTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
A07	Lib1-C78G	TCCGGAGGCACCCAGTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
B07	Lib1-T79A	TCCGGAGGCACCCACAGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
C07	Lib1-G80C	TCCGGAGGCACCCACTCCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
D07	Lib1-C81G	TCCGGAGGCACCCACTGGCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
E07	Lib1-C82G	TCCGGAGGCACCCACTGCGACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA

F07	Lib1-A83T	TCCGGAGGCACCCACTGCCTCCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
G07	Lib1-C84G	TCCGGAGGCACCCACTGCCAGCGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
H07	Lib1-C85G	TCCGGAGGCACCCACTGCCACGGCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
A08	Lib1-G86C	TCCGGAGGCACCCACTGCCACCCCGAAGAGTTGGGC TCTGTCAGCCGCGGGA
B08	Lib1-C87G	TCCGGAGGCACCCACTGCCACCGGGAAGAGTTGGGC TCTGTCAGCCGCGGGA
C08	Lib1-G88C	TCCGGAGGCACCCACTGCCACCGCCAAGAGTTGGGC TCTGTCAGCCGCGGGA
D08	Lib1-A89T	TCCGGAGGCACCCACTGCCACCGCGTAGAGTTGGGC TCTGTCAGCCGCGGGA
E08	Lib1-A90T	TCCGGAGGCACCCACTGCCACCGCGATGAGTTGGGC TCTGTCAGCCGCGGGA
F08	Lib1-G91C	TCCGGAGGCACCCACTGCCACCGCGAACAGTTGGGC TCTGTCAGCCGCGGGA
G08	Lib1-A92T	TCCGGAGGCACCCACTGCCACCGCGAAGTGTTGGGC TCTGTCAGCCGCGGGA
H08	Lib1-G93C	TCCGGAGGCACCCACTGCCACCGCGAAGACTTGGGC TCTGTCAGCCGCGGGA
A09	Lib1-T94A	TCCGGAGGCACCCACTGCCACCGCGAAGAGATGGGC TCTGTCAGCCGCGGGA
B09	Lib1-T95A	TCCGGAGGCACCCACTGCCACCGCGAAGAGTAGGGC TCTGTCAGCCGCGGGA
C09	Lib1-G96C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTCGGC TCTGTCAGCCGCGGGA
D09	Lib1-G97C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGCGC TCTGTCAGCCGCGGGA
E09	Lib1-G98C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGCC TCTGTCAGCCGCGGGA
F09	Lib1-C99G	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGG TCTGTCAGCCGCGGGA
G09	Lib1-T100A	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC ACTGTCAGCCGCGGGA
H09	Lib1-C101G	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TGTGTCAGCCGCGGGA
A10	Lib1-T102A	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCAGTCAGCCGCGGGA
B10	Lib1-G103C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTCTCAGCCGCGGGA

C10	Lib1-T104A	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGACAGCCGCGGGA
D10	Lib1-C105G	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTGAGCCGCGGGA
E10	Lib1-A106T	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCTGCCGCGGGA
F10	Lib1-G107C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCACCCGCGGGA
G10	Lib1-C108G	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGGCGCGGGA
H10	Lib1-C109G	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCGGCGGGA
A11	Lib1-G110C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCCCGGGA
B11	Lib1-C111G	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGGGGGA
C11	Lib1-G112C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCCGGA
D11	Lib1-G113C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGCGA
E11	Lib1-G114C	TCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGC TCTGTCAGCCGCGGCA
PLATE 4		
WellPosition	**Name**	**Sequence**
A01	Lib1-WT	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGCGGCTGA
C10	Lib1-T104A	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGCGGCTGT
D10	Lib1-C105G	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGCGGCTCA
E10	Lib1-A106T	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGCGGCAGA
F10	Lib1-G107C	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGCGGGTGA
G10	Lib1-C108G	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGCGCCTGA
H10	Lib1-C109G	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGCCGCTGA
A11	Lib1-G110C	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCGGGGCTGA
B11	Lib1-C111G	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTA CTTTGTTTTCCCCCGGCTGA

C11	Lib1-G112C	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTACTTTGTTTTCCGGCGGCTGA
D11	Lib1-G113C	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTACTTTGTTTTCGCGCGGCTGA
E11	Lib1-G114C	GTTGTTGTTGTTGTTTCTTTTGTTGTAAGTACTCCTTACTTTGTTTTGCCGCGGCTGA

CHAPTER III

Dissecting telomerase RNA structural heterogeneity in living human cells with DMS MaP-seq

(currently unpublished)

Abstract

Telomerase is a specialized reverse transcriptase that uses an intrinsic RNA subunit as the template for telomeric DNA synthesis. Biogenesis of human telomerase requires its RNA subunit (hTR) to fold into a multi-domain architecture that includes the template-containing pseudoknot (t/PK) and a three-way junction (CR4/5) folds. These two hTR domains bind the telomerase reverse transcriptase (hTERT) protein and are thus essential for telomerase catalytic activity. Here, we probe the structure of hTR in living cells using dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq) and ensemble deconvolution analysis. Unexpectedly, approximately 15% of the steady state population of hTR has a CR4/5 conformation lacking features required for hTERT binding. Mutagenesis demonstrates that stabilization of the alternative CR4/5 conformation is detrimental to telomerase assembly and activity. We propose that this misfolded portion of the cellular hTR pool is either slowly refolded or degraded. Thus, kinetic traps for RNA folding that have been so well-studied in vitro may also present barriers for assembly of ribonucleoprotein complexes in vivo.

Introduction

Telomeres are repetitive regions of DNA found at chromosome ends that are sheathed with a coat of protective proteins[1,2] . Telomerase is a ribonucleoprotein (RNP) enzyme that catalyzes telomere DNA repeat synthesis using a short segment of its integral RNA subunit as a template[3]. The catalytic action of telomerase counteracts the gradual erosion of telomeric DNA arising from incomplete replication of the lagging strand by the DNA replication machinery[4]. In the absence of telomerase, telomeres eventually shorten to critical lengths that can trigger cellular senescence or programmed cell death[4,5]. Telomere length maintenance by telomerase is thus an essential process for ensuring the genomic stability and replicative capacity of rapidly dividing or continually regenerating tissues[6,7].

Underpinning the action of telomerase is the multistep co-assembly of hTR, hTERT, and several additional proteins into the functional RNP complex. Human telomerase is structurally organized by the hTR subunit, which adopts a phylogenetically conserved multi-domain architecture and serves as the binding scaffold for telomerase proteins[8–10]. Prior chemical probing studies of hTR were instrumental in providing direct biochemical evidence of the proposed phylogenetically conserved secondary structure[11,12]. However, these studies came to differing conclusions on whether hTERT binds a pre-organized hTR or induces hTR remodeling. Central to this discrepancy is that these studies were technically limited to gathering structural information on the population average of what has been shown

to be a heterogeneous pool of hTR reservoirs[13,14]. Overexpression of telomerase has also historically been used to mitigate problems arising from its scarcity but has been shown to bypass the endogenous RNP assembly pathway[15].

Recent cryo electron microscopy (cryo EM) discoveries have illuminated the structure and molecular composition of the *Tetrahymena* and human telomerase enzymes[16–21]. While both enzymes share important structural features in the catalytic domain, human telomerase has a uniquely bilobed architecture[22,23]. A catalytic lobe contains the hTERT protein, the conserved RNA t/PK and CR4/5 domains, and a hTR-bound histone dimer (**Fig.1 A-B**). The biogenesis lobe includes two complements of a core tetramer of H/ACA proteins bound to a small Cajal-body RNA domain (ScaRNA) near the 3' end of hTR[20,22,24]. Assembly of telomerase involves a multistep cascade of protein binding events and nuclear trafficking of the immature telomerase RNP before finally arriving at telomeres as an active complex[25,26]. After transcription, assembly of the H/ACA RNP licenses hTR for maturation and prevents its degradation by the RNA exosome[27,28]. hTR accumulates in the nucleoplasm and transits through the nucleolar and Cajal body subnuclear compartments. Association with hTERT excludes hTR from the nucleolus while association with the TCAB1 permits transient association with Cajal bodies[29–31]. While the precise spatio-temporal details of hTERT association with immature telomerase RNPs remains unclear, it is well-established that mutations in telomerase components that disrupt RNP biogenesis can cause a wide range of inherited diseases[32,33]. Moreover, mutagenesis studies on hTR have reinforced the importance

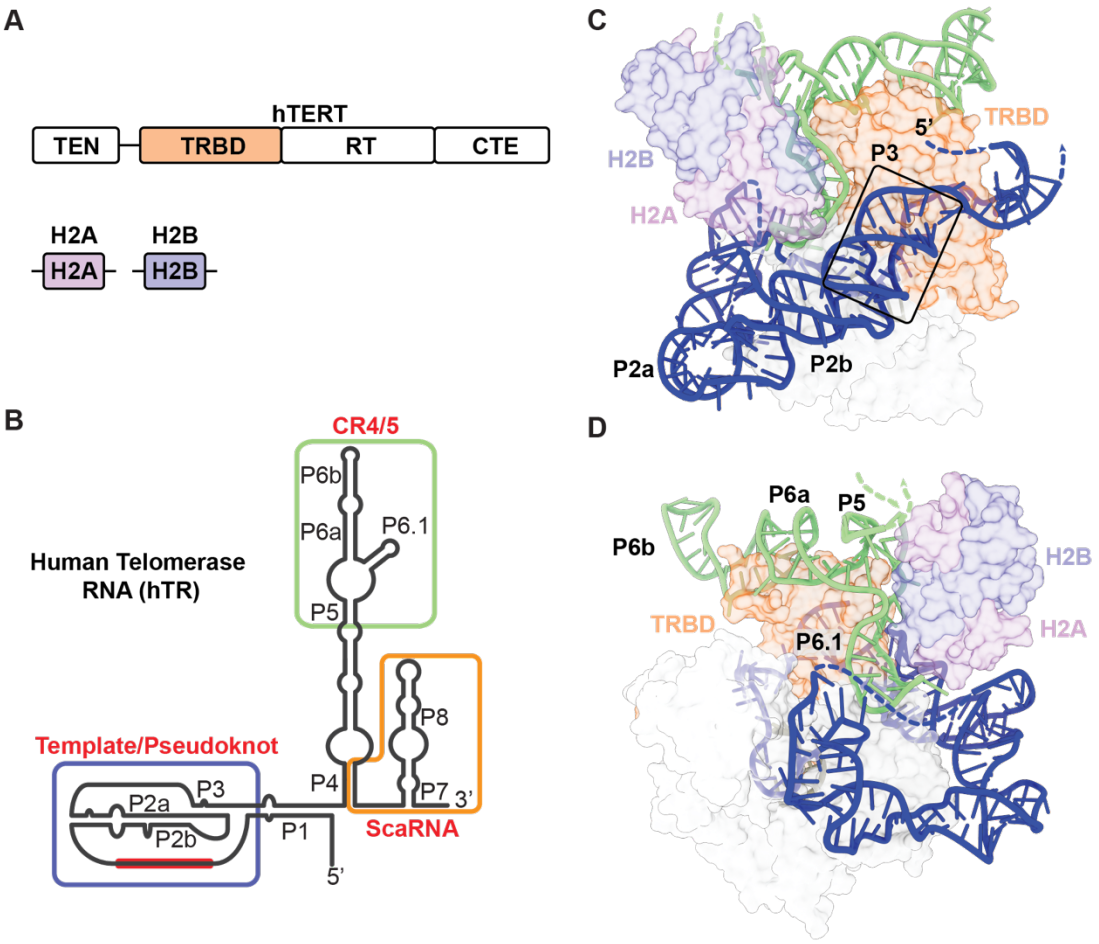


Figure 1: Overview of human telomerase components and hTR structure.

(A) The conserved domain organization of TERT includes the telomerase essential amino-terminal domain (TEN), RNA binding domain (TRBD), reverse transcriptase domain (RT), and carboxy-terminal extension (CTE). A H2A/B dimer comprises one H2A and one H2B protein. (B) The vertebrate-conserved architecture of hTR includes the template/pseudoknot (t/PK, template sequence shown as red bar), CR4/5, and ScaRNA domains. (C) The hTR template/pseudoknot domain (in blue) wraps around TERT and forms the pseudoknot helix P3. (D) CR4/5 (in green) adopts an 'L' shaped three way junction and sandwiches TERT TRBD between the P6.1 and P6a/b stems. A dimer of H2A/B binds P5 and P6.1.

of the t/PK and CR4/5 adopting precise conformations to achieve the assembly of a catalytically active telomerase RNP[34–36].

Here, we identify and characterize alternative conformations of endogenous hTR in living cells through a targeted DMS-MaPseq approach coupled with ensemble deconvolution[37,38]. Across the entire population of hTR molecules, DMS-reactive nucleotides are encoded as mutations in sequencing reads. Thus, each read represents a single-molecule measurement of hTR structure, permitting the use of bioinformatic clustering to identify groups of reads that are defined by distinct mutation patterns. We used DMS constraints derived from the clustered MaPseq results to guide RNA folding predictions, revealing that both the t/PK and CR4/5 domains exist in a structural ensemble that includes their canonically described conformations and an alternative conformation. Based on current cryo EM models of the assembled telomerase RNP these alternative t/PK and CR4/5 conformations are not representative of hTR in its hTERT-bound form[18–22,39]. CR4/5 mutants designed to destabilize the canonical three-way junction conformation while stabilizing the alternative conformation to varying degrees were misfolded in cells and exhibited catalytic defects. The decrease in activity is explained by impaired RNP assembly. Taken together, our results suggest that the in vivo hTR folding pathway is complex and includes alternative conformations that may represent dead-end misfolded states or possible intermediates along the telomerase RNP biogenesis pathway. In either case, our results show that assembly of catalytically active telomerase RNPs may be limited by a non-productive conformation of the CR4/5 domain.

Results

DMS-MaPseq captures endogenous hTR folding and reveals discrepancies with the established structure model

To test whether our MaPseq approach could capture structural features of endogenous hTR in living cells, we first generated and analyzed DMS reactivities of the ensemble average in HeLa cells. Our DMS reactivities of the t/PK domain (**Fig. 2A-B**) are in close accord with prior chemical probing studies[11,12] and recent cryo EM models of assembled telomerase[18,20] (**Fig. 2C**). We observed nucleotides with medium to strong degrees of DMS reactivity (0.3-0.8 and >0.8, respectively) in the template region (C46-C56), the internal loop joining the P2a.1 and P2a stems, the bulge joining P2a and P2b (J2a/b), and hinge region linking P2a.1 to P3 (J2a.1/3). Notably, we detected numerous nucleotides with low (0.1-0.3) and weakly reactive (<0.1) in the pseudoknot-forming P3 helix (C108, A11, C112, A174-176, C180, A181, C183).

DMS reactivity of the CR4/5 domain was consistent with a three-way junction architecture, with most of the reactive nucleotides residing within the internal loops that join the P4.2/P5 stems and the P6a/P6b stems, as well as the apical loop of P6 (C277-278) and the junction regions of the three-way junction motif (C247, C248, A252, C255, A301, C317, A318) (**Fig. 3A-B**). We detected DMS reactivity in the nucleotides comprising the P6.1 stem (A302, A304, C311, C313), a finding consistent with previous chemical probing studies[11,12,40]. In addition, we detected unexpected DMS reactivity at a trio of cytosines (C286-288) that contribute to the P6b helix.

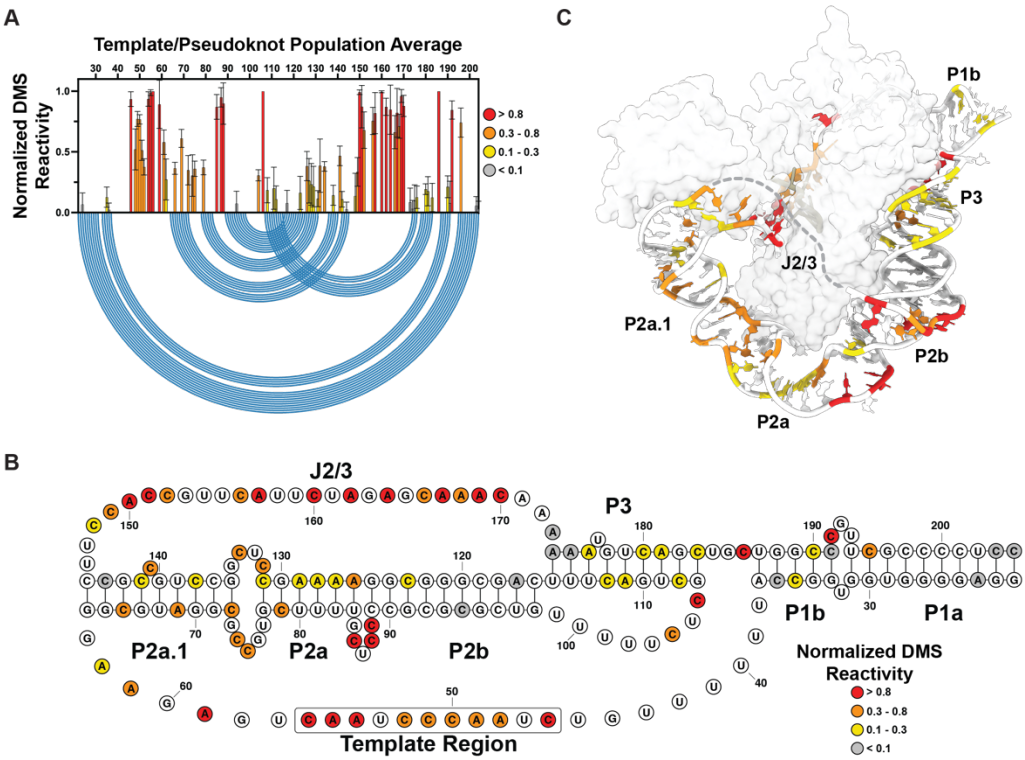


Figure 2: Population average DMS reactivity of the hTR t/PK domain.

(A) Normalized DMS reactivity of hTR t/PK domain (nt. 22-204). Intensity of DMS reactivity colored according to the provided legend. Blue arcs designate the base pairing pattern of the canonical t/PK conformation seen by cryo-EM of assembled telomerase. (B) Secondary structure of the canonical t/PK conformation with DMS reactivity of the hTR population average overlaid onto the nucleotides. Stem elements as well as the template sequence are labeled. (C) DMS reactivity overlaid onto the cryo-EM model of assembled telomerase with stem elements labeled.

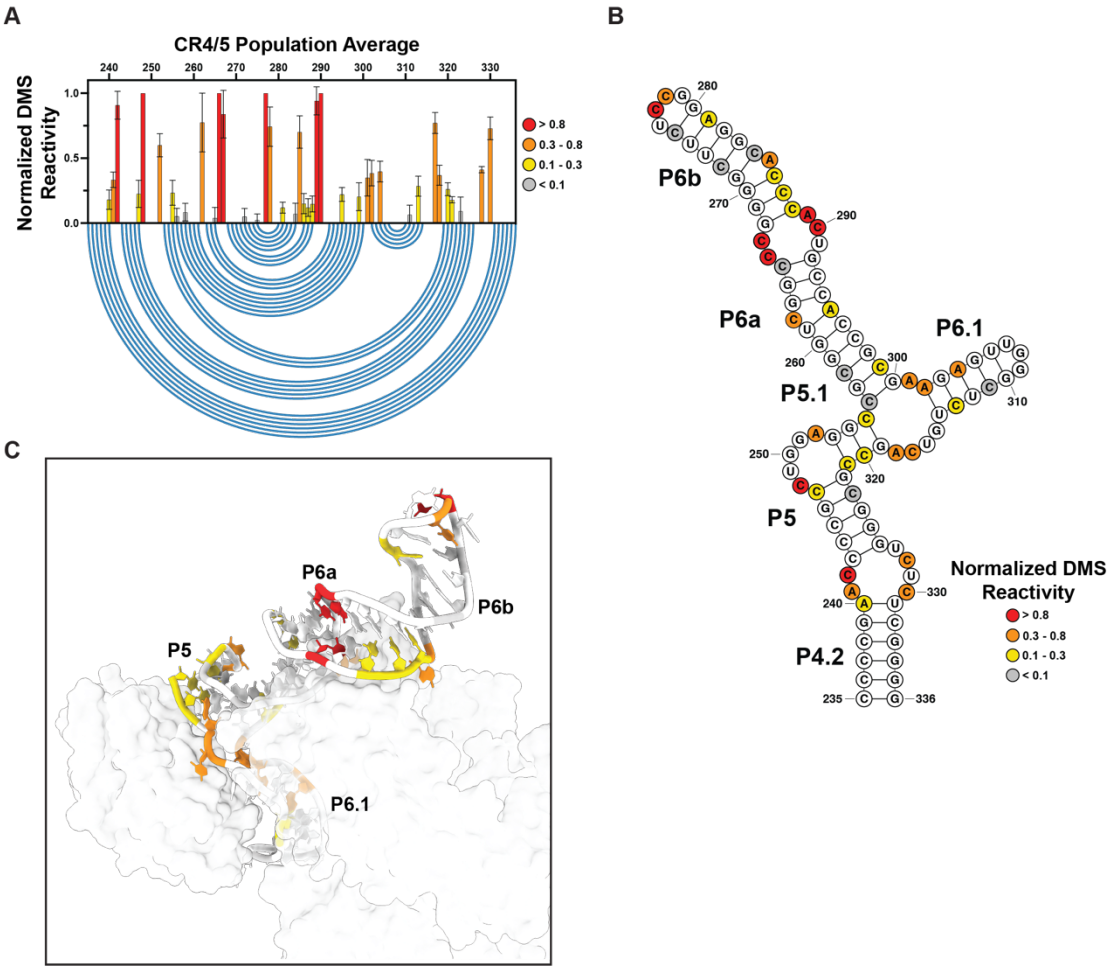


Figure 3: Population average DMS reactivity of the hTR CR4/5 domain.

(A) Normalized DMS reactivity of hTR CR4/5 domain (nt. -204). Intensity of DMS reactivity colored according to the provided legend. Blue arcs designate the base pairing pattern of the canonical t/PK conformation seen by cryo-EM of assembled telomerase. (B) Secondary structure of the canonical CR4/5 conformation with DMS reactivity of the hTR population average overlaid onto the nucleotides. Stem elements are labeled. (C) DMS reactivity overlaid onto the cryo-EM model of assembled telomerase with stem elements labeled.

Previously, we reported that uncharacteristic chemical reactivity of CR4/5 in vitro originates from its heterogeneous folding behavior[40]. Thus, we sought to determine whether alternative folding conformations of CR4/5 exist in the cellular environment and could explain our detected DMS reactivities.

Ensemble deconvolution reveals alternative conformations of the t/PK and CR4/5

Because MaPseq exploits the read-through behavior of multiple DMS modifications per RNA, it permits the use of bioinformatic deconvolution algorithms to identify distinct clusters of DMS reactivities that distinguish alternative RNA conformations[38,41]. We applied the Deconvolution of RNA Ensembles by Expectation Maximization (DREEM) algorithm[38] to our DMS reactivities of endogenous hTR. We then used DMS reactivities derived from the clusters to guide thermodynamic secondary structure predictions.

We found that the hTR t/PK domain forms distinct major and minor clusters across multiple biological replicates with relative abundances of ~80% and ~20%, respectively. Nucleotides responsible for forming the pseudoknot P3 helix (C108, A111, C112 and A174-A176, C180, A181, C183) are overall weakly DMS reactive in the major cluster (**Fig. 4A**), while in the minor cluster they partition into strongly reactive (C108, A111, C112, C180, A181, C183) and mostly unreactive (A174-176) groups . Cluster-guided thermodynamic structure prediction of the major cluster yielded a conformation similar to the canonical t/PK domain with a slightly truncated

pseudoknot P3 helix, while the predicted model of the minor cluster lacks the pseudoknot and is arranged as an extended stem-loop architecture (**Fig. 4B-C**).

The CR4/5 domain also forms major and minor clusters with relative abundances of ~90% and ~10%, respectively. Two cytosines that form the internal loop bisecting P6a and P6b (C266-267) are strongly reactive in the major cluster but are moderately reactive in the minor cluster (**Fig. 5A**). An additional three cytosines that form P6b (C286-288) are weakly reactive in the major cluster, while predicted to be strongly reactive in the minor cluster. Structure prediction of the major cluster yields the canonical three-way junction CR4/5 architecture (**Fig. 5B**). The minor cluster is predicted to form a register-shifted stem loop, with strongly reactive P6a nucleotides (A285-C290) forming the apical loop (**Fig. 5C**). Notably, this alternative conformation of CR4/5 lacks the P6.1 stem loop, which is essential for telomerase assembly and catalytic activity. This predicted alternative CR4/5 conformation is unlikely to arise from an artifact of DREEM clustering, as it is a predicted structure with similar free energy to the canonical three-way junction form even in constraint-free structure prediction runs. We note that while major and minor clusters are observed for both the t/PK and CR4/5 domains, our sequencing library preparation method splits the two domains prior to DREEM deconvolution, and therefore does not provide information regarding whether the two minor conformations reside in the same molecule.

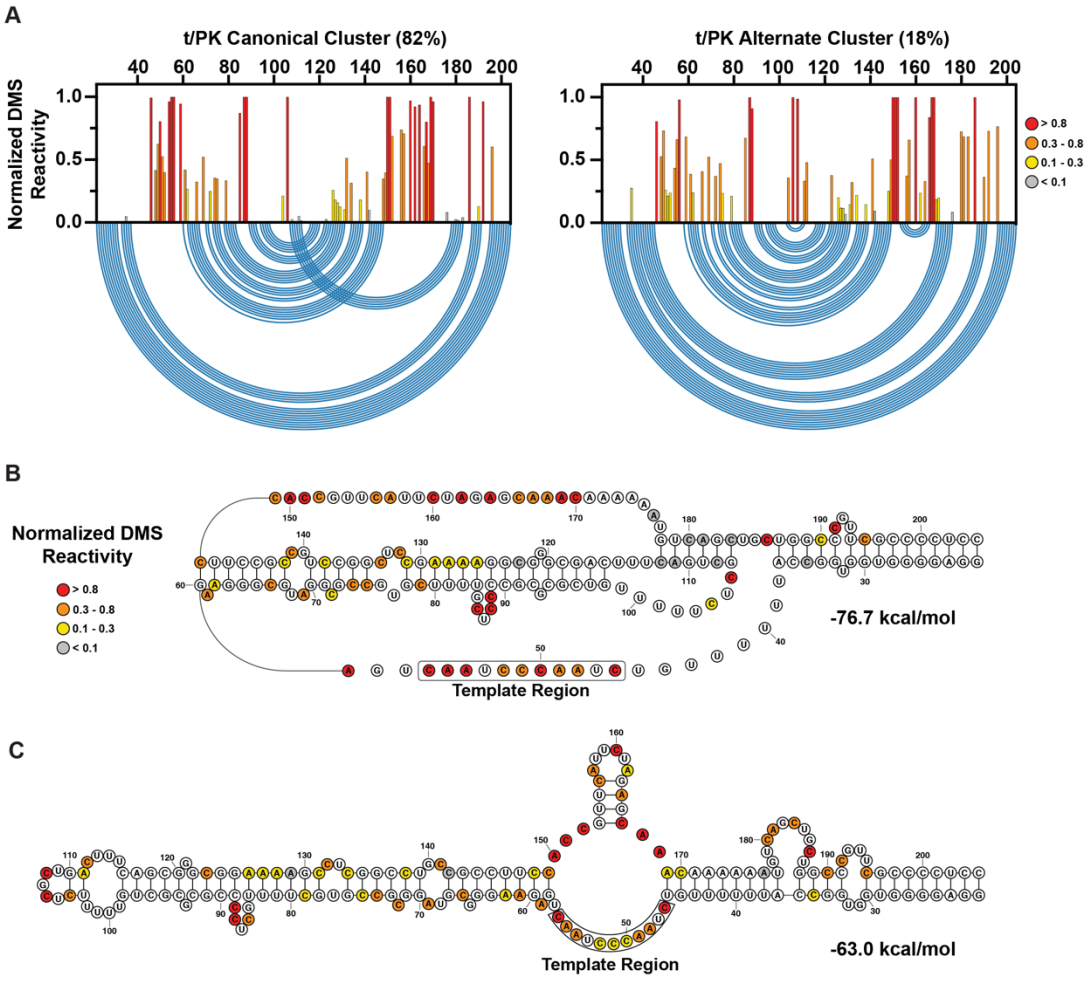


Figure 4: DREEM-deconvoluted DMS profiles and structure predictions of hTR t/PK.

(A) Normalized DMS reactivity profiles of the two clusters predicted by DREEM. Intensity of DMS reactivity colored according to the provided legend. Blue arcs designate the base pairing pattern of the data-guided predicted t/PK secondary structure. (B) Data-guided secondary structure prediction of the t/PK domain from the canonical cluster of DMS reactivities. (C) Data-guided secondary structure prediction of the t/PK domain from the alternative cluster of DMS reactivities.

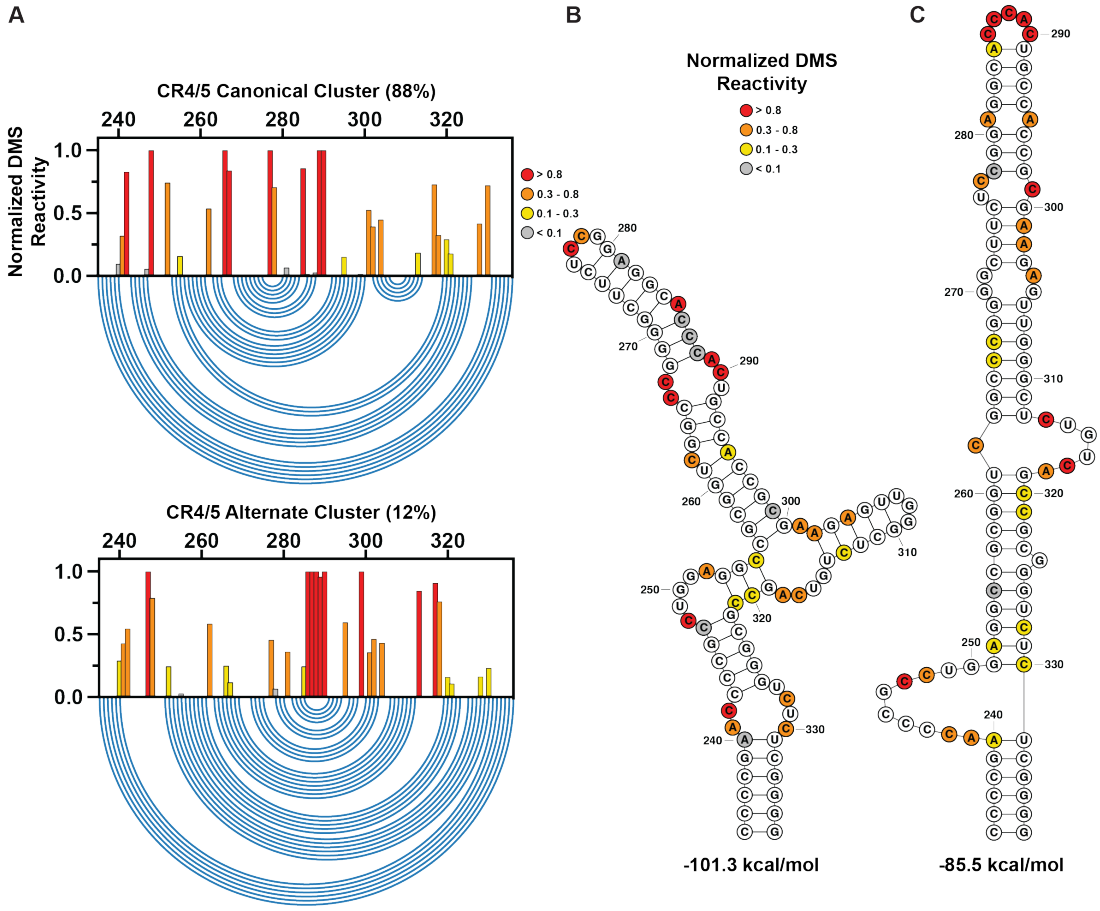


Figure 5: DREEM-deconvoluted DMS profiles and structure predictions of hTR CR4/5.

A) Normalized DMS reactivity profiles of the two clusters predicted by DREEM. Intensity of DMS reactivity colored according to the provided legend. Blue arcs designate the base pairing pattern of the data-guided predicted CR4/5 secondary structure. (B) Data-guided secondary structure prediction of the CR4/5 domain from the canonical cluster of DMS reactivities. (C) Data-guided secondary structure prediction of the CR4/5 domain from the alternative cluster of DMS reactivities.

The alternative CR4/5 conformation does not support efficient telomerase assembly and activity

To test the biological effect of the alternative CR4/5 on telomerase biogenesis, we aimed to bias the hTR folding landscape toward this conformation in vivo and query its effects on telomerase RNP assembly and catalytic activity. We identified three different nucleotide positions within the CR4/5 P6b stem that, when mutated, either disrupt base pairing interactions in the canonical conformation and/or introduce base pairs in the alternative conformation without directly disrupting interactions with hTERT (**Fig. 6A-B**). Our designed hTR mutant constructs contain either one (M1: G270C), two (M2: C277G, A281U), or three (M3: G270C, C277G, A281U) nucleotide substitutions in the CR4/5 domain. Each of the hTR variants were sub-cloned into a plasmid under the control of the U1 promoter and were individually transfected into HeLa cells together with a separate plasmid expressing hTERT with an N-terminal FLAG tag. Following each transfection, we performed the live cell DMS MaPseq procedure to analyze the folding properties of each transiently overexpressed hTR variant. The resultant DMS profiles of all mutants displayed elevated DMS reactivity at nucleotides C286-288, consistent with an increased relative abundance of the alternative CR4/5 conformation compared to the canonical conformation (**Fig. 6C**). Corroborating this finding, DREEM deconvolution followed by structure prediction revealed an increased representation of the alternative CR4/5 conformation in the ensemble. The CR4/5 structural ensemble of M1 and M2 is ~45% folded in the alternative conformation, while M3 adopts the alternative conformation

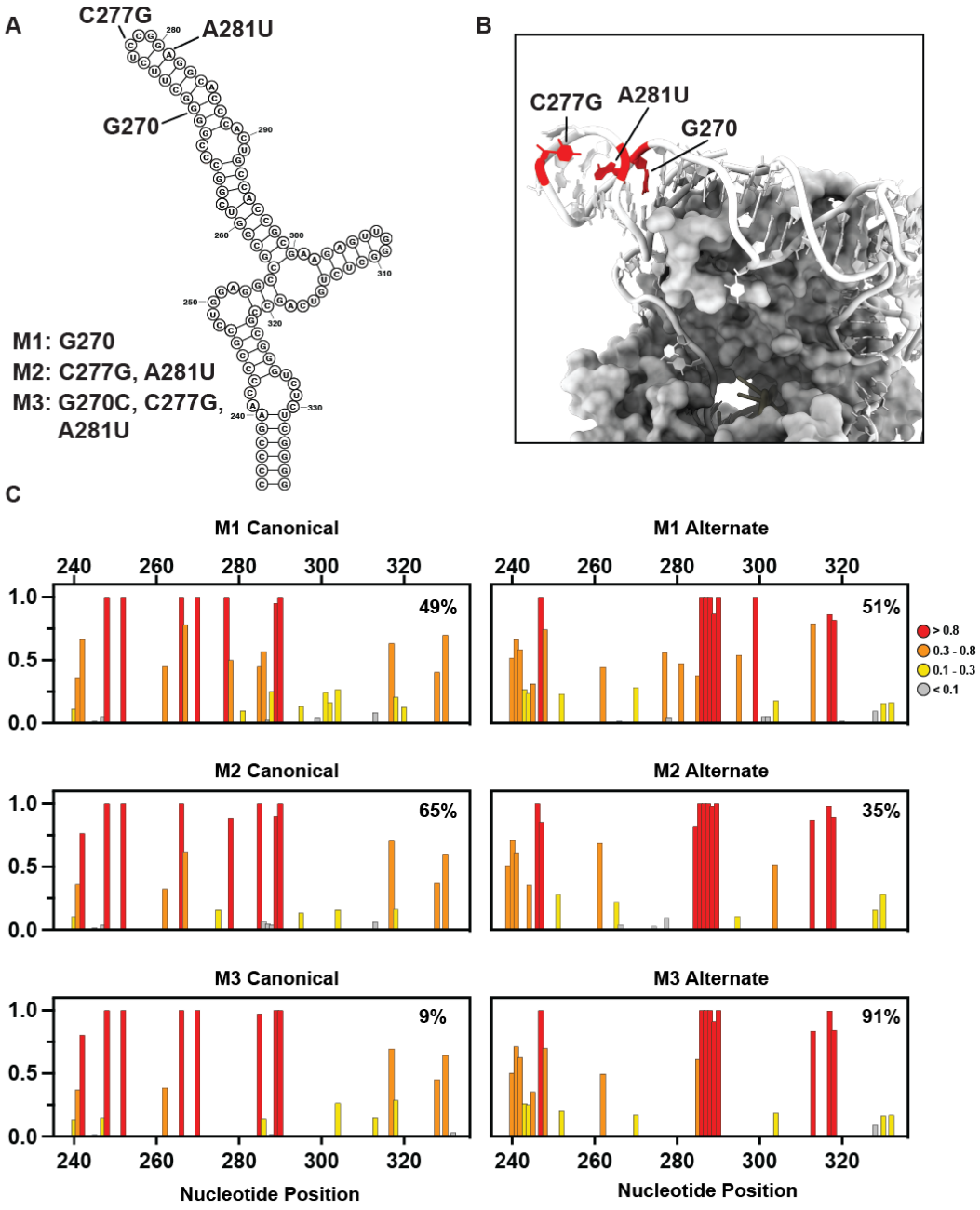


Figure 6: Design and MaPseq validation of CR4/5 mutants.

A) Canonical CR4/5 secondary structure with mutated nucleotides labeled. Three different mutants were generated: M1 (G270C), M2 (C277G, A281U), M3 (G270C, C277G, A281U). (B) cryo-EM structure of CR4/5 with mutated nucleotides highlighted. (C) Normalized DMS reactivity profiles of clusters predicted by DREEM for each CR4/5 mutant.

with ~90% abundance, representing a drastic redistribution of the CR4/5 structural ensemble away from the canonical conformation (**Fig. 6C**). We also found that compared to endogenous hTR overexpressed wild type hTR had an increased proportion of alternatively folded CR4/5 (~10% vs ~30%, respectively), potentially due to an effect of increased hTR concentration in the nuclear environment.

Having characterizing the increased abundance of the alternative CR4/5 conformation in cells transfected with the mutant hTR variants, we next set out to investigate the impact of the observed change in CR4/5 folding on telomerase RNP assembly and catalytic activity. For these experiments we transiently transfected either HeLa or HEK293T cells with the WT, M1, M2, or M3 hTR constructs together with FLAG-tagged hTERT. Following immunopurification of the telomerase RNPs from each transfection experiment via the FLAG-tagged hTERT subunit we performed direct primer extension activity assays on the IP fraction (**Fig. 7A**). When compared to the WT hTR construct, M1 and M2 each exhibited moderately reduced total activity compared to WT (~70-80% activity) while M3 was severely deficient (~5% activity) (**Fig. 7B**). In contrast, none of the telomerase mutants showed any apparent change in repeat addition processivity (RAP). We note that the residual catalytic activity observed for the M3 experiment may result from assembly of endogenously expressed hTR with FLAG-tagged hTERT. Importantly, Western blots for FLAG-hTERT show that perturbation of the CR4/5 folding landscape does not alter the amount of hTERT expression (**Fig. 7C**). Thus, the knock down in telomerase catalytic activity cannot be explained by reduced expression of the hTERT subunit.

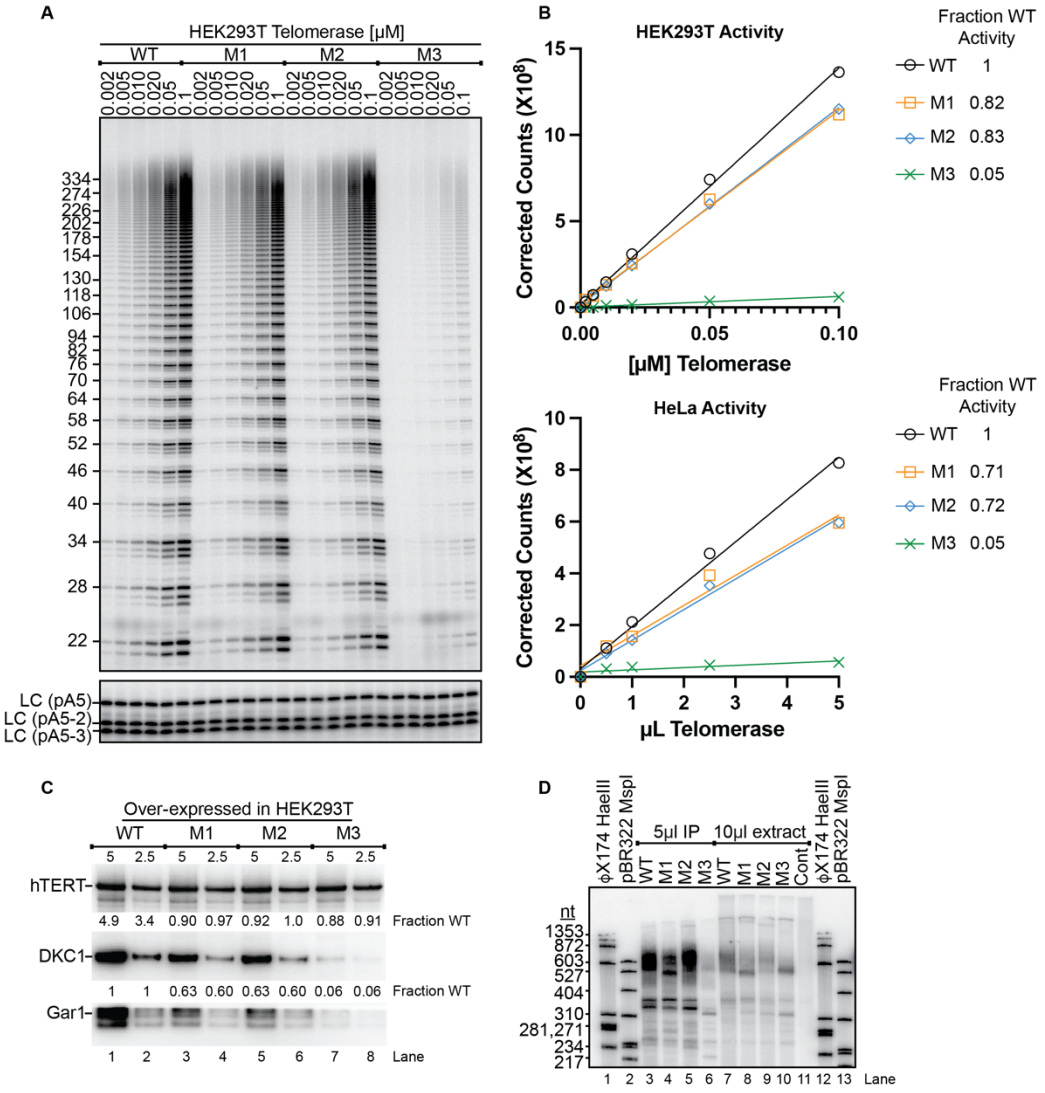


Figure 7: Telomerase activity assay and RNP assembly quantification for hTR mutants.

(A) Activity assay of FLAG-purified, HEK293T-derived telomerase, measuring incorporation of ^{32}P-GTP into a telomere primer. (B) Relative catalytic rates of HEK or HeLa-derived telomerase, calculated by summing the total lane signal and normalizing to WT. Fractional activity relative to WT was calculated comparing the slope derived from a linear regression fit. (C) Fold enrichment of reads corresponding to mutant hTR in purified telomerase RNPs vs cellular extract (input). Read abundances in RNPs and input were calculated as fractions of the total (WT plus Mutant), then compared between RNP and input fractions to derive fold enrichment.

We next asked if the reduced telomerase activity reflected less active assembled RNP enzymes or, rather, a failure to assemble RNP complexes. When telomerase was immunopurified by the FLAG epitope tag on the hTERT protein, Northern blotting for hTR in the IP and extract fractions revealed a significant enrichment of WT, M1 and M2 hTR variants. In contrast, the M3 hTR variant was not enriched in the IP, suggesting it has an impaired capacity to bind and assemble with hTERT (**Fig. 7D**). As a further test of whether each hTR variant was competent to assemble into a functional telomerase RNP complex, we performed Western blots for the dyskerin (DKC1) and Gar1 subunits of the H/ACA RNP complex. Our results show that expression of the M1 and M2 results in modest knock down in both DKC1 and Gar1 levels in the IP fraction, while M3 knocks down DKC1 and Gar1 levels much more drastically. Taken together, these results support the conclusion that expression of the M3 hTR variant that increases the abundance of the alternative CR4/5 fold in vivo can substantially decrease telomerase RNP assembly efficiency.

Discussion

Unlike the functional multi-state RNA folding properties of riboswitches[42], hTR structure is known to adopt a phylogenetically conserved secondary structure and therefore would not be predicted to adopt long-lived alternative conformations in vivo[8,10,34,43]. Therefore the observation that a portion of cellular hTR exists as a misfolded, catalytically incompetent conformer, was unexpected. The propensity of RNA to fall into kinetic traps during folding is well established in vitro; however, a

reasonable expectation would be that such misfolded RNAs would be subject either to refolding or to degradation in vivo[44–47]. Instead, we found that alternative conformations of the t/PK and CR4/5 – conformations that are inconsistent with the cryo EM models of telomerase – persist in the steady state at endogenous levels of expression. If the alternative hTR conformer is being refolded or degraded, the rates of such processes must be slow enough to give a substantial misfolded population in the steady state. On the other hand, it also remains possible that the alternative hTR conformer could exist in a different RNP with a non-telomerase function[48], though this notion is speculative and requires future exploration. Using structure-guided RNA mutagenesis, we identified nucleotide positions that, when mutated, can bias the hTR folding landscape away from the canonical fold and toward the alternative conformation. The hTR variants that modestly alter the hTR folding landscape to favor the alternative conformation (M1 and M2) support near wild type levels of telomerase assembly and catalytic activity. In contrast, the M3 hTR variant results in the majority of cellular hTR adopting the alternative conformation and consequently substantially disrupts RNP assembly.

In vitro biochemical and biophysical analyses of the hTR CR4/5 domain show that the RNA primarily folds into a heterogeneous ensemble of states[40]. Moreover, reconstitution of hTR with hTERT in rabbit reticulocyte lysates chaperones the RNA into a fold that is consistent with the recent cryo EM structures of the endogenously assembled RNP. In HeLa cells, the RNA subunit of telomerase is known to be in stoichiometric excess of hTERT[13]. Thus, our observation that the majority of hTR in

the cell, which is likely not in complex with hTERT, adopts a canonical CR4/5 conformation suggests that proteins other than hTERT promote functional hTR folding. The discovery of a histone H2A/H2B dimer directly bound to CR4/5 in recent cryo EM models[18,20,21] provides a strong candidate for such an RNA chaperoning activity. In fact, H2A/B could potentially perform a function analogous to that of the Tetrahymena telomerase holoenzyme protein, p65, which induces an RNA conformational change to promote RNP assembly[49–52]. Modest stabilization of the alternative CR4/5 fold potentially allows the M1 and M2 mutants to be refolded by nuclear chaperones, explaining why they have only small effects on telomerase assembly and catalytic activity. In contrast, the M3 hTR variant, which almost exclusively adopts the alternative CR4/5 conformation, presumably has a higher energetic barrier for RNA refolding and thus was unable to be rescued.

Our characterization of the alternative pseudoknot-lacking conformer of the t/PK domain builds upon previous studies of its heterogeneous folding in human[36,53,54] and ciliate telomerase RNA[55–57]. We observed high and moderate levels of DMS reactivity in template sequence (nt 46-57) of both of our predicted t/PK conformations. This result suggests that assembly with hTERT is primarily associated with the formation of the pseudoknot-forming P3 helix, and does not significantly reorganize the template region as was previously suggested[12]. Future experiments designed to manipulate the t/PK RNA folding ensemble will yield a richer understanding of how conformational changes of this domain promote telomerase biogenesis.

Overall, we demonstrate the efficacy of using in cell chemical probing in combination with bioinformatic deconvolution to parse individual RNA conformations of a noncoding RNA. Our findings lead us to speculate that heterogeneous hTR folding presents a barrier in the telomerase assembly pathway. This barrier may be overcome via the action of RNA folding chaperones that rescue improperly folded hTR molecules providing a possible mechanism for regulating telomerase biogenesis pathway and catalytic activity (**Fig. 8**). These same techniques could be used to explore the populations of other functional noncoding RNAs in cells. It will be interesting to determine if the substantial subpopulation of misfolded RNA seen for hTR is specific to telomerase or perhaps occurs more generally with long noncoding RNAs.

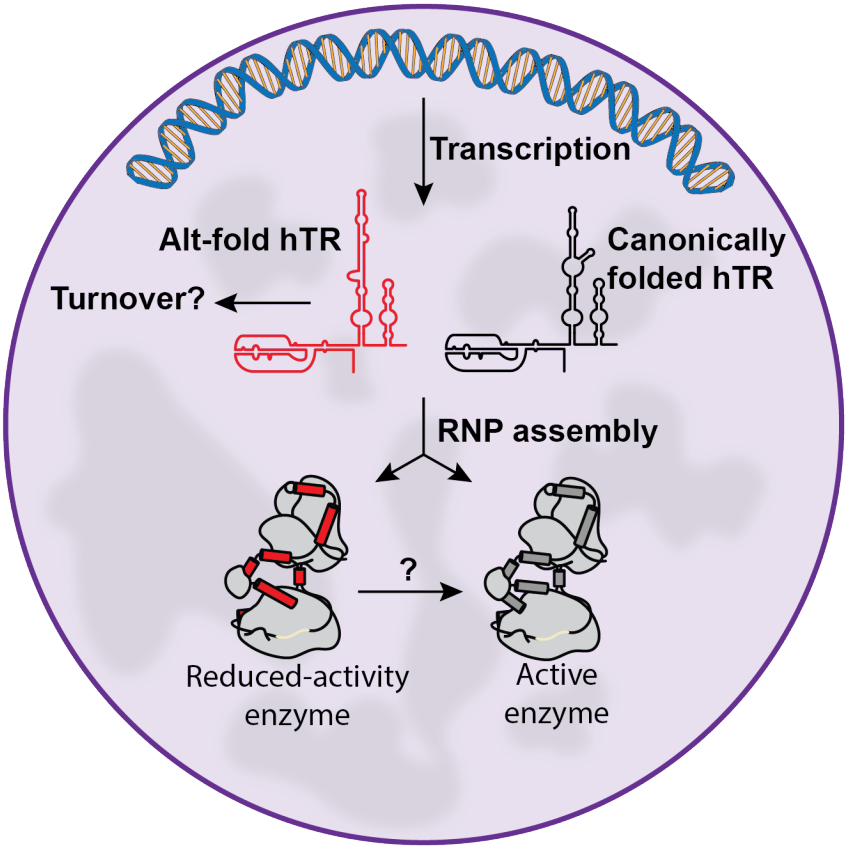


Figure 8: Model of telomerase RNP assembly with respect to CR4/5 conformation.

hTR exists as a structural ensemble in cells, with a relatively small population adopting alternative conformations in the t/PK and CR4/5 domains. hTR molecules possessing either the canonical or the alternative CR4/5 conformations are competent to assemble into telomerase RNPs, however the alternative CR4/5 fold greatly reduces telomerase catalytic activity.

MATERIALS AND METHODS

Cell culture and DMS modification

HeLa and HEK293T/17 cells (ATCC CRL-11268, lot:70040949) were cultured in DMEM containing 10% fetal bovine serum, 100 units/mL of penicillin, 100 mg/m of streptomycin, and 1X Gibco GlutaMAX at 37°C, 5% CO2. Cells were seeded into 6-well plates at 0.3 x 10^6 cells per well, and allowed to grow to 90% confluency prior to use. For DMS modification, plates of HeLa cells were divided into DMS and control groups (3 wells each). Media from the DMS group was aspirated and replaced with 2 mL of a 2% DMS (Sigma) solution in cell media (prepared by brief vortexing to incorporate DMS). Cells were incubated at 37°C for 4 min, media was aspirated from each well and replaced with 10 mL of ice cold DMS quench solution (30% beta-mercaptoethanol in phosphate-buffered saline). Plates were placed on ice and the cells were collected by scraping and brief trituration, then pelleted by centrifugation at 200 x g for 5 min. Cells were washed once with cold PBS and total RNA harvested with 1 mL Trizol (Invitrogen) following the manufacturer's protocol.

Transient transfection and purification of human telomerase

Plasmids containing hTERT (pvan107-3xFLAG) and WT or mutant hTR (pBSU1-hTR[58]) were transfected at a 1:3 molar ratio using lipofectamine 2000 (11668019, Thermo Fisher Scientific). The cells were further expanded 3-fold 24 h after transfection and then 24 h later either harvested or used in DMS probing reactions. To purify telomerase, the cell pellet was lysed with CHAPS lysis buffer (10 mM Tris–

HCl pH 7.5, 1 mM $MgCl_2$, 1 mM EGTA, 0.5% CHAPS, 10% glycerol, 5 mM b-mercaptoethanol,) for 45 min at $4^{O}C$ on a rotator. The lysate was then clarified by centrifugation at 13,000 x g at $4^{O}C$ for 30 min. Anti-FLAG resin (A2220, Sigma-Aldrich) was added to the clarified supernatant and the samples incubated on a rotator for 4 h (or overnight) at $4^{O}C$. The anti-FLAG resin was washed 3x with wash buffer (20 mM HEPES–NaOH pH 8.0, 2 mM $MgCl_2$, 0.2 mM EGTA, 0.1% NP-40, 10% glycerol, 1 mM DTT) before elution using wash buffer supplemented with 0.25 mg/ml 3xFLAG peptide (F4799, Sigma-Aldrich). Purified telomerase complex was verified by western blotting.

Preparation of mutant hTR plasmids

Starting with the pBS-U1-hTR plasmid[58], three hTR mutants (M1: G270C, M2: C277G & A281U, M3: G270C & C277G & A281U) were generated using the mega primer method with gel extraction of intermediate products after the first PCR step[59].

Preparation of hTR-targeted MaPseq libraries

10 µg of HeLa cell derived total RNA was DNAse treated in 1X TURBO DNAse buffer with 1 µL TURBO DNAse enzyme (Thermo Fisher Scientific), followed by column-purification with DNA Clean and Concentrator-5 columns (Zymo Research) following the manufacturer protocol. 5 µg of DNAse-treated RNA was mixed with 5 pmol hTR targeting primer containing a 4 nucleotide unique molecular identifier (UMI) in a volume of 11 µL, heated to 75°C for 3 min and annealed at 35°C for 15

min. Next, 4 µL of 5X M-MLV buffer (Promega), 1 µL 0.1M DTT, 1 µL RNAsin Plus (Promega), and 1 µL TGIRT III (Ingex) were added and the mixture was incubated at room temperature for 30 min. Then, 2 µL of 10mM dNTPs were added, and the reaction was mixed and incubated at 60 °C for 2.5 h. After RT, 1 µL of 5M NaOH was added directly to the cDNA and incubated at 95°C, followed by 2.5 µL of 2M HCl and purification by DNA Clean and Concentrator-5 columns (Zymo Research) using an 8:1 ratio of DNA Binding Buffer to cDNA volume. The cDNA was mixed into a 50 µL second strand synthesis reaction containing 10 µL 5X GC Phusion Buffer (Thermo Fisher Scientific), 1 µL 10 mM dNTP, 25 pmol of a second hTR targeting primer containing a 4 nucleotide UMI, and 1 µL Phusion polymerase. The second strand synthesis reaction was incubated in a thermocycler with the following program: 98°C 2 min, 60°C 2 min, 72°C 10 min. Second strand product was cleaned up with Ampure XP beads (Beckman Coulter) at a 0.8:1 bead to sample volume ratio, and eluted in 20 µL nuclease-free H_2O. Illumina sequencing libraries of two partially overlapping hTR regions (Pseudoknot: nucleotides 1-286 and CR4/5: nucleotides 193-451) were generated by two successive 50 µL PCR reactions. For the first PCR, 4 µL second strand product was added to a 50 µL PCR reaction containing primers for either the pseudoknot or CR4/5 amplicon and run with the following program: 10 cycles at 66°C annealing temperature, followed by 10 cycles at a 60°C annealing temperature. Amplicons were bead cleaned with a 0.8:1 ratio, and went into a second PCR of 10 cycles at an annealing temperature of 65°C to complete the sequencing adapters and multiplexing barcodes. Libraries were bead cleaned with a

0.8:1 ratio and quantified by Qubit (Thermo Fisher Scientific) and Tapestation (Agilent) for quality metrics before sequencing by 150 bp paired end reads on the iSeq100 (Illumina).

hTERT Western blots

Activity of the immunopurified human telomerase complex over-expressed in either HEK293T/17 (ATCC CRL-11268, lot:70040949) or HeLa cells was determined by a direct assay modified from a published protocol (Wang et al. 2007). The reaction mixture (20 mL) contained 1x human telomerase assay buffer (50 mM Tris-HCl at pH 8.0, 50 mM KCl, 75 mM NaCl, 1 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 1 mM spermidine), 0.05 mM telomeric DNA primer, 0.5 mM dTTP, 3.3 mM dGTP, 10 mM dATP, and 0.33 mM ^{32}P-dATP (3000 Ci/mmol, 1 Ci = 37GBq, PerkinElmer). Following a 1 hr incubation at 30^{O}C, reactions were stopped with the addition of 100 mL of 3.6 M NH_4OAc containing 20 mg of glycogen. Ethanol (500 mL) was added for precipitation. After incubating for 1 h at 80°C, samples were centrifuged for 15 min at 4°C. Pellets were washed with 70% ethanol and resuspended in 10 mL of H_2O followed by 10 mL of 2x loading buffer (94% formamide, 0.1x TBE, 0.1% bromophenol blue, 0.1% xylene cyanol). The heat-denatured samples were loaded onto a 10% polyacrylamide/7 M urea/1x TBE gel for electrophoresis (bromophenol blue run to bottom of gel). After electrophoresis, the gel was dried and quantified by using a PhosphorImager (Cytiva).

Telomerase activity assays

Activity of the immunopurified human telomerase complex over-expressed in either HEK293T/17 (ATCC CRL-11268, lot:70040949) or HeLa cells was determined by a direct assay modified from a published protocol (Wang et al. 2007). The reaction mixture (20 mL) contained 1x human telomerase assay buffer (50 mM Tris-HCl at pH 8.0, 50 mM KCl, 75 mM NaCl, 1 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 1 mM spermidine), 0.05 mM telomeric DNA primer, 0.5 mM dTTP, 3.3 mM dGTP, 10 mM dATP, and 0.33 mM ^{32}P-dATP (3000 Ci/mmol, 1 Ci = 37GBq, PerkinElmer). Following a 1 hr incubation at 30^{O}C, reactions were stopped with the addition of 100 mL of 3.6 M NH_4OAc containing 20 mg of glycogen. Ethanol (500 mL) was added for precipitation. After incubating for 1 h at 80°C, samples were centrifuged for 15 min at 4°C. Pellets were washed with 70% ethanol and resuspended in 10 mL of H_2O followed by 10 mL of 2x loading buffer (94% formamide, 0.1x TBE, 0.1% bromophenol blue, 0.1% xylene cyanol). The heat-denatured samples were loaded onto a 10% polyacrylamide/7 M urea/1x TBE gel for electrophoresis (until bromophenol blue runs to the bottom of the gel). After electrophoresis, the gel was dried and quantified by using a PhosphorImager (Cytiva).

hTR Northern blots

Deproteinized CHAPs extract or deproteinized immunopurified telomerase was run on a 6% polyacrylamide/7M urea/1XTBE gel. Nucleic acid was transferred to Hybond-N+ membrane at 2 amps in 0.5X TBE at 4C for 1.5hr. Nucleic acid was

crosslinked to the membrane using a CL-1000 ultraviolet crosslinker using 254 nm lamps with an energy level of 1200 X 100mJ/cm2. Membrane was blocked using rapid-hyb buffer for 30 min at 50C. 10 7 counts of a DNA probe labeled with 32 P-g-ATP and polynucleotide kinase was added and incubated overnight at 50C. Membrane was washed 3X with 2X SSC/0.1%SDS for 15min at 50C, followed by 1X for 15 min with 2X SSC/0.1% SDS. Membrane was wrapped in saran wrap and exposed to a phosphor screen (Cytiva).

DREEM analysis

Trimming of sequencing adapters and removal of reads smaller than 120 nucleotides was performed using TrimGalore with the following command:

```
trim_galore –paired –nextera R1.fastq R2.fastq –length 120
```

FASTQ files were aligned using Bowtie 2[60] to the hTR reference sequence using the following parameters: `--local --no-unal --no-discordant --phred33`

Alignment files were compressed using SAMtools[61] and analyzed with a local build of DREEM (Deconvolution of RNA Ensembles using Expectation Maximization). The start and end parameters for the Pseudoknot/Template and CR4/5 domains were 22-204, and 213-368, respectively. Folding predictions were performed using default parameters of the RNAStructure package[62]. Pseudoknot prediction was performed using the ProbKnots function of RNAstructure with default folding parameters.

Data visualization

Bar and scatter plots were generated with Graphpad Prism. RNA secondary structures were generated with VARNA[63] and modified with Adobe Illustrator. Arc diagrams were generated using R-chie[64].

Table 1: List of primers used in study

hTR RT, UMI, partial Nextera	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNGCATGTGTGAGCCGAGTCCTGGGTGC
hTR second strand, UMI, partial Nextera	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGGGTTGCGGAGGGTGGGCCTG
hTR PK amplicon, partial Nextera	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCGCCCCTCCCGGGGAC
hTR TWJ amplicon, partial Nextera	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCCTCCGGAGAAGCCCCG
Forward hTR cloning (pBS-U1-hTR)	TCTAGAACTAGTGGATCCCCCGGG
Reverse hTR cloning (pBS-U1-hTR)	CGAGGTCGACGGTATCGATAAGCTTG
M2 Megaprimer	GGGTGCCACCGCAGAAGCCCCGGGCCGAC
M3 Megaprimer	GCCACCGCAGAAGCGCCGGGCCGACCGCGGCC

Nextera_Primer_A1	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAGATG
Nextera_Primer_A2	AATGATACGGCGACCACCGAGATCTACACCTCTGTATTCGTCGGCAGCGTCAGATG
Nextera_Primer_A3	AATGATACGGCGACCACCGAGATCTACACGATCAGCATCGTCGGCAGCGTCAGATG
Nextera_Primer_A4	AATGATACGGCGACCACCGAGATCTACACACAGTATGTCGTCGGCAGCGTCAGATG
Nextera_Primer_A5	AATGATACGGCGACCACCGAGATCTACACGACTGGAGTCGTCGGCAGCGTCAGATG
Nextera_Primer_A6	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTCAGATG
Nextera_Primer_A7	AATGATACGGCGACCACCGAGATCTACACTTGCATGCTCGTCGGCAGCGTCAGATG
Nextera_Primer_A8	AATGATACGGCGACCACCGAGATCTACACCGAATCCTTCGTCGGCAGCGTCAGATG
Nextera_Primer_B1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGTGTAT
Nextera_Primer_B2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGTGTAT
Nextera_Primer_B3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGTGTAT
Nextera_Primer_B4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGTGTAT
Nextera_Primer_B5	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTGTAT
Nextera_Primer_B6	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGTGTAT
Nextera_Primer_B7	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGTGTAT
Nextera_Primer_B8	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGTGTAT

References

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

2. Lange, T. de. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).

3. Greider, C. W. & Blackburn, E. H. Tracking telomerase. *Cell* **116**, S83–S87 (2004).

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

5. Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).

6. Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866–868 (1990).

7. Allsopp, R. C. *et al.* Telomere length predicts replicative capacity of human fibroblasts. *Proc Natl Acad Sci U S A* **89**, 10114–10118 (1992).

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

9. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc Natl Acad Sci U S A* **101**, 10024–10029 (2004).

10. Chen, J.-L. & Greider, C. W. An emerging consensus for telomerase RNA structure. *Proc Natl Acad Sci U S A* **101**, 14683–14684 (2004).

11. Antal, M., Boros, É., Solymosy, F. & Kiss, T. Analysis of the structure of human telomerase RNA in vivo. *Nucleic Acids Res* **30**, 912–920 (2002).

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

13. Xi, L. & Cech, T. R. Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. *Nucleic Acids Res* **42**, 8565–8577 (2014).

14. Vogan, J. M. & Collins, K. Dynamics of Human Telomerase Holoenzyme Assembly and Subunit Exchange across the Cell Cycle. *J Biol Chem* **290**, 21320–21335 (2015).

15. Vogan, J. M. *et al.* Minimized human telomerase maintains telomeres and resolves endogenous roles of H/ACA proteins, TCAB1, and Cajal bodies. *eLife* **5**, e18221 (2016).

16. Jiang, J. *et al.* Structure of telomerase with telomeric DNA. *Cell* **173**, 1179-1190.e13 (2018).

17. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).

18. Ghanim, G. E. *et al.* Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* **593**, 449–453 (2021).

19. Wan, F. *et al.* Zipper head mechanism of telomere synthesis by human telomerase. *Cell Res* **31**, 1275–1290 (2021).

20. Sekne, Z., Ghanim, G. E., van Roon, A.-M. M. & Nguyen, T. H. D. Structural basis of human telomerase recruitment by TPP1-POT1. *Science* **375**, 1173–1176 (2022).

21. Liu, B. *et al.* Structure of active human telomerase with telomere shelterin protein TPP1. *Nature* **604**, 578–583 (2022).

22. Nguyen, T. H. D. *et al.* Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190 (2018).

23. Forino, N. M., Hentschel, J. & Stone, M. D. Cryo-EM structures tell a tale of two telomerases. *Nat Struct Mol Biol* **28**, 457–459 (2021).

24. Egan, E. D. & Collins, K. Specificity and Stoichiometry of Subunit Interactions in the Human Telomerase Holoenzyme Assembled In Vivo. *Molecular and Cellular Biology* **30**, 2775–2786 (2010).

25. Collins, K. Physiological assembly and activity of human telomerase complexes. *Mechanisms of Ageing and Development* **129**, 91–98 (2008).

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

27. Tseng, C.-K., Wang, H.-F., Schroeder, M. R. & Baumann, P. The H/ACA complex disrupts triplex in hTR precursor to permit processing by RRP6 and PARN. *Nat Commun* **9**, 5430 (2018).

28. Roake, C. M. *et al.* Disruption of Telomerase RNA Maturation Kinetics Precipitates Disease. *Mol Cell* **74**, 688-700.e3 (2019).

29. Venteicher, A. S. *et al.* A Human Telomerase Holoenzyme Protein Required for Cajal Body Localization and Telomere Synthesis. *Science* **323**, 644–648 (2009).

30. Laprade, H. *et al.* Single-Molecule Imaging of Telomerase RNA Reveals a Recruitment-Retention Model for Telomere Elongation. *Molecular Cell* **79**, 115-126.e6 (2020).

31. Klump, B. M. *et al.* TCAB1 prevents nucleolar accumulation of the telomerase RNA to promote telomerase assembly. 2021.05.27.445986 Preprint at https://doi.org/10.1101/2021.05.27.445986 (2023).

32. Vulliamy, T. *et al.* The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* **413**, 432–435 (2001).

33. Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nat Rev Genet* **13**, 693–704 (2012).

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

35. Robart, A. R. & Collins, K. Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked subunit variants. *J. Biol. Chem.* **285**, 4375–4386 (2010).

36. Deshpande, A. P. & Collins, K. Mechanisms of template handling and pseudoknot folding in human telomerase and their manipulation to expand the sequence repertoire of processive repeat synthesis. *Nucleic Acids Res* **46**, 7886–7901 (2018).

37. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nature Methods* **14**, 75–82 (2017).

38. Tomezsko, P. J. *et al.* Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**, 438–442 (2020).

39. He, Y. *et al.* Structures of telomerase at several steps of telomere repeat synthesis. *Nature* **593**, 454–459 (2021).

40. Palka, C., Forino, N. M., Hentschel, J., Das, R. & Stone, M. D. Folding heterogeneity in the essential human telomerase RNA three-way junction. *RNA* **26**, 1787–1800 (2020).

41. Morandi, E. *et al.* Genome-scale deconvolution of RNA structure ensembles. *Nature Methods* **18**, 249–252 (2021).

42. Serganov, A. & Nudler, E. A Decade of Riboswitches. *Cell* **152**, 17–24 (2013).
43. Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Current Opinion in Structural Biology* **16**, 307–318 (2006).

44. Herschlag, D. RNA Chaperones and the RNA Folding Problem. *J. Biol. Chem.* **270**, 20871–20874 (1995).

45. Cruz, J. A. & Westhof, E. The Dynamic Landscapes of RNA Architecture. *Cell* **136**, 604–609 (2009).

46. Woodson, S. A. Taming free energy landscapes with RNA chaperones. *RNA Biol* **7**, 677–686 (2010).

47. Huang, H. & Karbstein, K. Assembly factors chaperone ribosomal RNA folding by isolating helical junctions that are prone to misfolding. *Proceedings of the National Academy of Sciences* **118**, e2101164118 (2021).

48. Rubtsova, M. & Dontsova, O. Human Telomerase RNA: Telomerase Component or More? *Biomolecules* **10**, 873 (2020).

49. Stone, M. D. *et al.* Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).

50. Berman, A. J., Gooding, A. R. & Cech, T. R. Tetrahymena Telomerase Protein p65 Induces Conformational Changes throughout Telomerase RNA (TER) and Rescues Telomerase Reverse Transcriptase and TER Assembly Mutants. *Mol Cell Biol* **30**, 4965–4976 (2010).

51. Singh, M. *et al.* Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol Cell* **47**, 16–26 (2012).

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

53. Theimer, C. A., Finger, L. D., Trantirek, L. & Feigon, J. Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proceedings of the National Academy of Sciences* **100**, 449–454 (2003).

54. Hengesbach, M., Kim, N.-K., Feigon, J. & Stone, M. D. Single-Molecule FRET Reveals the Folding Dynamics of the Human Telomerase RNA Pseudoknot Domain. *Angewandte Chemie International Edition* **51**, 5876–5879 (2012).

55. Mihalusova, M., Wu, J. Y. & Zhuang, X. Functional importance of telomerase pseudoknot revealed by single-molecule analysis. *Proceedings of the National Academy of Sciences* **108**, 20339–20344 (2011).

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

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

58. Cristofari, G. & Lingner, J. Telomere length homeostasis requires that telomerase levels are limiting. *EMBO J* **25**, 565–574 (2006).

59. Tyagi, R., Lai, R. & Duggleby, R. G. A new approach to 'megaprimer' polymerase chain reaction mutagenesis without an intermediate gel purification step. *BMC Biotechnol* **4**, 2 (2004).

60. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

61. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

62. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).

63. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).

64. Tsybulskyi, V., Mounir, M. & Meyer, I. M. R-chie: a web server and R package for visualizing cis and trans RNA–RNA, RNA–DNA and DNA–DNA interactions. *Nucleic Acids Research* **48**, e105 (2020).

CHAPTER IV

Guide to targeted MaP-seq and data analysis

Motivation and Overview of Targeted MaPseq

A key determinant of RNA's function is its molecular structure; the arrangement of intramolecular interactions that comprise the overall shape of each RNA molecule. Given that we collectively lack structural information about the vast majority of known RNAs, there is an immense need for efficient RNA structure determination methods. Compared to cryo EM, X-ray scattering, and nuclear magnetic resonance (NMR), chemical probing of RNA molecules is a relatively cheap and straightforward technique that enables researchers to gather structural information from RNAs in solution, in the presence of proteins or small molecules, and also within living cells. Data gathered from chemical probing can also be used to guide RNA structure prediction software and increase its accuracy[1–4].

Chemical probing is the treatment of a target RNA or cells with an RNA-reactive chemical or 'probe' that selectively reacts with RNA nucleotides at single-stranded or flexible regions. The covalently-attached chemical probe is then detected by reverse transcription of the modified RNA, which either aborts at the site of modification, or 'reads-through' and incorporates a non-templated nucleotide[5,6]. Mutational profiling and sequencing (MaP-seq) is a recently developed technique that leverages next-generation sequencing (NGS) to quantify these chemical probe-induced mutations[7–9]. Because each read originates from a single RNA, MaP-seq can

be thought of as a ‘single molecule’ method, allowing researchers to deconvolute RNA structural ensembles, which may reveal subpopulations of molecules with distinct secondary structures that arise from the same primary RNA sequence[10–13]. While these recent technical advances have brought about a ‘revolution’ in RNA chemical probing, the method has become more complicated and multi-disciplinary with its use of bioinformatic tools.

This chapter serves as a practical guide and set of protocols for ‘targeted’ MaPseq library generation (focusing on an individual target RNA), RNA structure prediction, and RNA structure ensemble analysis. Targeted MaPseq can be broken down into the following steps:

1. Cell culture and RNA synthesis
2. Chemical probing
3. Library preparation
4. Sequencing
5. Bioinformatic analysis
6. Graphical visualization of RNA structures

It is important to thoroughly consider and plan out each step of the workflow for your target RNA because these steps will be tailored to your experiment. This chapter will provide an overview of each step and includes protocols for in-cell and in vitro MaPseq at the end of the chapter (See MaPseq Laboratory Protocols). Note, the

protocols provided here may not be the optimal strategies for every potential RNA target and can be used as starting points for the development of new library preparation workflows. Also included in this guide is a walkthrough of the bioinformatic steps for MaPseq data analysis, RNA structure prediction, RNA ensemble analysis, and structure visualization.

1. Cell culture and RNA synthesis

Protocols: 1.1 through 1.7

Applying MaPseq to RNAs within living cells has the obvious advantage of studying the target RNA in its native environment where it is in contact with its interacting partners. However, this adds significant labor to the workflow, as the cells need to be cultured and probed. Immortalized mammalian cell lines (HEK293T and HeLa) are amenable to in vivo probing, while experiments involving primary cells and cell types that may be sensitive to chemical perturbations (e.g. immune cells) could require optimization until the ideal probing workflow is identified. **Protocols 1.1 to 1.6** cover mammalian cell culture and preparation of chemical probing for HEK293T and HeLa cells.

Alternatively, in vitro MaPseq experiments enable high-throughput synthesis and testing of variants of the RNA of interest, as well as multidimensional chemical probing methods such as 'mutate-and-map'[14]. Synthesis of RNAs in vitro is most easily accomplished using primer assembly reactions designed by the Primerize tool[15] (primerize.stanford.edu). The downside of using primer assembly is that the PCR

reactions may require optimization if the RNA target is > 500 nt or contains repetitive sequences. **Protocol 1.7** details the synthesis of RNA in vitro using the primer assembly method.

2. Chemical probing

Protocols: 2.1 through 2.4

Chemical probing reagents fall into two categories; 1) modifiers of the RNA nitrogenous base moiety at single-stranded nucleotides, and 2) base-agnostic modifiers of the ribose 2' hydroxyl at flexible regions of the sugar-phosphate backbone, designated as Selective 2'-hydroxyl Acylation and Primer Extension (SHAPE) probes[16]. RNA structural information captured by these chemical probes are detected by reverse transcription of the modified RNA, where each modified site becomes encoded as a mutation in the cDNA. Depending on the chemical probe used, the reverse transcription reaction may require the use of specific enzymes and manganese cations supplemented in the reaction buffer. A breakdown of different chemical modifiers commonly used in the Stone Lab and their reverse transcription conditions is shown below (**Table 1**).

Table 1: List of RNA probing reagents currently used by the Stone Lab

(Group II RTs = TGIRT III, Induro RT, Marathon RT)

Probe	Mod. site	RT enzyme	Notes
Dimethyl sulfate (DMS)	Adenine (N1) Cytosine (N3) Guanine (N7)	Group II RTs SS II (w/ Mn^{2+})	Cell permeable. No Mn^{2+} required for RT with group II RTs. Highly toxic, use a fume hood. Recommended for in-cell MaPseq.
2A3	2'-OH (SHAPE)	Group II RTs SS II (w/ Mn^{2+})	Cell permeable. No Mn^{2+} required for RT with group II RTs. Recommended for in-cell or in vitro MaPseq.
1M7	2-OH (SHAPE)	Group II RTs (w/ Mn^{2+}) SS II (w/ Mn^{2+})	Recommended for in vitro MaPseq.

Dimethyl sulfate (DMS) has been used for decades as a robust methylation agent of unpaired RNA adenine (A, position N1) and cytosine (C, position N3) residues[5,17,18]. DMS also methylates Guanine at position N7 (the Hoogstein face) but is harder to detect as the N7 position does not efficiently induce mutations upon reverse transcription. This reagent has excellent cell permeability and is a gold-standard for in-cell MaPseq.

1-Methyl-7-nitroisatoic anhydride (1M7) is a well-validated SHAPE reagent commonly used for in vitro chemical probing studies[19,20]. 1M7 is an information-rich probe, providing signal at all four RNA nucleotides. This reagent is best suited for in vitro experiments due to its poor cell permeability.

A recently developed SHAPE probe, 2-aminopyridne-3-carboxylic acid imidazolide (2A3), was found to have a higher signal-to-noise ratio and cell

permeability than 1M7[21]. Whether the RNA chemical probing field adopts 2A3 over 1M7 will become clear in time as more studies become available.

MaPseq experiments involving protein-RNA binding should make use of both DMS and SHAPE probes. Because proteins can bind RNA at either the nitrogenous base or the phosphate backbone, a chemical probe may report a protein footprint in what could actually be a single-stranded region of RNA[22,23]. Therefore, it is best to employ a combination of chemical probes whenever possible.

Chemical modifications are detected by reverse transcription and encoded as mutations in the cDNA. The Group II Intron family of reverse transcriptases (TGIRT III[24], Marathon RT[25], and Induro RT (New England Biolabs)) are highly processive and have enhanced thermostability compared to traditional reverse transcriptases such as Superscript II. The 'read-through' behavior of modified RNA nucleotides by RT enzymes typically requires 'relaxed fidelity' conditions created by the addition of manganese ions (Mn^{2+}) to the reaction[26]. Notably, relaxed fidelity conditions are not required when reverse transcribing DMS and 2A3-treated RNAs with the Group II reverse transcriptases, yielding a higher signal-to-noise ratio in the resulting chemical probing data. **Protocols 2.1 to 2.3** covers the chemical probing of mammalian cells with DMS and 2A3, while **protocol 2.4** details probing RNA in vitro with 2A3.

3. Library preparation

Protocols: 3.1 and 3.2

Library preparation involves creating a DNA insert (the DNA to be sequenced) that is ~250-300 bp in length, which is flanked by DNA adapter sequences that enable Illumina sequencing. Libraries from different experiments can be pooled and sequenced together and their reads bioinformatically separated through a process called multiplexing. Multiplexed libraries contain a DNA barcode (typically 8 nt) on the 5' and 3' adapter regions of the DNA insert, referred to as the i5 and i7 indices, respectively (**Fig. 1A**). The Stone Lab is equipped with eight Nextera i5 index primers (designated as A primers) and eight Nextera i7 index primers (designated as B primers), enabling 64 unique dual-index combinations for PCR-based library preparation (**Table 2**).

Table 2: List of Nextera multiplex library oligos

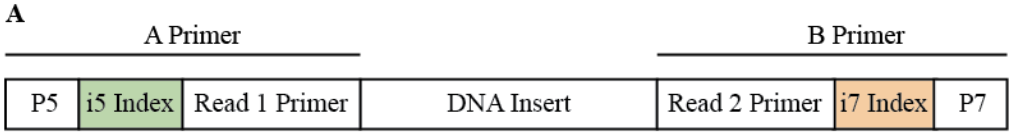
Primer Name	Type	Sequence
Nextera A Primer	i5	AATGATACGGCGACCACCGAGATCTACAC [i5]TCGTCGGCAGCGTCAGATG
A1 Index	i5	TAGATCGC
A2 Index	i5	CTCTGTAT
A3 Index	i5	GATCAGCA
A4 Index	i5	ACAGTATG
A5 Index	i5	GACTGGAG
A6 Index	i5	ACTGCATA
A7 Index	i5	TTGCATGC
A8 Index	i5	CGAATCCT
Nextera B Primer	i7	CAAGCAGAAGACGGCATACGAGAT [i7]GTCTCGTGGGCTCGGAGATGTGTAT

B1 Index	i7	TCGCCTTA
B2 Index	i7	CTAGTACG
B3 Index	i7	TTCTGCCT
B4 Index	i7	GCTCAGGA
B5 Index	i7	AGGAGTCC
B6 Index	i7	CATGCCTA
B7 Index	i7	GTAGAGAG
B8 Index	i7	CAGCCTCG

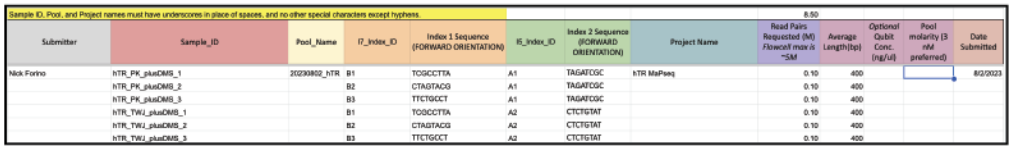
Kit-based multiplexing uses manufacturer-supplied i5 and i7 index primers (cat #E7600S). Refer to the product manual of the indexing primers used (found under Protocols, Manuals, & Usage) on the manufacturer website for index names and sequences.

Before beginning library preparation, choose the multiplexing barcode combination of each of your planned libraries and enter them into the Stone Lab iSeq 100 Sequencing Queue (**Fig. 1B**) (found in Google Drive, SRG Public Share > MaP-Seq > Stone Lab iSeq 100 Queue). Enter the i7 and i5 indices of each sample in the appropriate column in their 5' to 3' orientation (forward orientation). This prevents other researchers in the lab from mistakenly creating libraries with the same index combinations, preventing you from sharing a flow cell together when it's time to sequence. Information about the Nextera primers can be found in the SRG Queue under the 'Lab-specific Indices' tab.

A



B



Sample ID, Pool, and Project names must have underscores in place of spaces, and no other special characters except hyphens.

Submitter	Sample_ID	Pool_Name	i7_Index_ID	Index 1 Sequence (FORWARD ORIENTATION)	i5_Index_ID	Index 2 Sequence (FORWARD ORIENTATION)	Project Name	Read Pairs Requested (M) Flowcell max is ~5M	Average Length (bp)	Optional Qubit Conc. (ng/ul)	Pool molarity (3 nM preferred)	Date Submitted
Nick Forino	hTR_PK_plusDMS_1	20230802_hTR	B1	TCGCCTTA	A1	TAGATCGC	hTR MaPseq	0.10	400			8/2/2023
	hTR_PK_plusDMS_2		B2	CTAGTACG	A1	TAGATCGC		0.10	400			
	hTR_PK_plusDMS_3		B3	TTCTGCCT	A1	TAGATCGC		0.10	400			
	hTR_TWJ_plusDMS_1		B1	TCGCCTTA	A2	CTCTGTAT		0.10	400			
	hTR_TWJ_plusDMS_2		B2	CTAGTACG	A2	CTCTGTAT		0.10	400			
	hTR_TWJ_plusDMS_3		B3	TTCTGCCT	A2	CTCTGTAT		0.10	400			

Figure 1: Anatomy of a sequencing library and the Stone Lab Sequencing Queue.

(A) Schematic detailing the components of a sequencing library prepared with in-house Nextera primers. The P5 and P7 sequences anneal with oligos on the Illumina flow cell. The i7 (Index 1) and i5 (Index 2) barcodes enable sample multiplexing. Read 1 and Read 2 of paired-end sequencing are initiated from priming sites flanking the DNA insert (Read 1 and 2 Primers). The i5 and i7 indices are sequenced independently of Read 1 and Read 2. (B) Screenshot of the Stone Lab Sequencing Queue spreadsheet. Information about the MaP-seq library can be input into the relevant columns. i7 and i5 indices should be entered in their forward orientations (5’ to 3’) to minimize confusion when it is time to prepare the sample sheet for sequencing.

Targeted in-cell hTR library preparation

Protocol 3.1

hTR is targeted using specific primers that contain a partial Illumina sequencing adapter (Nextera) and a 4 nt unique molecular identifier (UMI). First, reverse transcription of the modified hTR proceeds from the reverse primer (with T_m > 60 C), generating a single-stranded cDNA. The sample is then treated with sodium hydroxide and heat to destroy all RNAs, and column purified to yield pure single-stranded hTR cDNA. The second DNA strand is generated using a forward primer that also contains the partial Illumina adapter and 4 nt UMI. This DNA template is used to generate two overlapping amplicons (Fragment 1: Nucleotides 1-286, Fragment 2: 193-451). A two-step PCR strategy yields libraries of each amplicon with dual-index barcodes and complete Illumina sequencing adapters (**Fig. 2**). Bead purification is used throughout the protocol to remove small DNA species such as primer dimers. This process is detailed in **Protocol 3.1**.

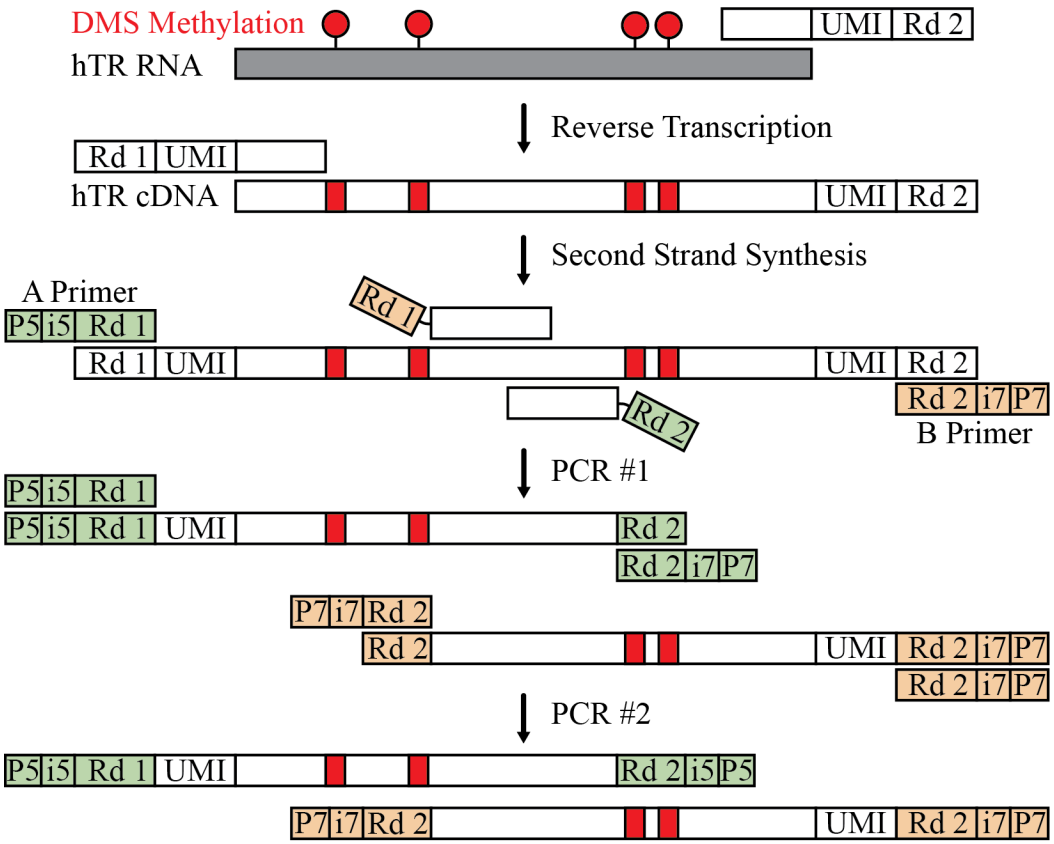


Figure 2: Library preparation from cell-derived hTR.

hTR from total RNA is reverse transcribed using a reverse primer containing a UMI and a partial Nextera i7 adapter. Methylated nucleotides originating from DMS modification (red circles) are encoded into the cDNA as mutations (red rectangles). The second strand is synthesized using a forward primer containing a second UMI and partial Nextera primer. The subsequent two PCR steps create two, overlapping hTR amplicons (orange and blue) in separate tubes. In PCR # 1, primers containing partial Nextera adapters targeting the interior of hTR are used to produce overlapping amplicons. PCR #2 completes the Nextera adapters.

In vitro RNA library preparation

Protocol 3.2

Pure RNAs in vitro are fragmented with heat and divalent cations into pieces 200-300 nt in length before reverse transcribing with random hexamer primers. The downstream second-strand synthesis and library generation steps are completed using kits (**Fig. 3**). Recommended kits are the NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module (cat #E6111S) and the NEBNext Ultra II DNA Library Prep Kit for Illumina (cat #E7645S) with compatible multiplexing oligos (cat #E7335S for single end multiplexing, and cat # E7600S for dual-end multiplexing). Library preparation for in vitro 2A3 MaPseq experiments is outlined in **Protocol 3.2**.

While kits streamline library preparation, they can be cost-prohibitive when many samples need to be processed. Library construction can also be completed without kits using strategies similar to **protocol 3.1**.

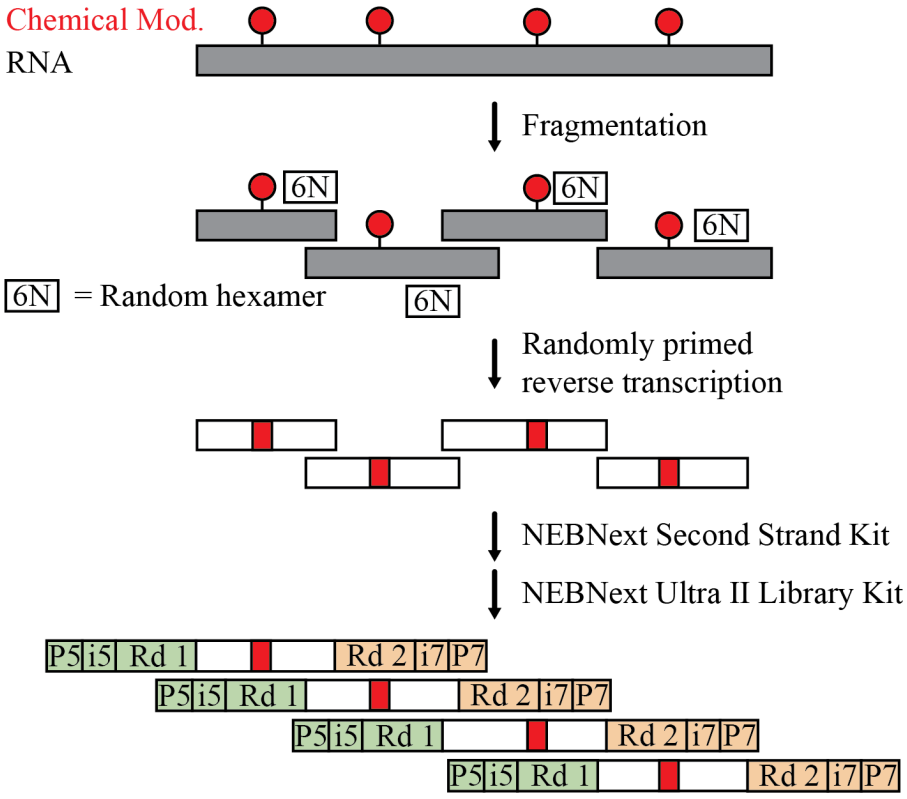


Figure 3: Library preparation from in vitro RNA.

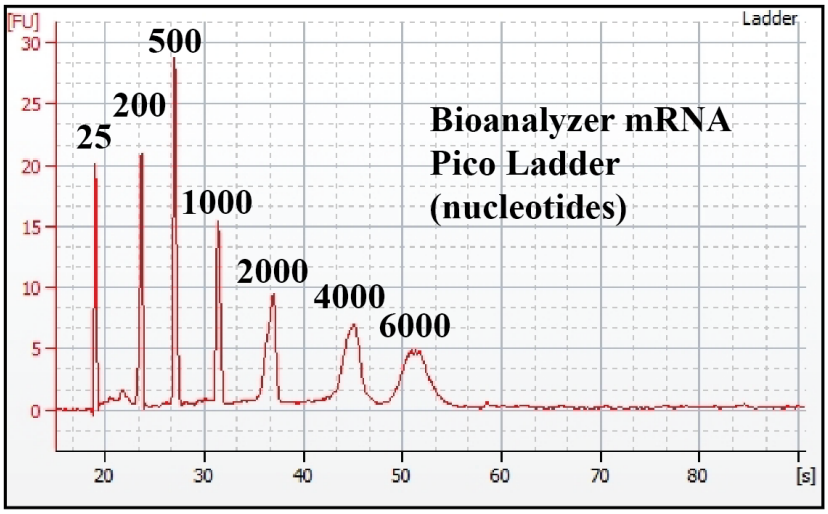
In vitro RNA is heat fragmented, then reverse transcribed using random hexamer primers to encode chemical modifications as mutation in cDNA. The cDNA is treated with two NEB kits to synthesize the second strand, and add sequencing adapters. Multiplexing is enabled through the selction of i5 and i7 indices.

Quality control and library pooling

If the MaPseq library generation method involves RNA fragmentation, the fragmentation time should be optimized for each target to yield mostly 200 nt long fragments (**Fig. 4**) Under-fragmenting the RNA will lead to loss of sequence depth at interior regions of the target RNA, while over-fragmenting will reduce the efficiency of library preparation.

Sample libraries must be quantified for DNA content and size distribution before sequencing. First, determine the DNA present in each sample library by Qubit, using High Sensitivity DNA reagents (cat #Q32851). Calculate the concentration of each sample using the molecular weight of either its exact sequence, or in the case of fragmented libraries, a general DNA molecular weight for fragments of average size in your library (https://nebiocalculator.neb.com/#!/dsdnaamt). Next, quantify the size distribution of each library by Tapestation analysis using the high-sensitivity D1000 reagents (part #5067-5587, #5067-5585, #5067-5603, #5067-5584). Load 1 μL of each sample per lane. Samples with DNA greater than > 0.5 ng/μL typically produce easily identifiable DNA peaks on the Tapestation trace. The appearance of smaller DNA fragments ~100-160 bp may indicate primer dimer contamination, and are unsuitable for sequencing (**Fig. 5A**). Perform another bead purification and re-quantify before sequencing. Unwanted DNA products can always be removed by additional rounds of bead purification provided they are not the dominant DNA species in the library (**Fig. 5B**). Consider gel extraction by PAGE as an alternative size selection strategy if necessary.

A



B

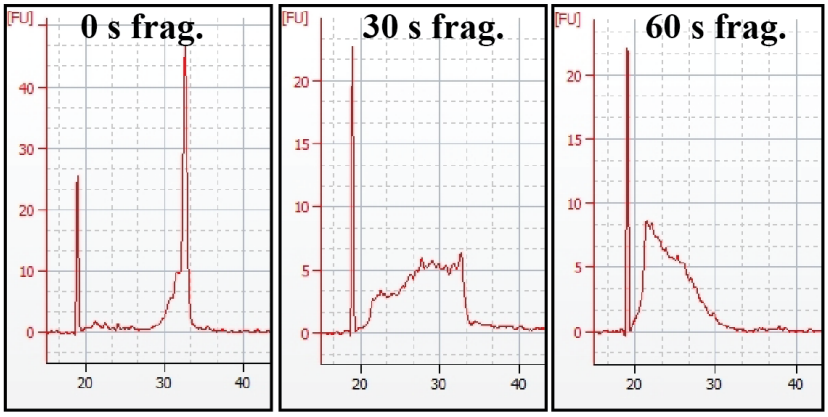
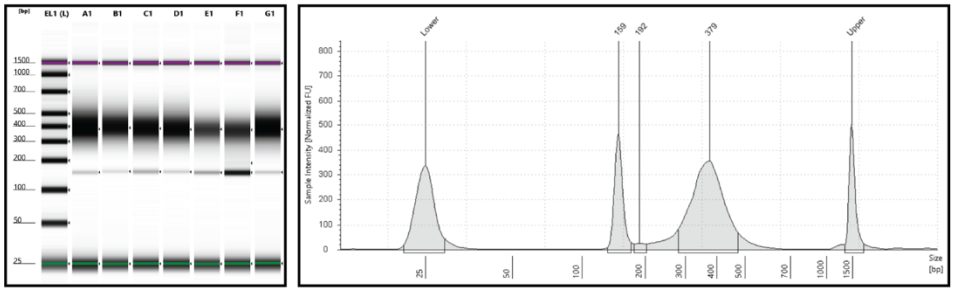


Figure 4: Heat fragmentation of RNA in vitro.

(A) Bioanalyzer trace of the mRNA Pico single stranded RNA ladder, with nucleotide lengths labeled. (B) A roughly 1 kb RNA fragmented at three time points (0, 30, and 60 s).

A



B

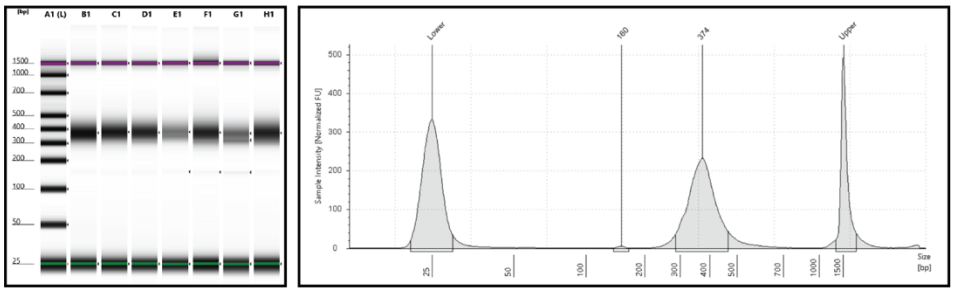


Figure 5: Tapestation analysis of MaP-seq libraries.

(A, Left side) Tapestation 'gel' traces of finished MaP-seq libraries. (A, Right side) Tapestation trace of Sample F1, detecting the majority of the library at ~380 bp and a primer dimer contamination at ~160 bp. (B) Tapestation gel traces of the same samples after an additional bead cleanup step.

Sequencing libraries passing quality control can be pooled together prior to sequencing. Within the pool, the more concentrated samples will receive the most reads. Use the Illumina Pooling Calculator (https://support.illumina.com/help/pooling-calculator/pooling-calculator.htm) to assist in pooling your individual libraries together. MaPseq libraries should be pooled to a total DNA concentration of 100 pM in DNA low-bind tubes. Illumina provides a PhiX control library for increasing the sequence diversity on the flow cell. It is recommended to spike in the PhiX control library at roughly 10% of the pooled library concentration prior to sequencing. (e.g. add 10 μL of 100 pM PhiX to 100 μL of 100 pM pooled library).

Most experiments will be pooled with all individual sample libraries being equimolar, thus receiving roughly the same number of reads. Each iSeq 100 flow cell run will produce ~4-5 M reads on average, so simply divide by the number of samples in the pool to calculate a rough estimate of the number of reads to expect each will receive. Most MaPseq experiments will require at least ~20,000 reads per sample. Greater sequencing depth (> 100,000 reads per sample) is beneficial for experiments involving structural ensemble deconvolution analysis (DREEM[10] or DRACO[11]). In general, a high number of reads for any particular sample is never detrimental, but care should be taken to ensure the sequencing is being partitioned efficiently. If desired, adjust the amounts of individual library components in the pool to increase their yield of reads.

4. Sequencing

Once the MaPseq library is pooled, create a sample sheet and set up a sequencing run on the Illumina iSeq 100 instrument. To create the sample sheet, navigate to the SRG Sequencing Queue (SRG Public Share > MaP-Seq > Stone Lab iSeq 100 Queue) and click the 'iS100_Sample_Sheet_Template' tab. Input the information about your library into the sample sheet, using what was previously entered into the SRG Sequencing Queue. The required fields to fill out are:

1. Experiment name
2. Experiment date
3. Sample_ID
4. Index

NOTE 1: Enter the REVERSE COMPLEMENT of the i7 index entered in the SRG queue.

NOTE2 : For multiplexed libraries constructed with a kit, for example NEBNext, refer to the product manual of the indexing primers used (found under Protocols, Manuals, & Usage) on the manufacturer website for index names and sequences.

5. I7_Index_ID (name of the i7 index)
6. Index2

NOTE: Enter the i5 index in FORWARD ORIENTATION, exactly as it is entered in the SRG queue.

7. I5_Index_ID

Download the filled-out sample sheet from the SRG Queue in comma-separated values (.csv) format to a USB drive.

On the iSeq 100 instrument, press the ‘Sequence’ button, log in with your BaseSpace credentials (make an account if needed), and navigate to the Local Run Manager. Upload the sample sheet to the instrument, which will proofread the sample sheet for any issues that need resolving before the run can be accepted.

To prepare for sequencing, follow the iSeq 100 recommendations for cartridge thawing and flow cell loading (https://support.illumina.com/downloads/iseq-100-system-guide.html).

5. Bioinformatic analysis

At this point, the workflow is entirely computational and makes use of a variety of software packages to transform sequence reads into chemical reactivities, RNA secondary structure predictions, and RNA ensemble predictions. The computational work can be completed on the Stone Lab remote server (recommended) or on a local MacOS machine. It is beneficial to be familiar with basic bash commands and the organizational concept of directories when analyzing MaPseq data.

Connecting to the Stone Lab Remote Server

The Stone Lab Remote Server is a computational environment with software useful for MaPseq analysis. Some of these programs and tools are listed below (**Table 3**). All MaPseq data storage and analyses are performed in the 'mapseq' directory (server path: /mnt/storage2/stonelab/mapseq). The mapseq directory is where sequencing data and all other necessary files are kept for the analysis of MaPseq data.

Table 3: List of software for MaPseq analysis and RNA structure prediction

Program	Purpose
Cyberduck (not on server)	Connect to the Stone Lab remote server, provides a user interface for file manipulation.
FastQC	Quality control metrics for sequence reads.
TrimGalore	Adapter and quality trimming of sequence reads.
Bowtie 2[27]	Read alignment to reference sequence.
Samtools[28]	Manipulations to read alignment files.
PEAR[29]	Merging of paired-end reads into longer single reads.
bamql[30]	Extraction of reads with nucleotide variants from alignment files.
RNA Framework[31]	Software suite for analysis of MaPseq data.
RNAstructure[32]	RNA secondary structure prediction.
ViennaRNA[33]	RNA secondary structure prediction.
DRACO[11]	RNA structural ensemble deconvolution.
DREEM[10]	MaPseq analysis and RNA structural ensemble deconvolution.
Biers[34] (not on server)	High-throughput chemical probing analysis, data-guided RNA structure prediction.
VARNA[35]	RNA secondary structure visualization.

(not on server)	
Adobe Illustrator (not on server)	Visualization and figure generation.

It is recommended to download a remote connection client such as Cyberduck. This program has a user interface similar to the 'Finder' application in MacOS that you can use to navigate the remote server and drag and drop files.

1. If connecting to the server from off campus, activate the UCSC Campus VPN.
2. In Cyberduck, click 'Open Connection', select 'SFTP'.
3. Enter the server, username, and password (ask the Stone Lab for the credentials).
4. Navigate to /mnt/storage2/stonelab/mapseq

To execute commands on the remote server, connect via the MacOS Terminal application.

1. In Terminal, enter:

```
ssh -t  stonelab@128.114.78.210 -p 22 "cd
\/mnt\/storage2\/stonelab\/mapseq && exec \$SHELL -l"
```

2. Enter the password (Ask the Stone Lab for the credentials).
3. Activate the software environment:

```
bash

conda activate rnaf
```

Quality control

First, check the quality of the sequence reads (files with extension .fastq.gz or .fastq). We will first generate quality control reports with FastQC. Check the FastQC manual for explanations on the different metrics. FastQC will throw warnings for sequence content, GC content, sequence duplication, and overrepresented sequences, for MaPseq libraries because they are amplicon libraries. Next, we will trim low quality bases from the read 3' end, and remove adapter sequences (such as Nextera sequences) that may be in the reads using TrimGalore. At this step, you can also add a read length threshold if desired.

1. In the directory containing the read files, generate FastQC reports:

   `fastqc yourfile_R1.fastq.gz yourfile_R2.fastq.gz`

   Tip: You can use the 'tab' key to autofill filenames you've partially typed if it matches a filename in the current directory.

2. Remove low quality bases from read ends, and trim adapter sequences:

   `trim_galore readfile_R1.fastq.gz readfile_R2.fastq.gz`

   Tip: Add `--length X` to the end of the command to remove reads under a length threshold of `X`.

3. Generate another FastQC report if desired.

Read alignment

Before alignment, reads are just strings of nucleotides. Alignment is the process of 'mapping' a read's location to a reference sequence, in this case the reference sequence is the DNA sequence that encodes our target RNA. Reference sequences are FASTA files (with extension .fasta or .fa) that contain a DNA sequence. The FASTA can be created from a simple text file (manually change the file extension to '.fasta'), and must follow the FASTA format conventions which can be found online.

Example reference sequence file (filename: hTRCR45.fasta):

```
>hTR_CR45_domain
CCTGCGGCGGGTCGCCTGCCCAGCCCCCGAACCCCGCCTGGAGGCCGCGGTCGGCCC
GGGGCTTCTCCGGAGGCACCCACTGCCACCGCGAAGAGTTGGGCTCTGTCAGCCGCG
GGTCTCTCGGGGGCGAGGGCGAGGTTCAGGCCTTTCAGGCCGCAGG
```

The reference sequence also needs associated index files to be usable by the alignment software. We will use the Bowtie 2 aligner to build those index files.

1. In the directory containing the reference FASTA file:

   `bowtie2-build reference.fasta reference`

   NOTE: The `reference` argument can simply be the name of the reference sequence file without the file extension (.fasta).

   NOTE 2: Indexing reference sequences only needs to be done once for a new reference sequence. The resulting reference files can be reused for future read

alignments. The reference FASTA file and index files (extension .bt2) together constitute the reference files the alignment software will use in the next step.

2. In the directory containing the quality trimmed read files, reference FASTA, and reference index files:

```
bowtie2 --local --no-unal --no-discordant --no-mixed --phred33 -x reference -1 readfile_R1.fastq.gz -2 readfile_R2.fastq.gz -S output.sam
```

NOTE: Replace `reference` with the name of your reference sequence file (without the .fasta extension).

The arguments supplied in the alignment command tell bowtie to do a local alignment, reject unaligned, discordant, and mixed reads, and to use the Illumina basecall quality standards (Phred scores). This alignment command is sufficient for most targeted MaPseq experiments. Although MaP-seq reads contain mismatches, it would take ~40 mismatches (assuming the rest of the read is perfect) for the software to reject a 150 base read. Thus, one can reasonably expect mapping to proceed normally even when an RNA sample has been 'overmodified' with the chemical probe. When Bowtie 2 is finished, it will print useful mapping metrics, such as percent of reads mapped, to the Terminal window. The output is a sequence alignment map file (SAM, with extension .sam), which we will use for MaPseq analysis.

MaPseq analysis

With a SAM file, we can now begin MaPseq analysis. This guide will cover the use of two different software packages. RNA Framework is a highly modular and customizable set of tools that can be run individually or chained together with scripts into a pipeline. DREEM (Deconvolution of RNA Ensembles by Expectation Maximization) is a standalone program that performs MaPseq analysis, RNA structure predictions, and RNA ensemble analysis. Other publicly available analysis software packages such as DRACO[11], ShapeMapper 2, and DANCE-Mapper are not covered by this guide.

MaPseq analysis with RNA Framework

Mutation counting with rf-count

Documentation: (https://rnaframework-docs.readthedocs.io/en/latest/rf-count/).

The reads within the SAM file contain mutations that originated from DMS or SHAPE probes reacting with the target RNA. In this step, we tally up the mutations at each nucleotide using the rf-count module of RNA Framework. If the MaPseq library was constructed using primers targeting your RNA of interest, the priming sequences will be ignored using a 'mask file'. Instructions on how to construct the mask file are included in the documentation.

1. In the directory containing the SAM file, reference FASTA file, and (if needed) the mask file:

```
rf-count -m -pp -nd -ni input.sam -mf maskfile.txt -f
reference.fasta
```

NOTE: If you plan on performing RNA ensemble analysis with DRACO, add the following arguments to the command to generate a Mutation Map file: `-mm -wl 1`

2. Next, convert the output RNA Count file (extension .rc) into a human-readable text file:

```
rf-rctools view input.rc >> output.txt
```

The contents of the output text file contains four lines of data:

1. The sequence name.
2. The primary sequence.
3. Mutation counts per nucleotide.
4. Sequencing depth per nucleotide.

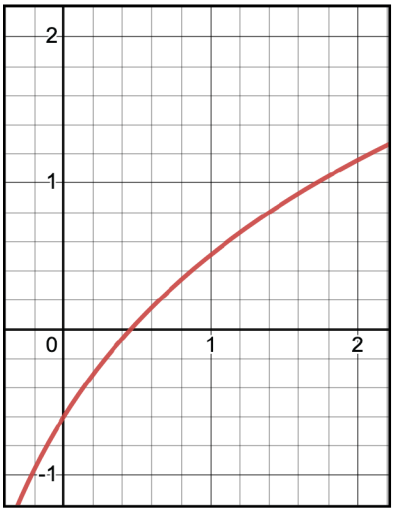
The mutation counts and sequencing depth are useful for generating bar plots in Excel or Prism. Before data normalization, it can be useful to inspect these 'raw' mutation counts and to ascertain whether any parts of the target RNA have low sequence depth. The data can be plotted as a bar graph in Excel or Prism. As previously stated, most MaPseq experiments aim for at least 20,000 reads. If large parts of the target RNA are drastically under-sequenced, consider deeper resequencing or optimizing the library generation.

Chemical reactivities

Documentation: (https://rnaframework-docs.readthedocs.io/en/latest/rf-norm/).

Chemical 'reactivities' of each RNA nucleotide are found using the mutation count and sequencing depth. While chemical reactivity is used to infer the 'single-strandedness' of an RNA nucleotide, it is not a perfect indicator of base-pairing status. Oftentimes chemical reactivity can appear within RNA double-stranded helices. It is useful to think of chemical reactivity of a nucleotide as a 'probability' of being single-stranded rather than a binary signal.

Reactivity values can guide RNA structure prediction algorithms that rely on free-energy minimization[1–4]. The reactivity value of a nucleotide is converted to a 'pseudoenergy' value according based on the mathematical function: ΔG_{SHAPE} = (m)ln(reactivity + 1) + b (**Fig. 6**).



$$\Delta G_{SHAPE} = (m)\ln(\text{reactivity} + 1) + b$$

Figure 6: Pseudoenergy function for chemical reactivity.

Chemical reactivity values from modified RNA nucleotides are converted into 'pseudoenergy' terms based on the above equation. The default values m = 1.6 and b = -0.6 were empirically derived to yield the best accuracy in a study of SHAPE-directed structure prediction of ribosomal RNA (rRNA)[36].

Where (m) and (b) describe the slope and intercept of a pseudoenergy function. These values are usually set to empirically determined default values (m = 1.6 and b = -0.6), but can be manually specified[36]. Depending on a nucleotide's degree of reactivity, it provides either an energetic bonus or penalty to a base pair, thus guiding the prediction software towards structures that are more in agreement with the chemical probing data. Low reactivities yield negative pseudoenergy which add to the stability of a base pair, while higher reactivities yield positive pseudoenergy that disfavor base pairs.

Reactivity calculation with rf-norm

To calculate reactivity values, first the rf-norm module calculates per-nucleotide mutation frequencies from the number of mutations divided by the nucleotide sequencing depth. Most MaPseq analyses will choose to do this by either the Siegfried[37] or Zubradt[8] methods, where mutation frequencies from the untreated control sample (no chemical probe) are considered or ignored, respectively.

Next, the mutation frequencies are normalized. The rf-norm module offers several different normalization methods that define and treat outlier nucleotides differently. It is recommended to try and compare different normalization methods. Consider using the '90% Winsorizing' approach if it is desirable to normalize the outlier nucleotides rather than to ignore them from normalization or remove them from the dataset. At this point in the workflow data from guanine and uracil nucleotides can be removed for DMS experiments if desired. Thresholds for per nucleotide mutation frequency and sequencing depth can be specified here as well, see the documentation for details.

1. In the directory containing the RNA count file of the probed sample (treated), (if desired) RNA Count file of the control experiment (untreated), and the XML_parse.py script:

```
rf-norm -u untreated.RC -t treated -sm X -nm Y -rb Z -
i index.rci
```

NOTE: The `-sm` argument (scoring method) takes an integer `X` from 1 to 4. Most experiments will use the Siegfried (3) and Zubradt (4) methods. The `-nm` method (normalization method) takes an integer `Y` from 1 to 3. The 90% Winsorization method (3) is appropriate for most experiments. The `-rb` argument specifies

reactive bases `Z`. ‘AC’ can be used for DMS experiments, and ‘N’ for SHAPE experiments.

2. Convert the output XML file (extension .xml) into a human-readable text file. Place a copy of the XML_parse.py script (located in /mnt/storage2/stonelab/mapseq/scripts_master/handy_utilities/) in the directory with the output XML file and run:

```
python3 input.XML output
```

Structure prediction with rf-fold and RNAstructure

Structure prediction can be performed with and without the incorporation of chemical reactivities. Performing this analysis both ways can be beneficial, as it highlights the effects of the chemical probing data on the predicted structure. RNA Framework can be used to generate the minimum free energy (MFE) structure of the target RNA with and without reactivity data.

1. In the directory containing the XML file containing reactivity data generated by rf-norm:

```
rf-fold input.xml -m X
```

NOTE: The `-m` takes an integer `X`, where 1 specifies ViennaRNA and 2 specifies RNAstructure.

While the MFE structure is the most energetically favorable secondary structure, it is not guaranteed to be representative of the actual RNA in a cellular or in vitro environment[38–40]. Thus, it is good practice to also study the suboptimal predicted structures that are energetically near the MFE, to grasp the scope of the secondary structures available to the target RNA. This can be accomplished by running RNAstructure directly on the server, using the reactivities generated by rf-norm. See the RNAstructure documentation for the 'Fold' command: (https://rna.urmc.rochester.edu/Text/Fold.html).

1. Modify the text file containing reactivity data to comply with SHAPE Data File Format as specified by the RNAstructure documentation: (https://rna.urmc.rochester.edu/Text/File_Formats.html#SHAPE).
2. In the directory containing the reference sequence (.fasta) and SHAPE file (.txt or .shape):

For DMS:

```
Fold input.fasta -dms dms_reactivities.txt output.ct
```

For SHAPE:

```
Fold input.fasta -sh shape_reactivities.txt output.ct
```

3. Draw secondary structure postscript images from the output CT (connectivity table) file. In the directory containing the CT file:

```
draw input.ct -s reactivities.txt output.ps
```

Ensemble analysis of MaPseq Data

For a given primary RNA sequence, each molecule is not guaranteed to fold into identical structures. In fact, one RNA sequence can adopt a diverse ensemble of structures that coexist across different molecules of the same sequence[41–44]. Historically, this folding behavior was thought to be an evolutionary drawback of RNA, and cells evolved RNA chaperoning mechanisms to ensure they didn't misfold. Today, we understand that RNA's ability to adopt a wide range of conformations enables hierarchical RNP assembly, ligand binding, and catalysis[45–48].

Recently developed bioinformatic approaches have capitalized on the single-molecule nature of the MaPseq method[12]. Because each RNA structure is unique in an ensemble, the encoded cDNA mutation patterns are also unique in the sequencing library. Tools like DREEM, DRACO, and DANCE-Mapper classify or 'cluster' reads based on their qualities of similarity, such as co-occurring mutation patterns that appear together (**Fig. 7**).

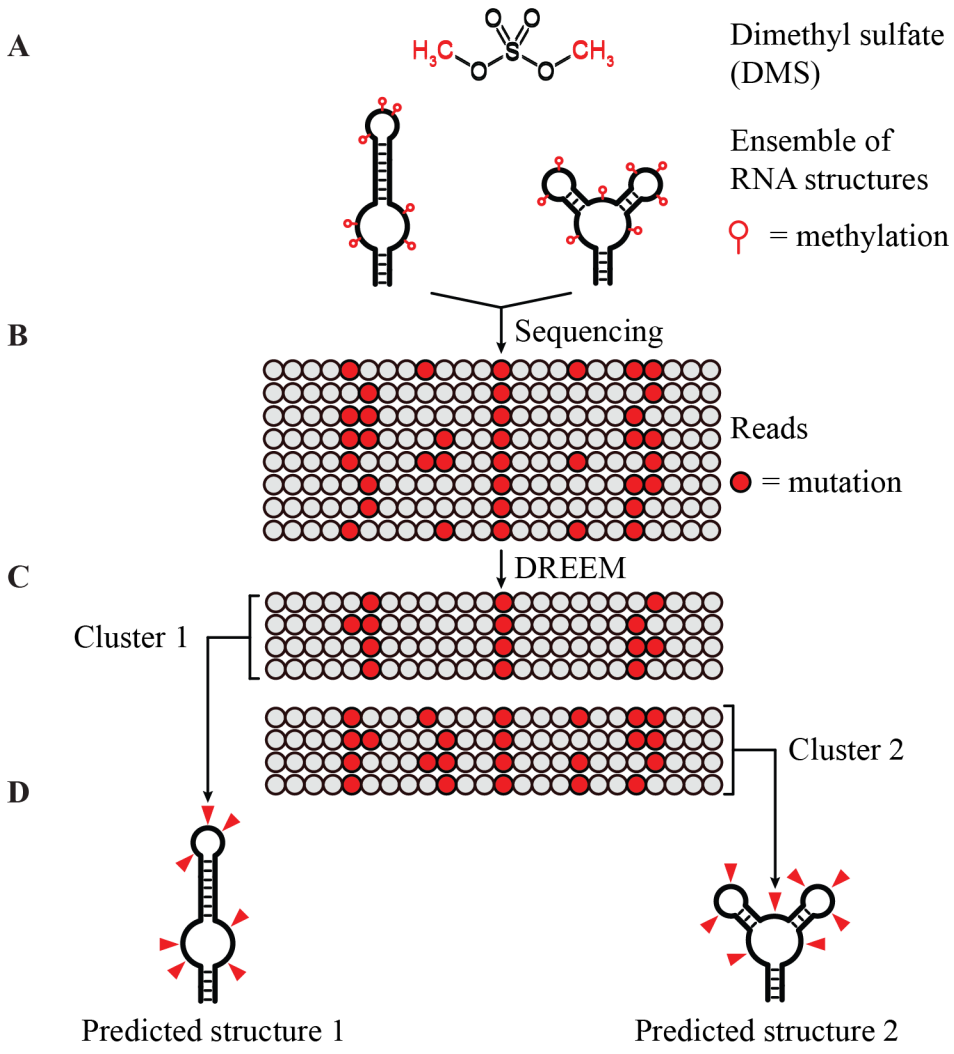


Figure 7: Ensemble deconvolution analysis of MaP-seq data.

(A) DMS (or any probing reagent) covalently modifies single stranded or flexible nucleotides of an RNA existing as a structural ensemble of two conformations. (B) The chemical signatures of both conformations in the ensemble are encoded as mutations into cDNA that are sequenced. (C) Ensemble deconvolution tools (such as DREEM) identify patterns of 'co-mutation', where certain nucleotide positions are more often mutated together, and cluster the reads into groups based on their similarities. (D) Reactivity values from the predicted clusters are used to guide RNA structure prediction software. The end result is that individual RNA conformations of the structural ensemble are deconvolved.

MaPseq and Ensemble analysis with DREEM

DREEM is a self-contained program that performs MaPseq and ensemble analyses from start to finish with one command. In its current version, running DREEM is simpler and less technically involved, but there are no ways to manually change analysis parameters without modifying the source code. Future releases of DREEM will likely add increased functionality.

DREEM converts each read into a 'bit vector' where mutated positions are changed to the integer 1, and wild type positions are changed to 0. This enables the algorithm to perform clustering by Expectation Maximization (EM).

First, convert the SAM file obtained from Bowtie 2 alignment to BAM (Binary Alignment Map). Then, rename the BAM file as such: The name of the output BAM file must contain exactly one underscore followed by the name of the reference sequence used for alignment. For example, the BAM filename 'HeLaCR45_hTR.bam' and reference sequence 'hTR.fasta' are acceptable.

1. Convert the SAM alignment file of the sample to BAM. In the directory containing the SAM alignment file:

```
samtools view -S -b input.sam > output_REF.bam
```

2. Run the batch_DREEM.sh script, located in: /mnt/storage2/stonelab/mapseq/scripts_master/DREEM. In the directory containing the BAM and reference FASTA:

```
bash
/mnt/storage2/stonelab/mapseq/scripts_master/DREEM/batch_DREEM.sh
```

3. When prompted, enter the name of the reference sequence, and the positions of the starting and ending nucleotide of the target RNA you desire analyzed by DREEM.

If the run is successful, DREEM will generate four directories:

1. EM_Clustering
2. BitVector_Plots
3. BitVector_Files
4. Mapping_Files

DREEM first assumes all the data forms one cluster (one RNA structure), then repeats the analysis with two and (if probable) three clusters.

The 'EM_Clustering' directory contains the data from the one cluster analysis (K1) and two or three cluster analyses (K2 and K3). In each of these directories is stored reactivity data, the predicted proportion of each cluster, the three lowest energy predicted structures in dot-bracket format and secondary structure image formats. BitVector_Plots contains HTML plots (viewable in an internet browser) of DMS reactivities, mutation rate per read, and sequence coverage. The bit vectors and

alignment files used in the analysis are found in BitVector_Files and Mapping_Files, respectively.

It is important to note that the 'reactivity' values DREEM returns in multi cluster analyses (K2 and K3) are actually predicted reactivities. Because predicted reactivities are not directly obtained from the raw data (sequencing reads), replicate experiments may not be plotted with error bars. When assessing the reproducibility of DREEM runs, it is important to have biological replicates and manually ascertain if their predicted reactivity patterns and predicted cluster abundances agree.

Bioinformatic purification of mutant RNA MaPseq reads

Perturbing the structure of an RNA through mutagenesis is an effective method for uncovering structure-function relationships in vivo. The primary challenge of this experiment is how to insert the mutant RNA in the cell. CRISPR modification of the endogenous gene is ideal as the RNA mutant will retain its normal transcription level, but it can be time consuming to ensure both loci of the gene are modified in the same manner. Transfection of an expression plasmid harboring the mutant RNA is a simple alternative. Importantly, in transfection experiments, or where the target RNA is a heterozygous mutant, the wild type RNA will be present alongside the mutant and therefore be present in the MaPseq data.

MaPseq reads originating from the wild type or mutant RNA can be bioinformatically filtered prior to DREEM analysis using the software tool 'bamql' (BAM query language). In this way, a DREEM analysis can be performed on only the

wild type or mutant RNA without contaminating sequence reads. This is accomplished by specifying the desired nucleotides at particular positions in the read or bamql to extract. The important caveat of this approach is that the specified nucleotides may be reactive to chemical probing reagents, and thus will be mutated in the resulting sequence read. While bamql filtering enables the 'purification' of reads coming from either the endogenous RNA or transfected mutant, it inherently selects for reads originating from RNAs that were not modified at any of the specified nucleotides.

1. Unzip the read files (.fastq.gz). In the directory containing the read files and reference sequence (.fasta):

```
gunzip -k readfile_R1.fastq.gz
gunzip -k readfile_R2.fastq.gz
```

2. Merge the paired-end reads into longer single reads using the Paired End Read Merger (PEAR):

```
pear -f readfile_R1.fastq -r readfile_R2.fastq -o
mergedreads.fastq
```

3. Compress the merged read file:

```
gzip mergedreads.fastq
```

4. Align the reads to the reference sequence of the mutant RNA with bowtie2 (see Read alignment section).
5. Convert the SAM to BAM:

```
samtools view -S -b input.sam > output.bam
```

6. Extract reads with specific nucleotides at specified positions using bamql. Use the name of the reference sequence (without .fasta) for `reference_name`:

```
bamql -f input.bam 'chr(reference_name) &
nt_exact(position,nucleotide)' -o output_REF.bam
```

NOTE: If multiple nucleotides are to be specified, modify the command as such:

```
bamql -f input.bam 'chr(reference_name) &
nt_exact(position1,nucleotide) &
nt_exact(position2,nucleotide)' -o output_REF.bam
```

The output BAM file contains only reads with the specified nucleotide at the specified position. Ensure the name of the BAM file is compatible with DREEM.

7. Run the batch_DREEM.sh script, located in: /mnt/storage2/stonelab/mapseq/scripts_master/DREEM. In the directory containing the BAM and reference FASTA:

```
bash
/mnt/storage2/stonelab/mapseq/scripts_master/DREEM/bat
ch_PEAR_DREEM.sh
```

6. Graphical visualization of RNA structures

Although several tools for drawing RNA secondary structures exist, none are as user friendly as VARNA (Visualization Applet for RNA secondary structure). This tool also has a MATLAB wraparound script called Biers (Basic Inference Engine for RNA Structure) that enables the quick generation of RNA secondary structures with overlaid chemical reactivity values. From VARNA, secondary structures can be saved as SVG (scalable vector graphics) images and imported into Adobe Illustrator for further modification.

Scripts useful for graphical visualization of RNA secondary structures are found in /mnt/storage2/stonelab/mapseq/scripts_master/MATLAB.

1. Download and install VARNA (https://varna.lisn.upsaclay.fr/index.php?lang=en&page=downloads&css=varna), MATLAB, and Biers (https://ribokit.github.io/Biers/install/) if not already installed. The Stone Lab office Mac desktop already has all three software tools installed.
2. Launch MATLAB from the Terminal app by typing: `matlab`

   It is critical to launch MATLAB from the terminal to enable the launching of VARNA with Biers scripts.
3. To generate a VARNA secondary structure, use the 'Biers_DREEM.m' or 'Biers_RNAFramework.' MATLAB scripts depending on which program your reactivity file came from.

Below is the ‘Biers_DREEM.m’ script. It requires a reference sequence of your RNA in FASTA format (do not use the DNA sequence, ensure it is in RNA) and a reactivity file from DREEM (a .txt file with ‘const’ in the file name). The script also allows the user to run Biers non-parametric bootstrapping to test the confidence of predicted helices of the RNAstructure MFE.

Example: Biers_DREEM.m script:

```
% ======== REFERENCE INPUTS ========

% Input reference sequence of RNA (.fasta file)
% First line of file: '>yourRNA'
% Second line of file: RNA sequence
input = importdata('input.fasta')
name = char(input(1))
name = extractAfter(name, '>')
sequence = char(input(2))
% Start and end nucleotide positions of reference RNA you want to plot in VARNA
sequence = sequence(START:END)

% ======== DREEM INPUTS ========

% DREEM DMS reactivity file input, converts G & U nucs to NaN.
data = importdata('reactivity_file.txt')
data = data(:,2)
% Convert G and U nucleotides to 'NaN'
data(data==-999.0000)=nan

% DREEM dot-bracket file with top three MFEs input
struc_input = importdata('input.dot')
struct1 = char(struc_input(3))
struct2 = char(struc_input(5))
struct3 = char(struc_input(7))

% ======== VARNA ========
```

```
% Make an RNA secondary structure drawing with VARNA
varna_fig('Name', sequence, structure, data, 2, 0)

% ======== BIERS BOOTSTRAPPING ========

% Run Biers on a DREEM dataset and open a VARNA representation.
%d_1D = data
%[structure_pred, bpp_pred] = rna_structure(sequence, d_1D, [], [], [], 100, 0)
%output_varna('Name', sequence, structure_pred, [], [],[], [], [], d_1D, bpp_pred)

% Print the BPP plot
%print_bpp_Z(bpp_pred,[],[],[])
```

MaPseq Laboratory Protocols

1.1 - Starting a culture from a thawed aliquot

Sterilize everything with 70% ethanol.

1. Thaw the cryovial of cells in a 37 C water bath (or your hands, not ideal). As soon as the last of the frozen cell pellet disappears into solution, pipette the cells into a 15 mL conical using a 2 mL serological pipette, not a P1000 pipette tip.
2. Very slowly (one drop at a time) add prewarmed media to the cells. After about 5 mL or so you can pick up the pace to a final volume of 10 mL.
3. Centrifuge at 200 x g for 5 min.
4. Aspirate the media, then resuspend with 5 mL fresh media and transfer to T25 flask.
5. Place T25 in incubator (37 C and 5% CO2) overnight.
6. The next day, remove dead cells by changing the media and check confluency. If ~80% confluent split out to T75s.

1.2 - Splitting Cells

1. Sterilize the biosafety cabinet and all tools/disposables with 70% ethanol.
2. Observe cells and note confluency under the microscope.
3. Aspirate media from cell flask (turn the T75 flask upwards and aspirate media that collects in the corner).
4. Was with 7 mL PBS twice (add PBS and gently rock the flask).

5. Add 0.75mL Trypsin, incubate at 37C ~2-3min at 37C, check to see cells lifting off surface. Can also tap the flask to dislodge cells.
6. Using 10 mL of fresh media, dislodge cells from the flask and pipette up and down about ten times. This creates a single cell suspension, devoid of cell clumps/
7. Add fresh media (containing no cells) to your new flask(s) and the desired amount of cells depending on your intended split ratio. 10-12 mL total media final volume.
8. Rock the flask gently in each direction (North, South, East, West) to ensure even distribution of cells over the flask surface.
9. Return plate/flask to incubator, 37C 5% CO2.
10. Dispose of any excess cells in your waste receptacle, add bleach to it for proper disposal.

Example splits:

1:5 – 8 mL media + 2 mL of cells

1:10 – 9 mL media + 1 mL of cells

HEK and HeLa cells grow very fast. A 1:10 or even 1:15 split will need splitting again in 3-4 days.

1.3 - Counting cells

(See https://www.stemcell.com/how-to-count-cells-with-a-hemocytometer.html)

1. Trypsinize cells and make a 10 mL suspension.
2. Sterilize the surface and coverslip of hemocytometer with ethanol.
3. Pipette 10 μL of cells into the hemocytometer.
4. Using a microscope, count the number of cells in each quadrant.
5. Average those numbers together and multiply by 10,000. The result is your cells per mL.

1.4 - Freezing Cells

1. 1 confluent T75 flask yields 3 cryovials of cells. Grow up the appropriate number of flasks depending on how many vials you want to freeze down. Before you start, have empty cryovials sitting on ice.
1. Trypsinize cells and spin down at 200 x g for 5 min.
2. Aspirate the media and resuspend with 3 mL freezing media (1 mL per vial).
3. Add 1 mL cells in freezing media per cryovial, keep on ice.
4. Place cells in a Mr. Frosty and put in -80C overnight. Alternatively, put them in a small Styrofoam container and place that in the -80C overnight.
5. The next day, transfer the cells to liquid nitrogen dewar (2^{nd} floor Biomed building) for long term storage.

1.5 - Preparing cells for DMS probing

In T75 flasks

Seed the flask with ~2 million cells and grow until 80-90% confluent. Overgrown cells will be stressed and are thus not ideal for DMS probing.

In 6-well plates

Seed each well with 300,000 cells in a 3 mL volume of media. Cells should ideally be 80-90% confluent, and should be ready for probing within 24-48 hours.

1.6 - Cell growth media formulation:

Ensure everything going into the biosafety cabinet is sterilized with ethanol.

1. 500 mL DMEM
2. 1 aliquot of GlutaMax (5 mL)
3. 1 aliquot of Penicillin/Streptomycin (P/S) (5 mL)
4. 50 mL fetal bovine serum (FBS)

In the biosafety cabinet, all ingredients can be directly added to a bottle of DMEM.

Invert three bottle several times to ensure mixing.

Store at 4 C for up to ~1 month with your initials on the bottle.

1.7 - Protocol for in vitro RNA synthesis (Primer assembly method)

1. Input the sequence of your RNA target or target subregion into the Primerize tool (primerize.stanford.edu). It will automatically add an upstream T7 promoter (with terminal guanine) for in vitro transcription.
2. Perform primer assembly reactions according to the primer assembly protocol.

NOTE: If primer assembly yields unwanted products, consider splitting the reaction into two separate PCRs, using primers intermediate from the absolute first and absolute last as new terminal primers. Those two PCR products can be used in a final PCR reaction to create one contiguous amplicon.

3. Purify the DNA product with a DNA cleanup kit.
4. Perform in vitro transcription (1 mL volume):
    a. 8 pmol DNA template
    b. 100 µL 10X RNAP Buffer
    c. 62.5 µL rNTP mix
    d. 22 µL 1M $MgCl_2$
    e. 90 µL 1M DTT
    f. H_2O to 1 mL
    g. 1 µL RNAse inhibitor
5. Incubate in vitro transcription reaction for 3h (or overnight) at 37C.
6. Spin down to remove inorganic pyrophosphate.
7. Add 5 µL DNAse and incubate 37C for 1 hr.
8. Split the sample into two 500 µL volumes.

9. Add 1 volume phenol (with pH appropriate for RNA, ~4.5) to each sample and vortex for 30 s.
10. Spin at max speed for 5 min and extract aqueous layer.
11. Add an equal volume of chloroform and vortex 30 s.
12. Spin at max speed for 5 min and extract aqueous layer.
13. Add 1/10 volume of 3M Sodium Acetate pH 5.2 and 3X volume cold ethanol. If needed, add glycogen as a coprecipitant.
14. Precipitate at -80C at least two hours or overnight.
15. Spin at max speed, 4C for 30 min.
16. Perform two ethanol washes with 70% ethanol.
17. Air dry the pellet briefly, avoiding overdrying.
18. Resuspend in desired amount of 1X formamide or urea loading buffer.
19. Gel purify by denaturing PAGE, using a gel percentage appropriate for your RNA size. Use a 40% 19:1 Acrylamide/Bisacrylamide solution.
    a. 5% for RNAs > 200 nt, 7% for RNAs < 200 nt.
20. Identify RNA bands by brief UV shadowing. Cut the gel pieces with a clean razor.
21. Extract RNA by soaking crushed gel slices in TE or 300 mM Sodium acetate pH 5.2.
22. Separate the RNA-containing supernatant from gel using a Costar SpinX column.
23. Add 3X volumes cold ethanol, 1 µL glycogen, and precipitate.

24. Wash the pellet twice with 70% ethanol.
25. Resuspend pellet in desired buffer and save aliquots.
26. Run a diagnostic PAGE to check RNA integrity.

2.1 - DMS probing of mammalian cells (T75 flasks)

NOTE: Work in a fume hood to prevent exposing yourself to DMS!

Separate solid and liquid waste into containers. DMS waste containers should be airtight, see DMS probing SOP.

Starting Materials:

Flasks containing cells of interest (80-90% confluent)

DMS-media (2%) *Wait to make this solution immediately before treating the cells*

1. 10 mL prewarmed DMEM (or cell-appropriate media)
2. 200 µL DMS (be careful, only work in the fume hood with DMS)

   Multiply by the number of flasks you are treating plus some extra for pipette error.

Quench Solution for DMS media (on ice)

1. 7 mL cold PBS
2. 3 mL BME (30% final)

   Multiply by the number of flasks you are treating plus some extra for pipette error. Store on ice.

PBS (on ice)

Trizol

Chloroform

Glycogen or linear acrylamide

Ethanol (100% and 70%)

Isopropanol

Small incubator set to 37C inside the fume hood

Tabletop centrifuge pre-chilled to 4 C

Protocol

Work with +DMS and -DMS samples separately. Perform the protocol for the +DMS samples first, stopping at Step 4, then work on the -DMS samples. Although -DMS samples do not get treated with DMS, they still get quenched.

1. Remove media from cell flask.
2. Create the DMS to media (skip for the -DMS sample). Add DMS to your aliquot of media and vortex until the DMS droplets incorporate. Quickly add 10 mL of DMS media to the recipient flask. Do this gently so as not to lift cells off the surface. (Can add to the bottom of the T75, then lay the flask horizontal gently). Skip this step for -DMS samples.
3. Incubate at 37 C for 5 minutes. Mix Quench solution while the flask is incubating.
4. Remove DMS media gently (to prevent dislodging cells) and dispose of in the appropriate liquid waste container.

5. Add 10 mL Quench solution and resuspend cells by pipetting or scraping. Transfer to 15 mL Falcon, place on ice.
6. Spin cells at 1000 x g for 5 min. Remove supernatant, being careful not to take cells, and add 10 mL of cold PBS to wash. Resuspend cells but limit pipetting.
7. Spin cells again. Remove supernatant and add 1 mL cold Trizol, mix by pipetting (1 mL serological is useful). The mixture should be goopy and pink. Incubate 5 min at RT.
8. Transfer to 1 mL Eppendorf, add 200 µL L chloroform. Vortex 20 sec and incubate 3 min RT.
9. Spin tube 12K x g, 15 min, 4 C. Extract the top aqueous phase being careful not to take cell debris, be conservative.
10. Add 1 µL glycogen or linear acrylamide and 500uL isopropanol.
11. Spin 12K x g, 10 min, 4 C. Remove the supernatant, you should have a pellet at this point. If no pellet appears, add an additional 1 µL of glycogen and spin for another 10 min.
12. Perform two washes with 700 µL fresh 70% EtOH, spinning 5 min at 7.5K x g, 4 C. If the sample still smells like BME, do more washes until it doesn't.
13. Air dry the pellet (don't overdry) and resuspend in 30 µL H_2O.
14. Nanodrop to get quality. 260/280 should be > 1.9-2.0. 260/230 should be >2.0.

2.2 - DMS probing of mammalian cells (6-well plates)

NOTE: Work in a fume hood to prevent exposing yourself to DMS!

Starting Materials:

Flasks containing cells of interest (80-90% confluent)

DMS-media (2%) *Wait to make this solution immediately before treating the cells*

1. 2 mL prewarmed DMEM (or cell-appropriate media)
2. 40 μL DMS (be careful, only work in the fume hood with DMS)

   Multiply by the number of wells you are treating plus some extra for pipette error.

Quench Solution for DMS media (on ice)

1. 2.1 mL cold PBS
2. 0.9 mL BME (30% final)

   Multiply by the number of flasks you are treating plus some extra for pipette error. Store on ice.

PBS (on ice)

Trizol

Chloroform

Glycogen or linear acrylamide

Ethanol (100% and 70%)

Isopropanol

Small incubator set to 37C inside the fume hood

Tabletop centrifuge pre-chilled to 4 C

Protocol

Work with +/- DMS plates separately. Perform the protocol for the +DMS samples first, stopping at Step 4, then work on the -DMS samples. Although -DMS samples do not get treated with DMS, they still get quenched.

1. Remove media from all wells. Add DMS to media (skip for the -DMS sample), vortex until DMS droplets appear to incorporate into media. Quickly add 2 mL of DMS media to each well. Do this gently so as not to lift cells off the surface. (Can add to the corner of each well, then lay the plate flat).
2. Incubate at 37 C for 5 minutes. Mix Quench solution while this is incubating.
3. Remove DMS media gently (to prevent dislodging cells) and dispose of properly.
4. Add 3 mL Quench solution and resuspend cells by pipetting or scraping. Transfer to 15 mL Falcon, place on ice.
5. Spin cells at 1000 x g for 5 min. Remove supernatant, being careful not to take cells, and add 10 mL of cold PBS to wash. Resuspend cells but limit pipetting.
6. Follow steps 6-13 of the T75 Flask DMS probing protocol (above).

2.3 - 2A3 SHAPE probing of mammalian cells

Adapted from Marinus et al[11].

Starting Materials:

Flasks or plates containing cells of interest (80-90% confluent)

NOTE: Carefully record and control the number of cells being probed. The optimal cell amount in a 2A3 probing experiment has not been explored.

1M 2A3 (50 μL per reaction)

1. (If using homemade 2A3) Before use, thaw and centrifuge at max speed for 2 min.
2. Take supernatant, and use for probing. Remaining powder is unreacted starting material.

1M DTT (500 μL per reaction)

PBS (on ice)

Trizol

Chloroform

Glycogen or linear acrylamide

Ethanol (100% and 70%)

Isopropanol

Small incubator set to 37C inside the fume hood

Tabletop centrifuge pre-chilled to 4 C

Protocol

1. Remove media from flask or well, wash with PBS.

2. Add trypsin (0.75 mL for T75, 0.5 mL per well of 6-well). Incubate cells for 1 min at 37 C.
3. Dissociate cells from plate with media and place in 15 mL conical.
4. Count the cells and determine optimal amount for experiment (see NOTE above).
5. Pellet the cells by spinning 180 x g for 1 min. Resuspend cells in 450 µL of PBS (pH 7.4) and transfer to a 1.7 mL Eppendorf.
6. Add 50 µL of 1M 2A3 (100 mM final conc.), gently mix by flicking the tube. I
7. Incubate the tube at 37 C for 15min with moderate shaking.
8. Quench the modification reaction by adding 500 µL 1M DTT.
9. Pellet cells by spinning 1,000 x g for 5 min at 4 C.
10. Remove supernatant and resuspend pellet in 1 mL cold Trizol.
11. Incubate 5 min at 56 C to break cell debris.
12. Add 200 µL chloroform. Parafilm tube and vortex 20 s. Incubate 3 min RT.
13. Follow steps 9-14 of T75 flask DMS probing protocol (see above).

2.4 - 2A3 SHAPE probing of in vitro RNA

Starting materials:

PAGE-purified RNA of interest

0.5 M Na-HEPES pH 8.0

1M DTT

10 mM $MgCl_2$

Glycogen or linear acrylamide

1M 2A3 (50 µL per reaction) in DMSO

1. (If using homemade 2A3) Before use, thaw and centrifuge at max speed for 2 min.
2. Take supernatant, and use for probing. Remaining powder is unreacted starting material.

RNA refolding

1. Make a solution containing 1.2 pmol of the RNA of interest in 65 mM HEPES pH8, with total volume 13.5 µL.
2. Heat the RNA at 95 C for 3 min.
3. Once cooled to RT, add 1.5 µL of 10 mM $MgCl_2$ (1 mM final).
4. Incubate at RT for 10 min.

2A3 probing reaction

1. Add 1.67 µL of 1M 2A3 (100 mM final).

NOTE: Consider doing a titration series to find the optimal probe concentration. 2A3 can be diluted in DMSO.

2. Incubate at 37 C for 2 min.
3. Quench the reaction with 20 µL of 1 M DTT.
4. Incubate the reaction for 5 min at RT.
5. Adjust the volume of the reaction to 100 µL with H_2O.
6. Clean the RNA using an RNA Clean and Concentrator-5 column.
7. Elute the purified RNA with 20 µL of H_2O heated to 60 C. Incubate the column at RT for 3 min prior to centrifugation.

3.1 - Protocol for Targeted hTR Reverse Transcription from total RNA

Starting materials

Total RNA derived from DMS probed or control cells

Turbo DNAse and 10X DNAse buffer

Reverse transcription primer NMF 109

Second strand synthesis primer NMF 110

TGIRT-III reverse transcriptase

5X M-MLV Buffer

0.1 M DTT

RNAsin Plus ribonuclease inhibitor

10 mM dNTPs mix (10 mM each)

5 M NaOH

5 M HCl

5X GC Buffer

Phusion polymerase

Ampure XP magnetic beads

Nextera A and Nextera B index primers

DNAse digestion

1. Digest 10 μg of total RNA with Turbo DNAse, following the manufacturer's protocol for a 'robust' DNAse treatment.
2. Purify the RNA with an RNA Clean and Concentrator-5 column. Elute with 12 μL H_2O heated to 60 C. Incubate the column ar RT for 3 min prior to centrifugation.

Reverse transcription with TGIRT-III

1. In a PCR tube, anneal reverse transcription primer to hTR.
    a. 5 μg DNAse-treated total RNA
    b. 1 μL of 5 μM NMF 109 (5 pmol)
    c. Adjust volume to 11 μL with H_2O
    d. In thermocycler, incubate at 75 C for 3 min, then 35 C for 15 min.
2. In the same tube, set up the reverse transcription reaction.
    a. 4 μL of 5X M-MLV Buffer
    b. 1 μL of 0.1 M DTT
    c. 1 μL of RNAsin plus

d. 1 µL of TGIRT-III

3. Incubate the reaction at RT for 30 min.
4. Add 2 µL of 10 mM dNTPs.
5. Incubate the reaction at 60 C for 2.5 hr.

Second Strand Synthesis

1. Add 1 µL of 5M NaOH.
2. Incubate at 95C for 3min.
3. Add 2.5 µL of 2M HCl
4. Purify the reaction with a DNA Clean and Concentrator-5 column.

NOTE: Use an 8:1 Binding Buffer to sample volume ratio.

5. Elute with 21 µL of H_2O heated to 60 C. Incubate the column at RT for 3 min prior to centrifugation.
6. In a new PCR tube, set up the second-strand synthesis reaction.
    a. 20 µL single stranded cDNA
    b. 10 µL 5X GC Buffer
    c. 1 µL 10 mM dNTPs
    d. 2.5 µL Forward Primer NMF 110
    e. 16 µL H_2O
    f. 0.5 µL Phusion polymerase
7. Run the second strand synthesis program on the thermocycler (hTR SSS).
    a. 1) 98 C for 2 min, 2) 60 C for 2 min, 3) 72 C for 10 min.

8. Clean the sample with Ampure XP beads, using a 0.8X bead to sample ratio.
    a. Add 40 µL of room temperature, resuspended beads, mix the sample.
    b. Incubate at RT for 3 min.
    c. Place on the magnetic rack and incubate for 3 min, or until all beads pellet on the side of the tube and the sample becomes clear.
    d. Remove the supernatant.
    e. Wash twice with 200 µL freshly prepared 70% ethanol.
    f. Allow the bead pellet to briefly air dry. The pellet should remain 'glossy'. Overdried pellets appear 'cracked'.
    g. Remove the tube from the magnet and resuspend the bead pellet with 22 µL of H_2O.
    h. Incubate for 10 min, and place back on the magnet.
    i. Carefully take 20 µL of DNA-containing supernatant, avoiding contamination with magnetic beads.

PCR #1 ‘Partial Adapting PCR’

This reaction creates two partially overlapping hTR amplicons. Each amplicon receives one complete Illumina adapter with multiplexing barcode on one side of the molecule.

hTR Fragment 1: Nucleotides 1-286

- NMF 113 (Forward primer) and Nextera B# (Reverse primer).
    - Where # is the multiplexing number of the B primer.

hTR Fragment 2: Nucleotides 193-451

- Nextera A# (Forward primer) and NMF 114 (Reverse primer).
    - Where # is the multiplexing number of the A primer.

1. In a PCR tube, set up PCR #1.
    a. 4 µL of cDNA
    b. 10 µL of 5X GC Buffer
    c. 1 µL of 10 mM dNTPs
    d. 2.5 µL of 10 µM Forward primer (NMF 113 or Nextera A#)
    e. 2.5 µL of 10 µM Reverse primer (NMF 114 or Nextera B#)
    f. 29.5 µL of H_2O
    g. 0.5 µL of Phusion polymerase
2. Place in the thermocycler, and Run the PCR #1 program.

a. 1) 98 C for 3 min, 2) 98C for 30 s, 3) 66 C for 30 s, 4) 72 C for 30 s, 5) Go to step 2, 10X, 6) 60 C for 30 s, 7) 72 C for 30 s 8) Go to step 6, 10X, 9) 72 C for 3 min, 10) 4 C hold

3. Clean the PCR product using Ampure XP beads, with a 0.7X bead to sample ratio.

PCR #2 'Final step'

This reaction completes the Illumina adapter and multiplexing barcode on the other end of the hTR amplicon, yielding a complete sequencing library. Both hTR Fragments 1 and 2 use Nextera A# and B# primers as the Forward and Reverse primers, respectively, according to your predetermined multiplexing scheme.

1. In a PCR tube, set up PCR #2.
    a. 10 µL of PCR #1 product
    b. 10 µL of 5X GC Buffer
    c. 1 µL of 10 mM dNTPs
    d. 2.5 µL of 10 µM Forward primer (Nextera A#)
    e. 2.5 µL of 10 µM Reverse primer (Nextera B#)
    f. 23.5 µL of H_2O
    g. 0.5 of µL of Phusion polymerase
2. Place in the thermocycler, and Run the PCR #2 program.

a. 1) 98 C for 3 min, 2) 98C for 30 s, 3) 65 C for 30 s, 4) 72 C for 30 s, 5) Go to step 2, 10X, 6) 72 C for 5 min, 7) 4 C hold

3. Clean the PCR product using Ampure XP beads, with a 0.7X bead to sample ratio.
4. Quantify libraries by Qubit (High Sensitivity DNA reagents) and Tapestation (High Sensitivity D1000 reagents) analysis.

3.2 - Protocol for in vitro RNA 2A3 probed library generation

RNA Fragmentation

NOTE: Fragmentation time will need to be optimized for your RNA of interest. See 'Quality Control' section of the guide. Use 1 µL of the final fragmented RNA for Bioanalyzer or Tapestation analysis to quantify fragmentation.

1. In a PCR tube, set up the RNA fragmentation reaction.
    a. 18 µL of fragmented RNA (or up to 250 ng total)
    b. 2 µL of 10X RNA Fragmentation Buffer
    c. Adjust volume to 20 µL with H_2O
2. Incubate the reaction in a thermocycler at 94 C for 1-5 min*.
3. Transfer the tube to ice.
4. Add 2 µL of 10X RNA Fragmentation Stop Solution.
5. Ethanol precipitate the RNA.
6. Resuspend the RNA pellet in 10 µL of H_2O.
7. Quantify fragmentation by Bioanalyzer or Tapestation analysis if needed.

Reverse transcription with SuperScript II

1. In a PCR tube, anneal random hexamer primers to fragmented RNA.
    a. 8 µL of fragmented RNA
    b. 2 µL of 20 µM random hexamers (cat #N8080127)
    c. 1 µL of 10 mM dNTPs
2. In a thermocycler, incubate at 70 C for 5 min.
3. Immediately transfer the tube to ice.
4. In the same tube, set up the reverse transcription reaction.
    a. 4 µL of 5X First Strand Buffer
    b. 1 µL of 120 mM $MnCl_2$
    c. 1 µL of RNAse inhibitor
    d. 1 µL of SuperScript II
5. Incubate the reaction at 25 C for 10 min.
6. Incubate the reaction at 42 C for 3 hr.
7. Incubate at 75 C for 20 min to heat inactivate.
8. Add 1.27 µL of 100 mM EDTA to chelate the $MnCl_2$.
9. Incubate at RT for 5 min.
10. Add 1.35 µL of 100 mM $MgCl_2$.

Second strand synthesis

1. Use the NEBNext Ultra II Non-Directional RNA Second Strand Synthesis kit (cat #E6111S), following the manufacturer's instructions.
2. Purify the cDNA with Ampure XP beads, using an 1.8X bead to sample volume ratio (144 µL of beads, to 80 µL of sample).
    a. Elute the cDNA from the beads using 53 µL of H_2O. Take 51 µL of eluant.
3. Quantify 1 µL of cDNA by Qubit analysis using the High Sensitivity DNA reagents.

Library preparation

Use the NEBNext Ultra II DNA Library Prep Kit for Illumina (cat #E7645S) following the manufacturer's instructions with the following recommendations.

1. NOTE: At Step 2 of the manufacturer's protocol, dilute the NEBNext adapters according to the input DNA quantification produced by Qubit analysis.
2. NOTE: Use Ampure XP beads for Step 3 of the manufacturer's protocol. Use the recommended bead volumes for a 200 bp approximate insert size distribution.
3. NOTE: The recommended number of PCR cycles for Step 4.1.3 of the manufacturer's protocol is 5-7 cycles.
4. NOTE: At step 5 of the manufacturer's protocol, purify the PCR reaction with a 0.8X Ampure XP bead to sample volume ratio (using 40 µL of beads).

5. Quantify libraries by Qubit (High Sensitivity DNA reagents) and Tapestation (High Sensitivity D1000 reagents) analysis.

References

1. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences* **106**, 97–102 (2009).

2. Cordero, P., Kladwang, W., VanLang, C. C. & Das, R. Quantitative DMS mapping for automated RNA secondary structure inference. *Biochemistry* **51**, 7037–7039 (2012).

3. Zarringhalam, K., Meyer, M. M., Dotu, I., Chuang, J. H. & Clote, P. Integrating Chemical Footprinting Data into RNA Secondary Structure Prediction. *PLOS ONE* **7**, e45160 (2012).

4. Lorenz, R., Wolfinger, M. T., Tanzer, A. & Hofacker, I. L. Predicting RNA secondary structures from sequence and probing data. *Methods* **103**, 86–98 (2016).

5. Peattie, D. A. & Gilbert, W. Chemical probes for higher-order structure in RNA. *Proceedings of the National Academy of Sciences* **77**, 4679–4682 (1980).
6. Weeks, K. M. Advances in RNA structure analysis by chemical probing. *Current Opinion in Structural Biology* **20**, 295–304 (2010).

7. Homan, P. J. *et al.* Single-molecule correlated chemical probing of RNA. *Proceedings of the National Academy of Sciences* **111**, 13858–13863 (2014).

8. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nature Methods* **14**, 75–82 (2017).

9. Mustoe, A. M., Weidmann, C. A. & Weeks, K. M. Single-Molecule Correlated Chemical Probing: A Revolution in RNA Structure Analysis. *Acc. Chem. Res.* **56**, 763–775 (2023).

10. Tomezsko, P. J. *et al.* Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**, 438–442 (2020).

11. Morandi, E. *et al.* Genome-scale deconvolution of RNA structure ensembles. *Nature Methods* **18**, 249–252 (2021).

12. Aviran, S. & Incarnato, D. Computational Approaches for RNA Structure Ensemble Deconvolution from Structure Probing Data. *Journal of Molecular Biology* **434**, 167635 (2022).

13. Olson, S. W. *et al.* Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP. *Molecular Cell* **82**, 1708-1723.e10 (2022).

14. Kladwang, W. & Das, R. A mutate-and-map strategy for inferring base pairs in structured nucleic acids: proof of concept on a DNA/RNA helix. *Biochemistry* **49**, 7414–7416 (2010).

15. Tian, S., Yesselman, J. D., Cordero, P. & Das, R. Primerize: automated primer assembly for transcribing non-coding RNA domains. *Nucleic Acids Res* **43**, W522–W526 (2015).

16. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610–1616 (2006).

17. Brookes, P. & Lawley, P. D. The reaction of mono- and di-functional alkylating agents with nucleic acids. *Biochem J* **80**, 496–503 (1961).

18. Stern, S., Wilson, R. C. & Noller, H. F. Localization of the binding site for protein S4 on 16 S ribosomal RNA by chemical and enzymatic probing and primer extension. *J Mol Biol* **192**, 101–110 (1986).

19. Mortimer, S. A. & Weeks, K. M. A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145 (2007).

20. Busan, S., Weidmann, C. A., Sengupta, A. & Weeks, K. M. Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry* **58**, 2655–2664 (2019).

21. Marinus, T., Fessler, A. B., Ogle, C. A. & Incarnato, D. A novel SHAPE reagent enables the analysis of RNA structure in living cells with unprecedented accuracy. *Nucleic Acids Res* **49**, e34 (2021).

22. Tijerina, P., Mohr, S. & Russell, R. DMS Footprinting of Structured RNAs and RNA-Protein Complexes. *Nat Protoc* **2**, 2608–2623 (2007).

23. Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA–Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).

24. Qin, Y. *et al.* High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA* **22**, 111–128 (2016).

25. Guo, L.-T. *et al.* Sequencing and Structure Probing of Long RNAs Using MarathonRT: A Next-Generation Reverse Transcriptase. *Journal of Molecular Biology* **432**, 3338–3352 (2020).

26. Kristen, M. *et al.* Manganese Ions Individually Alter the Reverse Transcription Signature of Modified Ribonucleosides. *Genes* **11**, 950 (2020).

27. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

28. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

29. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).

30. Masella, A. P. *et al.* BAMQL: a query language for extracting reads from BAM files. *BMC Bioinformatics* **17**, 305 (2016).

31. Incarnato, D., Morandi, E., Simon, L. M. & Oliviero, S. RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Research* **46**, e97 (2018).

32. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).

33. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).

34. Kladwang, W., VanLang, C. C., Cordero, P. & Das, R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat Chem* **3**, 954–962 (2011).

35. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).

36. Low, J. T. & Weeks, K. M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150–158 (2010).

37. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* **11**, 959–965 (2014).

38. Zuker, M., Jaeger, J. A. & Turner, D. H. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures

determined by phylogenetic comparison. *Nucleic Acids Research* **19**, 2707–2714 (1991).

39. Ding, Y. & Lawrence, C. E. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Research* **31**, 7280–7301 (2003).

40. Mathews, D. H. Revolutions in RNA Secondary Structure Prediction. *Journal of Molecular Biology* **359**, 526–532 (2006).

41. Herschlag, D. RNA Chaperones and the RNA Folding Problem. *J. Biol. Chem.* **270**, 20871–20874 (1995).

42. Bokinsky, G. & Zhuang, X. Single-Molecule RNA Folding. *Acc. Chem. Res.* **38**, 566–573 (2005).

43. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463**, 681–684 (2010).

44. Cordero, P. & Das, R. Rich RNA Structure Landscapes Revealed by Mutate-and-Map Analysis. *PLOS Computational Biology* **11**, e1004473 (2015).

45. Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Current Opinion in Structural Biology* **18**, 321–329 (2008).

46. Kim, H. *et al.* Protein-guided RNA dynamics during early ribosome assembly. *Nature* **506**, 334–338 (2014).

47. Haller, A., Soulière, M. F. & Micura, R. The Dynamic Nature of RNA as Key to Understanding Riboswitch Mechanisms. *Acc. Chem. Res.* **44**, 1339–1348 (2011).

48. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol* **20**, 474–489 (2019).

CHAPTER V

Investigations of hTR chaperones and disease-associated mutations

Histone H2A/B Dimers bind hTR and change its structural ensemble

Introduction

An unexpected finding of the cryo EM structures of human telomerase is the discovery of a histone H2A/H2B dimer associated with the catalytic lobe[1–4] (**Fig. 1A**). As higher resolution models of human telomerase were published, this finding proved robust and was corroborated by multiple laboratories, arguing that H2A/B is stably associated with the holoenzyme. While the role of histone proteins in forming the DNA-organizing nucleosome core is well studied, noncanonical functions of histone proteins as components of RNPs is an underexplored area of research with only a handful of published studies[5,6]. As of the time of writing, publications or preprints focusing on whether and how histone proteins contribute to telomerase assembly or catalytic function are nonexistent.

The telomerase-associated H2A/B dimer is located in the catalytic lobe of the RNP, nestled between the P5 and P6.1 stems of the hTR CR4/5 domain (**Fig. 1B)**. Given that CR4/5 forms a critically important interaction with TERT by sandwiching the TERT-RBD between the P6.1 and P6 stems, it is possible that H2A/B 'bookends' this protein-RNA contact and enhances its stability. In Chapter II, we demonstrated that CR4/5 is a structurally heterogeneous hTR domain, and adopts alternatively

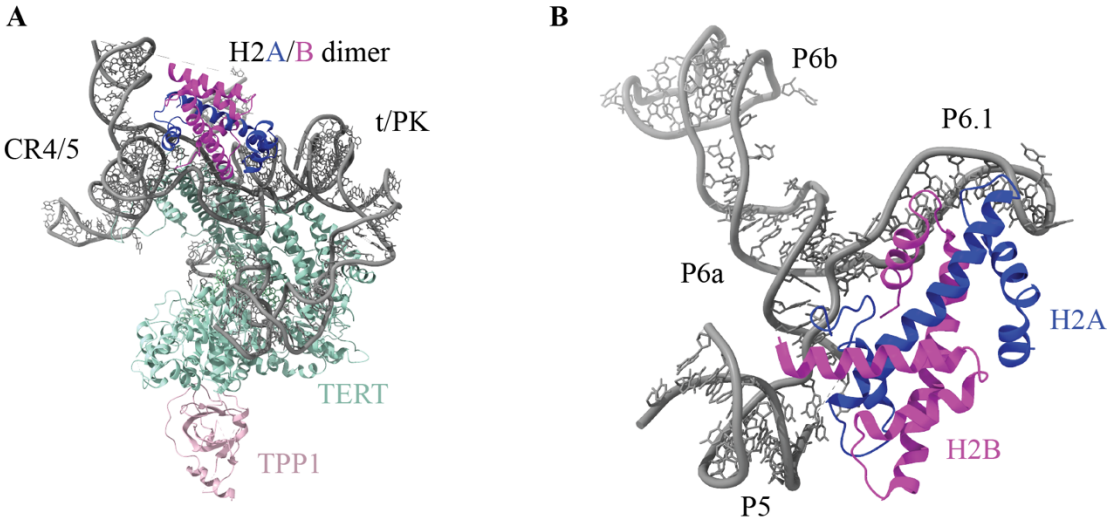


Figure 1: A histone H2A/B dimer bound within the human telomerase catalytic lobe.

Cryo EM structure adapted from Sekne et al[3]. (A) Overview of the catalytic lobe of human telomerase. Visible are the telomerase proteins TERT (light green), TPP1 (pink), histone H2A/B dimer (blue and purple), and the catalytic core domains of hTR CR4/5 and t/PK (gray). (B) Detailed view of the CR4/5 domain with the histone H2A/B dimer bound to a cleft formed by the P5 and P6.1 stems.

folded secondary structures in the absence of TERT protein[7]. In Chapter III, we showed that in the cellular environment, alternative CR4/5 conformations exist in the steady state and likely represent an unproductive obstacle in the telomerase biogenesis. It is interesting to speculate that H2A/B dimers could be a CR4/5 chaperone, contributing towards telomerase assembly in a manner analogous to that of the *Tetrahymena thermophila* La family protein p65 and Stem IV of TER[8–10]. Here, we aimed to characterize the binding of H2A/B dimers with hTR CR4/5 in vitro and to determine whether histone binding influences CR4/5 conformation. We find evidence that H2A/B dimers specifically bind CR4/5 and influence its structure. More work will be required to determine whether this phenomenon is of any significance in telomerase biogenesis or catalytic function.

Results

To study CR4/5-H2A/B interactions, we used an isolated CR4/5 RNA construct with FRET-coupled Cy3 and Cy5 dyes installed at positions U274 (P6 stem) and U306 (P6.1 stem), respectively (**Fig. 2A**). The construct was synthesized by splint ligating individually dye-labeled synthetic RNA fragments, as described previously[7]. Then, we queried the binding of H2A/B dimers to the dye-labeled CR4/5 RNA construct with a native electrophoresis mobility shift assay (EMSA) (**Fig. 2B**). In the presence of a nonspecific RNA competitor (yeast tRNA), we found that H2A/B binds dye-labeled CR45 in a concentration-dependent manner.

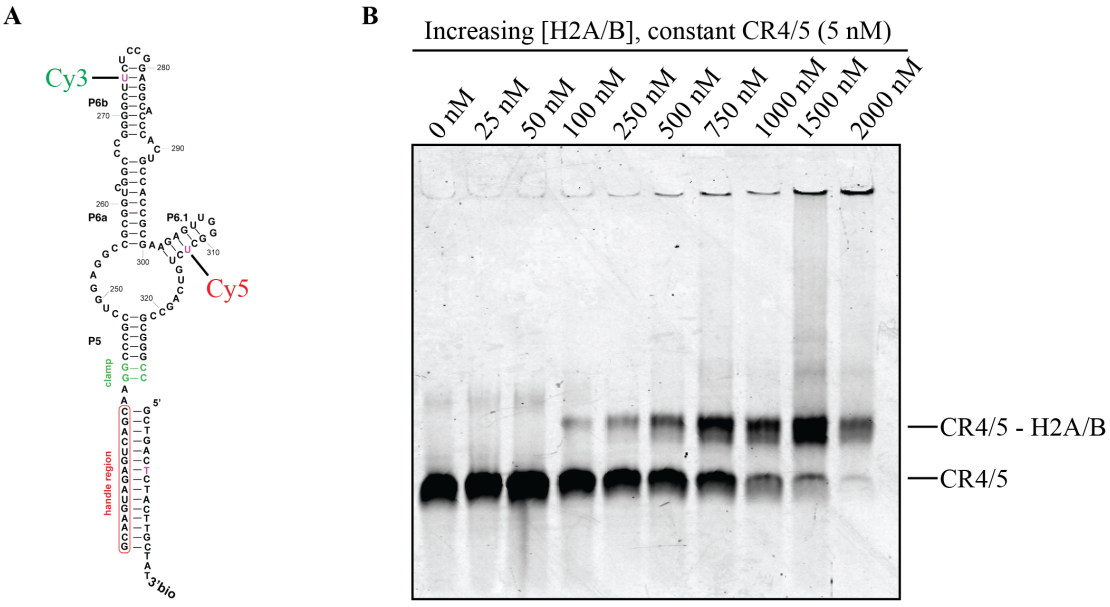


Figure 2: Gel electromobility shift assay of dye-labeled CR4/5 and H2A/B dimer.

(A) Synthetic CR4/5 construct with Cy3 labeled U274 (P6 stem) and Cy5 labeled U306 (P6.1 stem). And additional 5' single-stranded region was appended and can be hybridized with a biotinylated oligo for additional single-molecule studies. (B) Gel EMSA of CR4/5 and increasing amounts of H2A/B dimer.

Next, we employed a single-molecule FRET assay to ascertain whether H2A/B binding influences CR4/5 structure in regards to the helical arrangement of the P6 and P6.1 stems. CR4/5-histone dimer complexes were placed on the stage of a confocal microscope, and FRET values of free diffusing molecules were collected and binned into histograms. We found that in the absence of histone dimers, CR4/5 exhibits structural heterogeneity, with a FRET profile containing molecules ranging across the FRET scale, but forming mainly two populations roughly centered at ~0.3 and ~0.8 FRET (**Fig. 3, left panel**). In the presence of H2A/B dimer at two different concentrations (0.75 μM and 2 μM protein), the FRET distribution drastically shifts towards a broadened FRET distribution with molecules exhibiting roughly ~0.5 FRET (**Fig. 3, middle and right panels**). These data suggest that H2A/B binding does indeed alter the structural conformation of CR4/5 by altering the helical arrangement of the three-way junction, specifically by increasing the distance between the P6 and P6.1 stems. Notably, CR4/5-H2A/B complexes do not appear to adopt one major conformation but rather a broader ensemble when compared to naked CR4/5.

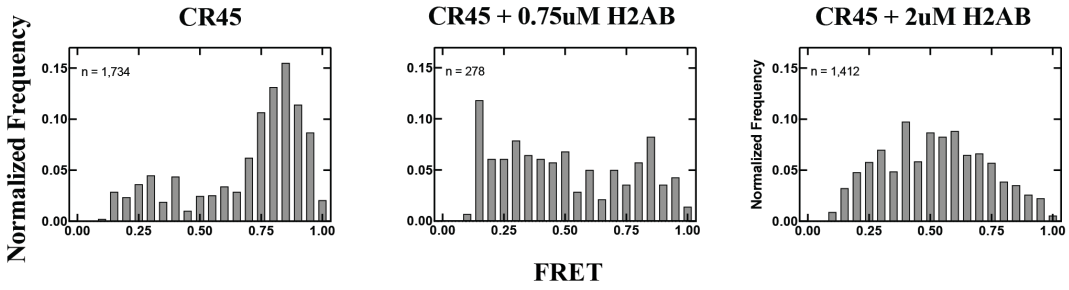


Figure 3: Single-molecule confocal FRET assay of CR4/5 and H2A/B.

FRET histograms depicting values extracted from CR4/5-H2A/B binding reactions. Binding reactions were directly placed on a blocked microscope slide prior to imaging. The number of molecules in the histogram analysis are labeled in the top left of each plot.

Discussion

We have shown that H2A/B specifically binds CR4/5 and remodels the relative orientation of the P6 and P6.1 stems. We previously demonstrated that TERT binding also is associated with a change in CR4/5 conformation, specifically by shifting the heterogenous TERT-free CR4/5 FRET distribution towards a more homogenous low FRET distribution[7]. Our findings represent a potentially fruitful lead into dissecting a role of H2A/B dimers in telomerase assembly. H2A/B dimers appear to shift the CR4/5 structural ensemble away from a mostly high FRET state and into a broad distribution of molecules with intermediate FRET values. With the available data, we can only speculate what the significance of this altered structural ensemble is regarding telomerase assembly. Given that CR4/5 appears to be structurally remodeled (or one of its conformations is selected) upon TERT binding, the histone

dimer-induced effect on CR4/5 structure may be beneficial for TERT recognition and telomerase assembly.

Future experiments should make use of CR4/5 mutants to better understand the sequence and structure determinants that underly the molecular recognition of H2A/B dimers. In vitro telomerase reconstitution assays performed in the presence of H2A/B dimers could shed light on whether the CR4/5 remodeling effect we observed is productive for telomerase assembly. Telomerase assembly in the presence of histone dimers could even be directly observed using a TIRF FRET assay, enabling kinetic measurements of the telomerase assembly steps. Finally, cell based assays making use of a CR4/5 mutant that does not bind H2A/B could provide information on whether histone dimers are bona fide hTR chaperones or play a role in the telomerase biogenesis.

MaP-seq studies of disease-associated hTR mutations

hTR mutations cause defects in telomerase assembly and catalytic function, leading to a spectrum of diseases with the hallmark of shortened telomeres[11–13]. A number of unique hTR lesions have been cataloged from patients with telomere syndromes[14] with etiologies manifesting at different steps of the telomerase biogenesis pathway. Mutations within the TERT-binding core domains (t/PK and CR4/5) reduce the binding affinity or TERT and decrease the overall production of active telomerase[11,15–17]. The Agarwal Lab has identified a CR4/5 mutation (G319A)

that reduces TERT binding[18], but underlying cause for this loss of binding affinity is not completely understood.

In a collaboration with the Agarwal Lab, we received patient-derived cells (primary fibroblasts and induced pluripotent stem cells (iPSCs)) harboring a similar heterozygous hTR mutation, A318C (**Fig. 4A**). The wild type A318 residue is located in the CR4/5 three-way junction, within the joint of the 'L' shape created by the P6 and P6.1 stems. In the assembled RNP, A318 is flipped outward towards TERT residue Cysteine 528, potentially forming a hydrogen bond through the 2' hydroxyl (**Fig. 4B**). Here, we propose to study the effect of the A318C mutation on hTR structure and to better understand the underlying mechanism that reduces this mutant hTR's affinity for TERT.

Proposed experiment

To study the impact of the A318C mutation on hTR folding in cells, we will employ a targeted DMS MaP-seq strategy as previously described in Chapter III. Because the A318C mutation is heterozygous and not expected to impact the processing of hTR, we expect that roughly half of all endogenous hTR molecules will be mutant. If mutant hTR adopts a secondary structure that is unique from the wild type, we hypothesize it will yield a distinct pattern of DMS modification. Furthermore, we hypothesize that RNA structure prediction guided by the DMS data will predict a secondary structure that is distinct from the wild type.

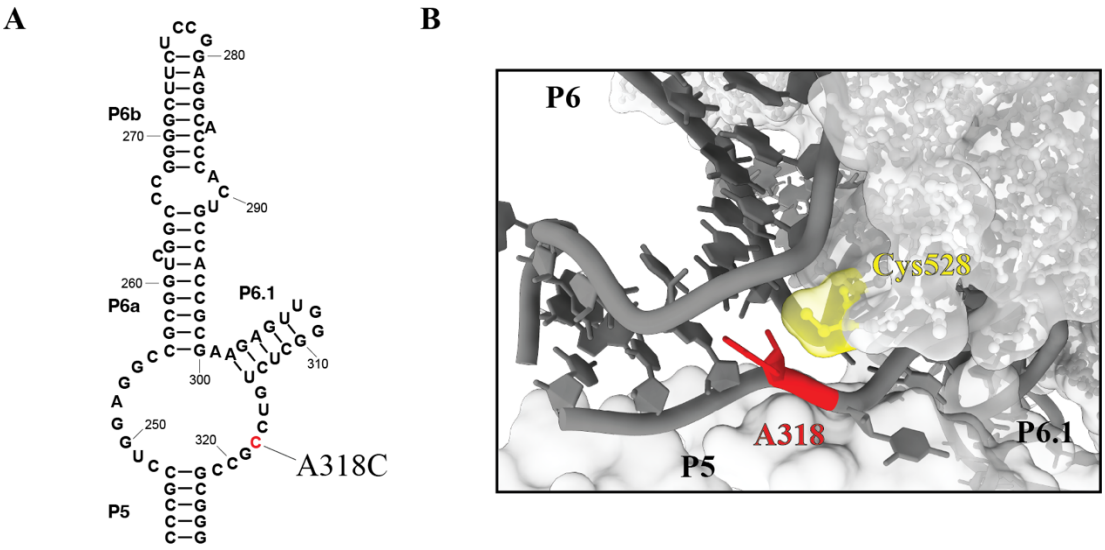


Figure 4: The telomere disease-causing CR4/5 mutation A318C.

(A) Secondary structure of the hTR CR4/5 domain with labeled A318C mutation. (B) Detailed view of A318 (red) flipped out of the CR4/5 three-way junction in the vicinity of TERT Cysteine 528 (yellow).

In the event that we detect structural differences of mutant hTR by DMS signals, we will also analyze the MaP-seq dataset with structural ensemble deconvolution tools. Because the biological significance of the wild type hTR structural ensemble is still poorly understood, it may be informative to compare it to the mutant hTR ensemble. If mutant hTR exhibits subtle differences in DMS signal, we can bioinformatically separate MaP-seq reads originating from wild type and mutant hTR. It will be interesting to compare the DMS profiles and structure predictions of wild type and mutant hTR alleles that are co-expressed in patient cells.

References

1. Ghanim, G. E. *et al.* Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* **593**, 449–453 (2021).

2. Wan, F. *et al.* Zipper head mechanism of telomere synthesis by human telomerase. *Cell Res* **31**, 1275–1290 (2021).

3. Sekne, Z., Ghanim, G. E., van Roon, A.-M. M. & Nguyen, T. H. D. Structural basis of human telomerase recruitment by TPP1-POT1. *Science* **375**, 1173–1176 (2022).

4. Liu, B. *et al.* Structure of active human telomerase with telomere shelterin protein TPP1. *Nature* **604**, 578–583 (2022).

5. Huang, R. C. & Bonner, J. Histone-bound RNA, a component of native nucleohistone. *Proceedings of the National Academy of Sciences* **54**, 960–967 (1965).

6. Soboleva, T. A. *et al.* A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLOS Genetics* **13**, e1006633 (2017).

7. Palka, C., Forino, N. M., Hentschel, J., Das, R. & Stone, M. D. Folding heterogeneity in the essential human telomerase RNA three-way junction. *RNA* **26**, 1787–1800 (2020).

8. Stone, M. D. *et al.* Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).

9. Berman, A. J., Gooding, A. R. & Cech, T. R. Tetrahymena Telomerase Protein p65 Induces Conformational Changes throughout Telomerase RNA (TER) and Rescues Telomerase Reverse Transcriptase and TER Assembly Mutants. *Mol Cell Biol* **30**, 4965–4976 (2010).

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

11. Yamaguchi, H. *et al.* Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. *Blood* **102**, 916–918 (2003).

12. Vulliamy, T. J. & Dokal, I. Dyskeratosis congenita: the diverse clinical presentation of mutations in the telomerase complex. *Biochimie* **90**, 122–130 (2008).

13. Savage, S. A. Human telomeres and telomere biology disorders. *Prog Mol Biol Transl Sci* **125**, 41–66 (2014).

14. Podlevsky, J. D., Bley, C. J., Omana, R. V., Qi, X. & Chen, J. J.-L. The telomerase database. *Nucleic Acids Res.* **36**, D339-343 (2008).

15. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H. & James, T. L. A molecular switch underlies a human telomerase disease. *Proceedings of the National Academy of Sciences* **99**, 16998–17003 (2002).

16. Robart, A. R. & Collins, K. Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked subunit variants. *J. Biol. Chem.* **285**, 4375–4386 (2010).

17. Errington, T. M., Fu, D., Wong, J. M. Y. & Collins, K. Disease-associated human telomerase RNA variants show loss of function for telomere synthesis without dominant-negative interference. *Mol Cell Biol* **28**, 6510–6520 (2008).

18. Boyraz, B., Bellomo, C. M., Fleming, M. D., Cutler, C. S. & Agarwal, S. A novel TERC CR4/CR5 domain mutation causes telomere disease via decreased TERT binding. *Blood* **128**, 2089–2092 (2016).